

**Fundamental validity issues of an English as a foreign language test: A  
process-oriented approach to examining the reading construct as measured by  
the DR Congo English state examination**

**By**

**Godefroid Bantumbandi KATALAYI**



**Thesis Presented in fulfillment of the requirements for the degree of Doctor of Philosophy  
in Language Education in the Faculty of Education at the University of the Western Cape,  
South Africa**

**Supervisor: Prof. Sivakumar SIVASUBRAMANIAM**

**February 2014**

## KEY WORDS

Validity

Construct validity

Reading construct

Reading context

Process-oriented approach

English state examination

Reading types

Processing levels

Text features

Item features



## DECLARATION

I, the undersigned, declare that “Fundamental validity issues of an English as a foreign language test: A process-oriented approach to examining the reading construct as measured by the DR Congo English state examination” is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

---

Godefroid Bantumbandi Katalayi

Date: 28 February, 2014



## ABSTRACT

The study aims to investigate the fundamental validity issues that can affect the DR Congo English state examination, a national exit test administered to high school final year students for certification. The study aspires to generate an understanding of the potential issues that affect the construct validity of a test within the epistemological stance that supports a strong relationship between test construct and test context.

The study draws its theoretical underpinning from three theories: the validity theory that provides a theoretical ground necessary for understanding the quality of tests needed for assessing students' reading abilities; the construction-integration theory that provides an understanding of how texts used in reading assessments are processed and understood by the examinees; and the strategic competence theory that explains how examinees deploy strategies to complete test tasks, and the extent to which these strategies tap into the reading construct. Furthermore, the study proposes a reading model that signposts the social context of testing; therefore, conceptualizing reading as both a cognitive and a social process.

As research design, the study adopts an exploratory design using both qualitative and quantitative data. Besides, the study uses protocol analysis and content analysis methodologies. While the former provides an understanding of the cognitive processes that mediate the reading construct and test performance so as to explore the different strategies examinees use to answer the English state examination (henceforth termed ESE) test questions, the latter examines the content of the different ESE papers so as to identify the different textual and item features that potentially impact on examinees' performance on the ESE tasks. As instruments, the study uses a concurrent strategies questionnaire administered to 496 student-participants, a contextual questionnaire administered to 26 student-participants, a contextual questionnaire administered to 27 teacher-participants, and eight tests administered to 496 student-participants.

The findings indicate that, the ESE appears to be less appropriate to the ESE context as the majority of ESE test items target careful reading than expeditious reading; on the one hand, and reading at global level than to reading at local level; on the other hand. The findings also indicate

that the ESE tasks hardly take account of the text structure and the underlined cognitive demands appropriate to the text types. Besides, the ESE fails to include other critical aspects of the reading construct. Finally, the findings also indicate that the ESE constructors may not be capable to construct an ESE with five functioning distractors as expected. Moreover, the inclusion of the implicit option 6 overlaps with the conceptual meaning of this option.

The entire process of the present study has generated some insights that can advance our understanding of the construct validity of reading tests. These insights are: (a) the concept of validity is an evolving and context-dependent concept, (b) reading construct cannot be examined outside the actual context of reading activity, (c) elimination of distractors can sometimes be a construct-relevant strategy, (d) construct underrepresentation is a context-dependent concept, and (e) a reading test cannot be valid in all contexts.

The suggested proposal for the improvement of the ESE requires the Congolese government through its Department of Education to (a) always conduct validation studies to justify the use of the ESE, (b) always consider the actual context of reading activity while developing the ESE, (c) revisit the meanings and interpretations of the ESE scores, (d) ensure the appropriateness of tasks to be included in the ESE, (e) ensure the construct representativeness of the ESE tasks, (f) revisit the number of questions to be included in the ESE, (g) avoid bias in the ESE texts in order to ensure fairness, (h) diversify the genres of ESE texts, (i) ensure the coherence of ESE texts through the use of transitions and cohesive devices, (j) ensure that the order of test questions is in alignment with the order of text information, (k) revisit the structure and length of the texts to be included in the ESE, (l) revisit the number of alternatives to be included in the ESE, and (m) reconsider the use of the implicit alternative 6.

## DEDICATION

I dedicate this thesis to my deceased father, Donatien Bantumbandi, who instilled in me the sense of hard work and the desire to achieve the highest accolades in life.



## ACKNOWLEDGEMENTS

*Almighty God, how happy are those who trust in you! (Psalms 84:12)*

I am humbly thankful to God, the Almighty Lord who has provided me with wisdom and divine guidance throughout this research endeavor. He has enabled me to get far in my academic itinerary despite moments of misfortune, doubt, confusion and challenges.

I am deeply grateful to my supervisor Professor Sivakumar Sivasubramaniam for his tutelage in the completion of this thesis. His insightful and constructive guidance have helped me to inform and shape the production of this thesis. I was extremely fortunate to have had a supervisor of such caliber and considerable skill. His deep humanity, unstinted care and support will certainly linger and serve to inform my academic development.

My sincere gratitude and heartfelt appreciation go to Professor Lorna Holtman, the Director of the Division of Post-Graduate Studies, for her heartfelt attention and support towards me during the entire process of this journey.

I owe to Professor Zubeida Desai, the Education Faculty Dean, enormous gratitude for her prompt support any time I have needed this. She was very considerate to me and she never hesitated to give me any fruitful opportunity necessary to sponsor this study.

I am deeply indebted to Professor Beverly Thaver, the Education Faculty Deputy Dean, for her valued advice and encouragement during some moments of doubt, confusion and challenges of my research journey. I am convinced that if I had missed such advice and encouragement, I would not have been able to complete my studies at the University of the Western Cape.

I am grateful to Dr. Caroline Kerfoot, former Head of the Language Education Department, for having accepted my application to study at the University of the Western Cape. Her support and advice were very valuable at the initial stage of my research journey.

I am highly appreciative of the financial support from The RYOICHI SASAKAWA YOUNG LEADERS FELLOWSHIP FUND (SYLFF). This support has largely contributed to the completion of this thesis in reasonable time.

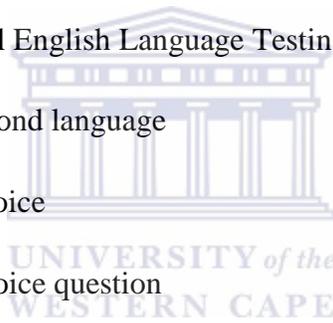
The production of this thesis owes much to my wife Angelique Kasokome Katalayi for her unrelenting support and never-ending unconditional love. It also owes much to all my children for their steadfast belief in me, their love, and confidence towards me. It particularly owes much to my mother Mbombo Katalayi, for her exceptional love and continuous encouragement towards me.

There are three people in the Democratic Republic of the Congo that I am sincerely grateful to. First, I am grateful to Mr. Stephane P. Lombela for initiating me to the basics of linguistics at the “Institut Supérieur Pédagogique de Kananga”. Secondly, I acknowledge the tutelage of Professor Tembwe Zembele for leading my first steps in research at the “Institut Supérieur Pédagogique de Bukavu”. Last, but not least, I am particularly indebted to Professor Andre Biselela Tshimankinda, for instilling in me the sense of research during my studies at the “Ecole Doctorale de l’Université Pédagogique Nationale”.

Finally, I wish to express my deepest gratitude to three friends. First, to Clement Basuayi who assisted me to find my way in the University of the Western Cape. His advice, support and encouragement have strongly contributed to the completion of this thesis. Then, to Dr. Justin Munyakazi who was at my side during hard times. I highly appreciate his advice and encouragement. Last, to Prosper Buabua Bakalowa who has always been helpful every time I travel to the Congo.

## ACRONYMS

<b>AERA</b>	American Educational Research Association
<b>ANOVA</b>	Analysis of variance
<b>APA</b>	American Psychological Association
<b>DR Congo</b>	Democratic Republic of the Congo
<b>ESE</b>	English state examination
<b>ID</b>	Item difficulty
<b>IELTS</b>	International English Language Testing System
<b>L2</b>	Foreign/Second language
<b>MC</b>	Multiple-choice
<b>MCQ</b>	Multiple-choice question
<b>NCME</b>	National Council on Measurement in Education
<b>TOEFL</b>	Teaching English as a foreign language



## LIST OF TABLES

Table 5.1: Demographics of students who participated in the study

Table 5.2: Demographics of teachers who participated in the study

Table 5.3: Number of schools and participants for each school category

Table 5.4: Description of tests used, participants' subject areas, number of schools selected and number of participants to which each test was administered

Table 6.1: Student-participants' self-rating of their proficiency levels in English

Table 6.2a: Reading frequency at school and at home for participants who were actually studying English 5 hours per week

Table 6.2b: Reading frequency at school and at home for participants who were actually studying English 2 hours per week

Table 6.3: Students' attendance to English classes

Table 6.4a: Preferences of the number of hours per week for the 420 student-participants who were actually studying English 5 hours per week

Table 6.4b: Preferences of the number of hours per week for the 76 student-participants who were actually studying English 2 hours per week

Table 6.5: Student-participants' experience with the ESE

Table 6.7: Student-participants' attitudes towards the ESE

Table 6.8a: Number of hours to write the test for groups who actually have 2 hours to write the ESE

Table 6.8b: Number of hours to write the test for groups who actually have 1 hour and half to write the ESE

- Table 6.9a: Participants' preferences for the number of questions to be included in the ESE for groups who actually had 9 questions
- Table 6.9b: Participants' preferences for the number of questions to be included in the ESE for groups who actually have 5 questions
- Table 6.10: Participants' preferences for the number of alternatives to be included in each test question
- Table 6.11a: Preferred number of parallel forms for groups who actually have 4 forms
- Table 6.11b: Preferred number of parallel forms for groups who actually have 2 forms
- Table 6.12a: Teacher-participants' qualification
- Table 6.12b: Teacher-participants' teaching experience
- Table 6.13: Sources of reading at school and at home
- Table 6.15: Frequency of classroom assessment methods
- Table 6.14: Frequency of use of different reading sources by the English teachers
- Table 7.1: Item frequency per test for the four reading types
- Table 7.2: Frequency of strategies the 496 student-participants used for answering the 56 test items
- Table 7.3: Mean item difficulty for the different reading skills included in each reading type
- Table 7.4: Comparing test focus and task difficulty between careful reading and expeditious reading
- Table 7.5: Comparing test focus and task difficulty between reading at global level and reading at local level
- Table 7.6: Processing levels required by the ESE tasks
- Table 7.7: Lower level processing versus higher level processing
- Table 8.1: Students' perceptions of text content difficulty and familiarity

Table 8.2: ESE texts classified in genres and related structural organizations

Table 8.3: Density of texts included in the study sample

Table 8.4: Average density of classroom texts

Table 9.1: ESE stem focus, completeness and orientation for the eight (8) tests

Table 9.2: Item difficulty index for test questions with negative stems

Table 9.3: Frequency of items per choice for the eight (8) tests

Table 9.4: Classification of the 56 test questions in categories of functioning distractors

Table 9.5: Frequency of guessing and elimination of distractors items for the 8 tests

Table 9.6: Frequency of participants' selection of alternative 6 in the 56 test items

Table 9.7: Frequency of participants' reasons for selecting alternative 6



## LIST OF FIGURES

Figure 3.1: Reading types (Source: Weir & Khalifa, 2008a)

Figure 3.2: Processing levels (Adapted from Weir & Khalifa, 2008a)

Figure 3.3: Embretson and Wetzel's (1987) Reading Comprehension Model

Figure 3.4: A multi-componential processing model of reading performance

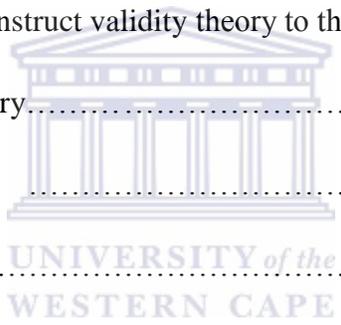


# CONTENTS

Key words.....	i
Declaration.....	ii
Abstract.....	iii
Dedication.....	v
Acknowledgements.....	vi
Acronyms.....	viii
List of tables.....	ix
List of figures.....	xii
Contents.....	xiii
<b>CHAPTER ONE: INTRODUCTION</b> .....	<b>1</b>
1.1 A point of departure .....	1
1.2 Aim and Scope of the Study .....	4
1.3 Context .....	7
1.3.1 The DR Congo: Sociolinguistic situation .....	9
1.3.2 The English state examination .....	10
1.4. Statement of the research problem.....	11
1.5 Objectives of the study .....	15
1.6 Main assumptions of the study .....	16
1.7 Research questions.....	16
1.8 Significance of the study.....	17
1.9 Organization of the chapters of the study.....	18
1.10 Clarification of key concepts.....	21

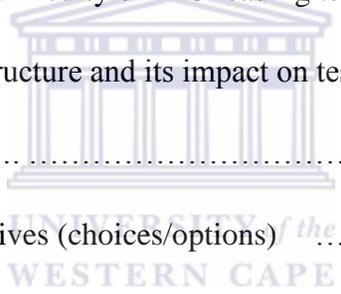


Summary of the chapter.....	25
<b>CHAPTER TWO: THEORETICAL FRAMEWORK.....</b>	<b>26</b>
2.1 Unified construct validity theory.....	27
2.1.1 Background.....	27
2.1.2 Unified construct validity theory: Content.....	29
2.1.3 Unified construct validity theory in language testing.....	37
2.1.4 Criticism on the unified construct validity theory .....	40
2.1.5 Responses to Lissitz and Samuelsen (2007)'s criticism....	43
2.1.6 Relevance of the unified construct validity theory to this study....	46
2.2 Construction-Integration theory.....	47
2.2.1 Constructivism in education .....	47
2.2.2 The schema theory.....	48
2.2.3 Construction-Integration theory.....	48
2.3 Strategic competence theory.....	52
2.3.1 Background.....	52
2.3.2 Strategic competence theory and reading test performance .....	54
2.3.3 Relevance of the theory to the study .....	55
Summary of the chapter.....	56
<b>CHAPTER THREE: DESIGNING A PROCESS-ORIENTED FRAMEWORK</b>	
<b>FOR EXAMINING THE READING CONSTRUCT.....</b>	<b>58</b>
3.0 Introduction.....	58



3.1	A process-oriented approach to examining the reading construct: Rationale .....	58
3.2	Weir and Khalifa (2008a)'s and Embretson and Wetzel (1987)'s frameworks to examining the reading construct: Insights and Issues.....	61
3.2.1	Overview of cognitive processing models .....	62
3.2.2	Weir and Khalifa (2008a)'s cognitive-processing approach to examining the reading construct: Description and insights.....	64
3.2.3	Embretson and Wetzel (1987)'s cognitive processing model of reading comprehension.....	72
3.3	Designing a framework for examining the reading construct as measured by the ESE .....	76
3.3.1	Naming the approach .....	76
3.3.2	How is the approach specific to the study context? .....	76
3.3.3	How does the present model conceptualize reading? .....	77
3.3.4	How does the present model relate to Weir and Khalifa's model? .....	78
3.3.5	Content of the model .....	78
3.4	Summary of the chapter.....	83
CHAPTER FOUR: LITERATURE REVIEW.....		84
4.1	Introduction.....	84
4.2	Reading and test taking strategies and their impact on test performance.....	85
4.3	Test context and its impact on the reading performance.....	89
4.3.1	Examinees' characteristics and their impact on reading performance: issues	

and insights from research.....	89
4.3.2 Social-contextual factors.....	96
4.3.3 Conditions under which the teaching and learning of reading take place.....	100
4.4 Reading types and processing levels as required by the test tasks.....	102
4.5 Text structure and its impacts on comprehension and test performance.....	107
4.5.1 Text content and its impact on comprehension.....	107
4.5.2 Text genres.....	111
4.5.3 Propositional density.....	117
4.6 Test variables as predictor of difficulty on MC reading tests .....	118
4.6.1 The multiple-choice stem structure and its impact on test performance.....	119
4.6.2 The multiple-choice key.....	122
4.6.3 The multiple-choice alternatives (choices/options).....	124
Chapter summary.....	130
<b>CHAPTER FIVE: RESEARCH DESIGN AND METHODOLOGY.....</b>	<b>132</b>
5.1 Statement of the study research questions.....	132
5.2 Research design .....	134
5.3 Research methods.....	135
5.3.1 Protocol analysis.....	136
5.3.2 Content analysis.....	137
5.4 Research site.....	137
5.5 Research participants.....	137



5.6	The study sample and sampling procedure.....	139
5.7	Research instruments.....	142
5.7.1	Strategies questionnaire.....	143
5.7.2	Contextual questionnaire administered to Grade 12 students .....	145
5.7.3	Questionnaire administered to Grade 12 English language teachers .....	145
5.7.4	Questionnaires validation process .....	146
5.7.5	Tests .....	149
5.8	Data collection procedure.....	151
5.9	Data analysis.....	154
5.9.1	Data coding .....	154
5.9.2	Data analysis procedures .....	155
5.10	Research ethics.....	158
	Summary of the chapter.....	160
	CHAPTER SIX: FINDINGS RELATED TO THE ACTUAL CONTEXT OF THE ENGLISH STATE EXAMINATION.....	161
6.0	Introduction.....	161
6.1	Grade 12 students' characteristics.....	162
6.1.1	Students' skills in English.....	162
6.1.2	Students' motivation for reading.....	164
6.1.3	Students' experience with the ESE.....	175
6.1.4	Students' attitudes towards the ESE.....	179



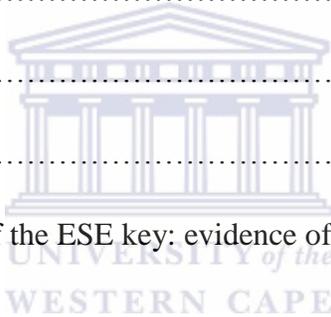
6.1.5	Additional insights on students' attitudes towards and motivation for the ESE.....	185
6.2	The Grade 12 English language teacher.....	196
6.2.1	Teacher qualification and experience .....	196
6.2.2	Teachers' in-service training .....	197
6.2.3	Teacher's motivation for teaching English... ..	199
6.3	Conditions under which the teaching/learning of reading takes place .....	201
6.3.1	Classroom size .....	201
6.3.2	Reading resources.....	202
6.3.3	The English curriculum.....	207
	Summary of the chapter.....	211
<b>CHAPTER SEVEN: FINDINGS RELATED TO THE VARIETY AND COMPLEXITY</b>		
<b>OF DIFFERENT READING TYPES AND PROCESSING LEVELS AND THE</b>		
<b>EXTENT TO WHICH ESE TASKS ARE APPROPRIATE TO THE ESE CONTEXT.....</b>		
7.1	Introduction.....	213
7.2	Frequency of items per test for different reading types and reading skills.....	215
7.2.1	The number of reading items included in each reading type .....	216
7.2.2	Reading skills necessary for answering different items.....	217
7.3	Frequency of strategies used for each of the four reading types.....	219
7.3.1	Strategies used for answering careful reading at global level items .....	220
7.3.2	Frequency of strategies used for answering careful reading at local level items... ..	224
7.3.3	Frequency of strategies used for answering expeditious reading at global	

level items.....	227
7.3.4 Frequency of strategies used for answering expeditious reading at local level items ...	233
7.4 Reading types and task complexity.....	234
7.4.1 Item difficulty for careful reading at global level test questions.....	236
7.4.2 Item difficulty for careful reading at local level test questions.....	236
7.4.3 Item difficulty for expeditious reading at global level test items.....	237
7.4.4 Item difficulty of expeditious reading at local level test questions..	237
7.4.5 Comparing task complexity for different reading types.....	237
7.5 The reading construct as investigated through processing levels required by the ESE tasks.....	239
7.5.1 The processing levels targeted by the test items.....	241
7.5.2 ESE focus: Lower level processing or higher level processing? .....	242
Summary of the chapter.....	244
 CHAPTER EIGHT: FINDINGS RELATED TO THE TEXTUAL FEATURES THAT POTENTIALLY AFFECT THE VALIDITY OF THE ENGLISH STATE EXAMINATION	
8.0 Introduction.....	247
8.1 The content of ESE texts.....	248
8.2 Identifying different genres included in ESE texts and the extent to which the ESE tasks tap into these different genres.....	262
8.2.1 ESE based on narrative texts.....	264
8.2.2 ESE based on expository texts.....	278

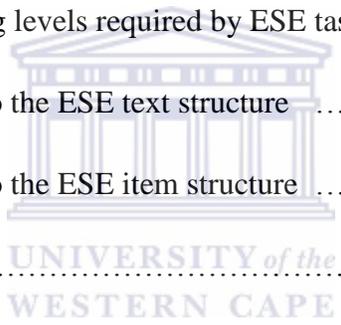
8.2.3 ESE based on practical writing texts .....	282
8.3 Examining the density of ESE texts and their appropriateness to the ESE context....	285
Summary of the chapter.....	289

**CHAPTER NINE: FINDINGS RELATED TO THE ITEM FEATURES THAT POTENTIALLY AFFECT THE VALIDITY OF THE ENGLISH STATE EXAMINATION.....292**

9.1 Introduction.....	292
9.2 ESE stem quality and its appropriateness to the ESE context.....	292
9.2.1 ESE stem focus.....	293
9.2.2 Stem completeness.....	297
9.2.3 Stem orientation.....	297
9.3 The ESE key.....	300
9.3.1 Investigating the structure of the ESE key: evidence of clues to getting the item right.....	300
9.3.2 Investigating the structure of the ESE stem: evidence of the distribution of the key..	304
9.4 The ESE distractors .....	305
9.4.1 Test items with no functioning distractors.....	307
9.4.2 Test items with only one functioning distractor.....	309
9.4.3 Test items with four functioning distractors .....	311
9.5 Investigating test wiseness through guessing and elimination of distractors strategies.....	314
9.5.1 Actual meaning of guessing strategy as investigated through a selected test item.....	316



9.5.2 Actual meaning of elimination of distractors strategy as investigated through a selected test item.....	318
9.6 Actual meanings of alternative 6.....	319
Summary of chapter.....	323
CHAPTER TEN: DISCUSSION OF STUDY FINDINGS .....	325
10.1 Introduction.....	325
10.2 The DR Congo ESE context and its impact on the validity of the ESE.....	327
10.3 Validity issues that pertain to the variety and degree of complexity of different reading types and processing levels required by ESE tasks .....	333
10.4 Validity issues that pertain to the ESE text structure .....	337
10.5 Validity issues that pertain to the ESE item structure .....	344
Summary of the chapter.....	349
CHAPTER ELEVEN: CONCLUSION.....	351
11.1 A Reappraisal .....	351
11.2 Relating the study outcomes to the research questions.....	353
11.3 New insights... ..	356
11.4 Proposal for ESE development.....	371
11.5 Limitations.....	377
11.7 Further research.....	379
REFERENCE LIST .....	382
APPENDICES.....	409



Appendix 1: State examination performance sheet .....	410
Appendix 2: Weir and Khalifa’s (2008) questionnaire.....	411
Appendix 3: Strategies codes.....	413
Appendix 4: Strategies questionnaire.....	414
Appendix 5: Contextual questionnaire administered to Grade 12 students.....	424
Appendix 6: Questionnaire administered to Grade 12 English language teachers.....	432
Appendix 7: Test materials.....	437
Appendix 8: Information sheet for research participants.....	459
Appendix 9: statement by the researcher.....	461
Appendix 10: Student-participants’ informed consent form.....	463
Appendix 11: Teacher-participants informed consent form.....	465
Appendix 12: Letter to the Provincial Education Services.....	466
Appendix 13: Letter to School Principals.....	468
Appendix 14: Teacher-participants report of variables that negatively affect reading classes and the students’ performance on the ESE.....	470
Appendix 15: Student-participants’ preferences for the number of hours per week for learning English in both general and technical schools.....	471
Appendix 16: Teacher-participants’ views on some statements related to English language teaching and assessment.....	472
Appendix 17: Strategies patterns and item difficulty for individual items.....	473
Appendix 18: Multiple-choice stem focus, completeness and orientation .....	480

Appendix 19: Distractors' weights across the 8 tests.....	483
Appendix 20: Frequency of option 6 across the eight tests.....	485
Appendix 21: Item difficulty and participants' perception of item difficulty for the 56 items across the eight tests.....	486
Appendix 22: Item difficulty for the 56 test items across the eight tests.....	488
Appendix 23: Student-participants selected responses to the contextual questionnaire.....	490
Appendix 24: Teacher-participants selected responses to the contextual questionnaire.....	504
Appendix 25: Selected student-participants profile.....	519
Appendix 26: Teacher-participants profile.....	522



# CHAPTER ONE

## INTRODUCTION

### 1.1 A point of departure

At the outset, I should mention that my background has been mainly in Applied linguistics, especially in the teaching of English as a foreign language, with limited experience in educational measurement in general and language assessment per se. My research background has been mainly in the issues related to the teaching of writing and grammar to students of English as a foreign language, particularly in disadvantaged contexts. As part of my professional experience, I have lectured TEFL methodology, reading and writing modules. My background therefore tends to limit my perspective to the validity concept to some extent. Nevertheless, I believe that divergent views can be informative. Keeping in mind this background information, I wish to discuss the educational and social concerns that have motivated this study. I hope that the discussion will act as an awareness-building exercise and a point of departure for this research.

In measurement research, educational tests are generally evaluated in terms of their validity and reliability. Simply defined, the validity of a test addresses the meaning and utility of the test scores, while the reliability of a test provides an indication of the consistency of the test scores (Mislevy, 2004; Moss, 1994; Messick, 1989; Kane, 2002, 2004). I hasten to state that this study deals with validity issues that potentially affect the Democratic Republic of the Congo (henceforth DR Congo) English state examination.

Based on my scrutiny of validity research as part of this research journey, I wish to mention three trends that appear to constitute the main concerns in validity research. The first trend focuses on test uses and consequences as main concern of validity (Bachman, 1990; Bachman & Palmer, 1996). Tenets of this trend argue that validity should mainly consist of providing evidence justifying the use of a particular test (Bachman & Palmer, 1996). The second trend seems to view validity as an argument; therefore, validation studies must be conducted from an argument-based framework perspective (Cronbach, 1988; Kane, 1992, 2002, 2011; Shephard, 1993). Studies that have used this framework have generally followed Toulmin's analysis of practical reasoning

(Toulmin, 1958), and they have generally suggested that, in validating tests, one needs to provide an explicit interpretive argument on claims he/she makes on the uses and interpretations of test scores (Kane, 1992, 2002, 2011; Shephard, 1993). The third and last trend appears to consider the analysis of test tasks and test context in order to justify the test use (Weir, 2005; Weir & Khalifa, 2008b; Weir et al., 2008a; Weir et al., 2008b; Sarojani & Krishnan, 2011). This trend suggests that validation studies should mainly focus on an extensive analysis of test tasks and make an evaluation of the appropriateness of these tasks to the test context. I wish to mention that this study is mainly situated in the third trend, although it is also informed by many insights from the second trend.

The educational and social concerns that motivate this study relate to the use of the English state examination as a high-stake large-scale student achievement test. As a high-stake test, students' scores on the ESE, combined to their scores in other school subject tests, are used to make some important decisions. One of these decisions is that the student gets the high school national certificate; therefore, he/she can further his/her studies at university or start a professional life. Another decision is that the student is denied of a certificate; therefore, he/she must retake the test one year later; with a possibility of being denied of a certificate many times. Furthermore, as a high-stake test, the results from the state examination in general, and the English state examination, in particular, can result in the labeling of schools in pejorative ways, and even the closing of some schools or the dismissal of the school principals. When state examination results are published, there is generally a range of reactions from pride to shame among students, school staff, and parents. There are some anecdotes that bring to light cases of suicides and scandalous behaviors.

As a large-scale test, the ESE is administered to all the final year high school students in the whole country on the same day, on the same time and in the same expected conditions. Furthermore, as a student achievement test, the ESE aims to evaluate the extent to which students have achieved the English curriculum standards.

The preceding clarifications on the ESE help to situate the educational and social concerns I propose to address in this study. Since the ESE scores, in combination with the scores from other school subject tests, are used to make important decisions, test developers and users need sufficient evidence to support their decisions. In light of this statement, as a multiple-choice

reading test, examinees' scores on the ESE are meant to provide an indication of their engagement with the test tasks. Such an engagement can be reflected in the way examinees work to understand the text and to understand the expectations of the test questions. It can also be reflected in the way examinees attempt to understand the meanings and implications of the different item options in light of the text, and how they select and discard item options based on the way they have understood the text. This argument appears to reflect Kane's (2002) conceptualization of validity as argument. Kane argues that validity should provide an argument on the evaluation of the plausibility of the proposed interpretations and uses of test scores. I hasten to state that this study also draws on this conceptualization of validity.

At this juncture, I wish to provide some data that illustrate how the ESE, in particular and the state examination, in general, constitute a critical social and educational concern that deserves attention. In 2013, the number of students who wrote the state examination was 545,569 according to statistics provided by the Ministry of Education and broadcast in national media (Radio Okapi, 22 June 2013; Le Potentiel, 21 June 2013). However, the national pass rate was 69 per cent with some disparities. In the province of Kinshasa, for example, the pass rate was 52 per cent while it was 17 per cent in the province of Maniema. Furthermore, while some schools recorded a pass rate of 100 per cent, other schools recorded a pass rate of 0.00% (Radio Okapi, 17 July, 2013; Le Potentiel, 17 July, 2013).

My scrutiny of the previous statistics suggests students' poor performances on the national test. Since the stakes I have described earlier are high, I can state that the state examination, in general and the ESE, in particular constitute an educational and social concern that should be addressed at all levels. However, from my experience as an English language educator for nearly twenty years, the public general opinion generally attributes students' poor performances on the state examination to the educational system that is characterized by rote learning, lack of appropriate resources, poor motivation on the part of the learners and the teachers, and the country's socio-economic and political environment that is characterized by unemployment, war, insecurity, poor salary, etc. However, less or no attention has been paid to the quality of the tests that are used to assess the students. Implicit in this observation is the belief that the state examination is assumed to be of good quality; and any student who does not pass it is presumed to have failed to achieve the national curriculum standards. At this juncture, I wish to caution

that this description should be viewed as a dynamic and discursive meaning structure I have constructed based on my continuous experience as an educator.

The preceding observation kindled my curiosity to ask myself this question: “Why do we not consider the quality of the state examination as one possible cause of students’ poor performances?” In my attempt to answer this question, I came up with these tentative answers: (1) The fact that the state examination is constructed by “experts” can be the reason why most people tend to believe that the test itself is good; (2) The fact that the test results are generally reported with some statistical sophistication that is hardly accessible to the teachers, the students and the parents may also be a reason why the majority of people generally believe that the test cannot be a problem; (3) The fact that there are *almost* no validation studies that ensue test administrations can also explain why it is hard to point to the state examination as a problem.

In light of the preceding interrogation, I realized that one way to understand the reasons that account for students’ poor performances on the state examination would be to interrogate the quality of this test. This could be achieved through validation studies that aim to ensure that the different subtests that are being used are valid; therefore, students’ results can provide an indication of their actual performances.

The preceding argument provides a rationale for this study. I wish to mention that, to the best of my knowledge, since the English state examination was instituted in 1967 as a subtest of the state examination, no validation studies have been conducted. Therefore, my position as Senior lecturer in the Department of English at the ‘Institut Supérieur Pédagogique de Kananga’ motivated me to address this educational and social problem. I wish to state that this study is a continuation of my two recent studies (Katalayi, 2011; Katalayi & Sivasubramaniam, 2013) that have provided a background understanding of the validity concerns I propose to investigate here. Having provided my motivation to conduct this study in this section, I propose to state the aim of the study and provide its scope.

## **1.2 Aim and Scope of the Study**

This research aims to investigate the fundamental validity issues that can affect the DR Congo English state examination, a national exit test administered to high school final year students for

certification. The study aspires to generate an understanding of the potential issues that affect the construct validity of a test within the epistemological stance that supports a strong relationship between test construct and test context.

Since the ESE is a multiple-choice text-based test, I am curious to find out if by examining the reading construct within the broader context of reading activity can help to understand:

1. If the actual context of the ESE can influence examinees' performances on the ESE tasks;
2. If there is a need to include specific types of tasks in order to ensure the ESE validity;
3. If the features of the ESE texts can impact on the examinees' execution of the ESE tasks;
4. If the features of the ESE item can impact on the examinees' execution of the ESE tasks.

In order to address the afore-mentioned purpose, I intend to tease out the theoretical underpinning of the study from these three theories: the validity theory, the construction-integration theory and the strategic competence theory. The validity theory is meant to provide a theoretical ground necessary for understanding the quality of tests needed for assessing students' reading abilities. The construction-integration theory is meant to provide an understanding of how texts that are used in reading assessments are processed and understood by the readers; and finally the strategic competence theory is meant to address how examinees deploy strategies to complete test tasks and the extent to which these strategies tap into the reading construct.

Concerning the conceptual scope, I believe that, since the ESE is a text-based test, examinees' attempts to process the text and comprehend it are significantly affected by the specific context under which they have been conducting reading activity (Bloome, 1985). By taking this conceptual stance, I view reading activity as a situated task that takes place in a specific context with specific task demands (Bloome, 1985; Weir, 2005). Consistent with this stance, I propose to advocate for a development of a reading model that signposts the social context of testing; therefore, conceptualizing reading as both a cognitive and a social process (Bloome, 1985). In order to develop this model, I premise that, since reading activity is a multi-componential process where a constellation of factors interact in order to impact on text processing and comprehension (Fletcher, 2006; Snow, 2003; Perfetti, 1997; Field, 2004), the assessment of reading should reflect not only the cognitive and social aspects of reading, but also the complex

and multiple factors that affect reading activity and their various interactions within the reading construct. Consistent with this premise, the model I intend to propose links together the reading construct, the test context, the features of the texts used as reading materials and the features of the test tasks included in the test. However, in order to operationalize the reading construct, the model I intend to propose purports adopts a four-cell matrix reading taxonomy with careful reading and expeditious reading that can be conducted at both global level and local level (Weir & Khalifa, 2008b; Weir et al., 2008a; Weir et al., 2008b; Sarojani & Krishnan, 2011; Katalayi & Sivasubramaniam, 2013).

Furthermore, I will argue that by examining the reading construct within the larger context of reading activity, the insights generated by this study can, not only be gainfully deployed in the educational practice of reading assessment, but also be used by reading assessment researchers to inform their validation studies. Therefore, in order to identify the gaps in reading assessment literature and relate these gaps to my study aim, I will scrutinize findings reported by studies that have investigated the reading construct by looking at the examinees' use of strategies while completing test tasks (Weir 2005; Hirano, 2008; Cohen & Upton, 2007; for instance). More specifically, I will gain understanding from studies that have evaluated how examinees' performance on reading tests largely depends on their capacity to deploy appropriate strategies (Cohen & Upton, 2007; Rupp, Ferne, & Choi, 2006). Furthermore, I will review studies that have investigated the impact of test context on test performance. More specifically, I hope to gain insights from studies that have investigated how the examinees' characteristics, the teachers' characteristics, and the conditions under which the teaching/learning of reading take place can impact on reading performance (Lee & Shute, 2010; Nassaji, 2003; Guthrie *et al.* 2007; Murray, Riazi & Cross, 2012; Pelayo & Brewer, 2010). Moreover, I will scrutinize studies that have investigated the different types of reading in relation to different processing levels in order to understand how the difficulty of the ESE can also be function of the complexity of ESE test tasks as well as the processing levels required by the ESE tasks (Weir & Khalifa, 2008b, Weir, Hawkey, Green & Devi, 2008; Weir, Hawkey, Green, Unaldi & Devi, 2008; and Sarojani & Krishnan, 2011). Besides, in order to understand the validity issues that pertain to the kind of ESE texts, I will review studies that have established the impact of textual features on comprehension difficulty and reading performance (Yousif & Shumaimeri, 2006; Chang, 2006; Shiotsu & Weir, 2007; Alidib, 2004; Yali & Jiliang, 2007). Last, but not least, since the ESE is a

multiple-choice test, and as research recognizes that this test format is sensitive to construction procedures, I will also gain insights from previous studies that have examined the impact of the structure of the multiple-choice test item on examinees' choice of strategies and their selection of test question answers (Downing & Haladyna, 2006; Ascalon *et al.*, 2007; Haladyna, Downing & Rodriguez, 2002).

In order to achieve my aim in this envisaged investigation, I will use an exploratory design using both qualitative and quantitative data (McMillan & Schumacher, 2006). Regarding the methodological orientation, I will use protocol analysis and content analysis methodologies. Protocol analysis methodology is meant to provide an understanding of the cognitive processes that mediate the reading construct and test performance (Ericsson & Simon, 1993; Park, 2009); and this will enable me to explore the different strategies used by examinees to answer the ESE test questions. Content analysis methodology will help me examine the content of the different English state examination papers and identify the different textual and item features that potentially impact on examinees' performance on ESE tasks.

Having provided the aim and scope of this study in this section, I propose to describe the context of this study in the following section. I hope an understanding of the context of the study can provide an understanding of the conceptual and methodological choices I have taken in this study.

### **1.3 Context**

This study is an investigation that I have initiated at the Department of Language Education, Faculty of Education, University of the Western Cape, South Africa. However, this investigation is situated in the Congolese (the Democratic Republic of) context. With an area of 2.34 million square kilometers and an estimated population of about 65.71 million in 2012, the Democratic Republic of the Congo is the largest country in sub-Saharan Africa in terms of size and the second largest in terms of population (The World Bank, 2013).

The study attempts to determine the centrality of test context in examining the test construct in a particular context characterized by social, political, economic and cultural challenges. This is to suggest that, the study intends to describe the kind of test tasks that are used to assess students'

reading abilities in English and evaluate the extent to which these tasks reflect the actual test context.

By considering an approach that acknowledges the centrality of the test context in examining the test construct, the present study attempts to determine whether this approach can address the malaise in the country with regard to the testing of English as foreign language. I wish to mention that the malaise referred to here is not only specific to English language testing, but it also relates to the whole Congolese educational system. This malaise manifests itself in many symptoms. One main symptom is a sharp decline in the education system after two decades of violent strife, large displacements of populations and prolonged economic deprivation. Today, a key feature of the Congolese education system is an almost complete lack of government provision and financing at all levels of education as it is documented that all types of educational institutions, public and private, are almost entirely financed by households (The World Bank, 2005).

Another symptom of the malaise that characterizes the educational system in the Congolese context is a lack of textbooks, laboratories, equipment and libraries (The World Bank, 2005). Besides, there are almost no system of regular in-service training and professional development for teachers; hence, this impacts on students' achievement (The World Bank, 2005; Kasanga, 2012). Furthermore, the education is outdated and it is of limited relevance since curricula and standards have not been officially revised for over twenty-five years (The World Bank, 2005). I believe that these symptoms, although they characterize the education system of Congo in general, they are also relevant in the context of English language assessment.

Nevertheless, there are some specific symptoms that can explain the malaise in the country with regard to English as foreign language testing. One main symptom is that the ESE has never been validated although it has been used for over than forty (40) years. Furthermore, the general public opinion is that the teaching of English, in general and its assessment, in particular occur in an environment where there is little or no inclination to read, abject paucity of reading materials, poor motivation on the part of the learners and the teacher as well as poor teacher quality.

These general and specific symptoms are seen as educational and socio-economic problems that the study hopes to investigate. Nevertheless, I wish to state that there is little evidence supporting

the previous description of the study context. However, I wish to caution that, this description should be viewed as a dynamic and discursive meaning structure I have constructed based on my continuous experience as an English teacher for nearly twenty years. Therefore, the study hopes to provide a more documented description of the context of the English state examination on the basis of which I will suggest a proposal for ESE development. At this juncture, in order to augment the understanding of the context of this study, I propose to present the sociolinguistic situation of the DR Congo in the following subsection.

### **1.3.1 The DR Congo: Sociolinguistic situation**

The sociolinguistic situation of the DR Congo is complex. It can be thought of as a three-level hierarchy in which French as the official state language occupies the top rung of the ladder. Then come the so-called four national languages (Ciluba, Kikongo, Kiswahili, and Lingala) that serve inter-group communication in their areas of influence. Finally, comes an estimated more than 200 “vernaculars” that serve intra-group communication (Kasanga, 2012). The 2006 Constitution of the DR Congo has recognized this hierarchy as it states that, in the first two grades of primary school, instruction should be provided in one of the four main national languages; while French becomes the language of instruction from grade 3 (Constitution de la République Démocratique du Congo, 2006).

However, this existing triglossic linguistic structure is changing in a quadriglossic structure with the extent to which world-wide globalization has spread English into a “global language” (Mufwene, 2010). In the DR Congo, much of this evolution can be attributed to the prescription of English as a school subject from grade 9 in secondary schools. Besides, the prestige of English has increased as a result of the influx of foreign multinationals and the United Nations linked organizations (Kasanga, 2010, 2012). In light of this emerging quadriglossic structure in the context of the DR Congo, English takes the top position as a ‘super-international language’ that is used for global business and regional and international communication. French keeps its status of a locally ‘higher’ language of official national institutions while the four national languages appear in the third position and the different vernaculars appear at the bottom of the hierarchy. In a study that aimed to investigate the visibility of English in the French-dominated linguistic landscape that is the DR Congo, Kasanga (2012, p. 198) found an infinitesimal presence of

English in intra-national transactions and in public display and he concluded that English “is still a creeper in a French-dominated linguistic space”; suggesting that there is very little or no sign of competition to rival French as a language of wider communication. Kasanga concluded that English in the DR Congo is a preserve of a privileged minority, the highly educated, the upwardly mobile, the widely traveled and the diasporic returnees.

The enormous linguistic diversity mentioned above, coupled with the educational challenges discussed prior to that complicate the task of the teaching/learning of English. Therefore, the study hopes to argue that, in order to understand students’ reading behaviours, we need to consider the specific situations in which reading occurs as well as the reading tasks that readers confront. Consistent with this argument, while validating the English state examination, the test tasks should be described in terms of cognitive processes and contextual parameters. In the following point, I provide a description of the English state examination. I believe this description can also contribute to an informed understanding of the context of the study.

### **1.3.2 The English state examination**

In DR Congo, final year students (grade 12) are required to take a national exit test in order to get a national certificate. This exit test is called the “examen d’etat” (state examination). Although the content of the state examination depends on the students’ main subjects, this test consists of four subtests: the science subtest, the cultural knowledge subtest, the main subject subtest, and the language subtest (Katalayi, 2011). Referring to Katalayi, each of the subtest has different parts, each part including questions on specific school subjects depending on the students’ main stream subjects. In light of this scope, the science subtest has four parts that include questions of mathematics, physics, chemistry and biology. The cultural knowledge subtest has three parts that include questions of geography, history and philosophy. The main subject subtest includes questions on the main subjects of the students. Finally, the language subtest has two parts that include the French and the English test. This study investigates validity issues that pertain to the English test.

The objectives of the state examination as stated by the government are to:

- check whether the curriculum was respected and completed entirely by each school;

- determine the students' level in the mastery of the materials foreseen in the curriculum;
- ensure that students who have obtained the national certificate are capable to successfully undertake their studies at the level of university either in the country or abroad;
- Ensure that the training received allows candidates to defend themselves at all solicitations of jobs when necessary (Revue de l'Inspecteur de l'Enseignement, 1984).

The ESE is a multiple-choice text-based test. In my previous study (Katalayi, 2011), I provided a description of the form and content of the ESE that I wish to summarize as follow: The ESE is printed on a four- page test booklet together with the French test. This test includes a text passage immediately followed by questions based on the text. The ESE includes nine (9) questions for candidates from general streams (and some specific technical streams) and five (5) questions for candidates from most technical streams. Candidates from each stream have a specific test that generally includes a text whose content relates to their subject knowledge, although it is common for candidates from two or more related streams to have the same test. In order to avoid collaboration between candidates, the ESE, like all other subtests of the state examination, is constructed in parallel/alternate forms. Students from general streams and some few technical streams where the test includes nine questions are given a four parallel form test, while students from most technical streams where the test includes five questions are given a two parallel form test. In the next section, I propose to state the research problem.

#### **1.4. Statement of the research problem**

One of the most important concerns with most large-scale assessments is the lack of capacity to clearly interpret evidence derived from examinees' performance. This usually results in having tests that hardly provide enough information to test developers and users; hence, the inferences drawn and decisions taken from the test results can be challenged.

The validity of a multiple-choice reading test like the ESE presents even more concerns. The first concern relates to the design of an approach that acknowledges the impact of test context on test construct. This is to suggest that, since reading is a situated activity that takes place in a certain

context with specific demands, the most relevant approach to examining the validity of a reading test would be the one that examines reading construct within the larger context of reading activity. Viewed from this stance, the validity of the English state examination should be examined in terms of the extent to which the test tasks not only target the different aspects of the reading construct, but also in terms of their appropriateness to the test context. This suggestion appears to concur with the concept of validity as “the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores” (AERA, APA, & NCME, 1985, p. 9).

In keeping with this stance, evidence suggests that examinees’ individual characteristics potentially impact on their performances on the reading test. Among these characteristics, I wish to mention the examinees’ skills (Nassaji, 2003; Barnett, 1986), their motivation for reading (Mucherah, & Yoder, 2008), their attitudes towards the test (Murray, Riazi & Cross, 2012; Rasti, 2009, Han, Dai & Yang, 2004), and their experiences with the test (Reeve *et al.*, 2009; Peña & Quinn, 1997). Furthermore, evidence suggests that the conditions under which reading instruction takes place can affect examinees’ performances on the reading test (Greaney, 1996; Lonsdale, 1997). Among these conditions, I wish to mention the availability and quality of the reading materials (Greaney, 1996; Lonsdale, 1997), and the availability and quality of the curriculum (Petrina, 2004). Moreover, evidence also suggests that teachers’ qualification and experience, their motivation for teaching, their participation in in-service trainings also contribute to students’ performance in reading tests (Harris & Sass, 2001; Samupwa, 2008; Jahangir, Saheen, & Kazmi, 2012; Conco, 2004). Picking upon these research insights, the first concern I propose to investigate in this study is to determine the extent to which the ESE tasks reflect the ESE context.

The second concern the study hopes to investigate pertains to the variety and degree of complexity of the different reading types examinees conduct while completing the ESE test tasks, and the extent to which the test tasks are relevant to the test context. Studies that have adopted a process-oriented approach to examining the reading construct that is based on a four-cell matrix reading taxonomy have suggested that, in order to be construct-valid, a reading test needs to include test tasks that require examinees to read the text carefully and expeditiously at global level and local level (Weir & Khalifa, 2008b; Weir *et al.*, 2008a; Weir *et al.*, 2008b;

Sarojani & Krishnan, 2011; Katalayi & Sivasubramaniam, 2013; Rouet *et al.*, 2001; van Steensel, Oostdam & van Gelderen, 2012; Song, 2008). Furthermore, these studies have indicated a difference in task complexity regarding the examinees' use of the different strategies in answering test questions. These studies have generally supported that, test questions that require examinees to read at local level appear to be easy as the examinees have to find relevant information to answer the test questions at the sentence level and recourse to their knowledge of grammar and vocabulary (Rouet *et al.*, 2001; van Steensel, Oostdam & van Gelderen, 2012; Song, 2008). On the other hand, questions that require examinees to read at global level appear to be difficult as they require examinees to read at paragraph/text level by attempting to generate more connections between their knowledge and text information (Rouet *et al.*, 2001; van Steensel, Oostdam & van Gelderen, 2012; Song, 2008). Furthermore, these studies generally support that test tasks that target careful reading appear to be more cognitively demanding than those tasks that target expeditious reading on the ground that careful reading tasks require a frequent combination of strategies while expeditious reading test tasks appear to require less frequent combination of strategies (Weir & Khalifa, 2008b; Weir *et al.*, 2008a; Weir *et al.*, 2008b; Sarojani & Krishnan, 2011; Katalayi & Sivasubramaniam, 2013).

In light of the preceding research insights, the second concern I propose to investigate is to determine the extent to which the ESE includes tasks that require examinees to read the text carefully and expeditiously at both local level and global level. Furthermore, since my epistemological orientation supports my intention to examine the reading construct within the framework of the context of reading activity, I also propose to investigate the extent to which ESE tasks are sensitive to the ESE context.

The third concern that the study proposes to investigate relates to the extent to which the different textual features of the ESE texts can impact on the complexity of ESE tasks. There is extensive literature that suggests the impact of textual features on text comprehension and test performance (Best *et al.*, 2006; Yousif & Shumaimeri, 2006; Chang, 2006; Alidib, 2004; Yali & Jiliang, 2007; Sheehan & Ginther, 2001; Embretson & Wetzel, 1987; Gorin, 2005; Kintsch, 1988, 1994, 1998; van Dijk & Kintsch, 1978). More specifically, evidence suggests that reader's knowledge of the text content and his/her degree of familiarity with this content can impact on his/her understanding of the text and his/her performance on the test (Best *et al.*, 2006; Snow,

2002; Yousif & Shumaimeri, 2006; Chang, 2006). Furthermore, research suggests that, since text genre functions as textual schemata and limits the meaning-potential of a particular text, reader knowledge of text genre can help to construct the text macrostructure, and this can facilitate his/her comprehension of the text (Alidib, 2004; Fowler, 1989; Yali & Jiliang, 2007; Sheehan & Ginther, 2001). Moreover, evidence indicates that the density of the texts examinees have to process can affect their comprehension of the text and their performance on the test tasks (Embretson & Wetzel, 1987; Gorin, 2005; Kintsch, 1988, 1994, 1998; van Dijk & Kintsch, 1978).

Picking upon the preceding research insights relating to the effect of text structure on reading comprehension, I propose to investigate the extent to which examinees' degree of familiarity with the ESE texts and their perceptions of the text content difficulty can affect their performance on the ESE. Furthermore, I propose to examine the extent to which the ESE includes texts that encompass the different genres and the extent to which the test questions tap into the specific structural organizations of the different genres. Finally, I propose to examine the density of ESE texts and evaluate the appropriateness of these texts to the ESE context.

The last concern the study hopes to examine relates to the item features that are predicted to affect examinees' performances on the ESE. Evidence suggests that, the different properties of a test item can influence examinees' scores by enabling some examinees to get an item right or wrong simply by the way it is constructed (Haladyna, Downing & Rodriguez, 2002; Lin, Chu, & Meng, 2010; Taylor, 2005). More specifically, different item writing guidelines (Haladyna, Downing & Rodriguez, 2002) advise test writers to construct tests with good stems since it is agreed that quality stems contribute to the quality of the test questions. Furthermore, these item writing guidelines also recommend the test writers to ensure that only one of the suggested alternatives is the right answer; that there are no clues to the right answer; and that the test question key is balanced (Taylor, 2005; Haladyna, Downing & Rodriguez, 2002). Moreover, evidence suggests that the different options to a multiple-choice item are good if they are equally plausible to examinees who have not mastered the content tested (Lin, Chu, & Meng, 2010); suggesting that quality test questions are those questions constructed with functioning distractors.

In light of the preceding research insights relating to the item structure, I propose to investigate the following issues that I believe can affect the quality of ESE test questions: First, I intend to

examine the extent to which the ESE stems clearly present the central idea of the test question. Besides, I propose to examine whether the stem is presented in a form of a complete statement or an incomplete statement, and also whether it is positively or negatively worded. Secondly, I intend to investigate if the ESE includes some alternatives that contain some hints that may enable some examinees to get the item right; and if there is some evidence of the distribution of the key. Thirdly, I propose to examine the quality of ESE distractors through an investigation of their plausibility.

### **1.5 Objectives of the study**

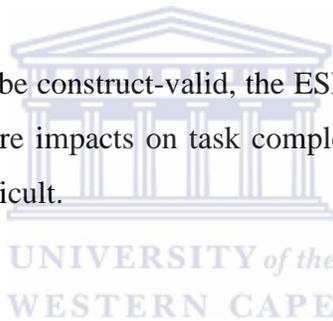
Having scoped in the previous section the four concerns this study hopes to investigate, I wish to state the following objectives this study purports to achieve:

- Describe the actual context of the ESE and determine the extent to which this context potentially influences examinees' performance on ESE tasks;
- Determine the variety and degree of complexity of different reading types and processing levels required by the ESE tasks and evaluate the extent to which ESE tasks are appropriate to the ESE context;
- Determine the textual features that potentially impact on the difficulty of ESE tasks and evaluate the extent to which the ESE texts are appropriate to the ESE context;
- Determine item features that potentially affect the validity of ESE tasks; and
- Design a proposal for the development of a more relevant and appropriate ESE.

### **1.6 Main assumptions of the study**

In light of the study aim, scope and objectives, I present the following statements as main assumptions that I believe can guide the entire process of this study:

- **Assumption 1:** The context of the ESE influences examinees' performances on ESE tasks; therefore, in order to be construct-valid, the ESE must include tasks that reflect this context;
- **Assumption 2:** In order to be construct-valid, the ESE must include tasks that require examinees to process the text carefully and expeditiously at global level and local level; and it must include tasks whose degree of complexity reflects the ESE context;
- **Assumption 3:** In order to be construct-valid, the ESE must include texts that cover different text genres; besides, the test questions must target the text information relevant to the structural organization of the specific genres, and the complexity level of these questions must be reflect the ESE context;
- **Assumption 4:** In order to be construct-valid, the ESE should *not* be constructed in such a way that the item structure impacts on task complexity by making the individual test questions either easy or difficult.



## 1.7 Research questions

Since this investigation aims to explore issues that potentially threaten the validity of the DR Congo English state examination, I wish to propose the following main question: *“What are the main issues that potentially threaten the validity of the DR Congo English state examination?”* In my attempts to answer this main question, I have addressed the following four sub-research questions as component parts of the main research question:

- What is the actual context of the ESE and to what extent this context potentially influences examinees' performances on ESE tasks?
- What is the variety and degree of complexity of the different reading types examinees conduct while completing the ESE tasks; and to what extent the ESE tasks are appropriate to the ESE context?

- What are the different textual features that affect the complexity of ESE tasks, and to what extent these textual features are appropriate to ESE actual context?
- What are the item features that affect the validity of ESE?

## **1.8 Significance of the study**

Although the study is primarily concerned about the educational and social problems in the specific context of the DR Congo, the findings of this study are also important in the general context of reading assessment research and practices. Therefore, the study is significant in the following respects:

- **In the context of the Democratic Republic of the Congo**

The study is believed to be of important social value at this time where there is a malaise in the general public opinion that is due to students' poor performances on the national test. The findings of the study are expected to help the Congolese government, academics and educators to critically interrogate the quality of tests that are actually being used in large-scale assessments.

For over forty years now the English state examination has been in use. But there have never been validation studies conducted in order to ensure its validity. Hence, the study is meant to address this educational concern. It is expected to create awareness and motivation on the part of stakeholders on the crucial need to validate tests that are being used;

The study is of practical importance to the DR Congo education officials and testing body as it is meant to serve as a point of departure from which the ESE validating agenda can be built where different stakeholders (test developers, government experts, English language teachers and inspectors, etc.) and researchers can work together in order to produce proposals for the improvement of the ESE.

- **In the general context of the assessment of reading construct**

Since the study assumes that reading construct needs to be examined within the wider context of reading activity, the insights to be generated by this study can be gainfully deployed in the

educational practice of reading assessment. Besides, they can be used by reading assessment researchers to inform their validation studies.

In light of these insights the study is expected to generate, the study opens up the possibility for further debate between academics, testing theorists, and test constructors on different validity issues pertaining to reading assessments.

## **1.9 Organization of the chapters of the study**

Given the variety and amount of data that this study is expected to use, and considering the variety and complexity of the issues that the study proposes to address, I believe it is better to depart from the traditional structural organization of dissertations by adopting an organization that can provide a clearer presentation of the research process. This proposed organization consists of the literature review in three distinct chapters (theoretical framework, conceptual framework and debate on previous studies). Furthermore, the results chapter is organised in four distinct chapters relating to the four study questions and objectives. I believe that such a different organization can be informative.

Having provided the rationale for the organization of the chapter of the study, I propose to present the eleven chapters that make up this study.

**Chapter one** serves as an introduction. It first discusses a set of educational and social concerns which act as an awareness-building exercise and a point of departure for this study. It then discusses the aim, scope and context of the study, the statement of the research problem, the objectives and main assumptions of the study, the research questions, the significance of the study, the structure of the chapters of the study and the clarification of key concepts.

**Chapter two** presents the theoretical framework of the study. In order to gain a theoretical understanding of the studies, three theories will be presented: the unified construct validity theory, the construction-integration theory, and the strategic competence theory. The discussion to be conducted on the validity theory aims to provide an understanding of the kinds of test tasks that are appropriate for inclusion in the ESE. The description of the basic assumptions of the construction-integration theory aims to evaluate the quality and relevance of text passages that

are included in the ESE. The insights from the strategic competence theory are meant to provide information necessary to evaluate different strategies used by study participants so as to decide which strategies are construct-relevant and which ones are construct-irrelevant. Since the theoretical insights presented in this chapter are meant to inform the conceptual framework that I intend to develop, the next chapter describes the designed framework so as to link the theoretical underpinning of the study to the conceptual scope of the study.

**Chapter three** describes a process-oriented framework that I have designed in order to explore the issues that threaten the validity of the ESE. The development of this framework starts with a rationale for examining the reading construct from the test takers' response processes; then I provide a description of two existing approaches that are expected to provide insights to the development of the framework, and finally I describe the framework content. The description of this process-oriented approach provides a ground to engage, in chapter 4, with relevant literature that has addressed the key concerns of validity as discussed in chapter 2; and to evaluate existing studies that have adopted different approaches to examining the reading construct as presented in chapter 3.

**Chapter four** relates to literature review where I present previous debate that is expected to provide relevant insights for this study. The debate is organized in five sections: in the first section, I review the studies that have investigated the reading construct from the examinees' strategies use perspective with a particular focus on the understanding of how examinees' strategies use impacts on their performance on reading tests. In the second section, I review studies that have investigated on the impact of test context on test performance. In the third section, I review studies that have investigated the variety and complexity of different reading types and processing levels examinees conduct while reading and the extent to which task complexity can be function of the cognitive demands required to complete test tasks. In the fourth section, I scrutinize studies that have signposted the impact of textual features on processing difficulty and therefore impact on reading performance. Finally, in the fifth section, I bring in review studies that have examined the issues relating to the construction of the multiple-choice item and the impact of this format on examinees' choice of strategies to read the text and select the item answers.

**Chapter five** presents the design and methodology of the study. In this chapter, I describe the selected design, the research methods to be used, the research site where the investigations are to be conducted, the research participants, the research population and sample, the instruments to be used to collect data, the procedures for data collection, the methods to be used to analyze data and finally the ethical considerations.

**Chapter six** presents findings related to the actual context of the English state examination. Three sections constitute this chapter: first, the description of Grade 12 students-participants' characteristics; then the description of teachers-participants' characteristics, and last the description of the actual conditions under which the teaching/learning of reading takes place.

**Chapter seven** presents findings related to the variety and degree of complexity of the different reading types examinees conduct while completing the ESE tasks; and the extent to which the ESE tasks are appropriate the ESE context. This chapter has three sections. In the first section, I present the item frequency per test for the four reading types so as to picture the frequency of items across the sampled tests as well as across the four reading types. In the second section, I present the frequency of strategies participants used for completing test tasks in order to provide an understanding of strategies participants deployed in answering test questions. In the third section, I present the item difficulty for each of the four reading types in order to capture the complexity of the different reading types. In the last section, I evaluate the extent to which the ESE tasks target the six processing levels in order to cover the reading construct.

**Chapter eight** presents findings relating to the textual features that are expected to impact on the difficulty of ESE tasks; and evaluate the extent to which the ESE includes texts whose features are appropriate to the ESE context. This chapter has three sections. In the first section, I investigate the extent to which the contents of ESE texts potentially affect examinees' capacity to comprehend the text and their performances on the test. In the second section, I investigate the extent to which the sampled texts encompass the different genres and the extent to which related test questions target text information relevant to the structural organization of each genre. In the last section, I investigate the extent to which the density of ESE texts can impact on examinees' performances on the test.

**Chapter nine** presents findings related to the different item features that are predicted to affect the validity of the English state examination. In this chapter, I first investigate the quality of the ESE stem in terms of their focus, completeness and orientation. Then, I investigate the quality of the ESE key. Next, I investigate the quality of the item alternatives; and finally, I examine the quality of the sixth option (option 6).

**Chapter ten** presents the discussion of the findings. In this chapter, I propose to organize the discussion in four sections corresponding to the four sub-research questions that underlie my main research question. In the first section, I discuss the findings that relate to the description of the actual context of the ESE and the extent to which this context can influence examinees' performances on ESE tasks. In the second section, I discuss the findings that relate to the variety and degree of complexity of the different reading types examinees conduct while completing the ESE tasks, and the extent to which the ESE tasks are appropriate the ESE context. In the third section, I discuss the findings that relate to the different textual features that can affect the complexity of ESE tasks, and the extent to which these textual features are appropriate to ESE actual context. Finally, in the last section, I discuss the findings related to the item features that can affect the validity of ESE.

**Chapter eleven** states the conclusions of the study. This chapter is organized into four sections. In the first section, I first recapitulate the principal parts of the study, explain what the study did and what ensued as results. In the second section, I relate the four sub-research questions to the outcomes discussed in Chapters 6, 7, 8 and 9 in order to answer in an informed way these questions. In the third section, I propose to elaborate on some research insights that the study is expected to generate. In the fourth section, I present a proposal for ESE development as my contribution to the improvement of the ESE. In the fifth section, I present the limitations to the study. In the sixth and last section, I present my agenda for the present study results dissemination, the research agenda on the English state examination, and the suggestions for further research.

### **1.10. Clarification of key concepts**

In this section, I propose to clarify the key concepts that are used in this study.

## 1. Test / testing

In EFL context, a language test is a device that attempts to assess how much has been learned in a foreign language course, or some parts of the course (Oller, 1979). Viewed from this perspective, a language test consists of specific tasks through which language abilities can be elicited (Davies, 1990).

Language testing, on the other hand, is a science that deals with issues pertaining to the design and use of tests as a means for determining language proficiency (Inbar-Lourie, 2008). It is a way to systematically measure a person's ability or knowledge, and it is formalized as a set of techniques or procedures (Yoshida, 2006). Viewed from the design and measurement perspective, language testing can be defined as a psychometric activity that is concerned with the development, analysis and use of language tests. This study adopts the latter conceptualization of language testing.

At this juncture, I wish to present some concepts that are related to the concept 'test'. These concepts are:

### **Tester versus testee, test constructor/developer versus examinee**

A testee is a student who takes a test while a tester is a person who tests. However, in the context of this study, I will use the concept 'test constructor/developer' to refer to the concept 'tester' and the concept 'examinee' to refer to 'testee'.

### **Test item/test question**

In the context of this study, a test item, also called test question is a specific ESE question which requires the examinee to choose one answer from a number of answers supplied. The test question consists of (a) the stem (the text of the question), (b) alternatives/options (the choices provided), (c) the key (the correct answer in the list of alternatives, and (d) the distractors (the incorrect answer in the list of alternatives). I wish to caution my readership that I will use the concepts 'test question' and 'test item' interchangeably; so will I use the concepts 'alternative', 'option' and 'choice' interchangeably.

## 2. Assessment

The concept “assessment” is sometimes used synonymously to the concept “test” although many researchers perceive the concept assessment to be an overarching term that is used to refer to all methods and approaches to testing and evaluation whether in research studies or in educational contexts (Davies, 2001). In language assessment literature, an assessment is the process of using tests in order to gather information and make interpretations of test results. Thus, on the basis of these interpretations, we make some decisions and eventually take action (Chapelle, 2005, Norris, 2009). Viewed from this process perspective, language assessment appears to be the entire procedure of test use; therefore, it helps us to be informed on the decisions we make and the actions we take in language education. In the context of this study, I will use the concept ‘assessment’ to refer to the process of gathering information; while the concept ‘test’ will be used to relate to the instrument used to gather information.

### **3. DR Congo state examination**

This is a national exit test that is designed to assess grade 12 high school students’ achievement. Successful students get a certificate that enables them either to go to university or to start a professional life.

### **4. DR Congo English state examination**

This is a subtest of the state examination. This subtest is designed to assess grade 12 high school students’ achievement in English. The test is a text-based test of multiple choice type.

### **5. Test validity**

The concept of validity has evolved over time. In the context of this study, I have adopted Messick’s conceptualization of validity that suggests that, validity is

an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment (Messick, 1989, p. 41).

In light of this conceptualization, I propose to define a valid test as one that includes tasks that target the reading constructs, and that are appropriate (proportional) to the reading context.

### **6. Test validation**

Test validation is an on-going process of gathering, summarizing, and evaluating relevant evidence concerning the degree to which that evidence supports the intended interpretations and inferences of test scores (Cizek, 2012; AERA, APA, & NCME, 2011). In the scope of this study, I propose a conceptualization of validation that implies a more objective attitude towards the evidence. Therefore, the concept validation relates to the evaluation of the extent to which the proposed scores interpretations are meaningful and appropriate.

## **7. Reading construct**

Reading is a macro-construct that can be best understood through an identification of the constructs that make it. There is a substantial degree of disagreement on what constitutes the constructs of reading (Alderson, 2000; Milanovic, 2011). Consequently, it has become hard to determine the constructs of reading and design tasks that can operationalize these constructs.

In order to overcome this issue, research has suggested that constructs of reading should be determined in light of the theory of reading and models which have been developed (or adopted) to assess various aspects of examinees' reading abilities (Milanovic, 2001). Since I propose to adopt in this study a process-oriented approach that conceptualizes reading activity on a four-cell matrix where students must process the text carefully and expeditiously at global level and local level, the constructs of reading will therefore consist of (a) reading skills that are required to read the text carefully at global level; (b) reading skills that are needed to read the text carefully at local level; (c) reading skills that are necessary to read the text expeditiously at global level; (d) reading skills that are necessary to process the text expeditiously at local level. However, since reading at global level and reading at local level are also constructs, I propose to operationalize them through the different traditional hierarchical processing levels that readers engage in while reading texts. These levels are: lexical access, syntactic parsing, establishing propositional meaning, inferencing, building a mental model and creating a text-level structure (Field, 2004; Wagenmakers *et al.* 2000; van Gompel, 2007; Kintsch & van Dijk, 1978).

## **8. Reading context**

In the context of English as a foreign language, reading activity generally occurs in the context of classroom instruction. Therefore, when one thinks about the context of reading instruction, one thinks mostly of the specific classroom contexts that influence students' comprehension of

texts (Brown, 1997). This is to suggest that, classrooms and schools themselves can reflect the socio-economic context of the larger society where learning takes place. In the scope of this study, the reading context is meant to be reflected in students' individual characteristics (their language/reading skills, their motivation for reading, their attitudes towards reading instruction), the availability and quality of the reading materials, the availability and quality of the curriculum, the characteristics of the English language teachers (their qualification, experience and motivation for teaching).

### **Summary of the chapter**

In this chapter, I have provided the introduction to the study. I have first discussed a set of educational and social concerns which act as an awareness-building exercise and a point of departure for this study. Then, I have discussed the aim of the study and its scope. Moreover, I have provided the context of the study with a particular focus on the sociolinguistic situation of the DR Congo and the description of the ESE objectives. Furthermore, I have provided the statement of the research problem, the objectives and main assumptions of the study, the research questions, and the significance of the study. Finally, I have presented the structure of the chapters of the study and the clarification of key concepts.

## CHAPTER TWO

### THEORETICAL FRAMEWORK

All research is founded on a question or premise. The theoretical framework of a particular study acts as the supporting mechanism that is developed from an existing theory or a combination of theories in order to understand the issues underlying the phenomenon to be investigated. This is very important because research, in essence, is a process of generating knowledge in order to find answers to problems or questions. The theoretical framework helps the researcher to identify and determine problem areas, research questions that need to be addressed, and the relevant design that can be used in conducting the study (McMillan & Schumacher, 2006). Implicit in the concept of theoretical framework is the concept of theory. Theory enables the researcher to connect a single study to the immersed base of knowledge to which other researchers have contributed, to determine gaps in scientific knowledge, and to provide suggestions for further research (McMillan & Schumacher, 2006).

In order to explore the multiple variables that potentially threaten the validity of a multiple-choice reading test like the DR Congo ESE, there is particular need to understand two constructs: the validity construct and the reading construct. Since validity is a concept, an epistemology and a theory, a sound theoretical framework must be able to describe the different developments and trends of the validity theory and show how these developments and trends can be rationally integrated in language testing enterprise. Viewed from this perspective, validity theory informs this study on the quality of tests that are needed in assessing students' reading abilities. Furthermore, in order to assess the extent to which the ESE tasks target the reading construct, there is a need to understand what is meant by reading activity, how students read the texts in both testing and non-testing situations; and how they deploy strategies to construct text meaning and complete test tasks. In light of this claim, the theoretical framework of this study also

examines two interrelated theories: the construction-integration theory and the strategic competence theory. The former theory provides an understanding of how readers process the text in order to construct its meaning, while the latter theory enables to gain insights into how readers use various strategies while reading the text and completing the test tasks.

In keeping with the above stated position, I articulate the theoretical framework of the present study around three theories: the unified construct validity theory, the construction-integration theory, and the strategic competence theory. In the following section, I describe the first of these three theories, the validity theory.

## **2.1 Unified construct validity theory**

Since the present study aims to explore various issues that threaten the validity of the DR Congo English state examination, the unified construct validity theory can serve as the core theory of this study. I propose to start the discussion of this theory by highlighting its background before I describe it, point its limitations and finally show its relevance to the present study.



### **2.1.1 Background**

From the outset, I wish to mention that validity is a measurement concept that has undergone many changes of meaning over time. Measurement experts and theorists have described the historical changes in the definition of validity (Cronbach, 1988, 1989; Goodwin, 1997, 2002; Kane, 2002, 2004, 2011; Messick, 1989).

An early definition of validity (in the 1940s) emphasized the test itself. This is to suggest that validity was considered as a static property of a measure (Goodwin & Leech, 2003). At that time, a test was considered valid if it measured what it was supposed to measure. Such a conceptualization of validity usually reflected the extent to which scores produced by a test correlated with some other external criterion measures (Goodwin & Leech, 2003).

Indeed, with the publication in 1966 of the first edition of *Standards for Educational and Psychological Testing* (Goodwin & Leech, 2003), the meaning of validity shifted its focus to test

use. At this time, validity was defined as the extent to which a test could produce information that was useful for a specific purpose. This edition of *Standards* also subdivided validity into three types: criterion-referenced validity, content validity and construct validity (Brown, 2000; Brualdi, 1999).

Criterion-referenced validity usually included any validity methods that focused on the correlation of the test being validated with some outside measures of the same objectives or specifications. Content validity was defined as the degree to which a test accurately sampled from the course of study or domain. This was termed the “real life” approach to validity in which a valid test was the one that included different tasks that were “a representative sample of the tasks from a well-defined target domain” (Bachman, 1990, p. 310). Finally, construct validity was traditionally defined as the experimental demonstration that a test is measuring the construct it claims to measure (Brown, 2000). In an attempt to clarify the initial conceptualization of construct validity, Moss (2007) argues that construct validity aimed to infer the degree to which an individual examinee possesses some hypothetical traits or quality (construct) that cannot be observed directly by determining the degree to which certain explanatory concepts or constructs account for performance on the test. Viewed from this perspective, construct validity could be investigated through studies that check on the theory underlying the test (Moss, 2007). I wish to mention at this juncture that this traditional conceptualization of construct validity is a post hoc validation process that attempts to establish test validity through empirical procedures after the test has been designed and administered to examinees. In light of this view, I am tempted to state that such a conceptualization of construct validity appears to fit well within the prevailing guiding scientific paradigm (functionalist paradigm and positivistic theory) which guided measurement research in the 1950s.

The revolution in the conceptualization of validity emerged during the 1980s and 1990s with the publication of Cronbach’s (1980) “Validity on Parole: How can we go straight?” and culminating with in Messick’s (1989) “Validity” in *Educational Measurement*. At that time, the concept validity was associated with the inferences and decisions that are made on the basis of test scores. Validity was therefore defined as “the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores”, and validation was defined as “the process of accumulating evidence to support such inferences (AERA, APA, & NCME, 1985, p.

9). I wish to mention that during this period, two important developments emerged: first, validity was conceptualized as a unitary concept. This is to suggest that the traditional view of validity theory that considered validity as three-faced (criterion-related, content, and construct) was challenged on the ground that it was fragmented (or discrete) and incomplete. Therefore, researchers started to see these three traditional discrete types of validity as three different facets of a single unified form of construct validity. Secondly, researchers started to direct much attention to the need for evidence about the social consequences of test use (Cronbach, 1988; Messick, 1989, 1994). In the present study, I also conceptualize validity as a unified construct. Therefore, the use of the concept ‘validity’ in this study points to ‘construct validity’.

Having provided a brief background of unified validity theory, I now wish to direct attention to the content of this theory by stressing its key assumptions.

### **2.1.2 Unified construct validity theory: Content**

In its contemporary conceptualization (see AERA, APA, ACNME, 2011; Kane, 2004, 2006, 2011; for instance), validity is a unitary concept that looks at multiple sources of evidence as described by Messick (1989) and embodied in the current *Standards of Educational and Psychological Measurement* (2011). This is to suggest that the unified construct validity theory is a general approach to validity that includes content and criterion-based evidence, reliability as well as a wide range of methods associated with theory testing (Messick, 1989).

Messick (1998, p. 41) argues that validity refers not only to the accuracy of score inferences, but also to the evaluation of the “appropriateness, meaningfulness, and usefulness of scores inferences”. He defines validity as

...an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment (p. 41).

From this definition, I wish to address the following two considerations that I believe are relevant to this study. First, validity is not a property of a test; rather, it refers to the use of a test for a particular purpose. In light of this statement, a test is not valid (or invalid) in itself; rather,

what is valid (or invalid) are the decisions that are taken on the basis of test scores. Second, in order to validate a particular test, one needs to gather multiple sources of evidence to justify the claims he/she makes. This is to suggest that enough evidence is needed in order to support or refute the meanings and interpretations assigned to test scores as well as the decisions taken (or to be taken) on the basis of test scores. In keeping with these two considerations, I agree with AERA, APA, NCME (1999, p. 9) when they define validity in practical terms as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests”.

Since we can now agree that a test is neither valid nor invalid in itself, and that what researchers and test developers have to do is to validate the use of test scores, I hasten to suggest that one important aspect in test validation is to investigate how test results are used and the way those results impact on the individuals as well as on the whole society. Given this, we need to think of how some individuals have been affected by decisions made on the basis of test scores. To take just an example, if a particular student cannot be admitted at university because of his/her score on high school matriculation examination, we need to ascertain that there is enough evidence to make such a decision on this student. Therefore, we need to validate the test that is used in order to ensure that the decision taken is justified. This process is called validation. It may refer to a formal test or other data gathering instruments. In both cases, validation process can include the description of examinee’s characteristics and the conditions of his/her preparation to test, the description of test administration procedures, the understanding of the quality of test tasks, the evaluation of the testing method, and so on.

Since the unified construct validity theory views all the three traditional types of validity (criterion-referenced validity, content validity, and construct validity) as three different facets of a single unified form of construct validity, one must expect to deal with some of the aspects the traditional validity while engaging in any validity study today. Shepard (1993) enumerates some of these aspects that I propose to summarize here. First, validity studies usually address both the internal structure of the test and the external relations of the test to other variables, and this is part of the “old orthodoxy”. The internal portion of construct validation may include gathering data on all of the traditional psychometric questions regarding item characteristics. It may also include questions about the appropriate weighing of different skills being tested. This is also part

of the “old orthodoxy”. Second, empirical evidence may include both correlational methodologies and experimental and quasi-experimental procedures, two methods that dominated the traditional conceptualization of validity. Last, but not least, convergent and discriminant correlations are still popular today as a tool used to address typical construct validity questions.

Messick (1998, p. 37) provides an explanation on the inclusion of traditional aspects of validity in the modern conceptualization of validity. He argues that the unified validity theory implies a single validity for test interpretation and use. Therefore, the three traditional types of validity are simply “complementary forms of evidence to be integrated into an overall judgment of construct validity”. Messick (1998, p. 37) therefore maintains that “what needs to be valid are the inferences made about score meaning, namely the score interpretation and its action implications for test use”. This new conceptualization of validity as superordinate concept to all other traditional validity types is considered as a major development in testing circles around the world today.

Since this research aims to explore factors that threaten the validity of the DR Congo English state examination, I wish at this juncture, to briefly describe two main threats to construct validity that have been reported in the literature. These threats include construct-irrelevant variance and construct underrepresentation variance (Young, 2008; Messick, 1989). I believe that these two main threats encompass the wide range of factors that will be identified and described in this study as those factors that appear to threaten the validity of the ESE.

**Construct-irrelevant variance** relates to factors (or variables) that influence students’ test scores but that are not directly related to the test construct (Bakker, *et al.*, 2008; Young, 2008; Messick, 1989; Haladyna & Downing, 2004). Such variables are expected to introduce irrelevant variance in the test scores. In view of this conceptualization, construct-irrelevant variance is undesirable because the scores obtained by an examinee or a group of examinees may be partly due to factors that are different from the construct the test is designed to measure. In order to better grasp the real threat of this concept on test validity, I would like to provide an example of a mathematics admission test administered to a candidate who has applied to study mathematics in a science faculty. If the test designer considers in the evaluation of this candidate the language problems the candidate has in providing answers to the test, and if part of the score is influenced

by the candidate's poor language knowledge in test writing, such an admission test can be said to be construct-irrelevant. This is because the candidate's reported score reflects not only his/her knowledge of mathematics, but also his/her knowledge of language, which in principle, should not be one of the criteria of evaluation.

The previous example implies that the notion of construct-irrelevant variance is a two-way notion that can be understood when one considers its two kinds: construct-irrelevant easiness and construct-irrelevant difficulty (Young, 2008; Messick, 1989). The former occurs when "extraneous clues in item or test formats permit some individuals to respond correctly in ways irrelevant to the construct being assessed", whereas the latter occurs when "aspects of the task that are extraneous to the focal construct make the test irrelevantly more difficult for some individuals or groups" (Messick, 1989, p.34). Briefly stated, construct-irrelevant difficulty leads to lower scores for some examinees while construct-irrelevant easiness leads to higher scores for some examinees.

By addressing the issue of construct-irrelevant variance in this study, I propose to determine the extent to which some examinees likely take advantage of the multiple-choice format and utilize test wiseness strategies (such as elimination of distractors, guessing, etc.) to get some test questions right. Furthermore, I propose to determine whether the examinees' difficulty in responding to MC reading test questions may also come from their difficulty to understand the test questions because of the way they are structured.

Furthermore, by addressing the issue of construct-irrelevant variance in the present study, I intend to highlight the crucial need for ESE test developers to design tests that yield scores that reflect the examinees' actual reading abilities. I hypothesize that, in order to construct-valid (or construct-relevant), the ESE must include tasks that require examinees to deploy strategies that reflect the fact that they have been actually actively working to understand the text, to understand the expectations of the test questions, to understand the meanings and implications of the different item options in light of the text, and to select and discard different options based on the way they understand the text.

On the other hand, **construct underrepresentation** variance occurs when the test is too narrow and fails to include some critical aspects of the construct (Messick, 1989; Young, 2008). To

avoid this validity threat, the test developer should ensure that the test is a representative sample of tasks that are likely to cover the entire course content.

By addressing the issue of construct underrepresentation in the present study, I propose to determine the extent to which the ESE test questions target all the critical aspects of reading construct. Since I propose to conceptualize reading construct on a four-matrix cell with careful reading and expeditious reading that can be conducted at global level and local level, I intend to evaluate the extent to which the ESE test questions require examinees to read carefully and expeditiously at global level and local level.

Validity theory is difficult to understand. By the same token, the concept of validity is no less difficult either. One best way to understand the validity theory is to understand what is meant by test validation, one of the main issues in validity enterprise. In the following section, I provide some issues and insights on test validation and its contribution to validity theory.

#### **2.1.2.1 Test validation**

The unified construct validity theory stresses the importance of test validation. In the preceding section, I have stated that test scores can provide specific information about examinees and this information can be used to make important decisions. Some of these decisions include, for example, the university admission, course placement, or job recruitment and promotion. Therefore, it is imperative that test users need to justify the appropriateness of the tests that are being used. In light of this statement, it is worth to mention that, if validity is about the different interpretations assigned to test scores, then validation is a process of evaluating, through a network of theory, hypotheses and logic, the appropriateness of these interpretations (Kane, 2004, 2011). More specifically, validation is an ongoing activity researchers undertake to find out whether a test has the property of validity (Borsboom, Mellenbergh & Van Heerden, 2004). Viewed from this perspective, it appears reasonable to state that, in validating a test, various sources of evidence must be gathered, synthesized, and summarized in order to achieve meaningful interpretations. This view has one relevant implication: the triangulation of data, instruments and methods while conducting a validation study remains as an appropriate route for

investigating the extent to which inferences and decisions made on the basis of test scores are appropriate.

At this juncture, I wish to mention that there are two conceptualizations of the concept validation, as articulated by Kane (2011). This distinction is worth mentioning in the scope of the present study as it helps to clarify which type of validation I as a researcher purport to undertake. First, the concept of validation can relate to the process of developing evidence to support the proposed interpretations. Kane argues that this conceptualization of validation implies an advocacy role for raising evidence; and this role is usually performed by test developers who want to justify the validity of their tests. However, by adopting such a stance on validation, test developers tend to demonstrate that the tests they have developed are valid and that inferences and decisions that can be made on the basis of test scores are appropriate. I believe that a test validated by someone who has developed it is likely to generate findings that embody some bias due to the test developer's effort to justify the test use. This study is not concerned with this type of validation.

Secondly, validation can relate to the evaluation of the extent to which the proposed interpretations are meaningful and appropriate (Kane, 2011). This usage implies a more objective attitude towards the evidence; and this role is usually performed by the researcher who wants to check the meaningfulness and appropriateness of proposed interpretations. In view of my quality of a researcher and as I come from a perspective of non-involvement in the ESE construction, the present study is concerned with this second usage of validation. It aims to identify and describe factors that potentially threaten the validity of the ESE and evaluate the extent to which the inferences and decisions made on the basis of the ESE test scores are relevant to the ESE actual test context.

However, although test validation heavily relies on the accumulation of evidence to support the proposed score meanings and interpretations, validation activity may be difficult to undertake if one does not understand what is meant by validity argument. In the following subsection, I scrutinize the notion of validity argument and its relevance to the validation process.

### 2.1.2.2 Validity argument

The unified construct validity theory views validity as an argument. This is to suggest that validity is never assumed; rather, it must be established through an argument that relates theory, predicted relationships and empirical evidence to suggested score interpretations and meanings in such a way that this relation is meaningful (Kane, 2002, 2004, 2006, and 2011; Downing, 2003). Viewed from this perspective, the validity argument provides a basis for an overall evaluation of the meaningfulness of the proposed interpretations and uses of test scores. Therefore, the aim to develop a validity argument is to come up with a “cogent presentation of all of the evidence to proposed interpretations” (Kane, 2002; p. 31).

A key notion in the validity argument is the concept of “score interpretations” as articulated by Kane (2002, 2004, 2006, and 2011) in his “interpretive argument”. Kane’s interpretive argument appears to be largely informed by Toulmin Model of Argumentation that proposes a layout containing six interrelated components for analysing arguments that are claims, ground, warrant, backing, rebuttal, and qualifier (Toulmin, 1958). According to Kane (2002, p.31), a proposed interpretation may be defined in some details by specifying it as “an interpretive argument, that is to say, a network of inferences and supporting assumptions leading from scores to conclusions and decisions”.

At this juncture, I wish to state that the interpretive argument is pivotal to any validation studies. This type of argument provides not only an explicit statement of a test proposed interpretation and/or use, but also a framework for developing a validity argument. As an example, by developing a text-based test like the ESE, test constructors expect test scores to provide some indications of student’s performance on non-test tasks (for instance, classroom tasks) involving the same kind of content, performance on other tasks, or future performance on the same or related tasks.

The interpretive argument involves four steps, as suggested by Kane (2002, 2011). The first step involves an evaluation of the examinee’s performance on each test question, resulting in the assignment of a score to the examinee’s response to that question. Then the examinee’s scores on all of the test questions are combined into a single observed score. Such an examinee’s single observed score supports a very simple kind of interpretation. It claims relatively little and

requires relatively little evidence for its support. At this stage, one would, for example, interpret an examinee's score on a reading test as his/her ability to comprehend the text and successfully answer the questions based on it.

The second step involves a generalization from statements about performance actually observed to conclusions about expected performance over a universe of possible performance on similar reading tasks performed under similar conditions (Kane, 2002, 2011). Such a generalization calls for a broader and sound interpretation. Considering the previous example, we have to show that the examinee's scores are generalizable over the tasks assigned to him/her. This is to suggest that, the examinee can perform equally well on similar tasks if the test is performed under similar conditions. Such a generalization is generally achieved through appropriate reliability and generalizability studies.

The third step involves extrapolation, and this would extend these conclusions to expected performance in reading tasks in general and performed under any condition. Such an extrapolation requires new kind of evidence for support, and this can be achieved through criterion-related studies or analyses of commonalities between test performance and performance in the wider domain (Kane, 2002; 2011). At this stage of extrapolation, other inferences can be added; for instance, the inferences that may lead to the theoretical constructs embedded in process theories and predictions of future performance. If we take into consideration the previous example, an examinee's score on a reading test may provide an indication that he/she can successfully/unsuccessfully undertake his/her studies at university by reading successfully/unsuccessfully subject-related or scientific materials. Such an inference requires evidence in support, and this can be achieved through predictive studies. Another plausible interpretation would be that the examinee's score can indicate that the goals of the decision program have/have not been attained.

The last step involves the decision that is taken on the basis of the student's score to determine his/her future. The test scores can be used to make a number of decisions about end-of-course grades, graduation from high school, university admission, or job recruitment. At this juncture, I wish to highlight one consideration I believe is relevant to the scope of this study. It is very unlikely that, in educational context, any single test score would serve a basis to take the previous decisions. This is to suggest that the decisions that are taken generally take account of

the student's combined scores from all subject discipline tests. In this perspective, we also expect these decisions to be appropriate so as to reflect the individual students' actual abilities. This requires on the part of test developers and users skills and knowledge. This is really a challenge. To cite an example of South African universities: many students have to pass through a foundation phase to start mainstream studies in a university. But it has generally been found out that some of these students who have been denied direct access to university because of their low scores on matrix examination were capable of performing better than those who scored better on the matrix test.

The discussion conducted on the notion of validity argument is relevant to the scope of this study. The insights from this discussion help to clarify the notion of score interpretations in the context where the ESE has been continuously used for more than four decades without any validation studies. Therefore, one can be tempted to question the validity of such a test, as well as the relevance of decisions taken on the basis of students' scores.

The validity theory as articulated by Messick (1989) and expanded by Kane (2002, 2004, 2011) is viewed as a general theory that is applicable to any research area. In the following section, I propose to describe this theory in relation to my research area of language testing.

### **2.1.3 Unified construct validity theory in language testing**

In the preceding section, I have described validity as a general measurement theory. However, the scope of this study requires a specific focus on the concept of validity. That is the reason why in this section, I describe the unified construct validity theory as it has been adopted by language education researchers.

In language education, some researchers have provided a historical account of the development of construct validity theory (Chapelle, 1999; Alderson & Banerjee, 2002; Shohamy, 1990). In the present section, I therefore propose to summarize this account.

The first language researchers to build on Messick's unified validity theory were Bachman and Palmer (1996) when they articulated a theory of test usefulness, which they considered to be the most important criterion by which language tests should be evaluated. Bachman and Palmer

incorporated the modern view of construct validity and they added some dimensions that affect test development in the real world. In order to understand the concept of validity, they provided the following qualities that a good language test must have: valid, reliable, positive impact, interactive, authentic and practical. However, as Alderson and Banerjee (2002) argue, what is less clear in the Bachman and Palmer account of test usefulness is how these various qualities should be measured and weighed in relation to each other. Shohamy (1990) alludes to the same consideration when she argues that a test validation agenda should be defined in terms of utility, feasibility, and fairness. For her, test utility relates to the extent to which a test serves the practical information needed of a given audience. Test feasibility relates to the ease of administration in different contexts; whereas test fairness relates to whether or not the tests used are based on the materials which the test takers are expected to know.

The unified construct validity theory as applied to language testing implies an emphasis on the centrality of construct. This is to suggest that test developers have to consider what is known about language knowledge and ability as well as the ability to use the language. In light of this suggestion, language testing must involve not only the psychometric and technical skills required to construct and analyze a test, but it must also involve the knowledge about language (Alderson & Banerjee, 2002). In keeping with this argument, I hasten to state that language test developers need to be applied linguists who are aware of the latest and most accepted theories of language description, language acquisition as well as language use. They also need to know how all these theories can be operationalized; that is, how these theories can be turned into ways of eliciting student's language use. This statement appears to concur with Alderson and Banerjee (2002)'s argument when they argue that language testing should not be confined to the knowledge of how to write test items that are capable of discriminating between the 'strong' candidates and the 'weak' candidates. Instead, it requires an understanding of what language is, and what it takes to learn and use the language. It is only by doing so that test developers might be able to establish appropriate ways of assessing students' language abilities.

Although Bachman and Palmer's (1996) articulation of test validity is sometimes unclear, it is important to acknowledge that Bachman (1990) was the first researcher to clarify the relationship between testing and language theories when he proposed an interactional model of test performance. This model was mainly based on a theory that considers language ability to be

multi-componential; and it included language ability and test method as two major components of validation. This model came to be known as the Bachman model (Bachman, 1990; Bachman & Palmer, 1996). This model has been extensively used for developing and validating language tests. Some researchers still use it today. It is a development of testing theories in conjunction with applied linguistic thinking by Hymes (1972) and Canale and Swain (1980). Although I do not use this model in this study because of its weaknesses and inappropriateness, it still provides useful insights to the present study on the ground that Bachman's model describes in a congruent way the different cognitive and metacognitive strategies examinees deploy to complete test tasks. Also, the model provides insights for understanding to what extent examinees' use of strategies associated with test taking should be considered as part of the test construct.

At this juncture, I wish to highlight some weaknesses of Bachman's model of language ability. This is in order to justify why I have not adopted this touted model. To start with, I consider Bachman's model to be more theoretical than practical. Chalhoub-Deville (1997) corroborates this view when she finds a lack of congruence between theoretical models and operational assessment frameworks, which necessarily define a construct in specific contexts. Chalhoub-Deville argues that, although the theoretical models are useful, there is an argument that is needed for an empirically based, contextualized approach to the development of assessment frameworks. Another criticism of Bachman model is elaborated by McNamara (1995), McNamara and Lumley (1997), McNamara and Roever (2006) and O'Sullivan (2000) when they argue that the model ignores the social dimension of language proficiency because it is based on psychological rather than socio-psychological or social theories of language use. McNamara and colleagues urge language testers to acknowledge the social nature of language performance and to examine much more carefully its interactional aspects.

The Messickian unified notion of construct validity has been well accepted in language testing circles. However, one important development in the validation research in language testing relates to the use of validation methods. Research in language testing indicates that initially, the validation studies mainly relied on the application of Item Response Theory (IRT) to language test validation (see McNamara, 1990, 1991; Hudson, 1991; for example), the use of multiple-faceted Rasch measurement models (see McNamara & Lumley, 1997; for example), the structural equation modeling approach (see Kunnan, 1998; Purpura, 1998; for example), the

factor analysis methodologies (Guthrie & Kirsch, 1987; Rosenshine, 1980; Schedl *et al.*, 1996; Young, 2008; for example), or again the linear progression methodologies (see Freedle & Kostin, 1993, 1999; for example). All these methods are mainly quantitative. However, nowadays, qualitative research techniques like introspection and retrospection are frequently used in validation research (see Cohen & Upton, 2006, 2007; Rupp, Ferne & Choi, 2006; for example). Furthermore, in order to understand what constructs are measured and how these constructs are measured, there is increasingly triangulation of research methods across the qualitative versus quantitative divide (see Phakiti, 2003; Rupp, Ferne & Choi, 2006; for example).

To conclude this section related to the content of the unified construct validity, I wish to state that this theory is not without critics. In the following section, I propose to summarize the basic criticisms made on the unified construct validity theory.

#### **2.1.4 Criticism on the unified construct validity theory**

The unified validity theory, as any scientific theory, has been challenged by researchers (Brennan, 1998; Fremer, 2000, Borsboom, Mellenbergh, & van Herdeen, 2004; Lissitz & Samuelsen, 2007, Embretson, 2011; to cite only some). Much of the dissatisfaction with Messick's conceptualization of validity is based on the view that the theory is impractical in an educational context, it does not clearly articulate, and its basic concepts are sometimes wrongly used.

Brennan (1998), Fremer (2000) and Borsboom, Mellenbergh, and van Herdeen (2004) center their criticism on the applicability of the unified construct validity theory. For Brennan (1998, p. 7), Messick's theory is difficult to be applied in education setting since it does not provide sufficient details on the basis of which validation evidence can be made. He says that:

In my experience those who are actually responsible for validation almost always require detailed and concrete guidance for conducting validation activities, and then 'unitary' notion is simply not helpful for them (p.7).

Fremer (2000, p. 1) echoes Brennan's view when he says: "We have elevated the concept of construct validation to so high a level that it seems an 'out of reach' goal"; and Borsboom, Mellenbergh, and van Herdeen (2004, p. 1061) add, "The concept that validity theorists are

concerned with seems strangely divorced from the concept that working researchers have in mind when posing the question of validity”.

However, the most serious criticism to the unified construct validity theory comes from Lissitz and Samuelsen (2007). I would like to comment in detail on this criticism because of the relevance of their argument and the reactions their argument has brought about in testing research circle.

At the outset, I would like to mention that Lissitz and Samuelsen are concerned about the practicality of the unitary approach to validity as articulated by Messick (1989). Their objective is to suggest a much narrower and more operational definition of validity so as to make the validation process more manageable. Lissitz and Samuelsen start their argument by pointing to some research that have questioned the unitary view of validity and that have suggested some proposals to shift from the current view of validity. They cite Kane (2004, p. 136) who recognizes that it is difficult to apply validity theory to testing programs, and who suggests an argument-based approach to validity as a way to overcome the ineffectiveness of testing bodies that are “exacerbated by the proliferation of many kinds of validity evidence and by the lack of criteria for prioritizing different kinds of evidence”. Lissitz and Samuelsen also cite Borsboom, Mellenbergh, and van Herdeen (2004, p. 1061) who contend that the current validity theory “fails to serve either the theoretically oriented psychologist or the practically inclined tester” and who advance a much more simplified version of the current theory.

At this juncture, I wish to summarize Lissitz and Samuelsen (2007)’s criticism in three main points: Firstly, they suggest a “deconstruction” of validity, that is, a change in terminology and emphasis regarding validity theory. Their basic question is: “Must a unitary conceptualization of validity be centered on construct validity?” Lissitz and Samuelsen argue against the idea that all validity is construct validity, as stated by Messick (1989, 1995, and 1996). They argue that validity is an issue of the content of a test, not the external definition of constructs. Therefore, according to Lissitz and Samuelsen, the best way to establish the validity of a test is by using appropriate operational definitions and item development procedures. My understanding of this argument suggests the following observations that I wish to highlight: (1) validity is not influenced by the way in which test scores are used; (2) validation studies must stress on test content rather than test construct; (3) the notion of validity is seen as whether a test measures

what it intends to measure; and this can be established by examining the internal properties of the test; (4) since content is internal to validity, therefore it is more relevant than construct; the latter can be established within a larger network in terms of external variables. Briefly stated, Lissitz and Samuelsen (2007) suggest that content validity is essentially the whole of validity; at least for educational tests.

Secondly, Lissitz and Samuelsen (2007) argue that the unified construct validity theory fails to provide adequate guidance for test validation. This argument raises the issue of practicality of many validation frameworks developed. In order to support this argument, they provide a “systematic view of the technical evaluation of test” (Lissitz & Samuelsen, 2007, p.437) in which test evaluation is separated into internal and external aspects. They argue that external aspects (e.g., relationships of test scores to some criterion variables) should be considered irrelevant to validity per se, and these can only be used in a larger framework for the utility and consequences of testing programs. In order to better provide practical guidance, Lissitz and Samuelsen argue that validity should be established through operational definitions and test development procedures. They support this argument by mentioning that, what educators and researchers want are practical tools for validation and not a theoretical framework to conceptualize validity. Lissitz and Samuelsen therefore suggest the procedures for investigating sources of validity evidence. Those procedures give weight to results from experts’ domain and task analyses and the completeness of test specifications regarding content representativeness.

Lastly, Lissitz and Samuelsen (2007) argue that validity is about test scores, not test interpretations. More concretely, they claim that validity is a property of the test, and it is independent of any proposed interpretation or use of the test scores. They suggest that any questions concerning validity must be answered by showing the extent to which the test provides a representative sample of tasks from some content domain.

Having provided the content of Lissitz and Samuelson’s criticism on the unified construct validity theory, I propose to present, in the following section, the responses to these criticisms. I believe such responses can add some clarify to our understanding of the validity theory.

### 2.1.5 Responses to Lissitz and Samuelson (2007)'s criticism

Lissitz and Samuelson's criticism on the unified construct validity theory provoked a wave of reactions by Messickian followers. Kane (2008), Sireci (2007), Gorin (2007) and Moss (2007) responded to Lissitz and Samuelson's criticism. They maintained that the unitary conceptualization of validity as stated by Messick (1989) pulled together all the strands traditionally included under the heading of validity and this conceptualization added (or at least emphasized) some new strands, particularly the role of consequence in validation (Kane, 2008). They also maintained that construct validity includes content relevance and content representativeness as well as criterion-relatedness, because such information about the content domain of reference and about the specific criterion behavior predicted by the test scores clearly contributes to score interpretations. Thus, when considered in a larger context of score meanings, correlations between test scores and criterion measures contribute to construct validity (Gorin, 2007).

Regarding Lissitz and Samuelson's criticism according to which validity is about internal content and not externally derived construct, Gorin (2007) argues that if we can define constructs with terms internal to the measurement system, then the difference between content and construct becomes less clear. Gorin (2007, p. 457) considers that the use of content validity as the whole validity theory "threatens to stifle many of the recent advances in test design resulting from construct-centric models of validity". Although Sireci (2007) agrees that much validity evidence can be gathered by evaluating test content, he strongly believes that validation cannot proceed without being referenced to the purpose for which test scores are being used. As a resolution of this criticism, Kane (2007, p. 278) simply interprets Lissitz and Samuelson's proposal as an effort to "relegitimize the most popular of the shortcuts in which an achievement test would be validated using a simple content-based model" as he asserts:

I do not think that Lissitz and Samuelson's proposals would improve the situation if they were implemented. They would either reduce validity to a very narrow concern about representativeness of test's content (...) or, more likely, take us back to the situation in the 1970s and 1980s, when we had a profusion of specialized validation methods (shortcuts), each designed for a specific kind of application" (p. 278).

As regards Lissitz and Samuelson's criticism of validity as established through operational definitions, Gorin (2007) argues that such an argument is narrow as it gives weight to results from experts' domain and task analyses and to completeness of test specifications as necessary and sufficient evidence of validity. Gorin maintains that since Lissitz and Samuelson place less emphasis on substantive examination of test taker's processing and cognition, the justification for their procedure is to restrict validity evidence to internal test properties. Therefore, Gorin is of the view that an expert review of the items' content may result in scores that reflect only the intended skills; and that correlations with external variables may be ignored. Yet, correlation provides a useful source of additional evidence to consider in the validity argument. Gorin asserts:

For the most part, the methods associated with test development based on operational definitions, like those proposed under the content classification of Lissitz and Samuelson, have led to empirically unsubstantiated claims regarding score meaning (p. 457).

Finally, in regard to Lissitz and Samuelson's criticism according to which validity is about scores, not interpretations, Sireci (2007), Gorin (2007) and Kane (2008) argue that it is essentially important to establish not only that a test is valid, but also that the test is measuring what we think it is measuring. Therefore, the aim of validating a test is to support its use for a specific purpose. Yet, this involves providing sufficient evidence to defend such a use. Gorin (2007, p. 460) argues that "the role of inference and interpretation has added complexity to validity examination". Sireci (2007, p. 478) considers Lissitz and Samuelson's proposal as a way to "evaluate aspects of a test independent of its application and to separate validation of the construct from validation of the test itself", while Gorin (2007, p. 460) echoes Sireci (2007) when she mentions that: "Validity evidence, if considered only in terms of the test itself, is meaningless". I wish to remind that this argument corroborates Messick's (1989) initial argument that states that validity is not a property of a test; rather, it refers to the use of a test for a particular purpose.

As my response to this dialogue between those who strongly support Messick's unified construct validity theory and those who amend it or suggest an alternative proposal, I wish to present the following observations that I consider relevant to this study. Firstly, the unified construct validity theory as articulated by Messick (1989) and expanded by Kane (2002, 2004, 2006, and 2011)

hardly offers practical guidance for its application; and this usually makes the theory not only difficult to understand, but also difficult to apply in specific validation studies.

Secondly, this theory has introduced methods (including those borrowed from psychology) and these methods have increased the level of description and the empirical nature of score description. One of these methods is the cognitive models that specify test takers' individual processes that describe item solutions (this study partly uses the cognitive processing approach, one of these models). These cognitive models provide a testable hypothesis regarding score meaning. Research indicates that the use of these cognitive models of test items has improved the quality of validity argument (Embretson & Gorin, 2001; Gorin, 2006; Yang & Embretson, 2007). They have also generated useful item development procedures (Embretson, 1999); and they have augmented score reports generated from tests (Briggs *et al.*, 2006; Huff, 2006).

Thirdly, the previous dialogue has given me the opportunity to understand that the complexity of validation research comes from the difficulty for the researcher to determine which evidence is most persuasive and how to gather such evidence. It is now that I understand that a sound evaluation of a test (like the DR Congo ESE) requires multiple sources of evidence, and that for each source of evidence, sufficient evidence must be provided. It is here that I locate a rationale to this study. One aspect of validity evidence (context validity) of the ESE was investigated in my previous study (Katalayi, 2011); and this study purports to augment the understanding of the ESE validity by investigating the encompassing aspect of the construct validity of the same test. Also, the triangulation of different methods (content analysis, protocol analysis, and descriptive methods) as well as different data (cognitive-metacognitive questionnaires, and ESE tests) in this study aims to provide sufficient evidence to my conclusions. Therefore, I expect the findings to be generated by this study to provide insights that can help augment our understanding of the concept of construct validity.

The preceding discussion on the unified construct validity will be meaningless if I do not clearly point out the relevance of this theory to the present study. I propose to undertake that in the following section.

### **2.1.6 Relevance of the unified construct validity theory to this study**

I view the unified construct validity theory as the core theory in this study. Its relevance to this study is manifold: First, the theory provides sufficient information to understand that test scores provide specific information about examinees and that this information is used to take strong decisions about university admission, course placement, promotion, recruitment etc. Therefore, it is desirable to justify the use of a test through validity studies by providing sufficient evidence for test scores use and interpretations. More specifically, investigating and improving the quality of the information yielded by the ESE is a core interest in this study, and this needs to be an essential concern for test users and consumers of test information.

Secondly, this theory can provide an informed understanding of score interpretations, a key concept in this study in particular and any validation studies in general. Most score interpretations (see Kane, 2002, 2004, 2006, and 2011) assume that the test scores should provide an indication of performance on non-test tasks involving the same kind of content, performance on other tasks, or future performance on the same or related tasks. If such inferences cannot be made on scores obtained by candidates on the ESE, there would seem to be little reason to administer the ESE to high school students as an exit test.

Finally, the unified construct validity theory appears to provide an understanding of the ESE test scores and different decisions that are taken on the basis of these scores (such as university admission, job recruitment, etc.). Although the ESE in itself would not help take such decisions (its scores must be combined to scores from other subtests), its appropriateness for any of these decisions can be questioned as its use requires sufficient evidence to ensure that the ESE does the job that it is expected to do.

Nonetheless, the unified construct validity theory, despite being the core theory of this study, does not provide sufficient ground to connect this study to the immersed base of knowledge to which other researchers have contributed, and to determine the gaps in existing validation studies on reading test. Therefore, I propose to describe, in the following section, another theory called the Construction-Integration theory. This theory, which is based on the understanding of how texts used in reading assessments are processed and understood by the reader, can augment our understanding of the concept of validity of reading tests, the matrix of this study.

## **2.2 The Construction-Integration theory**

Before I describe the Construction-Integration theory, I wish to explain two broad theories that I consider have influenced the construction-integration theory as applied to reading. These two broad theories are constructivist and schema theory.

### **2.2.1 Constructivism in education**

Constructivism is a macro-concept that refers to different things. It is a philosophy, an epistemology, a theory of communication and a paradigm in education. As a paradigm in education, constructivism is rooted in Piagetian and Vygotskian notions of cognitive and social constructivism. Cognitive constructivism stresses the cognitive development and individual construction of knowledge (Piaget, 1954, 1955, 1970; cited in Kaufman, 2004); while social constructivism stresses the social construction of knowledge that considers learning as socially constructed and that emerges from social interactions with the environment (Vygotsky, 1978).

Constructivism views reading as a complex process of social construction of meaning from print, where a reader actively interacts with a text and with the author of the text to construct the meaning (Stanovich, 1994; Donaldson, 2008). This constructivist stance to reading considers that meaning-making goes beyond simply knowing the meanings of words and then combining these words in grammatical sentences. Rather, meaning construction of texts is based on the background knowledge that readers bring to the text that is both internally formulated and socially constructed (Kaufman, 2004; Donaldson, 2008; Stanovich, 1994). Viewed from such a social process perspective, reading can be described in relation to the following three dimensions: all reading events involve a social context, reading is a cultural activity and reading is a socio-cognitive process (Bloome, 1985). Such a constructivist stance to reading can help understand how cognitive and social processes interact when a reader attempts to construct the meaning from the text and, eventually, complete the text tasks. In the following section I focus on the schema theory, a second theory that has influenced the construction-integration theory.

### **2.2.2 The schema theory**

Also called the theory of background knowledge of the reader, the schema theory is associated with cognitive psychology. According to Nassaji (2002), understanding and recall take place mainly in the context of past experience and with reference to the relevant information in the memory. Building on the semantics of the concept “schema” as articulated by Bartlett, Cook (1997) defines schema as “a mental representation of a typical instance which helps people to make sense of the world more quickly because people understand new experiences by activating relevant schema in their mind” (p. 86).

According to the schema theory, the process of text comprehension requires drawing information from both the text and the reader’s schemata until sets are reconciled as a single schema (Anderson *et al.*, cited in Hudson, 1982). This view is better clarified by Swales (1990) and Wallace (1992) when they argue that the reading process involves an identification of text topic, structure and genre, all of which activate schemata and allow the reader to comprehend the text. In line with this argument, the schema theory posits that the knowledge that a reader possesses influences not only what he/she remembers of the text he/she is reading, but also the way he/she processes it. This is to suggest that, the reader brings some knowledge to the text and this knowledge affects text comprehension. Viewed from this perspective, the schema theory is a theory about knowledge and about how knowledge is represented and how that representation facilitates text comprehension.

These two theories can provide an underpinning for the understanding of the construction-integration theory, a theory that informs this study in so far as it helps to understand how readers interact with text to construct meaning. In the following section, I describe the construction-integration theory by outlining its background, describing its content, pointing its application to reading and finally explaining its relevance to the present study.

### **2.2.3. Construction-Integration theory: Background**

Although the constructivist theory and the schema theory provide a useful ground for understanding the construction-integration theory, these two theories are general theories of

knowledge that describe how knowledge is constructed and represented. However, to better understand how texts used in reading assessments are processed and understood by the examinees, the Construction-Integration theory offers a more encompassing account of the role of knowledge and knowledge-based processes in L2 reading comprehension.

At the outset, I wish to mention that the Construction-Integration theory is informed by reading models that are influenced by research on human memory and recall as articulated in studies by Kintsch (1988, 1998, 2004), van Dijk and Kintsch (1983), Albrecht and O'Brien (1993), Gernsbacher (1995) and Rumelhart and McClelland (1986). These researchers describe in an explicit way the different cognitive processes involved in text comprehension. The most influential one is the theory of Kintsch (1988, 1998, 2004), and van Dijk and Kintsch (1983). In the following section, I outline the content of the Construction-Integration theory as articulated by these two reading research scholars.



### **2.2.3.1 Construction-Integration theory: content**

The Construction-Integration theory acknowledges that the reader's knowledge (knowledge about words, syntax, the world, spatial relations, in short, general knowledge about anything) influences text processing and comprehension. According to this theory, reader's knowledge provides part of the context within which a text is processed for understanding (Kintsch, 1988, 1998; and van Dijk & Kintsch, 1983). This assumption can help clarify the role of knowledge in text processing and comprehension, a view also acknowledged by most modern theories of reading.

Kintsch and van Dijk termed their model "Construction-Integration" because this model combines a construction phase in which a textbase that contains the text propositional meaning is constructed from the linguistic input, and an integration phase in which the textbase is integrated into the reader's knowledge in order to form a coherent mental representation of what the text is about (Kintsch, 1988, 1998, 2004; van Dijk & Kintsch, 1983).

In order to understand the construction process, Kintsch (1988, 1998, 2004) and van Dijk and Kintsch (1983) distinguish several steps that I propose to summarize here. First, text propositions

that correspond to the actual semantic properties of the text (that are called micro-propositions) are constructed directly from words and phrases in the text. Then, these propositions activate in the knowledge net other propositions and their associates so as to lead to a semantic network that includes both coherent and incoherent representations. Next, this semantic net is revised through elaboration and inference processes in which the textual propositions will be constrained by the reader's knowledge base. Finally, the textbase is organized and values are assigned to the different concepts and propositions. During this stage, the different propositions generated are then linked together and they become interconnected with both the previous and subsequent propositions, representing the microstructure (that is, local meaning relationships), and with higher-level concepts in the network, representing the macro-structure (that is, global relationships in the text). As a result, these connections create a kind of local coherence between and across the different propositions on the one hand, and between the different propositions and the overall text, on the other hand.

In order to understand the integration process, Kintsch (1988, 1998, 2004) and van Dijk and Kintsch (1983) explain that this process is a fine-tuning process that occurs at all levels of text processing (word, sentence, and discourse). According to Kintsch and van Dijk, the fine-tuning process occurs in short iterative cycles in which a new network of textual meaning is constructed, processed and integrated with what is retained in the working memory from the previous cycle. Kintsch (1988) and van Dijk and Kintsch (1983) explain that the integration process is automatic, and if this automatic process fails, the reader may engage in more strategic problem-solving processes.

Following the construction-integration epistemology, both the construction and the integration processes operate in a connectionist manner. This is to suggest that, "the knowledge that guides the comprehension system is not outside the text" as supported by schema theorists (Nassaji, 2002, p. 454); but this knowledge is generated through activation patterns initiated by textual information and different associations in the text.

The construction-integration theory suggests that textbase construction processes and the different principles that underlie the integration processes provide insights that can help understand and explain the impact of different knowledge-based processes. For example, research (Carrell, 1992; Horiba *et al.*, 1993; for example) has suggested that comprehension and

recall mainly depend on the efficacy of textbase and the encoding of the properties of the text. These studies have also indicated that comprehension is a process of creating a textbase as well as the knowledge-based interpretation of the text. Furthermore, research (Barry & Lazarte, 1995, 1998; Hammadou, 1991) has also suggested that the linguistic proficiency and prior knowledge make important but distinct contributions to reading comprehension.

Nevertheless, although the construction-integration theory explains how readers process and comprehend texts, a question worth asking at this juncture is to know how this theory sheds light to the present study. I wish to address this issue in the next section.

### **2.2.3.2 Relevance of the theory to the study**

The construction-integration theory provides to this study some useful insights that can help understand the complex mechanisms of how students process, understand and recall texts. This theory can help understand that students' text comprehension and recall depend on the extent to which they are capable to construct a textbase and to encode the different properties of the text. The theory also helps to understand that the key to successful comprehension may depend on the reader's ability to "efficiently, accurately and automatically" (Rupp, Ferne, & Choi, 2006, p. 446) extract and organize information from texts and to integrate it with his/her existing knowledge to form a text coherent mental representation. In the specific context of this study, it appears sound to state that, the construct validity of the ESE also depends on the extent to which examinees can process and comprehend the different ESE texts. This is to suggest that, in order to ensure its construct validity, the ESE texts must be carefully examined to ensure that students are capable, in light of their individual characteristics and the actual context of reading activity, to construct a textbase and to encode the different properties of these texts.

The unified construct validity theory and the construction-integration theory are not enough to design a theoretical framework for the present study. Since reading is also viewed as a strategic activity where the reader makes use of both cognitive and metacognitive strategies to construct text meaning, we need to understand how the knowledge and use of strategies by individual readers can impact on text processing and test performance. This understanding is provided

through the discussion of the strategic competence theory that I propose to address in the following section.

## **2.3 Strategic competence theory**

In this section, I offer the discussion of the theory of strategic competence in order to inform the present study on the kind of strategies readers engage in while processing the text. I believe that this discussion can help understand the extent to which the different strategies used by examinees tap into the reading construct. I first provide the background to strategic competence theory before I outline its main assumptions and its relevance to the present study.

### **2.3.1 Background**

In the discussion of the construction-integration theory I have offered in the previous section, I have highlighted that reading is a process of constructing meaning by interacting with a text. On the basis of this assumption, I have also mentioned that as individuals process texts, they use their prior knowledge along with some clues from the text to construct meaning. However, insights from research suggest that successful text comprehension also depends on the reader's ability to use different strategies to construct the textbase and integrate it in his/her knowledge in order to form a coherent mental representation of the text (Phakiti, 2003; Purpura, 1999). Viewed from this strategy use perspective, reading becomes a strategic activity as the reader adjusts, on the basis of the reading purposes, his/her reading to each purpose and for each reading task, suggesting that strategic reader uses a variety of strategies and skills as he/she constructs text meaning (Mannesa & Hoyesb, 1991; Phakiti, 2003; Purpura, 1999).

In language testing literature, the background to strategic competence theory can be traced back to the beginning of the 1990s when language testing researchers shifted their attention to the identification of the different characteristics that accounted for variation in performance on language tests (Bachman & Palmer, 1996; Purpura, 1999). These researchers found that one of the characteristics that accounted for variation in students' performance on reading tests was attributed to differences across individuals in their text processing strategies. Since then, the

notion of strategic competence was closely linked to the concept of metacognition (Bachman & Palmer, 1996; Alderson, 2000; Douglas, 2000; Purpura, 1999). Metacognition is a theory of “thinking about thinking” (Hacker, 1998), and it is defined as a “deliberate, planned, intentional, goal-directed and future oriented processing that can be used to accomplish cognitive tasks” (Flavel, 1977, cited in Phakiti, 2003, p. 56).

As a cognitive process, metacognition involves, as Phakiti (2003) suggests, active monitoring and consequent regulation and orchestration of cognitive processes to achieve cognitive goals. In light of this suggestion, three metacognitive processes have generally been identified; they are: monitoring, regulation and orchestration; and these processes generally take the forms of checking, planning, selecting and inferring, self-interrogation and self-introspection, interpretation of the ongoing cognitive processes and /or simply making judgments about what one knows or does not know to fulfill a task (Phakiti, 2003).

In light of the preceding identified metacognition processes, Phakiti (2003) suggests two facets of metacognition: knowledge of cognition and regulation of cognition. The knowledge of cognition refers to what individuals know about their cognition while the regulation of cognition refers to a set of activities that help individuals control their performance.

On the basis of the above semantic contours of metacognition, the strategic competence theory refers to a conscious and deliberate ability to use any strategies (cognitive and metacognitive strategies) to appropriately complete a language task at hand (Phakiti, 2003). Cognitive strategies (for example planning strategies, monitoring strategies) are those that are likely to be encapsulated within language competence (organizational and pragmatic knowledge) and world knowledge; whereas metacognitive strategies (for example self-evaluation, assessing situation, self-testing) are those higher order executive conscious processes that provide a cognitive management function to complete a language task (Phakiti, 2003).

In the following section, I propose to address the relationship between strategic competence theory and reading test performance.

### 2.3.2 Strategic competence theory and reading test performance

The main assumption of the strategic competence theory as it speaks to reading test performance is that understanding the process of reading is inevitably important in order to understand the nature of reading; therefore, to understand the nature of reading test performance.

The strategic competence theory posits that different individual readers' characteristics influence reading performance. This is to suggest that, depending on their individual capacities to use strategies, different readers may process the same text in different ways and they may perform differently to the same reading task.

Research suggests that successful readers have an awareness and control of the cognitive activities they engage in as they read, and that these readers know how to use appropriate metacognitive strategies to enhance the comprehension of the text (Phakiti, 2003; Chamot *et al.* 1989). In contrast, poor readers in general lack effective metacognitive strategies and they have deficiencies in their use of these strategies to monitor their understanding of the texts (Alderson, 2000; Phakiti, 2003).

At this juncture, I wish to discuss two concepts that are closely associated with strategic competence theory and which are the key concepts in this study. These concepts are reading strategies and test taking strategies. Reading strategies are those strategies (both cognitive and metacognitive) that the reader engages in processing and comprehending the text. These strategies are "conscious procedures" that are deliberate and purposeful (Cohen & Upton, 2007, p. 211). Therefore, since research has indicated that reading for test purpose may be a problem-solving process, reading strategy analysis can help understand how readers interact with the text and how their selection of different strategies influences not only their comprehension of the text, but also their performance on the test (Cohen & Upton, 2007; Rupp, Ferne & Choi, 2006).

Test taking strategies, on the other hand, are the test taking processes that the examinees select and which they are conscious of, at least to some degree (Cohen & Upton, 2007). Cohen and Upton make a distinction between test management strategies and test wiseness strategies. The former are those strategies that constitute an opting out of the reading task at hand (for example, read the question and consider the question options before going back to the passage, consider the question options and select preliminary option(s) because of lack of certainty), while the

latter are those strategies that test takers may use as a short-cut to arriving at answers ( for instance, select an option even though it is not understood, out of a vague sense that the other options could not be correct [elimination of distractors], select the option without reading the text [guessing]).

### **2.3.3 Relevance of the strategic competence theory to the study**

The main issues concerning the validity of an EFL test are the meaning, the relevance and the utility of test scores as a basis for action and the functional worth of scores in terms of the social consequences of their use (Messick, 1989). Investigating the validity of a text-based test like the DR Congo ESE from the point of view of strategies used by examinees through the process of taking the test can provide an understanding of the nature of reading and hence, the nature of reading test performance. Research on test taking strategies (Hirano, 2008; Cohen & Upton, 2007; Rupp, Ferne & Choi, 2006; Phakiti, 2003; for example) has indicated that a test is valid if it can elicit efficient strategies test constructors assumed examinees would employ to promote correct answers. In light of this claim, I believe that when examinees deploy a number of strategies, and that they use these strategies effectively in answering individual test items, the use of these strategies may indicate their genuine control over the test items.

In accordance with this claim, the theory of strategic competence is relevant to this study as it provides an understanding of the different strategies examinees use to complete the test tasks, and the theory helps to understand the extent to which these strategies associated with test taking can be considered as part of the test construct. This is to suggest that, the theory helps to understand whether or not the different strategies examinees use to read the text and answer MC questions based on it reflect the fact that these examinees actually engage with the reading test tasks in the manner desired by the ESE test constructors. Such an engagement with the text can be reflected in the way examinees work to understand the text and to understand the expectations of the questions. It can also be reflected in the way they attempt to understand the meanings and implications of the different item options in the light of the text, and how they select and discard different item options based on the way they have understood the text.

Furthermore, I consider the strategic competence theory to be particularly relevant to this study as it can guide me in the choice of the research instruments necessary to gather research data. This justifies my use of a strategies questionnaire in order to elicit different strategies the research participants deployed while taking the test (see Chapter 5, Section 5.7).

Finally, the strategic competence theory provides guidance in the choice of the process-oriented approach that is used as the framework of this study. This approach better suits this study as it looks at the relationship among respondents' use of test taking strategies, the test context, the structure of the text, and the features of the test items. Such a relationship appears to provide a sound argument to validate the ESE.

### **Summary of the chapter**

In this chapter, I have provided the description of the three theories that inform this study. I have organized this description in three sections. In the first section, I have described the unified construct validity theory by providing its historical background, and its main content. I have outlined the issues of test validation and validity argument and I have stressed their importance in any validity studies. Subsequent to this, I have provided a discussion on the adoption of unified construct validity theory in language testing before I provided the different criticisms formulated on the unified construct validity theory, and the subsequent responses articulated by those researchers who sustain the theory in its actual main conceptualization. In the second section, I have described the Construction-Integration theory, the second theory that sheds light on the study. In order to provide a background to this theory, I have first surveyed the schema theory and the constructivist theory in education, two epistemologies that have provided ground to the Construction-Integration theory. Then I have outlined the main assumptions of the theory and its content before I outlined its relevance to the present study. In the last section, I have described the strategic competence theory, the third theory used in this study. After providing the theoretical background, I have outlined its main assumptions so as to signpost its relevance to the present study.

In order to provide a better ground for identifying the factors that are a potential threat to the validity of the ESE, I need to design a framework that connects the above three theories to all

aspects of this study. This framework is meant to act like a map that gives coherence to my empirical inquiry. Therefore, I describe, in the following chapter, “**The process-oriented approach to examining the reading construct**”.

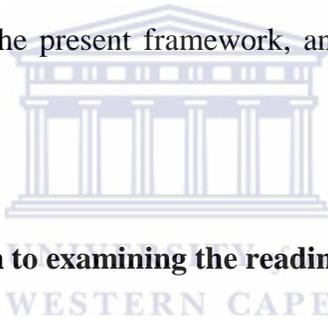


## CHAPTER THREE

# DESIGNING A PROCESS-ORIENTED APPROACH FOR EXAMINING THE READING CONSTRUCT

### 3.0 Introduction

In this chapter, I describe the approach I have used to examine the validity of the ESE. The development of this approach starts with a rationale for examining the reading construct from the examinees' response processes; then I describe the two existing approaches that have provided insights into the development of the present framework, and finally I describe the framework content.



### 3.1 A process-oriented approach to examining the reading construct: Rationale

Before I state the reason why I have developed this approach as my study framework, I wish to discuss existing approaches to examining the reading construct and state the reasons why I think these approaches are not appropriate in the context of this study.

At the outset, I wish to mention that from 1960s onwards the different approaches to examining the validity of reading tests have been strongly influenced by the issue of the divisibility of reading for testing purposes. This issue supports that reading is a complex and multidimensional process and the best way to describe reading process is to describe the subskills that underlie it (Song, 2008). This sub-skills approach to testing reading postulates that it is possible to link particular test items or test tasks to specific sub-skills that they are said to tap into (Weir & Porter, 1994; Alderson, 2005).

In light of these assumptions, it appears that this way of establishing the construct validity of reading tests reflects a product-oriented view of reading where students' performance on reading

tests is measured on the basis of their scores obtained from the test. Consistent with this subskills approach, researchers have often adopted purely quantitative approaches to establishing test validity. Such quantitative approaches have mostly relied on factorial analyses and linear regression methodologies. It is worth mentioning, at this juncture, that factor analysis is a method that investigates whether a number of variables of interest are linearly related to a smaller number of unobservable factors. In the case of reading tests, for example, results from factor analysis methodologies could tell whether or not the different test items included in the test load on the same factor. For example, through exploratory and confirmatory factor analyses, the result can indicate that the difficulty of a reading comprehension test can be accounted by such factors as passage content, text syntactic features, text vocabulary load, and so on.

There is substantial literature relating to the construct validity of reading tests using factor analysis methodologies (Guthrie & Kirsch, 1987; Rosenshine, 1980; Schedl *et al.*, 1996; Young, 2008; for example). Some of these studies have suggested that there are some factors that can account for variance across test items; suggesting that text comprehension may not be a unitary mental operation and that reading to comprehend, for instance, may be clearly differentiated from reading to locate specific information in the text (Guthrie & Kirsch, 1987; Rosenshine, 1980). Other studies have looked at the dimensionality of reading test items to see whether or not the hypothesis of divisibility of reading skills can be supported (Schedl *et al.*, 1996).

However, although studies using factor analysis methodologies have produced useful findings regarding the validity of reading tests, I agree with Weir and Khalifa (2008a, p. 2) when they believe that factor analysis approach is limited by its essence as it focuses on “the separability of the capabilities that a reader is assumed to need in order to tackle certain test items”. Moreover, this approach, since it is psychometrically driven, deals with the factors that statistically show their contribution to successful reading performance; thus, ignoring the actual processes the reader engages in reading the text in normal real life situations.

In line with the preceding stance, factor analysis methodology tends to model reading ability rather than the reading process. It views reading as a “set of theoretically distinct and empirically isolable constituents” (Hoover & Tunmer, 1993, p. 4). Another relevant shortcoming of factor analysis methodology is that some apparently different reading skills may behave in a similar statistical way; suggesting that there is a need to focus on a certain broad reading ability rather

than a range of skills that are involved in the reading activity. This finding appears to be in accordance with the findings reported by Gillis, DeFries and Fuker (1992), Doehring and Hoshko (1977) and Spearritt (1972), for example.

As for linear regression methodology, it attempts to model the relationship between two variables by fitting a linear equation to the observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, studies that aim to determine the relationship between the surface task features and performance on reading items generally identify a number of text and item variables, and they investigate the effect of these variables on item statistics (Freedle & Kostin, 1993, 1999). Or again, these studies indicate the specific variables that account for a substantial percentage of the variance in item difficulty (Kasai, 1997).

However, like studies which have used factor analysis methodology, the various studies which have used linear regression methodology to validate the reading construct present results that need to be interpreted with much caution. The reason is that, given their overreliance on text variables and item variables to account for difficulty on reading tests, these studies fail to reveal the different processes that are involved in solving the different individual test items. Most importantly, the statistic results from these studies may be oversensitive to the presence or absence of one individual variable (Kasai, 1997); hence, they may produce a wrong interpretation of the examinee's performances.

On the basis of the preceding argument, I believe that both factor analysis and linear regression methodologies do not appear to offer an appropriate route to examining the construct validity of the ESE. This is because these two approaches do not look at the reading process, and they ignore not only the range of types of reading examinees conduct but also the level of cognitive demands imposed on the examinees while processing different reading tasks. Since reading is a cognitive and social process, I believe that test developers will be better served if they attempt to design test items that cover the reading construct as appropriate to examinees' reading purpose and processing levels by ensuring that the processing demands that are required to complete the test tasks are proportional to reading processes required in non-testing situations as evidenced by reading research. I also hold with Weir and Khalifa (2008a) and Khalifa and Weir (2009) that, given the aim of evaluating the construct validity of reading tests, a quantitative approach solely

based on factor analysis and linear regression of reading items seems inappropriate. Weir (2005), Field (2010) and Sivasubramaniam (personal communication, 2011) caution against relying exclusively on a quantitative approach for construct validation as “statistical data do not in themselves generate conceptual labels” (Weir, 2005, p. 18), and that “such analyses by their nature tell us little about what is actually happening when a reader processes the text under test conditions” (Field, 2010, p. 37).

As a conclusion to this subsection, I believe that in order to investigate the construct validity of a text-based test like the ESE, and in considering the specificity of this test context (see Katalayi, 2011), I need to examine as extensively as possible the nature of reading activities and different response processes examinees engage during test taking. The adoption of this route suggests a need to understand the broad theory of what it means to comprehend the text. In this perspective, such a route to examining the reading construct must elicit the cognitive processing involved in not only in the text and test tasks, but also in contexts beyond the test itself; that is, in those contexts involved in performing reading tasks in classroom and real life settings.

However, to design a process-oriented framework for examining the reading construct requires an understanding of existing process-oriented frameworks and the identification of their main strengths and weaknesses. In the following section, I propose to describe two frameworks whose strengths have provided guidance to the development of my framework.

### **3.2 Weir and Khalifa’s (2008a) and Embretson and Wetzel’s (1987) frameworks to examining the reading construct: Insights and Issues**

After I have explained the reason why I have selected a process-oriented approach in order to examine the reading construct as measured by the ESE, I wish to describe, in this section, two process-oriented approaches that have provided insights and guidance to the approach I have developed in the present study. These two approaches are the “cognitive-processing approach to examining the reading construct” as developed by Weir and Khalifa (2008a) and the “cognitive processing model of reading comprehension” as developed by Embretson and Wetzel (1987). Prior to the discussion of these two approaches, I present an overview of different cognitive

processing models. Then I describe the main content, stress the relevance and highlight the limitations of each of these two approaches.

### **3.2.1 Overview of cognitive processing models**

In this section, I offer a description of existing processing models by first explaining the notion of componential models of reading before I describe the three widely used cognitive processing models: bottom-up, top down and hybrid models.

#### **3.2.1.1 Componential models of reading**

In viewing reading as a process, researchers have attempted to develop theories aimed to provide an understanding of the different processes that are involved while reading a text. Among the myriad of processing reading models, the componential models of reading deserve particular attention in the scope of this study. Following Urquhart and Weir (1998), componential models provide a description of the different components that are assumed to be involved in the process of reading and this with some explanations of how these different components interact with each other.

Among the different componential models of reading, a distinction is made between componential models with two components and those with three components. The componential models of reading with two components as developed by Carver (1992) and Guthrie and Kirsch (1987) are consonant with the bi-divisible view of reading where reading competence and vocabulary are considered as the two crucial reading aspects. Today, this way of viewing reading as bi-dimensional does not reflect recent applied linguistic developments that consider reading as a complex and multi-componential process.

The componential model of reading with three components as developed by Bernhardt (1991) includes three variables in reading process: language, literacy and knowledge. This model stresses the importance of metacognitive strategies that are goal-setting and comprehension monitoring. Although this model reflects some of the current trends in reading theory, it does not

appear to be effective in so far as reading needs to be described in more comprehensive and multi-componential terms rather than in a limited range of variables that interact during text processing. In light of the preceding argument on the componential models of reading with two or three components, I can conclude that these reading models do not appear to offer an appropriate route to examining the validity of reading tests although they provide some insights into the understanding of the reading process.

### **3.2.1.2 Bottom-up versus Top-down models**

Over the last two decades, cognitive theory in psychology has enormously impacted on the understanding of reading process. Among the cognitive processing models developed to understand the nature of reading, two groups of models deserve particular attention in the scope of this study; they are: bottom-up models and top-down models. Simply defined, bottom-up processing is an immediate left-to-right processing of the text data through a series of discrete stages (Ruddell, Ruddell, & Singer, 1994), whereas top-down processing is a processing in which the reader approaches the text with his/her existing knowledge and works it down to comprehend it (Ruddell, Ruddell, & Singer, 1994).

In bottom-up processing, reading occurs in a hierarchical order and the reader is expected first to process “the smallest linguistic unit, gradually compiling the smaller units to decipher and comprehend the higher units” (Dechant, 1991; cited by Weir *et al.*, 2008a). This is to suggest that, the reader is viewed as someone who passively and sequentially decodes the text meanings from letters, words and sentences. Consistent with this suggestion is the view that reading process itself is performed under the strict control of the text, and it has little to do with the reader’s prior knowledge or the reading context. LaBerge and Samuels’s (1974) automatic information processing model can serve as bottom-up processing model.

In the top-down processing, on the other hand, the reader is viewed as someone who contributes to the reading process over the incoming textual information. In this model, the reader is assumed to employ existing knowledge to predict text meaning. Two well-known top-down processing models include the psycholinguistic models and the schema-theoretic models. The psycholinguistic models (Smith, 2004) stress the interaction between language and thought,

especially the reader's capacity to infer information from existing knowledge. They view reading as an active, purposeful and selective process. The schema-theoretic models view reading as a process of activation of reader schemata. These models stress the centrality of reader's prior knowledge in reading comprehension.

Apart from the bottom-up and top-down divide, there are also hybrid reading models that have combined the relevant insights of both bottom-up and top-down models. I can mention here, for example, the interactive model (see Kintsch, 2004; for example) and the interactive-compensatory model (see Stanovich, 2000; for example).

In the context of this study, all these processing models, although they provide insights into the understanding of reading as a process, they do not offer an appropriate route to examining the construct validity of a reading test. Reading should not be viewed in term of bottom-up top-down divide or as an interactional process only. Instead, it must be viewed as a complex and multi-componential process where both cognitive and social processes interact in such a way that text comprehension depends on the reader's capability to make use of these multiple processes to construct the text meaning.

However, there are some processing models that have looked at reading as a multi-componential process and two of these models deserve particular attention in the scope of this study: they are Weir and Khalifa's (2008a) cognitive-processing model of reading and Embretson and Wetzel's (1987) cognitive processing model of reading. Since these two models offer guidance in the development of the model I have used in this study, I wish to discuss in the following two sections these two models by focusing on their content, highlighting their relevance to the present study and stressing their limitations.

### **3.2.2 Weir and Khalifa's (2008a) cognitive-processing approach to examining the reading construct: Description and insights**

In this section, I describe Weir and Khalifa's (2008a) model, the first model that provides guidance to the present study.

### 3.2.2.1 Weir and Khalifa's (2008) model: background and content

This model was developed by Weir and Khalifa (2008a), and it was first published in an article entitled “A cognitive processing approach towards defining reading comprehension” published in *Cambridge ESOL: Research Notes*; and it was subsequently published in a book by Khalifa and Weir (2009) entitled “*Examining Reading: Research and practice in assessing second language reading*”, published at Cambridge, by UCLES/Cambridge University Press.

Weir and Khalifa's framework conceptualizes reading as a multi-componential construct where reading can be broken down into underlying skill or strategy components for the purpose of both teaching and testing (Weir & Khalifa, 2008a; Khalifa & Weir, 2009; Weir *et al.*, 2008). This view is consonant with many process-oriented models that view reading construct as made up of more than one dimension. Implied in this conceptualization is the view that reading construct needs to be described in comprehensive and multi-componential terms. In keeping with this view, the model suggests that, in order to better examine the reading construct, we need to look at how readers engage constellations of cognitive processes to make sense of the text (Weir & Khalifa, 2008a).

In this framework, Weir and Khalifa hypothesize that difficulty in reading is a function of both the level of processing required by reading purpose and the complexity of the text (Weir & Khalifa, 2008a; Khalifa & Weir, 2009). Therefore, the model conceptualizes reading construct on multiple dimensions with careful reading and expeditious reading that can be conducted at both global level and local level. Such a conceptualization of reading is a development in the reading theory because many existing reading models stress on a careful reading model and ignore the impact of different purposes of reading on the choice of the types of reading. Yet, expeditious reading is usually conducted when readers skim the text for a gist in real life reading situations.

From this conceptualization of reading, it appears reasonable to state that Weir and Khalifa's model may provide a useful route for identifying the most appropriate combination of constellations of processes, skills and strategies that readers employ to construct text meaning. Figure 3.1 can help clarify this conceptualization. In this figure the goals that are open to the

reader and that characterize reading are either careful reading or expeditious reading. These two reading types may occur either at the global level or local level.

	<b>Global level</b>	<b>Local level</b>
<b>Careful reading</b>	<ul style="list-style-type: none"> <li>-Establishing accurate comprehension of explicitly stated main ideas and supporting details across sentences</li> <li>- Making propositional inferences</li> <li>- Establishing how ideas and details relate to each other in a whole text</li> <li>- Establishing how ideas and details relate to each other across texts</li> </ul>	<ul style="list-style-type: none"> <li>-Establish accurate comprehension of explicitly stated main idea or supporting details within a sentence</li> <li>- Identifying lexis</li> <li>- Understanding syntax</li> </ul>
<b>Expeditious reading</b>	<ul style="list-style-type: none"> <li>- Skimming quickly to establish: discourse topic and main ideas, or structure of text, or relevance to needs</li> <li>- Search reading to locate quickly and understand information relevant to predetermined needs</li> </ul>	<ul style="list-style-type: none"> <li>- Scanning to locate specific points of information</li> </ul>

Figure 3.1: Reading type (Source: Weir & Khalifa, 2008a)

Careful reading refers to different operations where the reader attempts to extract complete meanings within or beyond sentences right up to the level of the entire text so as to construct the text meaning (Weir & Khalifa, 2008a; Khalifa & Weir, 2009). Careful reading can be conducted at local level or global level. Careful reading at the local level involves processing a text until the basic meaning of a proposition is established whereas careful reading at the global level involves processing the text until its macro-structure is built. Such a type of reading appears to concord with van Dijk and Kintsch's (1983) and Kintsch's (1988) view when they argue that text comprehension is a slow, careful, linear and incremental process.

Expeditious reading refers to different operations where the reader quickly and selectively attempts to access needed information contained in the text. Weir and Khalifa's framework

posits that expeditious reading can be conducted at global level (through skimming and search reading) or at local level (through scanning).

Reading at global level is concerned with the understanding of text propositions beyond the level of microstructure. This is to suggest that, reading at global level is related to the relationships between different ideas as represented in different complexes of propositions. This is further to suggest that reading at global level refers to any macro-propositions that include the main ideas of the text and the way in which the different text micro-propositions elaborate upon them (Weir & Khalifa, 2008a; Khalifa & Weir, 2009).

Reading at local level is concerned with the understanding of propositions at the strict level of micro-structure; that is, the clause and sentence levels (Weir & Khalifa, 2008a; Khalifa & Weir, 2009). This is to suggest that, reading at local level is closely associated with reader's linguistic knowledge and that this comprehension level relates to text decoding process (word recognition, lexical access, and syntactic parsing) and to establishing propositional meaning at the clause and sentence levels.

Whether one reads carefully or expeditiously at either global level or local level, Weir and Khalifa believe that reading activity itself can be better understood when we examine different hierarchical processing levels that readers engage in while reading texts. These levels are: word recognition, lexical access, syntactic parsing, establishing propositional meaning, inferencing, building a mental model and creating a text-level structure. Figure 3.2 on the next page presents these processing levels as conceptualized by Weir and Khalifa.

Word recognition is concerned with the matching of the form of a word in a written text with a mental representation of the orthographic forms of the language (Adelman & Brown, 2008; Norris, 2009).

Lexical access is the retrieval of a lexical entry from the lexicon, containing stored information about the form or the meaning of a word (Field, 2004; Wagenmakers *et al.* 2000). Word form includes orthographic and phonological mental representations of a lexical item and possibly some additional information on its morphology. Word meaning includes information on word class and the syntactic structure in which the word can appear. It also includes information on the range of possible senses for a word. The model posits that text comprehension also depends on

the reader’s capacity to encode words. The following figure 3.2 can illustrate Weir and Khalifa’s processing levels.

<b>Careful reading</b>  <b>Local</b> Understanding sentence	Creating an inter-textual representation	Text structure knowledge
	Creating a text level representation	
<b>Global</b> Comprehend main idea Comprehend overall text Comprehend overall texts	Building a mental model	Meaning representation of text
	Inferencing	
	Establishing propositional meaning	
<b>Expeditious reading</b>  <b>Local</b> Scan/search for details	Syntactic parsing	Syntactic knowledge
<b>Global</b> Skim for gist Search for main ideas and important details	Lexical access	Lexicon
	Word recognition	

Figure 3.2: Processing levels (Adapted from Weir & Khalifa, 2008a)

Syntactic parsing relates to processes involved in constructing syntactic structures during reading (van Gompel, 2007). Weir and Khalifa’s model posits that fluency in syntactic parsing is important in the comprehension process. Once the meaning of words is accessed, the reader has to group these words first into phrases, and then into larger units at the clause and sentence level to comprehend the text (Weir & Khalifa, 2008a; Khalifa & Weir, 2009).

Propositional meaning is the “literal interpretation of what is on the page” (Weir & Khalifa, 2008a). Weir and Khalifa’s model posits that the reader, in his/her effort to process and comprehend a text, adds external knowledge to the text in order to turn the text into a coherent message that relates to the context in which it occurred (Weir & Khalifa, 2008a; Khalifa & Weir, 2009).

Inference in reading a text for comprehension, is also known as reading between the lines. In order to comprehend a text, the reader needs to integrate the information provided in the text with the information that he/she already knows. Such integration enables him/her to go beyond

explicitly stated ideas by linking these ideas to implicit ideas. In accordance with this view, inference appears to be a creative process whereby the brain adds information which is not stated in the text in order to impose coherence (Weir & Khalifa, 2008a).

Building a mental model relates to the reader's capacity to identify main ideas, to relate them to previous ideas, to distinguish between macro-propositions and micro-propositions and to impose a hierarchical structure of the information in the text (Weir & Khalifa, 2008a; Khalifa & Weir, 2009). This statement appears to chime in Kintsch and van Dijk (1978)'s argument when they note that the different propositions representing the meaning of a text are linked together to form a hierarchical textbase.

Finally, the text-level structure, sometimes called the discourse-level structure, is the final stage of text processing in which the reader attempts to build the discourse level of the text as a whole. According to Weir and Khalifa's approach to reading construct, a (skilled) reader attempts to recognize the "hierarchical structure of the whole text and determines which items of information are central to the meaning of the text" (Weir & Khalifa, 2008a, p. 7).

However, in processing a text for comprehension, the reader may choose to scan, skim, search read or read carefully in response to different task demands. Weir and Khalifa have elaborated on these different skills.

Skimming is a reading method of rapidly moving the eyes over the passage in order to get the text macro-structure (main ideas and an overall view of the text content). Weir and Khalifa (2008a) and Khalifa and Weir (2009) argue that the defining characteristic of skimming is that the reader attempts to read the text selectively by trying to build up the text macro-structure on the basis of as few details from the text as possible.

The cognitive processing approach posits that, in skimming the reader allocates his/her attention to propositions that seem to be macro-propositional. This is to suggest that skimming may be equivalent to gist extraction. Consistent with this view, Weir and Khalifa (2008a) are of the view that in skimming a text, propositions are committed to the long term memory on the hypothesis that they represent a macro-structure. This view appears to concur with Kintsch and van Dijk (1978)'s argument when they note that skimming requires the reader to create a skeletal text

level macro-structure by deciding on the superordinate macro-structure that encapsulates the meaning of the whole text.

Search reading is the type of reading where the reader samples the text in order to extract specific information. In this type of reading, the reader focuses on words, topic sentences or any important paragraph in his/her attempt to extract specific information. Weir and Khalifa (2008a) argue that in search the reader's attention is more focused as he/she looks for only information necessary to answer a given question or set of questions.

Weir and Khalifa (2008a) and Khalifa and Weir (2009) note that search reading starts with looking for the related vocabulary in the semantic field indicated by the test task. Once the required information to answer a test question has been quickly and selectively located, then the examinee recurses to careful reading by trying to establish the propositional meaning at the sentence level, trying to enrich the propositions through inferencing as well as trying to integrate information across sentences.

Scanning involves reading selectively by looking for specific words or phrases, names, dates, etc. Rosenshine (1980) defines scanning as a perceptual recognition process which is form-based and relies on accurate decoding of a word or group of words. In scanning process, a low level of attention is accorded to any part of the text that does not contain the pre-selected words. Besides, the reader is not bound to observing the author's sequencing by following the text in a linear way. To take just an example, a reader may scan a text forth and back while trying to find a reference to a particular author.

The cognitive-processing approach posits that scanning process requires only a limited amount of lexical access, little or almost no syntactic processing, no need to complete the reading of a sentence, no checking of coherence, and no attempt to build the text macro-structure. In brief, scanning does not involve any of the processes of meaning building.

The cognitive-processing approach posits that careful reading requires a demanding level of processing as the reader attempts to establish how ideas and details relate to each other in the portion of the text or the whole text. To achieve that, the reader carefully reads the text and rereads it until he/she not only understands the macro- and micro-proposition, but also finds out how they are interconnected (Weir & Khalifa, 2008a; Khalifa & Weir, 2009).

Based on the description of the cognitive-processing approach to reading construct as designed by Weir and Khalifa (2008a) and Khalifa and Weir (2009), I wish to point out the quintessence of this approach by stating that the difficulty of text processing is in large part a function of the cognitive load imposed by the complexity of the text employed in the test.

Although I believe that a cognitive processing approach can serve as a tenable and productive theoretical basis for establishing the construct validity of reading tests as it clarifies what comprehension really involves, I find this model incomplete. In the following section, I wish to discuss some of its limitations.

### **3.2.2.2 Weir and Khalifa's model: Limitations**

In this section, I describe the limitations of Weir and Khalifa's approach to examining the reading construct.

The most important limitation of this model is that it ignores the context of reading. Insights from research suggest that reading is a situated activity that does not occur on a vacuum; rather, it occurs in a specific context (Bloome, 1985). Consistent with this claim, reading context needs to be considered while developing testing models. This is to suggest that, examinees' capacities to process the text both carefully and expeditiously at either global level or local level and their ability to complete test tasks also depend on their individual characteristics (skills, motivation for reading, experience with the test tasks, attitude towards the test), the characteristics of reading materials (availability, quality, relevance), the conditions under which the teaching/learning of reading takes place (classroom size, reading resources), etc. This is to suggest that the best way to examine reading construct can be to investigate it within the larger context of reading activity.

Another limitation of Weir and Khalifa model is that the description of the main assumptions does not take into account the specificity of different reading testing methods. In reading testing literature, three main test formats are used: the multiple-choice format, the cloze format, the constructed response format. For the multiple-choice format (that is used for the ESE), the text characteristics and the constellation of processes required from the examinees to complete the test tasks as developed by Weir and Khalifa (2008a) are not sufficient to establish the construct

validity of a reading test. One additional variable is highly relevant: the test structure. Research (Ascalon *et al.*, 2007; Dudycha & Carpenter, 1973; Drum, Calfee & Cook, 1981) has established that the difficulty in responding to MC reading test items can come from either the difficulty in understanding the text or the difficulty in understanding the test items. Royer (1990), Davey (1988) and Embretson and Wetzel (1987) found that item features, not text features, can influence item difficulty; thus jeopardizing the construct validity of the test. In light of the preceding claims, Cohen and Upton (2006, 2007) and Rupp, Ferne and Choi (2007) found that the multiple-choice format influences some examinees to use reading and test taking strategies they would not use with other test formats or in non-testing situations. In light of these research insights, since Weir and Khalifa's approach does not take into account test/item variables from the analysis, I believe that this model cannot produce appropriate results since task completion on MC reading tests involves the text, the question, the options (alternatives) and the test takers' mental processes employed to answer the questions. I believe that a good model would be the one that provides results on how item stimulus properties influence the examinees' response processes.

Considering the preceding two limitations of Weir and Khalifa's model, and in the context of this study, there is a need to include the test context and the test structure as two additional components that encompass different variables that can account for variation in performance to MC reading tests. This argument justifies the need to develop an adapted model that can suit the specific context of the present study. But at this juncture, I wish to describe another model that also provides guidance to the model I propose to develop in the present study.

### **3.2.3 Embretson and Wetzel's (1987) cognitive processing model of reading comprehension**

Like the preceding model by Weir and Khalifa, Embretson and Wetzel's model provides useful insights that can augment my understanding on the various factors that interact in text processing and comprehension and that account for performance on reading tests. The following section provides a description of this model.

### 3.2.3.1 Embretson and Wetzel's model: background and content

In order to describe the processing difficulty of multiple-choice reading tests, Embretson and Wetzel developed a cognitive processing model of reading comprehension in 1987 in a study entitled "Component latent trait models for paragraph comprehension tests".

Embretson and Wetzel's (1987) model describes the sources of text processing complexity as deriving from two processes: text representation and response decision. Text representation process is the stage in which the text is comprehended. This process consists of two events: the lexical encoding (converting the visual stimuli of the text into meaningful representation) and coherence of the text for a set of items. The following figure 3.3 represents the model:

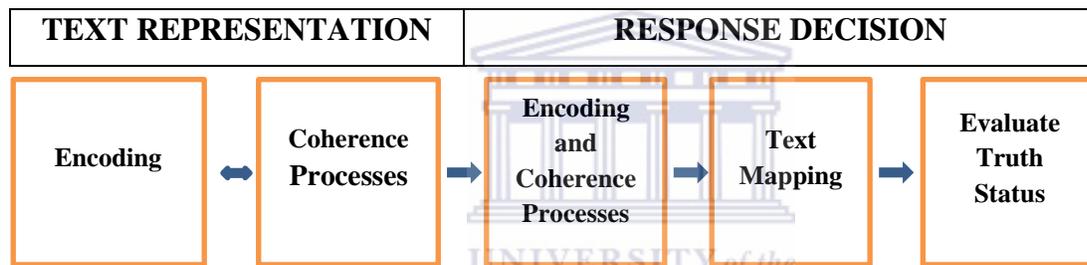


Figure 3.3: Embretson and Wetzel's (1987) Reading Comprehension Model

Embretson and Wetzel's (1987) model posits that the difficulty of lexical encoding is due to the linguistic features of the text, particularly vocabulary difficulty; a view previously supported by Drum, Calfee, and Cook (1981) and Graves (1986). On the basis of this assumption, Embretson and Wetzel postulate that passages with high level of vocabulary are difficult to encode and retrieve when responding to comprehension questions. Text coherence, on the other hand, is a process of connecting word meanings and different text propositions into a meaningful representation of the text. This view appears to match with Kintsch and van Dijk's (1978) theory of text comprehension where they argue that the reader processes the text in cycles during which a network of coherent propositions is constructed from the text base, and then integrated into a coherent whole from the reader's prior knowledge. Following this view, Embretson and Wetzel's model posits that the difficulty of coherence processes is strongly influenced by the propositional

density of the text. Therefore, it is assumed that propositionally dense texts are difficult to process and integrate for later recall and comprehension.

Response decision, or decision process is a stage in which the question stem and the options/alternatives are compared for accuracy of the text. Embretson and Wetzel's (1987) model describes three decision processes: encoding and coherence processes, text mapping and evaluation of the truth status of response alternatives. Encoding and coherence processes of responses options/alternatives are the same as in text representation, except that they apply to test questions and response options/alternatives.

Text mapping is a process of relating the information in the item question and response options/alternatives to the information in the text. Therefore, the model assumes that difficulty in text mapping is influenced by the amount of information needed from the text to answer individual test questions (Embretson & Wetzel, 1987). This is to suggest that, if the relevant information necessary for answering an individual test item is contained in a single short sentence, text mapping will be expected to be easy. Conversely, if the relevant information necessary for answering an individual test item is spread throughout the text, text mapping will be expected to be difficult.

The third decision process, evaluating the truth status of response options/alternatives, is a two-stage process of falsification and confirmation of response alternatives (Embretson & Wetzel, 1987). These two decision processes describe the extent to which information given in the text can be used to make decision regarding the response options/alternatives. In the falsification stage, the examinee attempts to falsify as many response options/alternatives as possible since it is generally agreed that in many multiple-choice questions, the distractors are not wholly false or their truth status cannot be determined (Embretson & Wetzel, 1987). In light of this assumption, the model posits that a question item is difficult if few of its response options/alternatives are falsifiable. In the confirmation stage, the remaining response options/alternatives are compared to the text passage to determine if the text confirms the response as correct. Therefore, a response option/alternative that is confirmed is selected as the key (correct answer). The model therefore assumes that the examinee's ability to falsify and confirm response options/alternatives depends on her/his capacity to convert the wording of the response option/alternative to that of the text and her/his capacity to compare propositions between the response alternatives/options and the

text. This is to suggest that, a test item will be difficult if the alternatives contain many propositions where each of these propositions must be compared to many propositions in the text.

Embretson and Wetzel's (1987) approach to examining reading comprehension provides useful insights that are specific to multiple-choice reading tests, and these insights inform the approach I propose to develop in this study. Nevertheless, this approach cannot be solely used to examine the reading construct as measured by the ESE because of some limitations. In the following section I outline these limitations.

### **3.2.3.2 Embretson and Wetzel's model: Limitations**

The description of Embretson and Wetzel's model in the previous section has clarified some useful insights that can be used to construct a framework for examining the reading construct as measured by the ESE. However, Embretson and Wetzel's model is not appropriate to be solely and fully used in the specific context of the present study. The main reason is that the model is psychometrically driven and it conceptualizes reading as a cognitive activity where the reader draws on various encoding and decoding sub-processes to construct text meaning. Therefore, Embretson and Wetzel's model ignores the social dimension of reading. Yet, as I have stated earlier, reading should not be conceptualized as a cognitive activity only; rather, it needs to be conceptualized as both a cognitive and a social activity.

Another important limitation of Embretson and Wetzel's (1987) approach is that it relies mainly on the use of regression analysis methodologies; a psychometric approach that I believe may produce results that tell little about the specific context of testing. Yet, such a specific context provides information that can explain the reading behaviors of individual examinees by explaining why some examinees can/cannot comprehend the text and successfully complete test tasks.

In summing up these two models, I believe that Weir and Khalifa (2008a) and Embretson and Wetzel (1987)'s models, although they provide useful insights that shape my understanding of how the reading construct can be examined, they nevertheless cannot be solely used in the

context of the present study. Therefore, I propose to design a model that I believe can articulate around the key assumptions of the aforementioned two models, but that includes other key assumptions that are appropriate to the specific context of the present study. I propose to do that in the following section.

### **3.3 Designing a framework for examining the reading construct as measured by the ESE**

In this section, I wish to describe a framework that I have developed and that I propose to use in the present study in order to examine the reading construct as measured by the ESE.

#### **3.3.1 Naming the approach**

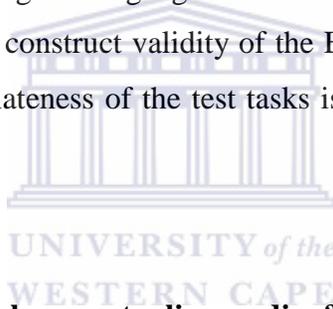
The present approach is a multi-componential processing approach that assumes to describe reading skills and activities in a complex and multi-componential target language domain terms. The model suggests that in reading for testing purpose, the reader engages different constellations of processes in order to comprehend the text and complete the test tasks.

#### **3.3.2 How is the approach specific to the context of the study?**

The present model proposes to signpost the social context of testing. It hypothesize that cognitive processing always occurs within and is significantly affected by the test specific context. Following this hypothesis, the model conceptualizes reading as a situated cognitive activity that takes place in a specific context with specific task demands. Consistent with this conceptualization of reading activity, strategy use depends not only on the reader's processing skills, but also on the knowledge domain and the tasks involved. This is to suggest that, in order to understand an individual's reading behaviors, one needs to consider the specific situations in which reading occurs as well as the reading tasks that readers confront. Therefore, the present model recommends that the test tasks should be described in terms of cognitive processes and contextual parameters.

Furthermore, the present model supports that the test to be used should take into account the social parameters of the examinees such as where and how English, in general and reading, in particular is learnt by examinees. For the specific context of the present study, it is worth mentioning that English is used *almost* only in the instructional milieu; that is, the classroom where formal instruction takes place. In light of this argument, the present model acknowledges that students' exposure to English language is limited and such a limited exposure must be reflected through the design of test tasks.

Moreover, the model supports that in developing the ESE, attention must be paid to the impact of learning supports (appropriate aids and materials), the socio-economic and political environment where the English language instruction takes place (dominated by war, insecurity, unemployment, poor salary, little/no inclination for reading, etc.), and the learners' language skills and their motivation to the English language. The model supports that all these contextual parameters are capable to alter the construct validity of the ESE. In brief, the model purports to ensure that the contextual appropriateness of the test tasks is integrated within the scope of the test construct.



### **3.3.3 How does the present model conceptualize reading?**

The present model conceptualizes reading as both a cognitive and a social process. As a cognitive process, the model views reading as a complex and multi-componential activity where the reader uses various cognitive processes and sub-processes in order to successfully comprehend the text (Hudson *et al.*, 2009; Perfetti, 1985; Biancarosa, & Snow, 2006; Breznitz, 2006). As a social process, the model views reading as socially learned and a set of social practices (Freire & Macedo, 2013) where a reader actively interacts with the text in order to construct its meaning (Bloome, 1985, Knoester, 2010; Nunan, 1991). The social aspect of reading is conceptualized in this approach by looking at the environmental factors that influence reading, the characteristics of reading materials, and the interest and motivation of the readers.

### **3.3.4 How does the present model relate to Weir and Khalifa's model and Embretson and Wetzel's model of examining the reading construct?**

The present model builds from Weir and Khalifa's model by operationalizing reading construct on a four cell-matrix with both careful reading and expeditious reading that can be conducted at both global level and local level (Weir & Khalifa, 2008a). In light of this operationalization, the model proposes to investigate global comprehension by looking at how examinees understand propositions at the level of macrostructure. On the other hand, the model proposes to investigate local comprehension by looking at how examinees understand propositions at the strict level of micro-structure. Furthermore, the model proposes to investigate careful reading by examining the different operations examinees use to extract complete meanings within or beyond sentences right up to the level of the entire text so as to construct the text meaning (Weir & Khalifa, 2008a; Khalifa & Weir, 2009). On the other hand, the model proposes to investigate expeditious reading by examining the different operations examinees use to quickly and selectively access to desired information contained in the text (Weir & Khalifa, 2008a; Khalifa & Weir, 2009).

Furthermore, the present model builds from Embretson and Wetzel (1987)'s model by considering the textual features and item features as two important components that account for the difficulty of a multiple-choice reading test. In light of this consideration, the present model proposes that the text linguistic structure (grammar and vocabulary), the text structural organization and the text density impact on the examinee's ability to process and comprehend it, and to complete test tasks. Furthermore, the model suggests that the phrasing of the test question, the location of information in both the text and the test questions, and the plausibility of item options/alternatives impact on examinee's performance on the test.

### **3.3.5 Content of the model**

- **Main assumption**

The present model assumes that performance on a multiple-choice reading test depends on the extent to which test tasks are appropriate to the test context, the types and levels of

comprehension that are being assessed with test questions, the characteristics of the texts and the characteristics of the test questions.

In light of this assumption, I wish to suggest the following components of the models that I believe need to be considered while investigating the construct validity of a multiple-choice reading test. These components are: the reading construct (or test theory), the test context, the reading types and the processing levels, the text features and the test features.

In light of the main assumption of the model and in order to provide an understanding of the essence of the model, the underlying idea is that, in order to design valid MC reading tests, test constructors need to design appropriate tasks that require the examinees to deploy skills and strategies that reflect the fact that they actually engage with the reading test tasks in the manner the test constructors desire, and that examinees actively work to process and comprehend the text, to understand the expectations of each test question, to understand the meaning and implications of the different item options in the light of the text, and to select and discard options based on the way they understand the text. The following figure 3.4 illustrates the model:

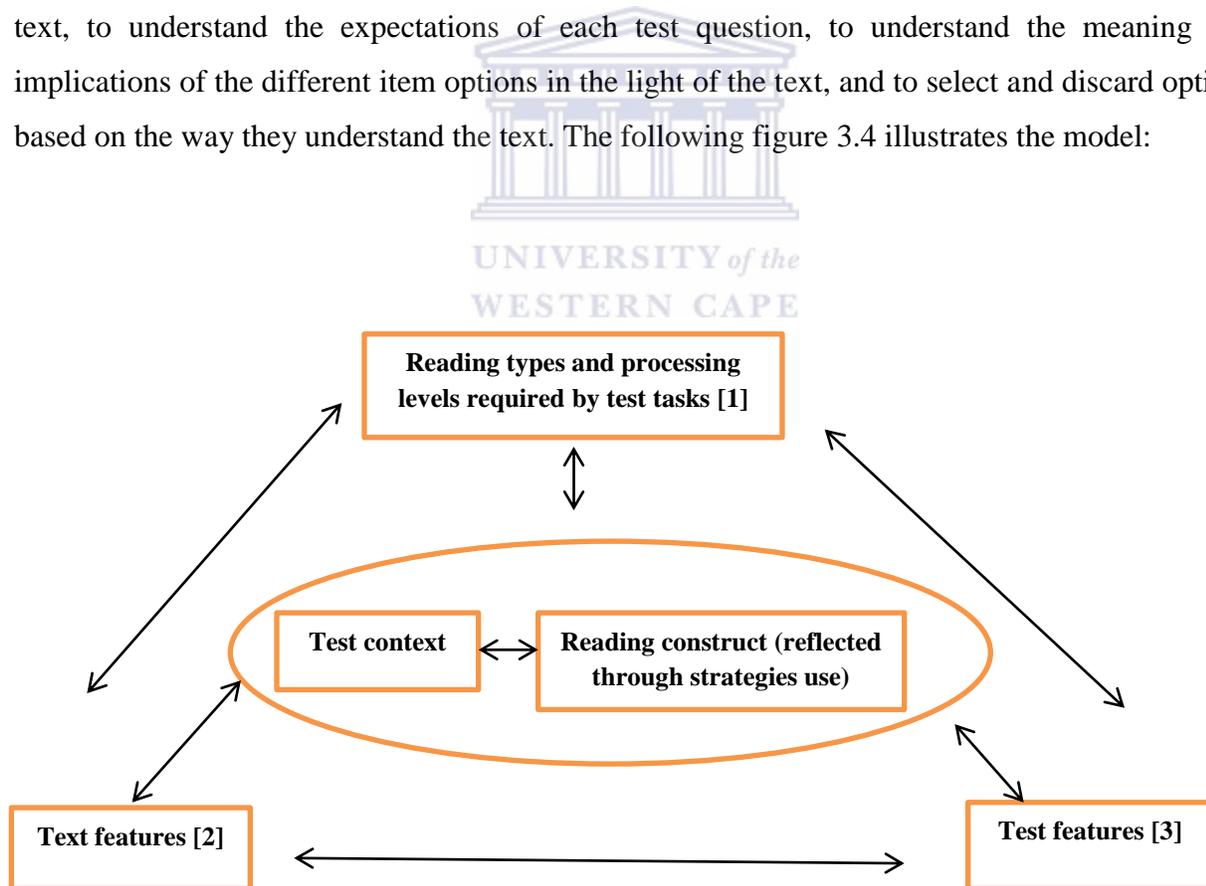


Figure 3.4: A multi-componential processing model of reading performance

The multi-componential model presented in the above figure presents four broad components that interact to account for test performance.

- **The reading construct – Test context component**

The model hypothesizes that reading construct and test context are the core elements in validating reading tests; suggesting that reading construct cannot be examined outside the appropriate context of reading activity. This hypothesis is consistent with my conceptualization of reading as both a cognitive and social activity.

In light of this main hypothesis, the model suggests that performance on reading tests depends on the examinees' capacity to consciously and deliberately use different strategies (cognitive and metacognitive strategies) to appropriately complete the test, a view that has been supported in previous research (Phakiti, 2003). However, the present model proposes that examinees' skills to deploy appropriate strategies to read the text and complete the test tasks depend on the extent to which the test tasks are appropriate to their characteristics. Therefore, when test tasks are relevant to examinees' characteristics, examinees likely deploy appropriate strategies to complete them; hence they better perform on the test.

In keeping with the aforementioned position, the model suggests that appropriate tasks are those that reflect examinees' characteristics (their skills, their motivation for reading, their experience with the test tasks, and their attitudes towards the test), and the conditions under which the reading instruction takes place (reading resources, class size, teacher's characteristics). This proposition is reflected in the interaction between test context and reading construct in the figure.

- **Reading types and processing levels required by the test tasks**

The present model hypothesizes that the examinees' decision about the types of reading (careful reading or expeditious reading; reading at global level or reading at local level) determines their choice of strategies to comprehend the text and complete the test tasks (Katalayi & Sivasubramaniam, 2013). Also, the model acknowledges the impact of processing levels on the

reading construct by proposing that when examinees process the text at different levels, their selection of strategies is therefore affected by the level of processing. This relationship is illustrated in the interaction between component 1 and the central (core) component.

The model also proposes that the intensity and quality of reading comprehension processes that examinees engage in can vary considerably with different levels of reading comprehension assessed by individual test items. Consistent with this proposition, test questions that require examinees to read at global level may be more difficult than those that require them to read at local level; a view that appears to relate to a previous view articulated by Weir and Khalifa (2008a) and Katalayi and Sivasubramaniam (2013). Besides, test questions that require examinees to process the text at a higher level may be more difficult than those that require them to process the text at lower level. This proposition is reflected in the interaction between component 1 and 3.

The model also suggests that the features of the text may determine the examinees' decision to read carefully or expeditiously or both. In light of this suggestion, when texts are dense, examinees may choose to skim them (due to time pressure); whereas examinees may choose to carefully read less dense texts. This suggestion is reflected in the interaction between component 1 and 2.

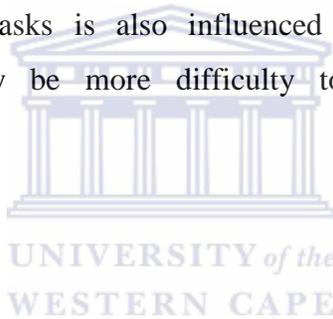
In order to cover the reading construct, the present model suggests that a test must include questions that require the examinees to process the text at the six hierarchical levels. These levels are: lexical encoding, syntactic parsing, establishing propositional meaning, inferencing, building a mental model, and creating a text level representation. This taxonomy appears to concur with Weir and Khalifa (2008a)'s taxonomy of processing levels.

In light of this taxonomy, test items that require lower level processing (lexical encoding, syntactic parsing and establishing propositional meaning) may be perceived to be easier than test items that require higher level processing (inferencing, building a mental model and creating a text level representation).

- **Text features**

The model postulates that the features of the text impact on the examinees' capacity to deploy relevant strategies to process it. Consistent with this postulate, when texts are not appropriate to examinees' characteristics, examinees may have difficulty in comprehending them. This assumption is reflected in the interaction between text features and the core component (test context-reading construct) in the figure.

In keeping with this statement, the model suggests that: (1) the topic and content of the text influence processing difficulty. Following this suggestion, easy texts are those where the examinee can easily activate his/her schemata in order to construct meaning; (2) examinees' use of strategies to construct text meaning and complete the test tasks is also affected by the structural organization of the text; (3) examinees' selection of strategies to construct the text meaning and execute the test tasks is also influenced by text density; suggesting that propositionally dense texts may be more difficult to process and comprehend than propositionally less dense texts.



- **Item features**

The present model acknowledges the impact of item features in the examinees' selection of strategies to construct the text meaning and complete the test task. The model hypothesizes that different features of the multiple-question stem, the correct answer, and the distractors also determine the intensity with which examinees engage different strategies to complete the test tasks. This hypothesis is illustrated in the interaction between component 3 and the core component.

The model hypothesizes that if the relevant information to answer a particular test question is contained in a single short sentence of the text, text mapping will be expected to be easy, suggesting that the examinees are likely to get the item right. Conversely, if the relevant information is spread throughout in the entire text, text mapping will be expected to be difficult; suggesting that examinees might find the item to be difficult. This hypothesis, that is reflected in

the interaction between components 2 and 3 in the figure, appears to chime with Embretson and Wetzel (1987)'s model of reading.

Consistent with the preceding suggestion, the present model hypothesizes that test questions will appear to be difficult to examinees if the alternatives contain many propositions where each of these propositions must be compared to many propositions in the text. This hypothesis appears to be in accordance with Embretson and Wetzel (1987)'s model of reading.

Among the item features that are expected to exert influence on test performance, the model enumerates the structure of the stem, the quality of the key, the number of alternatives and their plausibility and the meanings examinees have on the implicit alternative 6.

### 3.4 Summary of the chapter

In this chapter, I have described an approach for examining the construct validity of the ESE. In the first section, I have provided the rationale to the approach by first outlining existing approaches to examining the validity of reading tests, and then by stressing on the need to design a process-oriented approach to validate the test reading construct. In the second section, I have described two approaches that have provided insights in the development of my approach. The first is Weir and Khalifa (2008)'s cognitive processing approach to examining the reading construct. I have provided a thorough description of this approach and mentioned the reasons why this approach cannot be fully used in the context of the present study. The second is Embretson and Wetzel (1987)'s cognitive processing model of reading comprehension. As with the first model, I have also provided a full description of this model and I have mentioned its limitation in the context of the present study. In the third and last section, I have described my approach by first showing how it is specific to the present study, and mentioning how it conceptualizes reading. Finally I have provided a full description of the content of the approach by pressing its main components. In the following chapter, entitled **Literature Review**, I synthesize relevant studies that have discussed the issues I propose to investigate in this study.

## CHAPTER FOUR

### LITERATURE REVIEW

#### 4.1 Introduction

This study hypothesizes that reading is a cognitive and a social process. Consistent with this hypothesis, reading is viewed as a multi-componential process where a constellation of factors interact in order to impact on reading process and comprehension, thus affecting examinees' performance on reading tests. In light of this claim, the assessment of reading should reflect not only the cognitive and social aspects of reading, but also the complexity and multiplicity of factors and their interaction with the reading construct.

The literature review to this study examines first the studies that have investigated the reading construct by looking at the examinees' use of strategies while completing test tasks; as well as the studies that have evaluated how examinees' performance on reading tests depends on their capacity to deploy appropriate strategies.

Since the proposed model supports that reading construct cannot be examined outside the appropriate context of reading activity, I also propose to look at the social aspect of reading by reviewing studies that have investigated the impact of test context on test performance.

Since the proposed framework also supports that the difficulty in reading is function of both the level of processing required by reading purpose and the complexity of test tasks, the review will also look at studies that have investigated the different types of reading in relation to different processing levels in order to examine how the difficulty of reading tests can be influenced by the complexity of test tasks as well as the processing levels targeted by the reading tasks.

Furthermore, this review will scrutinize studies that have established the impact of textual features (text content, text genres, and propositional density) on processing difficulty; and therefore, on reading performance.

Last, but not least, since the ESE is a multiple-choice test, and as research recognizes that this test format is sensitive to construction procedures, the present review will also focus on studies that have examined the multiple-choice item and how the construction of this format strongly impacts on examinees' choice of strategies and their selection of item answers.

#### **4.2 Reading and test taking strategies and their impact on test performance**

In viewing reading as a strategic activity, the theoretical framework to the present study has highlighted the importance of examinees' use of strategies while constructing text meaning as well as while completing the test tasks.

Research indicates an increased interest in the strategies investigation as part of construct validation of reading tests. As Cohen (2006) notes, studies on test taking strategies have increasingly provided useful insights for the validation of the construct being assessed. These studies (Weir 2005; Hirano, 2008; Cohen & Upton, 2007; for instance) have highlighted the importance of establishing as clearly as possible what examinees are actually doing while taking the test because what a task actually tests is what is central to validity. In light of this argument, Cohen and Upton (2006) believe that the analysis of examinees' strategies provides insights as to how readers interact with the text and how their choice of strategies influences their comprehension of the text and their performance on the test tasks.

A considerable number of studies have indicated that in the area of second or foreign language reading tests, there are certain types of strategies that are used by examinees while taking the test (Cohen & Upton, 2006, 2007; Hirano, 2009). Among these studies, I wish to present a study by Cohen and Upton (2006) because of the insights it provides to this study. Cohen and Upton investigated the reading and test taking strategies used by examinees to complete the reading tasks in the 'Reading' section of the TOEFL. The study aimed to determine if there was a variation in the types of strategies used when answering the different categories of item types. The study participants were non-native speakers of English (Chinese, Japanese, Korean, and "other") and the data were collected through concurrent verbal reports. The analysis of data focused on the patterns of strategy use that characterized the responses for each item type. The main reported finding was that, examinees approached the TOEFL reading section as a test-

taking task. This is to suggest that, they actually read the text, attempted to comprehend it before they could answer the test questions. In light of this finding, Cohen and Upton concluded that the Reading section of the TOEFL is construct-valid as the different strategies used by respondents were consistent with the TOEFL's claims that success to this test requires academic reading ability, that is, a local and a general understanding of the text.

The findings from Cohen and Upton (2006) are relevant to this study. They provide insights for determining which strategies are construct-relevant and which ones are construct-irrelevant while investigating the construct validity of a reading test from examinees' strategies use perspective. Furthermore, Cohen and Upton (2006)'s taxonomy of reading strategies, test taking strategies and item types can provide a useful lead for adapting Weir and Khalifa (2008a)'s taxonomy to the context of the present study.

Research on test taking strategies has primarily aimed at collecting pieces of evidence for the construct validity of the reading test. In these studies, one concern has been the use of appropriate instruments to gather credible data. Two instruments have been frequently used: the verbal reports and the retrospective report questionnaire, a checklist strategy questionnaire used to facilitate the process of collecting the protocols. These two instruments have been used either alone or in combination. Among the studies that have used these instruments, I propose to present an investigation made by Salehi (2011) because of its particular importance in the context of this study. In order to collect pieces of evidence for the construct validity of the reading section of a high-stakes test, Salehi (2011) analyzed the test taking strategies of 40 examinees by using a checklist of strategies consisting of 28 strategies tapping into examinees' behaviors while taking some reading comprehension items. The specific objective of the study was to see whether there was concordance between the type of strategies and the item types in the reading comprehension passages. The main finding reported was that the different strategies employed by the participants as elicited via a check list of strategies spoke to the validity of the test in question. This is to suggest that, the right type of strategies were used for the right type of item types and that the more frequent the proper strategies were used on the right type of item types, the more valid test inferences were. Another finding was that any contributory strategies (such as guessing) could not be validly used for items that were difficult in nature, a finding that replicates a study by Nevo (1989).

The previous findings from Salehi (2011) are relevant to the present study in two respects. First, the use of a checklist of strategies as an instrument to collect data can provide guidance to the design of the strategies questionnaire I propose to use as the main research instrument of this study. Secondly, the analysis of the concordance between the type of strategies and the item types as investigated by Salehi provides an understanding of how examinees make use of specific strategies in order to answer specific types of questions; a concern I propose to investigate in my attempt to answer the second research question.

One important issue that threatens the validity of a multiple-choice reading test pertains to the variability in reading comprehension situations and the format by which reading comprehension is tested. Rupp, Ferne and Choi (2006) are of the view that asking examinees to respond to text passages with MC questions induces response processes that may not be similar to those that examinees would draw on reading in non-testing situations. In light of this view, research has suggested that some examinees tend to use strategies like the elimination of distractors or guessing, and they sometimes may not read and comprehend the text, but still get some test items right (Sheeham & Ginter, 2001; Rupp, Ferne, & Choi, 2006; Ozuru *et al.*, 2008; Katz *et al.*, 1990; Freedle & Kostin, 1993). Picking upon this suggestion, I am tempted to state that reading comprehension appears to be a complex construct and the process of constructing the text macrostructure in non-testing situations is “a complex, fluid and purpose-driven process” (Rupp, Ferne & Choi, 2006, p. 442). A relevant conclusion is that reading for test purpose can impact on the types of strategies readers utilize as readers adjust their strategies and engage in the type of comprehension processes that most suit their reading purpose. Rupp, Ferne and Choi (2006) argue that since test takers aim to get MC questions right, there are likely to select their strategies accordingly to maximize their chance of success.

Having looked at the preceding myriad of studies that provide evidence on how the reading construct is shaped through the use of MC test format, I wish to focus on specific findings from a study by Rupp, Ferne and Choi (2006) because of their relevance to the context of the present study. The goal of Rupp, Ferne and Choi (2006, p. 455)’s study was to “illustrate how reading comprehension is a construct with many fine-grained differentiations and shades that become operationalized differently across passages and questions”. Their participants were 10 students from different linguistic backgrounds to whom they administered the CanTEST, a standardized

English proficiency test used to determine if a candidate is able to meet admission requirements of Canadian university institutions. Data were elicited using verbal reports as the participants were asked to think aloud while responding to the MC reading test. The results indicated that: (a) examinees developed both macro-level and micro-level strategies for responding to test questions. The macro-level strategies (skimming, reading carefully, for example) were utilized to construct the text macro-structure whereas the micro-level strategies (scanning, word matching, highlighting, etc.) were used to respond to individual test questions; (b) while it was expected that examinees could read and understand all the responses choices before they could select the correct choice, it was observed that an answer choice was quickly selected when the examinees perceived the item question to be easy; (c) the examinees drew on their prior knowledge not only to aid in text comprehension, but also to eliminate incorrect answers; (d) among the variables that accounted for the difficulty of test items, the following were of particular relevance: the semantic similarity and plausibility of distractors, the length and the vocabulary of the stem and the choices, and the difficulty of the text.

These findings from Rupp, Ferne and Choi are relevant to this study. They provide useful insights for understanding the types of reading and the related/relevant test taking strategies examinees use in the process of reading the text and answering to test items. Furthermore, Rupp, Ferne and Choi's findings shed light on the two distinct types of strategies examinees use while completing test tasks: those that are construct-relevant and those that are construct-irrelevant. Besides, these findings help understand how the structural format of an item can alter item difficulty, thus accounting for score variations.

In summing up this subsection, I wish to mention that the findings of all the aforementioned studies provide insights for investigating the construct validity of the ESE. These studies help understand what examinees are actually doing while they are taking the test, and this helps to see whether or not the intention of the ESE constructors is realized when students are taking the ESE.

The examinee's capacity to effectively use strategies to complete the test tasks is also influenced by the test context. In the next section, I review studies that have established the impact of contextual factors on reading test performance.

### **4.3 Test context and its impact on reading performance**

The framework I have developed in this study to examine the reading construct hypothesizes that reading construct cannot be examined outside the appropriate context of reading activity; a hypothesis that is consistent with my conceptualization of reading as both a cognitive and a social activity. In keeping with this hypothesis, examinees' capacity to deploy appropriate strategies to read the text and complete the test tasks also depends on the extent to which the test tasks are appropriate to examinees' characteristics as well as the conditions under which reading instruction takes place. In this section, I propose to review studies that have addressed this issue.

#### **4.3.1 Students' characteristics and their impact on reading performance: issues and insights from research**

One of the value-based orientations that inform this study is constructivism. The constructivist epistemology conceptualizes learning as a process of constructing personal meanings in social contexts (Vygotsky, 1978). A related implication of this conceptualization is that learners' emotional investment affects both the learning and the assessment processes. Therefore, an understanding of learner's characteristics and their influence on both learning and assessment is likely to provide a useful ground for understanding how examinees' individual characteristics affect their performance on reading tests.

Although research presents a plethora of individual factors that impact on text comprehension (see Lee & Shute, 2010 for a detailed taxonomy of these factors), student skill and motivation, his/her experience with the test, and his/her attitude towards the test appear to be the characteristics that highly impact on comprehension difficulty and performance on reading tests. This study also considers the importance of these factors in examining the reading construct.

#### 4.3.1.1 Student skills

It is generally agreed that students have different linguistic abilities and this is usually reflected in some students who have the skills to comprehend a text with more ease than others. In reading literature, the issue of the student's skill has generally been addressed in terms of the distinction between two dichotomous concepts: good readers versus poor readers. Although these two concepts are used throughout this study, other synonymous dichotomous concepts are widely used in the literature as well. These include skilled/unskilled readers, proficient/less proficient readers, successful/unsuccessful readers, or fluent/non-fluent readers.

There is substantial literature on good and poor readers. In the scope of this research, I examine the distinction between good and poor readers with respect to three levels: linguistic, cognitive and metacognitive levels. The linguistic level relates to the formal linguistic knowledge (vocabulary, syntax and discourse) a reader needs to comprehend a text. The cognitive level relates to the reader's capacity to recourse to his/her background knowledge and different strategies to comprehend the text. Finally, the metacognitive level relates to the reader's ability to monitor and control reading strategies he/she has deployed.

Research studies that relate to reader language knowledge and its effect on L2 text comprehension have generally pointed to the positive association between these two variables. In a study of higher-level text processing skills in reading comprehension in English as a second language, Nassaji (2003) reported that lower-level processes (word recognition) and higher-level processes (syntax and semantic) could contribute significantly to the distinction between good readers and poor readers. Nassaji found that poor readers were slower in word recognition and weak at rapid and automatic syntactic parsing; a finding that appears to concur with Barnett (1986) when he reported that readers' recall significantly increased in accordance with their vocabulary and syntactic proficiency levels.

Other studies that relate to the impact of the reader language knowledge on reading performance have addressed the issue of inference generation during text processing and comprehension. Hammadou (1991) reported that L2 readers with good linguistic proficiency were better to making inferences and identifying causal structures in the text than L2 readers with low language proficiency; a finding that was later replicated in a study by Lu (1999).

Research studies that relate to the reader's cognitive ability and its effect on L2 text comprehension have recognized that good readers are those who are strategic readers; that is, those who use reading strategies to comprehend the text. Block (1986) investigated the use of reading strategies with L2 readers and found that four characteristics differentiated good readers from poor readers. These characteristics are integration and recognition of different aspects of text structure, use of general knowledge, personal experiences and finally response in extensive versus reflexive modes. Block explains that, in a reflexive mode, readers tend to shift their attention away from text information towards themselves in an affective and personal way. In an extensive mode, on the other hand, readers tend to focus on the author's ideas expressed in the text instead of relating the text to themselves personally and affectively. Therefore, Block argues that good readers react to a text in an extensive mode by integrating information from the text and monitoring their understanding in a consistent and effective way; and this is what poor readers fail to do.

As to the impact of reader metacognitive ability on text processing and comprehension, Pang (2010) argues that comprehension monitoring competence is particularly crucial in L2 context. Since metacognition is the control readers execute on their ability to understand a text, it involves thinking about what one is doing while reading (Block, 1992). Klein, Peterson and Simington (1991) found that good readers attempt the following strategies while reading: identifying the purpose of reading before reading, identifying the text genre, projecting the author's purpose for writing the text (while reading it), choosing scanning or reading in details, and finally making continuous predictions about what will occur next, based on information obtained earlier, and conclusions obtained within the previous stages. Klein, Peterson and Simington also found that good readers attempt to form a summary of what they have read. They concluded that metacognitive knowledge requires good readers to classify information obtained in a text, to sequence it, to establish the relationships between the whole text and its parts, to compare and contrast text details, to determine cause-effect relationships, to summarize ideas contained in the text, to make predictions on the basis of available information, and to draw conclusions on the basis of information contained in the text.

On the other hand, poor readers do not possess knowledge of reading strategies and they are often not aware of how and when to apply the knowledge they possess. For Alderson (2000),

these readers fail to infer meaning from surface level information. Their knowledge about how reading system works is inadequate; thus, they have difficulties to evaluate text for clarity, consistency and plausibility. For these readers, reading includes nothing than verbatim. Yet, following Kintsch and van Dijk (1983) when they describe the theory of text comprehension, the complete reading process goes beyond simple verbatim. Reading process starts from recognizing words until constructing a representation of the meaning of the text. In line with this argument, when someone reads a text, an understanding of the text is created in the reader's mind. Kintsch and van Dijk believe that in the process of reading, a good reader builds three different mental representations of the text: a verbatim representation, a semantic representation and a situational representation. The verbatim representation, also called propositional representation, consists of a list of propositions derived from the text. When a reader processes a complete text sentence, he/she transforms the list of propositions into a network of propositions, and if the text is coherent, all nodes of the network are connected to each other (Kintsch & van Dijk, 1983). The semantic representation describes the meaning of the text that a good reader should get through text processing. Finally, the situational representation relates to the situation to which the text refers.

In addressing the issue of reader skill in the scope of this study, I want to highlight the need to consider the appropriateness of the complexity of test tasks to be included in a reading test. This is to suggest that, a valid reading test needs to also take into account examinees' actual reading skills. In the scope of this study, I intend to investigate this issue by identifying examinees' proficiency levels in English and evaluate the extent to which the complexity of the ESE tasks is proportional to examinees' skills. The main assumption is that, in order to be construct-valid, the ESE needs to include tasks whose degree of complexity also reflects the reading skills of the majority of the examinees.

#### **4.3.1.2 Student motivation for reading**

Student motivation for reading is another variable that has been found to impact on examinee's capacity to process and comprehend a text and successfully complete test tasks (Guthrie *et al.* 2007; Retelsdorf *et al.* 2011; Taboada *et al.* 2009). Guthrie and Wigfield (2000, p. 405) cited in

Guthrie *et al.* (2007), define reading motivation as “the individual’s personal goals, values, and beliefs with regard to topics, processes, and outcomes of reading”.

At this juncture, I wish to outline a distinction that is generally made between intrinsic reading motivation and extrinsic reading motivation. Intrinsic reading motivation relates to the feelings of enjoyment a person gains from performing a reading task; and this kind of motivation is guided by the idea that reading is driven by some interest (Eccles, 2005). Therefore, two types of reading are distinguished in relation to intrinsic reading motivation: reading for enjoyment and reading for interest. In reading for enjoyment, one reads because he/she experiences reading itself as enjoyable. Such a kind of reading generally produces high reading activity. Reading for interest relates to personal topic interest, that is, an individual’s relatively stable orientation towards a certain topic (Schiefele, 1991, cited in Retelsdorf *et al.*, 2011). This is for example the kind of reading conducted by researchers when writing theses or papers. Extrinsic motivation, on the other hand, refers to action due to external values and demands such as competition, the desire to outperform or better perform on a task (Ryan & Deci, 2000, Wigfield & Guthrie, 1997). This is for example the reading students conduct in order to write a test.

Research on reading motivation has generally attempted to establish a positive (or negative) association between reading motivation and text comprehension and reading test performance. Mucherah and Yoder (2008) examined sixth- and eighth-grade students' reading motivation and its relations to their performance on a standardized test of reading. A reading motivation questionnaire was administered to the participants after they had completed the test. Results indicated that students who had high motivation in their reading successfully read challenging materials, while those students who were not highly motivated had difficulty approaching the same materials. Similar findings were reported by Taboada *et al.* (2009) when they found that reading motivation explained a substantial proportion of variance in reading performance, and Guthrie *et al.* (2007) who reported that reading motivation significantly contributed to the prediction of reading comprehension growth.

However, some other research studies reported a negative association between reading motivation and reading and/or test performance. Becker, McElvany and Kortenbruck (2010) found a negative effect of extrinsic reading motivation on reading amount and reading performance. More importantly, Morgan and Fuchs (2007) reported that failure in reading

motivates poor readers to read only when they must do it, and this may result in poor reading skills.

Investigating reading motivation as an isolated factor that accounts for reading performance can shade many factors that relate to reading motivation. In light of this statement, some studies have examined the relationship between reading motivation, perceptions of reading instruction and reading amount. Lau (2009) examined this relationship with Hong Kong secondary school students. As instrument, Lau used the Chinese Reading Motivation Questionnaire (CRMQ) to measure participants' self-efficacy, intrinsic motivation, extrinsic motivation and social motivation in reading. In order to measure students' perceptions of reading instruction, Lau used a modified version of the Survey of the Classroom Goal Structure (SCGS), and in order to measure reading amount, he used the Reading Activity Inventory (RAI). After the analysis of data using multiple correlations, Lau found that: (a) students' intrinsic motivation was mostly strongly related to their reading amount; suggesting that the more one reads the more motivated one becomes; (b) students' perceptions of reading instruction they had received in language class significantly related to their motivation; suggesting that those students with positive perceptions of reading instruction are mainly those who are motivated in reading; (c) reading amount was not correlated to reading motivation, but it was indirectly correlated to students' perceptions of reading instruction; suggesting that those students who are reading more in classroom tend to have positive perceptions of reading instruction.

These studies on reader's motivation assume particular importance in the scope of this study. They provide insights for understanding that successful completion of reading tasks may also depend on the examinee's degree of motivation for reading. From the methodological perspective, these studies provide guidance to operationalize the construct of motivation by designing appropriate statements that are expected to elicit participants' motivation for reading. Furthermore, considering the context of the DR Congo where English is a foreign language that is used as a subject only at school, and facing the multiple challenges relating to reading instruction, students' motivation to read in English might be very low. Therefore, the ESE developers need to ensure that the test includes tasks that also reflect the level of students' motivation for reading.

#### **4.3.1.3 Student's attitudes towards the test and its impact on test performance**

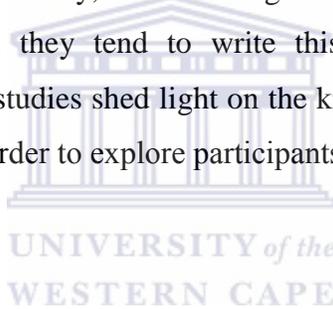
One measure of the impact of high-stakes assessments is the attitude that the students hold towards the assessment. Although this issue seems not to be extensively investigated, it has been suggested that positive attitudes towards the test can produce beneficial effects while negative attitudes can erode confidence and potentially impact negatively on performance (Murray, Riazi & Cross, 2012; Rasti, 2009; Han, Dai & Yang, 2004).

There is a paucity of studies that have addressed the issue of the impact of students' attitudes on test performance. Most of these few studies have reported a high incidence of negative attitudes although these studies could not establish a relationship of these negative attitudes and students' test performance. Han, Dai and Yang (2004) found that the majority of their participants had negative attitudes to the CET (China English Test), a finding that was later replicated by Gu (2005) in a study that included administrators, teachers and students.

However, one study attempted to establish a relationship between attitudes towards the test and performance on the test. Murray, Riazi and Cross (2012) investigated participants' attitudes by exploring their opinions, beliefs and feelings on a professional gate-keeping test. Their objective was to determine whether or not there was a correlation between examinees' attitudes towards the test and the examinees' demographics and experiential factors. The participants were 105 candidates who were enrolled in a preparation course for the Professional English Assessment for Teachers, a test that aims to determine the English proficiency level of oversea trained teachers wishing to teach in Australian state schools. The participants were asked to complete a written survey questionnaire that comprised three sections: the first section included questions that sought to determine the nature of participants' attitudes towards the test. The second section included questions that aimed to explore the relationship of attitudes and the demographic data, whereas the third section included questions that intended to identify participants' perceptions of sources of their attitudes. The results indicated a slight predominance of negative attitudes, particularly among candidates who had unsuccessfully attempted the test. The other finding was that the main reported sources which correlated with negative attitudes were personal experiences and feelings as well as the impact of other people, notably teachers and other candidates.

However, some other studies did not report negative attitudes to the test. Rasti (2009) investigated the attitudes of Iranian candidates who had attempted IELTS, a large scale test used for admission at British universities. His objective was to look at the relationships between examinees' characteristics (sex, age and educational background) and their attitudes towards the test. As instrument, Riazi used a questionnaire and interviews while correlations were used to determine different relationships among variables. Rasti found that 80 % of the 60 participants had overall positive attitudes towards the IELTS. However, he did not find a significant relationship between participants' attitudes and individual characteristics; a finding that suggests that the IELTS scores might not have affected participants' attitudes.

The findings reported in these studies are relevant to the present study. They provide insights for understanding why students' attitudes towards the test can also be considered while developing language tests. In the scope of this study, these findings also suggest that when students have positive attitudes towards a test, they tend to write this test with confidence. From the methodological perspective, these studies shed light on the kinds of statements I need to include in the contextual questionnaire in order to explore participants' attitudes towards the ESE.



#### **4.3.2 Social-contextual factors**

Social-contextual factors involve those factors that originate from outside the student but that influence her/him in the successful completion of the test task (Lee & Shute, 2010). Among the myriad of social-contextual factors that exert an influence on examinees' performance on test as outlined by Lee and Shute (2010), I wish to mention only those that I find are closely related to successful reading. These factors are: (a) the school climate which is the organizational characteristics of the school under which the students learn and prepare for the test (such as reading resource supports, policies adopted by the school in terms of discipline, teachers' collaboration, demographics of the class size etc.), (b) the teacher's qualification, his/her motivation and commitment for teaching, how he/she interacts with learners and the kind of support system provided to him/her; (c) the curriculum content and quality, (d) parents' involvement in school activities, their attitudes towards education, their assistance with child's classwork. All these factors and many others not listed here have been found to strongly impact

on learning; therefore test performance (cf. Goddard *et al.* 2000; Hoy *et al.*, 2002; Hill & Craft, 2003; Taylor, Clayton, & Rowley, 2004; for example). In the scope of this study, I wish to review some studies that have addressed some of these issues.

#### **4.3.2.1 Teacher quality, qualification and experience**

One of the issues related to students' achievement is the identification of specific teacher characteristics that predict effectiveness. This issue is usually reflected in education policies that usually state the qualities and qualifications to promote in teachers' recruitment. Research provides some evidence that teacher quality is one factor that also predicts student's achievement.

In education literature, teacher quality has usually been defined in terms of a combination of observable and unobservable characteristics (Pelayo & Brewer, 2010; Goldhaber & Anthony, 2007). Among the observable teacher characteristics, Pelayo and Brewer mention the degree level and the years of experience; two variables that are also typically the basis for teacher salary in many education institutions. Besides these two factors, Pelayo and Brewer mention some more refined measures such as the degree major (a measure generally used in attempt to determine the importance of content knowledge or pedagogical knowledge), the quality of the university or college attended by the teacher or again the teacher's results on tests of teacher ability. Unobservable characteristics, on the other hand, refer to characteristics not easily measured. Among these, Pelayo and Brewer mention pedagogy used by the teacher, his/her classroom management skills, or the philosophy of teaching. Research indicates that 97 % of teacher effects are attributable to unobservable characteristics (Pelayo & Brewer, 2010; Goldhaber & Anthony, 2007).

There is evidence of a positive relationship between teacher quality and students' performance. Darling-Hammond (2000) equated teacher quality with specific qualifications and their expected effects on students' performance and found that the effect of well-prepared teachers on student achievement could outweigh student background factors such as poverty, language background and minority status. In an earlier study, Rivkin, Hanushek and Kain (1998) reported similar

conclusions although they described teacher quality in terms of overall differences between teachers in students' gains.

The notion of teacher qualification has usually been addressed in terms of teacher certification. Studies that have attempted to establish a relationship between teacher certification and student performance have come up with conflicting results. While Goldhaber and Brewer (2000) found that students of teachers with standard certification had higher achievement than students of teachers who were not credentialed in their subject area, Glazerman *et al.* (2006) did not find such a relationship.

Some studies have addressed the issue of teacher qualification in terms of the degree level and subject matter preparation. While the majority of these studies have suggested that teacher degree level and subject matter preparation do not significantly impact on student achievement (Hanushek, 1994; Hanushek, *et al.*, 2005; Rivkin, *et al.*, 2005), Goldhaber and Brewer (2000) found a positive relationship between teacher degree level and subject matter knowledge.

As to the relationship between teacher experience and student achievement, many education studies that have addressed this issue did not reach similar conclusions. Hanushek (1986) and Rivkin, *et al.* (2005) found that a teacher's years of experience positively impacted on student achievement although they acknowledged that beginning teachers generally perform worse than more experienced teachers. However, Greenwald, *et al.* (1996) and Eide, *et al.* (2004) found that teacher experience was related to students' achievement.

The findings from these studies on teacher quality, qualification and experience are relevant to this study. Since in DR Congo, English language is learnt at school and its use is limited to classroom settings, it sounds reasonable to investigate the quality and qualification of those who teach this language and eventually examine how these two variables can also impact on reading instruction as well as on reading test performance. Besides, the methodological suggestions offered in these studies provide guidance in the design of the contextual questionnaire that I propose to administer to English language teachers in order to elicit data relating to teacher characteristics.

#### **4.3.2.2 Teacher's in-service training and motivation for teaching**

There seems to be a consensus that in-service training programs for the teachers tend to increase the qualities possessed by good teachers; hence these programs positively affect teachers' performance. Harris and Sass (2001) studied the effects of teachers' in-service training programs on the teacher quality and concluded that teachers' in-service training was positively related to teacher productivity.

Similarly, Samupwa (2008) found that teachers who were regularly subjected to in-service trainings showed significant changes in behavior in both classroom and administrative work. Samupwa concluded that in-service trainings had to be considered as one of the indicators of good teacher and the roles and capacities of teachers could be improved through teacher regular participation in training sessions.

Among other studies that have suggested that in-service training is a contributory factor influencing teacher's performance, a study by Jahangir, Saheen and Kazmi (2012) indicated that teachers' in-service training brought significant changes in their perceptions of the teaching profession. This is to suggest that, some teachers with negative perceptions of teaching profession changed their views after they had participated to different training sessions. Jahangir, Saheen and Kazmi also reported significant increase in positive behavior (personality, knowledge, and commitment skills) on the part of these teachers; a finding that appears to concord with Conco (2004)'s finding.

As to teacher motivation for teaching, some studies have revealed a positive association between high motivation and teacher commitment and a positive relationship between teacher motivation and student achievement. Bishay (1996) found that teachers with high motivation for teaching had students whose self-esteem was high. Among factors that are believed to promote teacher motivation, Sylvia and Hutchinson (1985) found that pay incentives slightly related to teacher motivation while those factors that derived from the satisfaction of higher-order needs (social relations, esteem, and actualization) strongly related to teacher motivation.

These findings are relevant to this study. In order to develop valid reading tests, the test should include tasks that also reflect the characteristics of the majority of the English teachers. If we can agree that teacher participation in in-service training and his/her motivation for teaching add

positive value to teacher quality, we can also acknowledge the difficulty students may have to read in a context where their teachers are not exposed to any in-service training programs; and that due to socio-economic challenges, teachers motivation for teaching is low. This assumption justifies my need to explore these two variables in order to understand the context of the ESE. Moreover, from the methodological perspective, these studies provide light in the design of the contextual questionnaire to be administered to English teachers.

### **4.3.3 Conditions under which the teaching and learning of reading take place**

In English as a foreign language learning where instruction takes place essentially in the classroom setting, the conditions under which reading activity is conducted play a crucial role in determining the efficacy of reading. Among these conditions, I wish to present studies that have addressed the issue of the availability and quality of reading resources.

Reading activity supports an interaction between a reader and a text. The presence of a text implies the issue of availability and quality of reading resources. This issue is relevant especially in developing countries where budgets and donors' constraints sometimes limit the possibility to access to reading resources that are relevant to students' needs.

The main reading resource in developing world is the textbook. Textbooks generally aim to provide an orderly introduction to a discipline or subject area that may assist the students to learn independently (Crossley & Murby, 1994). The issue of the development and provision of textbooks appears to be one of the most investigated issues in education research conducted in developing countries. Most of the studies that have addressed this issue have drawn attention to the economic and social imperatives of strategies designed to improve the quality of teaching and learning (Crossley & Murby, 1994; Lockheed & Verspoor, 1991). Crossley and Murby (1994) argue that access to reading materials is limited in many African countries, and these reading materials may be of poor quality. This view is echoed by DFID (2010, p. 17) when it states that: "It is still relatively rare to find countries, particularly in Africa, the Indian sub-continent and the countries of the former Soviet Union, where textbooks are made available regularly, reliably, predictably and in sufficient quantities to meet curriculum requirements".

In the difficult economic crisis of many developing countries where they are sometimes subjected to budgetary restraints by lending agencies (such as the World Bank, the International Monetary Funds), the provision of more and better quality textbooks has emerged as an issue that has been addressed by researchers. Jones (1992) and Lewin (1987) reported that lending agencies such as the World Bank influence many countries' policies in terms of textbook design, quality and content. On their part, Heyneman, *et al.* (1981) reported that since textbook provision is generally conducted by international agencies, their contents may sometimes be unrelated to students' needs and backgrounds; a view that appears to concur with Crossley and Murby (1994) when they found that in many developing countries, textbooks were in poor quality and in short supply.

Consistent with the previous viewpoint, the issue of textbook development suggests that the process of textbook writing is much more than a writing process; instead, it involves professional skills and processes that may be a challenge in developing countries due to the abject paucity of outstanding forms of independent publishing infrastructure. As a result, there is for many developing countries a tendency to rely upon former colonial powers for the supply of textbooks, as the case of the DR Congo where most textbooks used for teaching English (Cartledge series, English for Africa, Today's English, Go for English) are designed and published by Belgium editors and publishers. There is evidence in the literature that has mentioned the dangers of cultural imperialism and continued economic dependency generated by such reliance (Altbach & Kelly, 1988; Scrase, 1992, for example). In case of reading, one way to solve the problem of textbook content would be to adapt existing textbooks in order to fit the students' cultural and social background characteristics. Crossley and Murby (1994) are of the view that such a process of adaptation can be faster, easier and cheaper. However, I believe that such an adaptation process can appear to be a tough enterprise if it is conducted by people not so much skilled to beneficially influence both textbooks' content and approach.

Research confirms the central role of an adequate supply of good quality textbooks in improving student performance (Bukonya, 2005; Kalibala, 1999). In a study that aimed to investigate the levels of competency in literacy of primary school pupils in Uganda, Bukonya concluded that the impact of textbooks is greatest in the poorest countries; and that the provision of textbooks of good quality can counter-balance the problems of poorly trained teachers and lack of basic

facilities in schools. This conclusion appears to chime with Lewin and Stuart's (2003, p. 73) view when they mentioned that "The impact of textbooks is greatest in the poorest countries where teacher quality may be low and where facilities and resources are scarce and generally of poor quality".

Research evidence substantiates a positive relationship between textbooks quality and availability and student achievement in reading. Greaney (1996) found a strong relationship between classroom collections of primary reading books and student performance on reading tasks. This finding appears to concord with Elkin and Lonsdale (1997)'s finding on reading achievement when they reported that students who were in schools with a library outperformed students who were enrolled in schools without a library.

My aim in addressing the issue of availability and quality of reading resources in this study is to augment the understanding of the actual context of the ESE. I hypothesize that, when the reading resources are available, and that they are of good quality, students can be motivated to read both inside and outside the classroom; therefore, their performance in reading can be enhanced. In light of this hypothesis, I am inclined to state that, in order to ensure its validity, the ESE must include tasks that also reflect the conditions under which reading instruction is conducted in terms of availability, quality and relevance of reading resources. In the next section, I propose to review studies that have addressed the issue of the variety of complexity of reading types examinees conduct while reading, and the cognitive demands imposed by the complexity of test tasks.

#### **4.4 Reading types and processing levels as required by the test tasks**

The model I have developed in this study conceptualizes reading construct, like Weir and Khalifa (2008a)'s model, in a four-cell matrix with both expeditious reading versus careful reading on the one hand, and local reading versus global reading on the other hand. The model also recognizes a hierarchy of processing levels that reflect the cognitive demands imposed by the test tasks. These hierarchical levels are lexical encoding, syntactic parsing, establishing propositional meaning, inferencing, building a mental model and creating a text level structure.

In light of this proposed model, there is a scarcity of studies that have investigated the variety of complexity of reading types examinees conduct while reading, and the cognitive demands imposed by the complexity of test tasks. To date, the few studies that have addressed this issue are mainly the result of investigations made by Weir and colleagues (Weir & Khalifa, 2008b, Weir, *et al.*, 2008a; Weir, *et al.*, 2008b) that have been essentially conducted in order to test their model. Nevertheless, two other studies that have adopted this four-cell matrix taxonomy of reading construct include a study by Sarojani and Krishnan (2011) and a study by Katalayi and Sivasubramaniam (2013).

In adopting Weir and Khalifa's model of reading comprehension, test questions are generally classified on the basis of whether the examinee has to read the text carefully or expeditiously at global level or local level in order to answer individual test questions (Weir & Khalifa, 2008b, Weir, *et al.*, 2008a; Weir, *et al.*, 2008b; & Sarojani & Krishnan, 2011; Katalayi & Sivasubramaniam, 2013).

These studies have shown differences in the use of strategies between test items that require examinees to read at global level and those that require examinees to read at local level. They have indicated that test items that target reading at local level require examinees to focus on text micro-propositions and quickly search for text information through a browsing of some parts of the text (Weir, *et al.*, 2008a; Weir, *et al.*, 2008b; & Sarojani & Krishnan, 2011). On the contrary, test items that target reading at global level require examinees to read the text at global level by focusing on text macro-propositions and by attempting to generate more connections between their knowledge and text information (Weir, *et al.*, 2008a; Weir, *et al.*, 2008b; & Sarojani & Krishnan, 2011). Furthermore, these studies have shown differences in the use of strategies between test items that require a careful reading of the text and those test questions that require an expeditious reading of the text. Katalayi and Sivasubramaniam (2013) have reported that, careful reading items appear to require a frequent combination of strategies for reading at both global and local level; while expeditious reading items appear to require less frequent combination of strategies for reading at both global and local level.

Validation studies that have used Weir and Khalifa's (2008a) framework have generally sought to determine whether the test mainly includes tasks that require the examinees to read the text carefully or expeditiously.

In an attempt to test their model, Weir and Khalifa (2008b) applied Weir and Khalifa (2008a)'s cognitive processing model to the Main Suite Reading papers, a component of Cambridge ESOL (English for Speakers of Other Languages) examinations. The study aimed to determine the variety and degree of complexity of the reading types required by the test tasks and the cognitive demands imposed by these tasks. This would determine whether the cognitive processes elicited by the test resemble those expected from the reader in non-testing situations. The main finding was that Main Suite Reading papers seem to follow the order of difficulty in cognitive processing as suggested in the model. Some of the specific findings reported include: (a) regardless the examinees' L1, the decoding processes relied upon recognizing not only letters but letter clusters and whole words; (b) the effort of decoding made considerable cognitive demands on the less skilled readers as examinees often failed to employ comprehension processes; (c) textually implicit questions appeared to be difficult for examinees as these questions required examinees to find the answer at text level structure whereas textually explicit questions appeared to be easy as examinees had to find information in a single sentence; (d) certain inferencing questions required the examinees not only to report on information from the passage, but also on what that information entailed; (e) only skilled readers were able to cope with questions requiring them to make an overall text representation. These findings appear to concord with findings reported in a previous study by Oakhill and Garnham (1988) when they found that less skilled readers could hardly cope with questions that targeted implicitly stated information because of their failure to make inferences in comprehension. On the basis of the range of reading types and cognitive demands required by the task as measured by the Main Suite, Weir and Khalifa (2008b) concluded that Main Suite Reading papers is construct-valid although their subjects could hardly well perform on the test. These findings are relevant to the present study. They provide useful insights to understand how text complexity and task demands are crucial in developing a valid comprehension test. More specifically, Weir and Khalifa (2008b)'s findings suggest that task complexity can be function of the processing level targeted by test questions. In light of this suggestion, in order to be construct-valid, one would expect the ESE to include test questions that mainly target the processing level(s) that corresponds to the examinees' individual characteristics.

In order to provide grounded insights into the congruence between the construct measured by a reading test and academic reading practices, Weir, *et al.*, (2008a) used a retrospective

questionnaire to identify the range of cognitive processes that students employ when they perform the various tasks in an IELTS reading test, a test used to assess the English language proficiency of foreign students wanting to study in Great Britain universities. Participants included 352 students to whom one of the six IELTS test parts was administered and the students were instructed to fill in the retrospective questionnaire immediately after answering each individual test question. After the descriptive statistics have been generated, the results indicated that: (a) although for university students expeditious skills and strategies are just as critical for academic study, the data indicated that the major focus of the IELTS was on careful reading; (b) for most participants across the different task types, expeditious reading (the majority of respondents chose to read the text quickly and selectively before reading the question) was the type of reading chosen to answer the questions. These two findings suggest that expeditious reading and careful reading appear to be integrated reading types. Other findings were: (c) although there was evidence that for some test items, the answers could be found within one sentence, the majority of respondents reported finding the information necessary to respond to questions by putting information together across sentences; and (d) there was evidence of relationship between the selection of certain strategies and success on various items although this relationship was not straightforward. On the basis of these findings, Weir and colleagues conclude that the IELTS is construct-valid because the picture of reading in response to IELTS test items “is consistent with the general approach to academic reading” (Weir *et al.*, 2008a, p. 178).

In a similar study, Weir, *et al.* (2008b) conducted a study on the relationship between the reading activities in IELTS test and the reading experiences of students in their first year at university. They asked their expert judges to apply a descriptive framework of expeditious and careful reading strategies to each item in each testlet. They reported that the major focus of IELTS test was on careful reading, yet the survey data conducted indicated that for academic study, university students needed expeditious skills and strategies. They also found that individual characteristics of IELTS texts did not match those typically identified with academic texts.

In a replicative study, Sarojani and Krishnan (2011) used a combination of qualitative and quantitative methodologies to investigate whether or not the strategies employed in responding to recent IELTS reading tests can validate the reading construct, and also whether or not the IELTS

reading items test reading ability comprehensively. Their subjects consisted of two examinees to whom they administered the 14 IELTS reading test under testing conditions. Sarojani and Krishnan found that there was an imbalance of item distribution in the sense that the majority of test items targeted careful reading and that expeditious reading was less targeted by the test items. They concluded that the IELTS test overemphasizes careful reading; yet expeditious reading is crucial in academic reading. Sarojani and Krishnan also found that the majority of test items focused on basic comprehension at local level (within the sentence), while items that target constructing the overall text representation were almost inexistent, suggesting that the IELTS includes few items that require higher level processing. As a conclusion, Sarojani and Krishnan (2008, p. 31) argue that the IELTS “may not test comprehensively the reading ability, hence their product may not reflect the actual ability of the students”.

In order to provide an understanding of the strong relationship between the test construct and test context in validating reading tests, Katalayi and Sivasubramaniam (2013) used Weir and Khalifa (2008a)’s model of reading construct to examine the task complexity of 50 reading test items from the English state examination, a multiple-choice reading test used in the DR Congo for certification. They found that the test included more test items that targeted careful reading than those test items that targeted expeditious reading. They concluded that the ESE appear to have weak construct validity as it included tasks that were less appropriate to the test actual context.

The aforementioned studies assume particular relevance in the scope of the present study. They provide a useful argument on the importance of examining the validity of a reading test through an investigation of the different types of reading and the different processing levels in order to understand how the difficulty of the ESE can also be function of the complexity of ESE test tasks as well as the processing levels required by the ESE tasks. Since at present these studies are limited in quantity, the present study aims to augment the evidence on the benefit of validating reading construct through a four-cell matrix model of reading construct. In the following section, I propose to review studies that have suggested the impact of text structure on text comprehension and test performance.

## **4.5 Text structure and its impacts on comprehension and test performance**

Text comprehension is an interaction between a reader and a text. The multi-componential model of reading I have proposed in this study hypothesizes that text variables are factors that impact on processing and comprehension difficulty, and therefore affect reading performance. In this section, I scrutinize the studies that have investigated the impact of textual features on text comprehension and test performance. My motivation to consider textual features as a highly critical aspect in examining the reading construct is justified by my belief that, in order to develop construct-valid reading tests, test developers should provide sufficient evidence that establishes that the test questions actually measure examinees' comprehension of the passage and their ability to think and reason about the content of the passage.

Research on text comprehension indicates different textual features that potentially affect text comprehension and test performance (Kintsch, 1988, 1998, 2004; van Dijk & Kintsch, 1983; Graesser *et al.*, 1991; Graesser *et al.*, 2004). However, in the scope of this study, I propose to review studies that have investigated the text variables that most likely impact on comprehension difficulty and performance on reading test. These text variables include: the text content, the text genres and the propositional density. In the following subsection, I review studies on the text content and its impact on comprehension and reading performance.

### **4.5.1 Text content and its impact on comprehension**

The reader's knowledge of the text content plays a crucial role for successful comprehension. The construction-integration model that is used as theoretical framework in the present study acknowledges the role of the reader's knowledge in text comprehension. One main assumption is that readers generally use their prior knowledge to integrate meanings of individual sentences into a coherent representation of events contained in the text in order to construct the text macrostructure (Kintsch, 1988, 1998, 2004; van Dijk & Kintsch, 1983). In light of this assumption, easy texts appear to be those where the reader can easily activate his/her schemata in order to construct meaning whereas difficult texts appear to be those where the reader has problems to activate his/her schemata. This view appears to chime with Best *et al.* (2006) when they found that the readers with high level of background knowledge comprehended the text

better than readers with low level of background knowledge; a finding that also appears to concur with Snow (2002) who reported that readers' difficulty to comprehend expository text was a result of their lack of background knowledge needed to process the texts.

Indeed, some researchers have addressed the issue of text content and its impact on comprehension difficulty in terms of reader's familiarity with text content. Such a familiarity with text content has usually been examined in terms of reader's capacity to comprehend discipline-related/unrelated texts or texts with contents that relate to the reader's background. It was generally predicted that when readers read materials that relate to the situations that are familiar to them or to their subject-knowledge, or subject discipline, they more easily comprehend these materials than when they read materials that are not related to their subject-knowledge (Hock, 1990; Yousif & Shumaimeri, 2006; Chang, 2006).

The issue of text content familiarity has also been addressed in terms of authenticity of reading materials. Some researchers suggest that more emphasis should be placed on the use of authentic texts because these texts reflect a rich source of linguistic input for the students (Devitt, 1997; Bacon & Finnemann, 1990; for example). These authors are of the view that the use of authentic materials in testing contexts suggests that text comprehension should not be viewed as a function of reader's capacity to understand single words, sentences and paragraphs to build the text macrostructure; rather, it should be viewed as a process of developing strategies for selecting and identifying multiple verbal and non-verbal cues. Therefore, these strategies reflect the use of language in natural and authentic ways.

At this juncture, I wish to raise a concern that is linked to the use of authentic materials in reading assessments. This concern relates to the place of literature in foreign language teaching and testing programs. I wish to mention that, many researchers have acknowledged the linguistic, academic, intellectual, and cultural benefits of the use of literary texts in both teaching and testing situations (Widdowson, 1983; Swaffer, 1985; Sivasubramaniam, 2004). The interest in the use of literary texts as recommended by these researchers can be perceived as a result of the need for resources that take foreign language readers beyond the elementary level to a level which can enable them to function in the target language contexts. Viewed from this perspective, the use of literary texts in testing students' comprehension gives examinees a real opportunity to contrast the cultural expectations of situations and events in their own culture with the cultural

schema of the target language. In support of this view, a study by Atai and Soleimany (2009) which used literary texts to investigate the effect of text authenticity on the reading performance indicated that subjects who were given literary texts performed better than those who were given non- authentic texts.

However, in reading as a foreign language context, like the context of the present study, the use of literary texts in teaching situations is interesting, but its use in testing situations can be a source of serious problems for many students since these students may be confronted with the most obvious problem of vocabulary. This view finds support in a study by Ostyn and Godin (1985), when they found a significant difference between the vocabulary learned by foreign language readers and that which is required for reading a literary text. Therefore, although I acknowledge the richness of literary texts in foreign language reading contexts, I believe that their use in testing situations should be taken with much caution. I advise that a prior thoughtful study of examinees' contextual characteristics must serve as an indicator of whether or not to use literary texts in testing.

The issue of text content and its impact on text comprehension has also been investigated in terms of text vocabulary and grammar. It has been generally agreed that the vocabulary and grammar of a text exert influence on reader's capacity to comprehend the text. This is to suggest that texts with vocabulary that is not familiar to readers and that include sentences readers find difficult to parse are potentially difficult to be comprehended.

However, the main issue that relates to the impact of vocabulary and grammar of the text on text content is the issue of relative contribution of vocabulary breadth and syntactic knowledge on reading comprehension. Research has consistently attempted to address this issue (Shiotsu & Weir, 2007; Yalin & Wei, 2011). At present, there is an inconclusive debate on this issue. Some studies have claimed that L2 reading requires little syntactic processing but much lexical-conceptual processing (Ulijn, 1984; Brisbois, 1995; Yamachita, 1999), while others have reported that syntactic processing is more useful for text comprehension than vocabulary breadth (Alderson, 1993; Shiotsu & Weir, 2007; Yalin & Wei, 2011). Still, other studies have established that both vocabulary breadth and syntactic knowledge equally contribute to text processing and comprehension (Barnett, 1986; Bosser, 1992).

There is empirical evidence on the great contribution of vocabulary to text comprehension. Brisbois (1995) separated grammar and vocabulary as independent predictor variables and she found that that her vocabulary measure correlated more strongly with her reading comprehension measure than did her grammar measure. This finding appears to replicate a finding in an earlier study by Haynes and Carr (1990). Furthermore, in a study using regression analysis methodology, Yamachita (1999) found a greater contribution of vocabulary knowledge to reading comprehension performance than the contribution of grammar knowledge.

Other studies have found that both grammar knowledge and vocabulary knowledge contribute significantly to text comprehension and reading performance. Barnett (1986) reported that both syntactic knowledge and vocabulary knowledge affect reading comprehension in almost the same way; a conclusion that was also reached in a study by Bosser (1992) when he concluded that both grammar and vocabulary were significant predictors of reading comprehension performance.

However, other studies have come out with contrasting conclusions. These studies reported that grammar was a better predictor of reading comprehension performance than could be vocabulary. Shiotsu and Weir (2007) investigated the significance of syntactic knowledge and vocabulary knowledge through three successive studies employing separate measures of vocabulary knowledge and syntactic knowledge. Using the structural equation modeling of reading items to analyze data, they found that although both syntactic knowledge and vocabulary knowledge were contributing factors to the prediction of reading comprehension performance, the syntactic knowledge was a better predictor of text comprehension than vocabulary knowledge. This conclusion was replicated by Yalin and Wei (2011) who found a positive linear correlation between the students' performance on reading comprehension test and their vocabulary breadth and syntactic knowledge, with the syntactic knowledge outperforming vocabulary breadth in predictive power.

The above studies on the significance of vocabulary and grammar on text processing and reading test performance are relevant to this study. Beyond their contradictory findings, one conclusion I can offer is that both vocabulary knowledge and syntactic knowledge are relevant to the text comprehension and these two variables account for item difficulty.

In addressing the issue of text content familiarity and difficulty in the scope of this study, I intend to investigate the validity of the ESE by also examining the extent to which the contents of the ESE texts are familiar to the examinees; and the extent to which the text grammar and vocabulary is proportional to examinees' linguistic knowledge. I hypothesize that when examinees are given familiar texts whose vocabulary and grammar appear to relate to their linguistic knowledge, they tend to easily comprehend such texts; therefore, they are likely to better perform on the test.

Text content is not enough to account for readers' comprehension difficulties; one other variable that relates to text structure and that can affect comprehension is the reader's knowledge of text genres. In the following section, I review some studies that have addressed this issue.

#### **4.5.2 Text genres**

Text genre is another variable that has been found to impact on reading comprehension. Cohen and Upton (2007) classify texts into narrative, argumentative or expository genres; and each of these genres has at least one or more major structural organizations, such as description, classification, comparison/contrast, cause/effect, and problem/solution, with the information presented from more than one perspective or point of view.

Whether one reads at global level or local level, constructing text meaning is likely to be affected by text genre. Alidib (2004) argues that text genre functions as textual schemata; therefore, it limits the meaning-potential of a particular text. This is to suggest that, the knowledge of text genre can help a reader to construct the text macrostructure. Supporters of genre theory (see Fowler, 1989; Alidib, 2004; for example) claim that genre provides an important frame of reference that helps readers to identify, select and interpret texts. Viewed from this genre theory perspective, readers need some exposure to a variety of text genres in order to familiarize themselves with, and be able to recognize the shared features of each genre (Fowler, 1989; Alidib, 2004).

Researchers have been interested in understanding how different text genres may be associated to processing and comprehension difficulty. Alidib (2004) investigated the effects of text genre

on French language reading comprehension at two levels of college French instruction. His participants consisted of college students enrolled in the second quarter of elementary French class and the second quarter of intermediate French class at The Ohio State University. The participants were rated based on their level of academic performance in their respective levels of French study. Two excerpts from twentieth century French literature were given as reading task to the two groups. Data were analyzed using the analysis of variance (ANOVA) procedure with some design variables as whether the text genre was a play or a novel, the participants' level of instruction and the teacher's rating. The results from the two-way ANOVA revealed a significant effect for text genre and level of class on reading performance. Similar results were reported by Brantmeier (2005) when she found a significant effect of genre on the cloze-test performances.

Yali and Jiliang (2007) investigated the effects of text genre on L2 reading comprehension test performance. The subjects consisted of 30 non-English major freshmen randomly selected from 150 candidates. The results indicated that the participants performed significantly better in summary writing task for the narrative text than the expository text. Also, for the same narrative text, the results indicated that summary writing task tended to elicit significantly better general test performance than the multiple-choice task, while for the same expository text there was no significant difference between the test performances elicited by the two tasks.

Other studies have compared the students' reading performance on texts of different genres. Best *et al.* (2006) found that their subjects could better comprehend narrative texts than expository texts. They also found that the text genre interacted with the question type. They found that for global questions, their subjects performed better in narrative texts than in expository texts. This finding replicates earlier studies by Olson (1985) who reported that readers encountered more difficulty to process expository texts than narrative texts, and Snow (2002) who found that for expository texts, readers had more difficulty to answer global questions than local questions.

The findings from the aforementioned studies are relevant to the context of the present study. They clarify the need for test developers to be mindful of the different text genres while developing reading tests. These findings suggest that performance on reading tests can also depend on test developers' knowledge of the different genres and the kind of questions associated with these genres.

In addressing the issue of the impact of text genre on text comprehension and test performance, my intention is to justify why the selection of ESE texts must also consider the students' degree of familiarity with different text genres and the kind of tasks that are specific to these genres.

In text genre literature, it is usually assumed that each genre is associated with a specific structural organization. This structural organization serves as a frame that can guide and help readers identify important information and logical connections between ideas (Seidenberg, 1989; Alidib, 2004).

The most familiar and most studied text structure is the narrative. Also called a story, a narrative usually depicts events, actions, emotions, or situations of people in a cultural experience (Graesser *et al.*, 1991). A story is written to excite, inform, or entertain readers and it may report true or fictitious experiences (Graesser *et al.*, 1991). Furthermore, the concept 'story' is usually used to refer to the sequence of events described in a narrative. In a narrative, the descriptions of different events are made in such a way that the readers can anticipate the outcome of the story. In light of this argument, readers are expected to deduce motives of different story characters and connect the different relations among events.

Alidib (2004) provides the major components of a narrative. He describes narratives as involving (a) animate beings as characters with goals and motives; (b) temporal and spatial placements usually presented at the beginning of the story; (c) a problem or goal faced by the main character that imitates a major goal; (d) plots or a series of episodes that eventually resolve the complication; (e) impacts upon the reader's emotions and arousal levels; and (f) points (e.g., justice, honesty, loyalty), morals, or themes.

Since the narrative structure is based on sequence of events, it is expected that the order of information in both the passage and the test question can impact on test performance. Urquhart (1984) and Beck *et al.* (1982) found that, when a narrative text was organized according to a sequence of events, it was more easily comprehended than when its sequencing was distributed. This is to suggest that, narrative texts with a consistent spatial organization as well as distinct chronological sequencing appear to be easier to understand and recall than narratives without specific organization. This point of view is more clarified by Shin (2002) when he found that examinees poorly performed on most test items that could not follow the story line of the text as

such test items did not call for operations examinees were likely to use to construct the narrative macrostructure. Yet, as Shin notes, there is evidence that narrative texts typically have a hierarchical structure, and that readers are sensitive to such a structure so that when it is used to guide comprehension and recall, both are facilitated and comprehension is enhanced.

Whether it is a narrative, an argumentative or an expository text, the sequencing of information in both the text and test questions can affect examinee's text comprehension and test performance. Urquhart (1984) and Sheehan and Ginther (2001) reported that the location of relevant information within a passage was associated with the difficulty of MC questions on the TOEFL- 2000 reading test. They found that when two pieces of connected and relevant information were not presented in close proximity, their location in the representation could also be distant. This finding appears to rhyme with two earlier studies by Freedle and Kostin (1992) and Kintsch (1998). Freedle and Kostin found that an individual's expectations regarding the location of relevant information in the text affected the activation of information previously stored in the knowledge representation. On his part, Kintsch reported that the order in which information was encoded related to the manner in which it was built into the knowledge representation. All these findings relating to the relationship between the order of information in the text and in the test questions suggest that it can be difficult for some students to construct a coherent representation of a text from which comprehension questions relate to information not presented in the expected order. This view is clarified by Kintsch (1994) and Gorin (2005) when they found that changing the order in which the information in the text is presented impacted on the construction of coherent text representation, and this was found to increase the difficulty of test questions pertaining to the newly ordered information.

By addressing the issue of text genre and their related different organizational structures in this study, I wish to highlight the importance of examining the validity of reading tests by also considering how the test questions reflect the specific structural organizations of the different genres. In light of this argument, I expect the ESE test questions that are based on expository texts organized through a description to mainly focus on the identification of the main descriptors in the text. Likely, I expect those test questions that are based on narrative texts to be ordered in a manner to reflect the order in which text information is presented, with a focus on spatial and chronological markers that are meant to help describe different text events. From the

methodological perspective, I propose to feature this expectation in the qualitative content analysis of individual texts included in the study sample by evaluating the extent to which the test questions are sensitive to the structural organizations of the different texts, and the ordering of these test questions relates to the way information is ordered in the text. The working hypothesis is that, in order to be construct-valid, a reading test must also include test tasks that call for cognitive operations that relate to the specific genre used in the text. Furthermore, the test questions must be ordered in such a way to target the information in the order it is presented in the text. By adopting this conceptual stance to examining the reading construct, I expect the present study to generate insights that can augment our understanding of the construct validity of reading tests.

One important variable related to text organization is text coherence. Text coherence, has attracted the attention of researchers of text comprehension (McNamara, 2001, Lightman *et al.*, 2007; for example). Clarifying the concept of coherence, McNamara and colleagues argue that a coherent text is a text where the different text elements provide more explicit clues to relations within and across sentences (McNamara *et al.*, 1996). In light of this conceptualization, in order to understand how a text as discourse should be coherent, one has to understand the nature of the relations that exist between different text segments, and how these relations are made explicit in the text. This is what is called coherence relations. Sanders and Noordman (2000) argue that coherence relations are made explicit by linguistic markers, and these markers serve to connect two text segments, such as cause-consequence, problem-solution, comparison-contrast, etc. Viewed from this structural organization perspective, an understanding of coherence relations in a text is helpful because the construction of a text coherent representation requires that these coherence relations be established between different segments of a text, rather than between the different representations individual readers have of text segments (Sanders & Noordman, 2000). As a logical conclusion, if we agree that coherence relations play a crucial role in text comprehension, we can expect that linguistic marking of coherence relations impacts on text processing and comprehension.

In light of the preceding epistemological stance to text coherence, several studies have investigated the impact of linguistic signaling devices to find the extent to which these devices can impact on text processing and comprehension. Sanders and Noordman (2000) investigated

the impact of connectives and lexical markers of relations and they found that the different kinds of signaling devices in an expository text affected readers' text recall. An earlier study by Briton *et al.* (1982) also indicated that a marked text was easier to process than an unmarked text because the former required less cognitive processing demands than the latter. As an explanation to this claim, Briton and colleagues found that readers of the unmarked text had to use many inferences in order to establish different relations between ideas. This finding suggests that relational markers can guide the construction of text representation by providing explicit information about the relations between different text segments.

The issue of text coherence has been also addressed in terms of linguistic simplification. Research has indicated that, although the use of simple and/or simplified linguistic structures aims to make the text readable, linguistic complexity cannot be detrimental to comprehension; rather, syntactic simplicity can decrease text cohesion and thereby hinder comprehension (Bernhardt, 1991). Graesser *et al.* (2004) addressed this issue and they found that, in the texts included in their corpus, no explicit markers were used to achieve text coherence; instead, the different interclausal relationships were expressed simply through juxtaposition of clauses. Graesser and colleagues found that linguistic devices such as 'therefore', 'however', 'nonetheless', as well as conjunctions such as 'although', 'since' etc... were not used. Yet, the inclusion of such devices in a text is crucial as these devices can act as guiding cues that can assist readers to understand how an idea in one clause relates to other ideas in adjacent clauses. This viewpoint is consonant with arguments articulated by Degand and Sanders (2002), Graesser *et al.* (2004) and van Dijk and Kintsch (1983) when they acknowledge that an understanding of interclausal relationships is central to reading comprehension.

By addressing the issue of text coherence in this study, I want to investigate the extent to which the ESE include texts that are coherently organized through the use of appropriate connectives and linguistics markers that can make coherence relations between the different sentences included in the paragraph and the different paragraphs included in the text more explicit. I hypothesize that, when foreign language readers process texts where relations between sentences and paragraphs are made explicit through the use of appropriate connectives and linguistic markers, they are likely to comprehend these texts better than when they process texts where relations between different parts are not made explicit. By considering this critical aspect of

reading construct in validating reading tests, I expect the findings of this study to augment our understanding of the complex nature of reading assessment.

The examinees' knowledge of text content, their degree of familiarity with different text genres, and their understanding of text structure are not sufficient to account for their capacity to process and understand a text. The examinees must also be able to approach the text in terms of its density. In the following section, I review studies that have acknowledged the impact of propositional density on text comprehension.

### **4.5.3 Propositional density**

The construction-integration theory that is the theoretical framework of the present study describes text comprehension as a cyclic propositional processing (Kintsch, 1988, 1998). This is to suggest that, during each processing cycle, the different propositions are constructed from the textbase and arranged into a network. Then, in the integration phase, activation process spreads throughout the network and accumulates at points of high interconnectivity (Kintsch, 1988, 1998). In light of this statement, the difficulty in processing a text is influenced by the propositional density of the text. Kintsch (1994) and van Dijk and Kintsch (1978) argue that propositionally dense texts are more difficult to process and comprehend than less propositionally dense texts. This is to suggest that, texts with sentences that contain more words are potentially more difficult to comprehend than those texts that contain sentences with few words. Likely, texts with paragraphs that include more sentences tend to be difficult for readers, and texts organized around many paragraphs can be perceived more challenging to readers than those texts organized around few paragraphs.

Picking upon the above argument, some studies have attempted to examine the impact of propositional density on text comprehension. Embretson and Wetzel (1987) found that it was difficult to answer comprehension and inference questions on propositionally dense text, especially when the text contained modifier propositions. This finding appears to concur with a similar finding by Just and Carpenter (1971), cited by Gorin (2005) when they found that processing a longer sentence placed greater demands on working memory. Kintsch and Keenan (1973), cited by Gorin (2005) also found that increased propositional density impacted on text

processing and correlated to item difficulty. However, a study by Gorin (2005) using linear logistic latent trait model parameter estimates revealed no effect of propositional density on processing difficulty. Gorin (2005, p. 368) explains the lack of consistency with other previous studies as a consequence of methodological design. She thinks that the lack of significant results “might be a consequence of weak experimental manipulations arising from constraints of the passages”.

These findings on text density and its impact on comprehension are relevant to this study. They suggest that foreign language readers may have some difficulty processing longer texts than shorter texts; texts that include paragraphs with more sentences than those with few sentences; and texts that have longer sentences than those with shorter sentences. In light of this suggestion, I intend to examine the extent to which the density of ESE texts is appropriate to examinees’ characteristics. From the methodological stance, I propose to achieve this through a comparison of the density of texts used in the ESE with the texts planned in the teaching manuals. I hypothesize that, when the density of ESE texts relates to the density of classroom texts, examinees may feel more comfortable to read these texts than when they are confronted with texts whose density is not proportional to the classroom texts density.

The process-oriented approach I have proposed in this study conceptualizes reading construct at a multi-componential basis, where reading types, processing levels, test context, different features of the text and the features of the items interact and impact on the examinees’ selection of strategies to construct text meaning and complete test tasks. In the preceding sections, I have reviewed the relevant literature that pertains to the first four of these components. In the next section, I propose to review studies that have addressed the issue of the impact of test features on the difficulty of multiple-choice reading tests.

#### **4.6 Test variables as predictor of difficulty on MC reading tests**

Multiple-choice tests are no doubt the most popular type of objective tests administered in educational settings. The multiple-choice test has the advantage to provide a better coverage of the content and processes to be assessed (Rogers & Harley, 1999), and it is scored quickly and objectively. A traditional multiple-choice question is a question which requires the student to

choose one answer from a number of answers supplied. The multiple-choice question consists of (a) a stem (the text of the question), (b) the alternatives/options (the choices provided), (c) the key (the correct answer in the list of alternatives, and (d) the distractors (the incorrect answers in the list of alternatives).

However, one main challenge of the multiple-choice test is the difficulty encountered by test constructors to develop high-quality test items. This challenge usually results in test items that are poorly written and that are susceptible to test wiseness, and this threatens the test validity (Downing & Haladyna, 2006). However, research has indicated that the properties of an item can influence examinees' scores as some examinees who do not know the answer to a test item but are more sensitive to the item characteristics may get the item right simply because of the way it is constructed (Dudycha & Carpenter, 1973; Drum *et al.*, 1981; Ascalon *et al.*, 2007; Haladyna *et al.*, 2002).

Since each of the components of the multiple-choice question is capable of impacting on examinees' performance on test, I propose to review in this section, the studies that have looked at (a) the impact of the multiple-choice stem on examinees' performance on MC tests; (b) the impact of the structure of multiple-choice key on examinees' performance on MC tests; and (c) the impact of the quality of distractors on examinees' performances on the MC test. In the following section, I review studies that have investigated the impact of the stem structure on examinees' performance on MC tests.

#### **4.6.1 The multiple-choice stem structure and its impact on test performance**

The structure of the multiple-choice stem can influence examinees' performance on MC tests. My review of empirical studies outlines three striking characteristics of stems: focus (how clearly the stem presents the central idea of the test item), completeness (whether the stem is presented in a form of a complete statement or an incomplete statement), and orientation (whether the stem is positively worded or negatively worded).

A study by Downing *et al.* (1991) revealed that unfocused stems produced scores that could hardly discriminate high performing students from low performing students. More specifically,

Downing and colleagues found that unfocused stems did not contain enough information to allow the examinees to know what was being asked and that examinees had to first review the different item alternatives in order to understand the expectation of the question. Haladyna, Downing and Rodriguez (2002)'s study also replicated these findings.

The issue of stem completeness has been extensively investigated in the literature. The key question that has been usually addressed is whether test constructors should state the stem as a complete sentence or as an incomplete sentence where the test alternatives complete the stem. Although results are inconclusive, the main trend is that stems written in the form of a complete sentence appear to be easier to comprehend than those written in the form of incomplete sentences. An early study by Board and Whitney (1972) indicated that test items with incomplete stems were found to be significantly more difficult than those with complete stems. They also found that the validity as measured by the product moment correlation was lower on test items with incomplete stems than on test items with complete stems. Dudycha and Carpenter (1973) also found that incomplete stems were more difficult to process than complete stems although they did not find a significant effect of stem format on item discrimination (the proportion of those students who get the item right and those who get it wrong). Dudycha and Carpenter's findings were later replicated by Schmeiser and Whitney (1975), Violato and Harasym (1987), and Violato and Marini (1989).

However, other studies (Crehan & Haladyna, 1989b; Ascalon *et al.*, 2007; Schaefer, 2009) did not find any significant difference in the effect of stem completeness/incompleteness on the item difficulty although Crehan and Haladyna advised the use of complete stem format in large-scale assessments. On the contrary, Sireci, Wiley and Keller (2002) reported that incomplete stems were found to be significantly easier than complete stems although they could not find a significant difference in item discrimination between the two stem formats.

The findings of the aforementioned studies assume particular prominence in this study. Beyond the contradictory nature of their findings, these studies highlight the importance of considering the stem format to ensure the quality of the test. The guiding principle is that, examinees should not find an individual test item difficult simply because the stem is worded in a format that does not facilitate comprehension. Since I propose to inspect how each stem format can be associated

with ESE item difficulty; the findings reported in these studies are expected to provide useful insights.

Another characteristic of the multiple-choice stem that has been found to impact on test performance is the stem orientation. This refers to whether the stem is positively worded (positive orientation) or negatively worded (negative orientation). Dudycha and Carpenter (1973) found that examinees performed better on positively worded stems than on negatively worded stems. They also found an interaction of orientation and completeness. This is to suggest that test items having complete stems interacted with positively oriented test items as examinees highly scored on these test items; whereas test items with incomplete stems interacted with negatively oriented test items as examinees poorly scored on these test items.

In order to investigate a component of Haladyna, Downing and Rodriguez (2002)'s taxonomy of item-writing guidelines and an extension of Smith and Smith's (1988) exploratory work into the differential use and impact of item characteristics on test performance, Ascalon *et al.* (2007) investigated the effects of stem orientation on item difficulty. The participants were students from three California high schools to whom they administered 800 test items. The results indicated that, the test items that contained stems written as questions were perceived to be easier than those test items with stems written as open-ended statements; a finding that replicates an early study by Dudycha and Carpenter (1973). Tamir (1993) came up with a similar finding when she found that negatively worded stems were more cognitively demanding than positively worded stems. Haladyna, Downing, and Rodriguez (2002) came up with a similar finding and recommended test constructors to develop test items with positively worded stems. However, although neither Violato and Harasym (1987), Violato and Marini (1989), nor Schaefer (2009) could find significant differences in scores yielded by these two types of items, most item writing guidelines (see Haladyna, Downing, & Rodriguez, 2002; Haladyna, 1999) recommend test constructors to develop items with stems that are positively oriented.

The aforementioned studies relating to the multiple-choice stem structure are relevant to the present study in that they help understand that the validity of a multiple-choice reading test can be threatened by the quality of the item stems as performance on a reading test may be influenced by examinees' failure to comprehend the test questions simply because of the way the item stem is worded. These findings provide insights for understanding that constructing a valid

multiple-choice reading test is a laborious task as the format of the items has pronounced effects on the test difficulty. I am tempted to state that, if the structural format of an item is capable to account for test difficulty, then we have some reasons to bother of the validity of the tests used as such tests can fail to measure student's actual reading ability.

The multiple-choice stem structure is not the only variable that impacts on the difficulty of a multiple-choice reading test, one other variable worth looking at is the multiple-choice key. In the following section, I review studies that have indicated the effect on MC key on test performance.

#### **4.6.2 The multiple-choice key**

The key is the correct option in a multiple-choice question. Different item writing guidelines (see Haladyna, Downing & Rodriguez, 2002; Haladyna, 1999) recommend test writers to ensure that only one of the alternatives is the right answer. These guidelines advise that test writers should be careful not to provide clues to the right answer. Also, these guidelines advise that the location of the key must be varied. This last guideline has attracted the attention of researchers, and they have termed it balancing the key (Taylor, 2005; Haladyna, Downing & Rodriguez, 2002).

Balancing the key is a procedure in which each test option should be correct to an approximately equal proportion of time (Taylor, 2005). This procedure has been suggested first as a conventional wisdom before researchers attempted to validate it in empirical studies. Conventional wisdom suggests that when the correct answers to a multiple-choice test are distributed in different options, this distribution can minimize the effect of guessing. Haladyna (1994) suggests that any negligence to balance the distribution of the correct answer in a multiple-choice test can lead the higher-performing examinees to deduce the pattern of correct answers. Therefore, this constitutes a threat to test validity since these high-performing examinees cannot actually read and understand the expectations of the questions, the meanings and implications of the different item options in the light of the text, and they can simply ignore to select and discard options based on the way they understand the text. Rather, these examinees receive credit on test items simply by guessing from the response patterns of the test.

In a theoretical study Haladyna, Downing and Rodriguez (2002) reviewed the literature on item writing guidelines and they strongly recommended varying the location of the correct answer according to the number of options, a procedure that would balance the key. Other theoretical studies have also advocated balancing the key on the assumption that test wiseness examinees can use a response set for guessing (Sechrest, Kihlstrom & Bootzin, 1993; Ory & Ryan, 1993; Ellsworth, Dannel & Duell, 1990). However, because of their theoretical nature, all the recommendations from the above-mentioned studies can be questioned because they are not supported by empirical evidence.

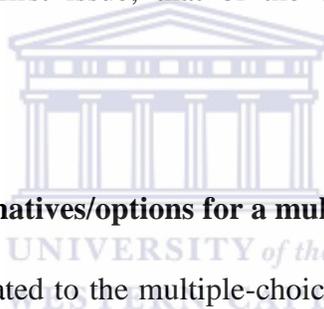
Among the few studies that have empirically addressed this issue, a study by Taylor (2005) deserves particular importance in the scope of this study. Taylor empirically examined whether a balanced key would make difference in test performance. Her subjects consisted of 80 introductory psychology students whom she divided to two equal groups (40 students each group) to whom she administered three versions of a 30- items test. The first version contained a balanced key, such that each of the four options was correct for about 25 per cent of all the 30 items; the second version of the test contained items where 80 per cent were constructed with A (40 per cent) and B (40 per cent) as correct responses, and about 20 per cent with C (10 per cent) and D (10 per cent) as correct responses. The third version of the test skewed correct responses in the opposite direction so that A and B were each correct for 10 per cent while C and D were each correct for 40 per cent. After an analysis of data using different statistics (chi-square, ANOVA test), the main finding reported was that balancing the key was not critically important factor impacting on test performance. Furthermore, the results indicated that performance on the three versions of the test was not significantly different. The students' selection of response options reflected the selection of plausible distractors, not the response bias pattern. This finding suggests that suggestions previously made by Haladyna, Downing and Rodriguez (2002), Sechrest, Kihlstrom and Bootzin (1993), Ory and Ryan (1993) and Ellsworth, Dannel and Duell (1990) are not based on empirical evidence and should therefore be taken with much caution.

In the scope of this study, Taylor (2005)'s findings can help understand that the response pattern cannot be taken as a variable that must be investigated when we address the issue of the impact of item structure on reading test performance. Nevertheless, this is not to suggest that we should ignore the suggestions made in the theoretical studies that advise test developers to balance

response patterns while designing multiple-choice tests. In light of this view, I propose to investigate the distribution of the response pattern in the sampled tests so as to find out the extent to which the ESE developers are aware of the need to balance the key between the six alternatives.

### **4.6.3 The multiple-choice alternatives (choices/options)**

The third component of a multiple-choice test is the alternatives. Also called options or choices, they are a highly critical component of any MC test. In the scope of the present study, I address the issue of the impact of MC alternatives on test performance in terms of the debate on the optimal number of alternatives to be included in the test and the issue of test wiseness. In the following section, I address the first issue; that of the optimal number of alternatives for inclusion in an MC test.



#### **4.6.3.1 How many optimal alternatives/options for a multiple-choice item?**

One of the most studied issues related to the multiple-choice alternatives/options is the issue of the number of alternatives. It is worth to mention that most educational large scale tests have four or five alternatives. Some tests, like the ESE, have six alternatives among which five alternatives are explicitly given and the sixth one is implied; that is, examinees select it in case the test question does not include the correct alternative among the five suggested alternatives. In light of this high number of alternatives included in the ESE, the issue of the number of alternatives becomes a concern that needs to be addressed in the present study.

The key underlying question related to the issue of the number of alternatives is to know if it is possible to construct as many functioning alternatives as possible as it is generally agreed that the different alternatives to a multiple-choice item are good if they are equally plausible to examinees who have not mastered the content tested (Lin, Chu, & Meng, 2010). There is a plethora of empirical studies on this issue. Most of these studies are experimental, comparing the effects of three- option test to four- option test, three- option test to four- and five- option test or again four- option test to five- option test. At this juncture, I hasten to state that my review of

these studies suggests that the three-option test presents psychometric qualities that can be substantially superior to the four- and five-option tests.

Haladyna and Downing (1988) addressed the issue of the optimal number of alternatives by studying the quality of a five- option test. The test consisted of a large scale test of reading comprehension with 200 multiple-choice items. Haladyna and Downing found that only 5.5 percent of the test items had four functioning distractors. In a subsequent study aimed to replicate these findings, Haladyna and Downing (1993) found the number of functioning options per item to be only about one and that only 1 to 8 per cent of items had two or three functioning distractors. In light of these findings, Haladyna and Downing (1993) suggested that the most appropriate number of alternatives per item would be three alternatives. These findings from Haladyna and Downing (1988; 1993) appear to question the validity of many multiple-choice tests using four, five or six options.

Owen and Froman (1987) addressed the issue of number of options in a multiple choice item with regards to examinees' preferences. They synthesized empirical results from studies that compared the effects of three- versus five- alternative test format on item difficulty, item discrimination, test score validity and reliability. Using the multiple correlation and covariance statistics, they found that the three- alternative test had better psychometric properties than the five- alternative test. Furthermore, they noted that examinees had preference for three- alternative test, especially in regarding to the test taking time.

The most elaborated argument for the use of three- alternative item is provided in a study conducted by Rodriguez (2005) in which he reviews existing empirical studies that have addressed the issue of the optimal number of alternatives to be included in a multiple-choice test for a period of eighty years. Using a meta-analysis approach, Rodriguez synthesizes empirical findings in order to provide a validity-related argument to support the use of a three- alternative test. The main finding was that three alternatives are optimal for multiple-choice items in most settings. The specific findings Rodriguez reported were: (a) moving from five- alternative items to four- alternative items reduces slightly item difficulty (the percentage of the total number of correct responses to the test item), item discrimination (the difference between the percentage of examinees in the upper group who obtained the correct response, and the percentage of those in the lower group who obtained the correct response) as well as item reliability; (b) moving from

five- alternative items to three- alternative items reduces greatly item difficulty, but does not affect item discrimination and item reliability; (c) moving from four- alternative items to three- alternatives items reduces significantly item difficulty, increases item discrimination and increases item reliability slightly. These findings support the argument of the use of a three- alternative test in educational settings.

Although the majority of educational tests in use include four or five alternatives, there is no substantial evidence to support such a practice. Also, there is no substantial evidence against the use of a three alternative test. Rodriguez (2005, p. 10) reported that “the vast majority of authors who studied this rule recommended using 3-option items”. Echoing Haladyna and Downing (1988; 1993), Rodriguez contended that “using more options does little to improve item and test score statistics and typically results in implausible distractors’ (Rodriguez, 2005, p. 11). Therefore, Rodriguez concluded for the use of only three alternatives. This is to suggest that, a multiple-choice item should consist of three alternatives, one correct (the key) and two incorrect (the distractors). This conclusion appears to concur with a view articulated by Vyas and Supe (2008) when they concluded their literature review on the optimal number of options by stating that “there is no significant change in the psychometric properties of the 3 options test when compared with 4 or 5 options MCQs.” In this review, Vyas and Supe further gave additional reasons related to test preparation and administration when they mentioned that “MCQs with 3 options had a higher efficiency because fewer distractors needed to be prepared, took up less space and required less reading time, decreased the time required to develop the items and the time to administer, and more items could be administered in a given time thus increasing the content sampled” (Vyas & Supe, 2008, p. 133).

Before I outline the relevance of the findings of the aforementioned studies to the present study, I wish to discuss the related issue of test wiseness as reflected in the comparison of three- versus four- alternative test.

#### **4.6.3.2 Three- alternative item versus four-/five- alternative item and the issue of test wiseness**

One issue related to the debate for or against the use of three- alternative item is the issue of test wiseness. Test wiseness is the examinee's capacity to utilize the characteristics and formats of the test to receive a high score (Sarnaci, 1979). The most striking element of test wiseness is guessing. Guessing in testing context is always considered as a threat to test validity because it tends to lead to the overestimation of the examinee's test scores, resulting in a construct-irrelevant variance. Viewed from the practical point of view, construct-irrelevant variance as a result of test wiseness suggests that guessing brings in the examinee's true score (that reflects his/her knowledge of the test content) a variance in that examinee's actual score is inflated with another score due to guessing among different alternatives.

There is a relationship between the examinee's capacity to guess answers from the different alternatives and the number of alternatives in a test item. Guessing as a test wiseness strategy is used by the examinees to eliminate one or more of the item alternatives as incorrect. If such elimination is successful, it increases the probability that the alternative selected is the correct one. The element of elimination that is present in guessing strategy suggests that test wiseness is a function of the examinee's ability to determine when it is profitable to guess because an examinee rarely guesses blindly (Ebel, cited in Sarnaci, 1979).

In light of the previous argument on guessing and elimination of alternatives as basic strategies of the test wiseness examinee, a related strategy worth considering is the deductive reasoning. Like guessing, deductive reasoning is a test wiseness strategy that allows the examinee to get some additional points beyond those obtained through direct knowledge of the test content (Sarnaci, 1979). Sarnaci notes that successful deductive reasoning is dependent on some knowledge of the tested materials, but the correct answer is not known without the presence of the other alternatives which serve as cues to the reasoning process. Therefore, in the deductive reasoning process, in order to select the correct alternative from a list of alternatives included in a test item, the examinee goes through the different alternatives and starts to eliminate all alternatives that he/she finds are virtually incorrect because these alternatives are not consistent with the information contained in the item stem. At this juncture, I wish to provide an example to illustrate this:

The political capital of South Africa is:

1. Cape Town
2. Bloemfontein
3. Pretoria
4. Johannesburg
5. East London

In this example, a test wiseness examinee will start by eliminating alternatives 2 and 5 because these two alternatives are not among the three cities that are either political capital (Pretoria), legislative capital (Cape Town) or economic capital (Johannesburg). Then, this examinee will therefore inspect the three remaining alternatives. If he/she lives in Cape Town, he /she might also eliminate alternative 1 because of his/her knowledge that the President of the Republic and all the ministers do not have their offices in Cape Town. Then the examinee will have to look at only two outstanding alternatives 3 and 4 and his/her chance of getting the item correct without necessarily knowing the capital of South Africa becomes very high.

Referring to insights from logic, the aforementioned example is a deductive reasoning that uses the absurd option strategy. Simply defined, an absurdity is something unreasonable, nonsensical. The absurd option strategy has been frequently used in test wiseness research (see Gibb, 1964; Bajtelsmit, 1975b; Sarnaci, 1979). Most of these studies have concluded that test wiseness examinee recurses to deductive reasoning when test items are written in such a way that they provide some cues to examinees.

My primary aim in addressing the issue of test wiseness in this section is to explore research insights relating to the number of alternatives and to evaluate how this number can result in examinees relying on test wiseness strategies to complete test tasks. However, most large-scale examinations prefer to include four or five alternatives on the ground that a reduced number of alternatives can result in examinees easily guessing the correct answer. This issue was addressed by Rogers and Harley (1999) when they investigated whether or not there are changes in the influence of test wiseness on performance when the number of alternatives is reduced from four to three. Rogers and Harley's instrument consisted of a 40 four- option multiple-choice reading items they administered to grade 12 students in Alberta, Canada. Half of these items contained one or two test wiseness elements and the remaining half was not susceptible to the application

of test wiseness. In order to compare the effect of four- alternative format versus three- alternative format on the examinees' use of test wiseness strategies, Rogers and Harley used two forms where the first form consisted of items in their original four- alternative format and the second form consisted of a three- alternative format where one distractor was deleted from the original form through existing rules for eliminating non-functioning distractors. The main finding reported was that the influence of test wiseness was lessened when the number of alternatives was reduced from four to three. Therefore, Rogers and Harley (1999, p. 245) recommended the use of a three- alternative format as this format is "less susceptible to the influence of test wiseness; thereby, leading to a more interpretable score in terms of content, behaviours, and skills assessed". One other finding reported was that the time needed to complete the test was lesser for the three- alternative test than for the four- alternative test, a finding that was replicated later by Haladyna, Downing and Rodriguez (2002).

At this juncture, I wish to outline the relevance of the discussion I have so far conducted on the issue of the number of alternatives to be included in an MC test and show the contribution of such a discussion to the present study. Since validity is about the meanings and interpretations of test scores, there is no doubt that the preceding discussion has helped me to understand how the number and quality of the alternatives included in the test items can impact on test validity. Since the ESE uses a six- alternative format of which five alternatives are explicitly given and the sixth alternative is implied, the preceding discussion can provide me with a basis to interrogate the quality of these alternatives in terms of how well they can be plausible or functioning. Besides, there is the issue of the content of the sixth alternative that states "*If there is no good answer, write 6*". By giving the opportunity to examinees to select alternative 6 in case there is no good answer, the test constructors assume that for some test items, there is not a correct alternative among the five listed. This leads me to interrogate their conceptualization of multiple-choice question. In light of this argument, I wish to pose these questions: when do examinees actually select alternative 6: is it when they fail to identify the correct answer from the suggested five alternatives or is it when they find that there is no good answer provided among the five alternatives? How does the test developer ensure that alternative 6 does not influence the examinees in their effort to read and comprehend the five alternatives and find the correct answer to the test item? All these questions will deserve attention in the present study, and my attempt to

provide empirical evidence to the validity of the implicit alternative 6 will also be one of my contributions in the body of testing research.

By addressing the issue of the number of alternatives in the present section, I intend to look at the quality of distractors included in the test in terms of how functioning they appear in the test item. Researchers have attempted to analyze the distractors quality from different perspectives, including observational, item analytic and item similarity perspectives (see Bruno & Dirkzwager, 1995; for example). These studies have used test quality results to identify and remove (or replace) non-functioning distractors from the tests. Among the various evaluation methods used, I can mention the expert judgment (Cizek & O'Day, 1994; Taylor, 2005) and the observation of option characteristic curves (Haladyna & Downing, 1993). The judgmental approaches involve a group of experts interpreting the degree of each distractor's plausibility, and the average of assigned values is used to evaluate each distractor. Although this method does not provide empirical evidence on the quality of distractors, it can be adequate in some testing contexts where for some test items, results from data produced by examinees' responses can be misleading. I think of a group of examinees who might not select a given distractor probably because they may have not understood the information contained in the item question. Expert judgments can, in this case, correct statistical results reported by other empirical methods. Some other methods have focused on evaluating distractors quality through the observations of how option characteristic curves are presented (Haladyna & Downing, 1993). These methods compute weights of each distractor and look at the mean scores of examinees selecting each distractor. In the scope of this study, I intend to evaluate the quality of the distractors used with the ESE by using these two methods.

### **Summary of the chapter**

In this chapter, I have reviewed relevant literature that pertains to this study. On the basis of the research objective, theoretical and conceptual frameworks, I have organized the literature in five themes. In the first theme, I have reviewed studies that have established the impact of reader's use of cognitive and metacognitive strategies on text processing and comprehension and reading test performance. In the second theme, I have reviewed studies that have outlined current debates

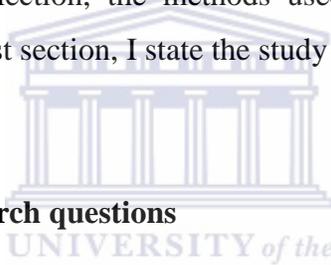
on the impact of contextual variables on students' capacity to read and comprehend texts. The different issues addressed in these studies include the students' characteristics including their skills and motivation, and the social-contextual environment under which both teaching and assessment take place. In the third theme, I have reviewed studies that have addressed the issues of the variety and complexity of reading types and processing levels and the appropriateness of related test tasks. In the fourth theme, I have surveyed studies that have debated on the impact of text features on text processing difficulty and performance. Studies reviewed here are those that have addressed the issues of the impact of the following textual features on text processing and comprehension: text topic and content, text genres and their related structural organizations, and the text propositional density. In this theme, I have also outlined the debate on the relative impact of vocabulary breadth and syntactic knowledge on text processing and comprehension. In the fifth and last theme, I have discussed in this chapter was the impact of item features on item difficulty. In this section, I have outlined the current debates on the MC stem structure (focus, orientation and completeness) on examinees' comprehension of test questions, the number of alternatives and the related issues of test wiseness and psychometric properties, the quality of MC distractors, and the quality of the MC key.

In order to answer the research questions, an appropriate design that reflects the present study objectives, questions, scope and framework is necessary. I propose to address this concern in the next chapter entitled “**Research Design and Methodology**”.

## CHAPTER FIVE

### RESEARCH DESIGN AND METHODOLOGY

In this chapter, I propose to describe the design and methodology I have used to conduct this study. Before I present this description, I first state the study questions and provide the rationale for each question. Then I explain the research design selected and the research methods used. Next I describe the research site where investigations were conducted, the research participants, and the research population and sample. Finally, I describe the instruments used to collect data, the procedures used for data collection, the methods used to analyze data and the ethical considerations. In the following first section, I state the study research questions.



#### 5.1 Statement of the study research questions

The unified construct validity theory I have used in this study conceptualizes test validity as an evaluation of the appropriateness, meaningfulness, and usefulness of test scores inferences. In view of this conceptualization, a reading test is valid if test scores can be interpreted as evidence of examinees' abilities to comprehend a text and answer the test questions based on it. Consistent with this view, one can expect test scores on a multiple-choice reading test to provide an indication of examinees' engagement with the reading test tasks by actively working to understand the text, to understand the expectations of the questions, to understand the meanings and implications of the different item options in light of the text, and to select and discard options based on the way they understand the text.

Indeed, the process-oriented framework I have proposed in this study is predicated on the investigation of examinees' strategies in order to validate the reading construct. This is to underscore the importance of establishing as clearly as possible what examinees are actually doing while taking the test as a concern that is central to validity. I have also mentioned in the

introductory chapter that the aim of the present study is to explore variables that potentially affect the ESE validity so as to come up with useful test construction suggestions.

In light of the aforementioned scope, this study addresses the following main research question:

*“What are the main issues that potentially threaten the validity of the DR Congo English state examination?”*

In my attempts to answer this main question, I believe that it will be useful to address the following four sub-research questions as its component parts:

1. What is the actual context of the ESE and to what extent this context potentially influences examinees’ performances on the test?
2. What is the variety and degree of complexity of the different reading types examinees conduct while completing the ESE tasks; and to what extent the ESE tasks are appropriate to the ESE context?
3. What are the textual features that affect the complexity of ESE tasks, and to what extent these textual features are appropriate to ESE actual context?
4. What are the item features that affect the complexity of ESE tasks, and to what extent these item features are relevant to ESE actual context?

The answer to the main research question aims to identify and describe the various factors that are predicted to affect the validity of the ESE.

The first sub-research question aims to describe the actual context of the ESE and evaluate the extent to which this context potentially affects examinees’ performance on the ESE. By actual context, I mean the context that is lived or experienced by the students and not the context that is assumed or articulated by/in government policies. This actual context purports to provide evidence on how the students’ characteristics, the conditions under which the teaching/learning of reading takes place, and the characteristics of the English language teachers are factors that can impact on the examinees’ completion of the ESE tasks.

The second sub-research question aims to provide evidence of the variety and complexity of different reading types and the cognitive demands required by different text processing levels to complete the ESE tasks. Furthermore, this sub-research question aims to evaluate the extent to which the ESE tasks are appropriate to the ESE actual context.

The third sub-research question aims to inspect the textual features that can be responsible to make the ESE tasks either difficult or easy. The answer to this question will help evaluate the extent to which text content, text genres and related structural organizations and text density exert influence on examinees' abilities to comprehend the text and complete test tasks.

The last sub-research question aims to scrutinize the different features of the ESE test item that account for the complexity of the ESE tasks. The answer to this question will establish evidence on the extent to which the quality of ESE stems, alternatives, and key impacts on examinees' performance on the test. In the following section, I describe the research design.

## **5.2 Research design**

In this section, I wish to describe the design I have selected by defining it, outlining what it consists of and highlighting the underlined question that backs it up.

The research design is generally defined as the researcher's overall plan or blueprint for obtaining answers to the questions that guide the study (McMillan & Schumacher, 2006). Viewed from this perspective, the research design becomes an important tool that helps the researcher to plan and implement the study in such a way that he/she can obtain the intended results. In light of this statement, the research design therefore enables the researcher to ensure that the evidence obtained helps answer the research questions as unambiguously as possible.

Since the process-oriented framework, the model I have proposed to use in this study, finds ground on the investigation of examinees' strategies in order to validate the reading construct, a descriptive design appears to be an appropriate avenue for establishing as clearly as possible not only the extent to which the different strategies examinees use to complete test tasks reflect their understanding of the text, but also the extent to which the test tasks are appropriate to the test

context. Therefore, the descriptive design enables me as the researcher to explore the different variables that are a potential threat to the ESE validity.

From an epistemological perspective, the descriptive design describes the current state of a situation. It is the response to the question “*what is going on?*” In the context of this study, because no validation studies have been conducted since the ESE was used by the Congolese government as a national test in 1967, I believe that the use of a descriptive design in this study can help to explore the current state of the English state examination by identifying and describing the multiple variables that are likely to exert influence on test performance. Viewed from this angle, the present study uses an exploratory descriptive design by attempting to explore new insights, understandings and meanings associated with the strategies used by readers in completing test tasks as well as the appropriateness of test tasks in relation to the test context.

One important characteristic of descriptive design that I wish to highlight is that, a descriptive design does not fit neatly into the definition of either quantitative or qualitative research methodologies; instead, it usually utilizes elements of both, often within the same study (McMillan & Schumacher, 2006). This observation is worthwhile in the scope of this study because I do not approach this study in terms of a qualitative-quantitative divide; instead, I consider that it is better to explore the plethora of factors that are a threat to the validity of the ESE by using both qualitative and quantitative data as a means to an end. I believe that such a choice can provide useful sources to describe these factors.

### **5.3 Research methods**

In order to explore the different issues that threaten the validity of the ESE, I have used two methods: the protocol analysis and the content analysis. In the following section, I describe the protocol analysis by providing the rationale for its use, its underlying assumption as well as its content.

### 5.3.1 Protocol analysis

Protocol analysis has become popular as a methodology used to uncover psychological processes that a person goes through while performing a task (Faerch & Kasper, 1987; Ericsson & Simon, 1993). This method is based on the assumption that people have “privileged access to their experiences” (Ericsson & Simon, 1993: xii), and that the information in their verbal reports is trustworthy (Park, 2009). Therefore, it is possible to verbalize their thoughts in a way that does not alter the sequence of thought that mediate the completion of task (Ericsson & Simon, 1993). This methodology has gained a strong ground in validating reading tests (see Rupp, Ferne, & Choi, 2006; Sasaki, 2000; Yamashita, 2003; Nevo, 1989; Hudson & Park, 2002; Phakiti, 2003; Weir & Khalifa, 2008b; for example).

The rationale for employing protocol analysis in the present study is my desire to understand the cognitive processes that mediate the reading construct and test performance. Viewed from this perspective, the use of protocol analysis in this study can help identify and describe the different strategies examinees employ in order to comprehend the text and locate text information necessary to answer the test questions.

In conducting protocol analysis, the underlying hypothesis is that the way examinees search for text information, evaluate item alternatives, and choose the best option can be registered through their verbalizations; and this can be later analyzed to discover their decision processes and patterns (Kuusela & Paul, 2000). In light of this hypothesis, there is a need to investigate examinees’ strategies use by collecting data during the task completion process. This argument provides the basis for validating verbal reports as voiced by Ericsson and Simon (1993). Based on their theoretical analysis, Ericsson and Simon argue that the closest connection between thinking and verbal reports is found when subjects verbalize thoughts generated during task completion.

At this juncture, I wish to provide the operational definition of protocol analysis as utilized in this study. This is a methodology used for collecting and analyzing data about text processing. Such an operational definition can help understand why the present study uses the concurrent verbal protocols in which participants’ verbalizations were collected during task completion.

### **5.3.2 Content analysis**

Content analysis is a research method used to identify the presence of certain words or concepts within texts or sets of texts. Stemler (2001) defines it as a method for making inferences by objectively identifying specified characteristics of texts. As a method of analyzing the content of documents, content analysis allows the researcher to test theoretical issues to enhance understanding of the data (Elo & Kyngas, 2007).

In the scope of this study, I have used this method in order to examine the content of the sampled English state examination papers and identify the different textual features that can impact on examinees' performance on reading tests. Viewed from this perspective, content analysis is expected to help identify and describe, through a scrutiny of ESE test paper contents, the relevant textual features that characterize both the text passages and the test questions.

### **5.4 Research site**

The research data were collected in Kananga, the capital city of Kasai- Occidental province. Kananga is one of the five top most populated towns in the DR Congo, with a population of over a million people in 2010 (about 1 130 100 inhabitants <http://monusco.unmissions.org>, 2010). According to some educational sources (Revue de l'Inspecteur, 2011; for example), in 2011, there were 86 secondary schools in Kananga that were providing both general and technical education.

### **5.5 Research participants**

I conducted this investigation with two categories of participants: the Grade 12 students and the English language teachers. The Grade 12 secondary school students, also called student-participants, were the main participants of the study. These were students who were in the final year of secondary school and who were preparing for writing the national test. These student-participants spoke French and almost all the four national languages (Ciluba, Swahili, Lingala, and Kikongo) with different proficiency levels. The information provided in Table 5.1 suggests

that the majority of the student-participants reported to have the highest proficient level in Ciluba (68.96 %) and the lowest proficient level in Kikongo (0.8 %). Furthermore, the majority of student-participants (74.60 %) reported to have good proficiency level in French, the language of education and administration. They were both males (59.9 %) and females (40.1 %) and they were aged between 18 and 24; with the majority of them being 19 (30.0 %), 20 (25.6 %) and 18 (21.6 %). Table 5.1 provides information on participants' age, gender and linguistic profile. It is worth noting that concerning the proficiency level of participants' languages spoken, five levels are represented on a scale where level 5 refers to the highest proficient level and level 1 refers to the elementary/basic level. Appendix 25 provides extensive details on student-participants profile.

Table 5.1: Demographics of students who participated in the study

Variables		N			%		
<b>1. Age</b>	18 years	107			21.6		
	19 years	149			30.0		
	20 years	127			25.6		
	21 years	63			12.7		
	22 years	29			5.9		
	23 years	14			2.8		
	24 years	7			1.4		
	<b>Total</b>	<b>496</b>			<b>100.0</b>		
<b>2. Gender</b>	Male	297			59.9		
	Female	199			40.1		
	<b>Total</b>	<b>496</b>			<b>100.0</b>		
<b>3. Languages spoken</b>	<b>Proficiency level</b>	<b>Level 5</b>	<b>Level 4</b>	<b>Level 3</b>	<b>Level 2</b>	<b>Level 1</b>	<b>Total</b>
	Ciluba	69.0	10.3	14.9	4.2	1.6	100.0
	Swahili	18.1	14.7	13.5	27.2	26.5	100.0
	Lingala	13.5	22.0	41.1	18.6	4.8	100.0
	Kikongo	0.8	1.8	5.7	13.5	78.2	100.0
	French	3.4	74.6	22.0	0.0	0.0	100.0

In order to gain more insights on the actual context of the learning of English in general and the learning and assessment of reading in particular, I also included English language teachers to participate in this study. As study participants, these were secondary school teachers who were teaching English in Grade 12. These teacher-participants had no/some qualification and experience in teaching English. Table 5.2 on the next page provides demographic information on teachers who participated in the study. However, the complete teacher-participants profile is provided in Appendix 26.

Table 5.2: Demographics of teachers who participated in the study

Qualification		Experience	N Participants
1	B. Ed (Hon.)	Less than 3 years	2
		Between 3 years and 5 years	4
		More than 5 years	3
		<b>Subtotal</b>	<b>9</b>
2	B. Ed	Less than 3 years	4
		Between 3 years and 5 years	5
		More than 5 years	4
		<b>Subtotal</b>	<b>13</b>
3	No higher degree	Less than 3 years	-
		Between 3 years and 5 years	1
		More than 5 years	4
		<b>Subtotal</b>	<b>5</b>
<b>TOTAL</b>			<b>27</b>

## 5.6 The study sample and sampling procedure

The present study was conducted with four hundred and ninety-six (496) Grade 12 students to whom I administered eight (8) tests and a contextual questionnaire, twenty-seven (27) English language teachers to whom I administered a contextual questionnaire, and twenty-six (26) Grade 12 students to whom I administered a contextual questionnaire only.

Since the main participants of the present study were Grade 12 students currently registered in different secondary/high schools, the first step in sampling was to group these schools into three strata corresponding to three categories: (1) high achieving schools, (2) average achieving schools, and (3) low achieving schools. My motivation to design a sample that could include the participants from these three categories was meant to ensure that the sample included data that reflected the actual reality on the ground so as to reduce some bias. In order to classify the different schools in these three categories, I used the state examination performance forms (from 2008 to 2011). These are forms released by the national education services immediately after the publication of the state examination results in which each school's performance on the national

test is reported for all subjects. These forms are sent to each school via the provincial education inspection, and the school principals are therefore required to adjust their pedagogic and administrative strategies in regard to the results obtained by their students on the national test. The state examination performance form contains for each school these main details: the individual school details (name, code, location, and number of participants), the number of participants who have passed the test, the percentage of success, the school average percentage of success on school marks, and the school average percentage of success on the national test. I wish to mention that these details, when considered for a certain period (more or less five years) for each individual school, can provide a definitive picture of that school's actual performance. Educational authorities also rely on these details to classify schools in different categories. I have furnished a copy of the state examination performance form in Appendix 1.

After I had grouped the different schools into three categories (strata), I then used the quota sample to assign a quota to each stratum in order to reflect the overall population. Since I had intended to conduct the investigation in 13 schools, I assigned a quota of 4 to high achieving schools, 5 to average achieving schools and 4 to low achieving schools. I wish to mention that this distribution was informed by insights from reliability theory (Feldt & Brennan, 1989; Parkes, 2007).

The last step for designing the sample of participants was to randomly include participants in schools from each stratum. Table 5.3 provides information on the number of schools included in each school category and the number of participants selected in each school category.

Table 5.3: Number of schools and participants for each school category

School categories		N schools	N participants
1	High achieving	4	133
2	Average achieving	5	217
3	Low achieving	4	146
<b>TOTAL</b>		<b>13</b>	<b>496</b>

As to the selection of Grade 12 English language teachers, I first grouped these teachers in three strata related to their qualifications and experience in teaching Grade 12 students; and then after, I selected them randomly.

The third sampling step consisted in the selection of the ESE test papers to be used as test materials. As with the sampling of participants, I first grouped the collected test papers in two main categories corresponding to two strata (general education and technical education). Then within each general education stratum, I designed four sub-strata corresponding to the four major subjects in general education (Mathematics-Physics, Bio-chemistry, Pedagogy and Literary), and four other sub-strata corresponding to the four major subjects in technical/specialized education (Commercial-Administration, Secretary-computing, Agriculture, and Dressmaking). This made a total of eight (8) groups. Then I randomly selected one testing paper per each group. Therefore, eight (8) test papers constituted the sample of test materials. Table 5.4 presents the test codes, the group of participants to which each test was administered and the number of participants included in each group. In this table, each of the eight tests is given a code. For example, the code T1 means the 2008 edition test used to test students from the general subject, Mathematics-Physics stream.

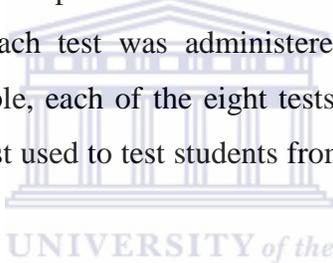


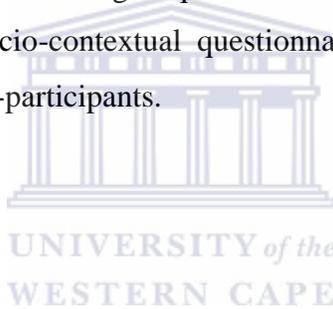
Table 5.4: Description of tests used, participants' subject areas, number of schools selected and number of participants to which each test was administered

Education category	Subject areas	Test used	N Schools	N participants
<b>General Education</b>	Math-Physics	T1	2	75
	Bio-chemistry	T2	2	64
	Pedagogy	T3	2	82
	Literary	T4	1	57
	<b>SUB-TOTAL</b>			<b>7</b>
<b>Specialized Education</b>	Com.-Admin.	T5	3	107
	Secretary-Computing	T6	1	35
	Agriculture	T7	1	48
	Dressmaking	T8	1	28
	<b>SUB-TOTAL</b>			<b>6</b>
<b>TOTAL</b>			<b>13</b>	<b>496</b>

However, in order to gain deeper understanding on the actual context of the ESE, I randomly selected one additional group of Grade 12 students from a shortlist of ten (10) average achieving schools. This group consisted of twenty-six 26 student-participants to whom only a contextual parameters questionnaire was administered. Since no test was administered to this group of participants, I could not include it in the main group of student-participants; instead, I treated it separately. Nevertheless, this group of participants shared the same characteristics with the major group in terms of age, gender and languages spoken.

## **5.7 Research instruments**

In order to explore the different issues that appear to affect the validity of the ESE, I used two main instruments to collect the data: a strategies questionnaire and eight (8) tests. I supplemented these two instruments by two socio-contextual questionnaires that were administered to 26 student-participants and 27 teacher-participants.



### **5.7.1 Strategies questionnaire**

The aim of this questionnaire was to elicit information on student-participants' use of strategies during test writing. In this questionnaire, I instructed the 496 student-participants to select from a list of suggested strategies, the different strategies they had used to answer individual test items. To make the participants' selections reflect the kind of strategies they had actually used, I used a concurrent questionnaire where I instructed them to select individual strategies they had used to answer each test question immediately after they had answered that individual test question. My preference for the use of concurrent strategies questionnaire instead of a retrospective strategies questionnaire finds some justification in Wilson (2002)'s argument that states that the more time that elapses between the occurrence of cognitive processes and the participants' reports about these processes, the more likely the reports are to be distorted or fabricated.

### **5.7.1.1 Weir and Khalifa (2008a)'s questionnaire**

In this study, I have used Weir and Khalifa's (2008a) questionnaire that I have adapted to the study context. The rationale for adapting Weir and Khalifa's (2008a) questionnaire is motivated by my desire to align the study design and methodology on the approach I have used in this study which is itself an adaptation of Weir and Khalifa's framework. I believe, together with Saw and Ng (2001), that researchers need to adapt existing questionnaires to their own study contexts as the known validity and reliability of these published questionnaires should not be assumed in all contexts. Weir and Khalifa's (2008a) questionnaire has two parts: the first part gives instructions to the participants, and the second part is mainly based on the strategies used by participants to answer the test questions. The description of Weir and Khalifa's questionnaire is provided in Appendix 2.

### **5.7.1.2 Adapting Weir and Khalifa's questionnaire to the study context**

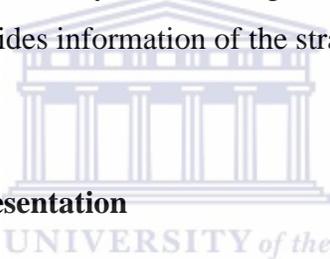
In order to adapt Weir and Khalifa's (2008a) questionnaire to the context of the present study, I tested it with a small group of students who shared the same characteristics with those included in the survey. The rationale for testing this questionnaire was to ensure its validity and reliability in the context of the present study.

#### **▪ Procedure for adaptation**

The first step in adapting Weir and Khalifa's questionnaire was to test it through the use of verbal reports conducted with three students. My primary objective to use verbal reports to test the validity of this questionnaire was to ensure that it included the most strategies that the study participants could use to complete the test tasks. Therefore, the verbal reports appeared to be a useful technique to elicit actual strategies that the study participants were likely to use. I would like to caution my readership that the use of verbal reports in this study should not be confounded to a research instrument; rather, it should be simply seen as a technique I have used to test the validity of the questionnaire, as evidenced by research (Presser, *et al.*, 2004).

Therefore, a detailed procedure of protocol administration, coding and analysis appears to overlap the scope of this section.

In order to elicit the actual strategies the study participants could use to complete the test tasks, I instructed these three participants to report aloud about their thinking while completing the test task. Then I analyzed the three reports to make an inventory of strategies utilized during test task completion. Although the three reports were not similar, some patterns of strategies actually used emerged from the participants reports, and I then made a list of these strategies that I compared to Weir and Khalifa's (2008a)'s list. The comparison between these two lists entailed two relevant considerations: first, not all strategies included in Weir and Khalifa's list were actually used by the three participants; secondly, some strategies that were actually used by the three participants were not included in Weir and Khalifa's list. As an implication, I decided to develop an adapted strategies list that included only those strategies that were actually employed by the three participants. Appendix 3 provides information of the strategies list and their codes.



▪ **Adapted questionnaire: presentation**

On the basis of results obtained from testing Weir and Khalifa's questionnaire, I developed an adapted version that I believed could suit the context of this study. The adapted questionnaire was administered to the 496 student-participants who wrote the eight different tests. This questionnaire had three parts: the first part included the contextual questions that aimed to provide participants' demographic information and the general questions on test context. The contextual questions related to participants' proficiency level in English language, their experience with the ESE, their motivation for reading, their attitudes towards the ESE, their views on the time allocated to complete the test tasks, the conditions under which testing takes place and the number of test questions included in the ESE. The second part was the questionnaire proper that aimed to elicit strategies the participants used to complete test tasks. In this part, the 496 student-participants were instructed to select, from a list of suggested strategies, all the strategies they had used to respond to individual questions. They were requested to order the strategies in the manner they had used them. Finally, the third part of the questionnaire related to the participants' perceptions of the test tasks. In this section, the 496 student-

participants were asked, after they had responded to all test questions and filled out the second section of the questionnaire, to answer some questions related to their perceptions of the test difficulty. Appendix 4 presents the adapted questionnaire I have used in this study. I wish to mention that I translated this questionnaire into French to enable the student-participants correctly respond to it.

### **5.7.2 Contextual questionnaire administered to Grade 12 students**

- **Objective:**

This second questionnaire aimed to gain richer qualitative information on the ESE context. The questionnaire included the same questions contained in the first section of the main questionnaire; but the participants were instructed to provide an explanation of each (or some) of their choices.

- **Participants:**

I administered this questionnaire to twenty-six (26) Grade 12 students randomly selected from one school. I wish to mention that these students did not participate in the main questionnaire that was administered to the 496 students who wrote the different tests. Appendix 5 presents this questionnaire.

### **5.7.3 Questionnaire administered to Grade 12 English language teachers**

- **Questionnaire objective**

The third questionnaire was administered to Grade 12 English language teachers. The objective of this questionnaire was to elicit additional information that could help explore the socio-contextual factors that are predicted to influence students' successful completion of the reading test tasks. More specifically, this questionnaire aimed to elicit contextual information on the testing methods English language teachers use in classroom tests, the reading resource supports they use, their perceptions of students' motivations and attitudes towards both teaching and

testing, as well as many other factors that negatively affect the reading classes in particular and the English classes in general.

#### ▪ **Participants**

This questionnaire was administered to 27 Grade 12 English language teachers; some of them randomly selected from the schools investigated.

#### ▪ **Questionnaire development**

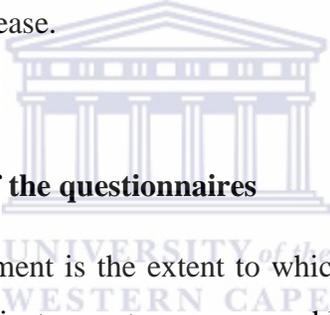
In order to determine the content of the questionnaire, I designed a pre-questionnaire of the open-ended type that I administered to a small sample of 4 English language teachers. This pre-questionnaire included the following questions: (1) Which methods do you use to evaluate your students? (2) Which sources of materials do you use for teaching reading classes? (3) What are the main difficulties you encounter while teaching English in general and reading classes in particular? (4) Which advice would you give for the improvement of the teaching of English in general and the teaching of reading in particular? (5) What are your views about the English state examination?

After the analysis of the responses from this pre-questionnaire, I then used all the responses suggested to build the questionnaire. Appendix 6 presents this questionnaire.

#### **5.7.4 Questionnaires validation process**

The three questionnaires were tested through a pilot study with a small number of participants (37, 7 and 2 for the first, second and third questionnaires respectively) sharing the same characteristics with the study participants, but who were not included in the main study. The main objective of piloting the three questionnaires was to ensure that they were valid and reliable. Thus, my aim in piloting the strategies questionnaire was to elicit most possible strategies the participants could check from the context of each question, and evaluate if the choice of these strategies reflected not only the fact that they had read and understood each individual question, but also the kind of strategies I expected from these questions.

My aim in piloting the two contextual questionnaires administered to Grade 12 students and Grade 12 English language teachers was to ensure that it included most possible answers that reflected the context of English language teaching and testing in both classroom and testing contexts. Therefore, during the piloting stage, I took detailed notes on how the participants were reacting to both the general format of the questionnaire and the specific questions. I was very attentive to see if the participants could show any confusion or surprise at a particular question. During the piloting stage, I also inspected the questionnaire statements to ensure that each statement was valid enough to measure a specific knowledge. In order to achieve this goal, I looked at the phrasing of the questionnaire statements to ensure that the participants were able to understand the meaning and implication of each statement in the same way. Of utmost importance were the instructions I had provided to the participants. The piloting stage enabled me to ensure that all the participants understood the instructions I had provided and that they could approach the task with some ease.



#### **5.7.4.1 Validity and reliability of the questionnaires**

The validity of the research instrument is the extent to which inferences and uses made on the basis of data produced by the instrument are reasonable and appropriate (McMillan & Schumacher, 2006). On the other hand, the reliability of the research instrument refers to the consistency of the measurement; that is, the extent to which the data generated by an instrument can be replicated by another research using the same instrument in other environments (McMillan & Schumacher, 2006).

##### **▪ Questionnaires validity**

In order to validate the three questionnaires, I used three types of validity: face validity, content validity, and construct validity. I established the questionnaire face validity during the piloting process through a careful revision of the questionnaire form and content to ensure that the questionnaire included only statements that measured the target construct. In order to ensure that I had accurately revised the questionnaire, I showed the questionnaire to my supervisor who in turn inspected it by looking at the phrasing of individual questions in order to ensure that it was face valid.

I established the content validity of the strategies questionnaire during the piloting process through a careful inspection of the questionnaire statements to ensure that the questionnaire sampled most strategies that reflected the different strategies the participants were likely to use in order to complete the test tasks. I also ensured that there could not be redundant statements or other strategies actually used by the participants but that were not included in the questionnaire statements. On the other hand, I established the content validity of the two contextual questionnaires administered to Grade 12 students and English language teachers through a careful inspection of all questions and statements to ensure that the questionnaires elicited most possible information that enabled to explore the ESE context.

Finally, I established the construct validity of the strategies questionnaire during the piloting stage when I ensured that all the questionnaire statements aimed to elicit participants strategies used during task completion and that no questionnaire statement was unrelated to the questionnaire construct.

#### ▪ **Questionnaires reliability**

In order to ensure that the data collected from the strategies questionnaire were reliable, I ensured that, after the participants had answered an individual test item, they were instructed to immediately fill out the questionnaire after answering that individual test question so that they were capable of remembering what they actually did to answer that particular question. The reliability of the strategies questionnaire was also established through clear instructions I had given to the participants in order to ensure that they had understood how to fill out the questionnaire form because failure to correctly fill out the questionnaire could result in data that could not be credible.

Among the three kinds of reliability indices reported in the literature, I used the Cronbach coefficient of reliability to establish internal consistency of the three parts of the cognitive-metacognitive questionnaire as well as the internal consistency of the individual statements within each questionnaire section. My choice of the internal consistency reliability was motivated by the assumption that, in order to confirm that the participants' strategies reported were reliable, there needed to be a certain consistency in their report of strategies on individual test items. I also computed Cronbach coefficient of reliability to establish the internal

consistency of the different questions in the two contextual questionnaires administered to Grade 12 students and English language teachers.

Since the reliability indices obtained from Cronbach coefficient were .71, .79 and .82 respectively for the strategies questionnaire, the contextual questionnaire administered to Grade 12 students and the questionnaire administered to teachers, I had evidence to conclude that the three questionnaires were reliable. In the next section, I describe the test as the second instrument I have used in this study.

### **5.7.5 Tests**

The second instrument I used to collect data was the test. In order to identify and describe the different strategies the 496 participants used while completing the test tasks, I used eight (8) different ESE tests that I administered to thirteen (13) different groups of participants under normal testing conditions. I wish to mention that in Table 5.4 provided on page 141, I have described the composition of the different groups and the different tests used.



#### **5.7.5.1 Test materials**

In this section, I wish to present the 8 ESE papers that I used as test materials to collect research data. Each test paper was made of two parts, the first part included a text passage examinees were requested to read, and the second part included test questions mainly based on the text.

- **T1**

This test was used with two (2) groups composed of 75 participants from Math-Physics subjects. This test was taken from the 2008 state examination edition test. This test material is presented in Appendix 7, Section 1.

- **T2**

This test was used with two (2) groups composed of 64 participants from Bio-Chemistry subjects. This test was taken from the 2010 state examination edition test. This test is presented in Appendix 7, Section 2.

- **T3**

This test was used with two (2) groups composed of 82 participants from Pedagogy subject. This test was taken from the 2011 state examination edition test. This test is presented in Appendix 7, Section 3.

- **T4**

This test was used with one (1) group composed of 57 participants from Literary subject. This test was taken from the 2010 state examination edition test. This test is included in appendix 7, Section 4.

- **T5**

This test was used with three (3) groups composed of 107 participants from Commercial-Administration subject. This test was taken from the 2011 state examination edition test. It is presented in Appendix 7, Section 5.

- **T6**

This test was used with one (1) group composed of 35 participants from Secretary-Computing subjects. This test was taken from the 2009 state examination edition test. It is presented in Appendix7, Section 6.

- **T7**

This test was used with one (1) group made of 48 participants from Agriculture subject. This test was taken from the 2008 state examination edition test. It is presented in Appendix7, Section 7.

- **T8**

This test was administered to one (1) group of 28 participants from Dressmaking subject. This test was taken from the 2010 state examination edition test. This test is presented in Appendix7, Section 8.

### **5.7.5.2 Validity and reliability of the test materials**

In this section, I explain how I ensured the validity and reliability of test materials.

- **Validity**

The validity of the eight ESE papers was assumed because these were authentic test materials that had been actually used for testing purpose, and not for research purpose. By using these materials to collect data in the scope of the present study, I expected the data to generate score properties similar to those generated by the English state examination. Also, the fact that these ESE papers covered the period of the four years (from 2008 to 2011) by including ESE papers sampled from different school subjects ensured the content validity of the test materials.

- **Reliability**

The reliability of the test materials was established when I used each ESE papers with two or three different groups of participants and found that the results obtained across the different groups did not relate to group differences, but they related to participants' individual abilities in reading. Therefore, the test materials were reliable as the results across the different groups reflected that the participants' individual reading abilities accounted for their performances on the test.

## **5.8 Data collection procedure**

This section describes the procedures I used to collect data as well as the process I used to ensure the validity and reliability of collected data. I collected the data in two phases: the first phase took place between 28 May 2012 and 7 July 2012, and the second phase took place between 17

December 2012 and 20 January 2013. During the first phase, I administered the eight tests to the 496 participants selected from 13 schools. During the second phase, I administered the contextual questionnaires to 27 Grade 12 English language teachers and to 26 Grade 12 students.

During the first phase of data collection, I strictly followed my schedule and moved from one school to another to administer the test. This was facilitated by prior contacts I had made with the school principals and the English language teachers. When I arrived in school at the set time, I first introduced myself to the students and explained the objective of the investigation as well as the importance for students to participate in the study. I assured the students that they were free to participate in the enquiry, and that those who would consent to participate were guaranteed anonymity of their responses. After the introduction, I gave to all the students who had agreed to participate a consent form, and I requested them to sign it to attest their free consent. Then I collected all consent forms after I had ensured that students who had agreed to participate had signed their forms. After this step, I arranged the participants in alphabetical order from the class list provided earlier to me by the school principal. In most cases, I was assisted by one or two school teachers to whom I had instructed not to tolerate any collaboration and cheating during test writing.

Then I told the participants that the test was accompanied with a questionnaire that had three sections. Section 1 was to be completed prior to test taking, Section 2 was to be completed while taking the test, and Section 3 was to be completed after having answered all the test questions. Then after, I requested the participants to complete **Section 1** of the questionnaire and this generally took between 5 to 7 minutes. After they had completed this section, I asked them to turn to the test and instructed them to ensure that the test had to be completed in exactly two hours following the time I had finished explaining test instructions, and that no extended time had to be granted to any participants.

After I had made sure that all the participants were ready to write the test, then I distributed the test materials and the research questionnaire. I invited the participants to carefully read the instructions before looking at the test. After five minutes, I also loudly read the instructions and explained them to the participants, trying to ensure that all the participants had clearly understood these instructions.

During test administration, my assistants and I circulated in the test venue to ensure that the participants were completing the test task as instructed. I repeatedly reminded them that they had to fill out the questionnaire each time immediately after answering an individual test question. I also reminded them that they had to carefully go through the list of suggested strategies in the questionnaire and check all strategies they remembered they had actually used to find the answer to each individual question. Then I instructed them that after they had answered all the test questions, they had to move on to Section 3 of the questionnaire and complete it. This section took between 3 to 5 minutes to be completed.

During test administration, my assistant(s) and I made sure that the participants, not only were working on their own and that no other support was available to them, but also they did not attempt to collaborate or chat. After one hour time, my assistant(s) and I started to remind the participants on the remaining time each ten minutes; and when only twenty minutes were left, did we remind them the remaining time every five minutes. When two hours were completed, we collected all the copies especially from the participants who were still busy writing the test. In general, most participants completed the test between twenty and ten minutes before the time.

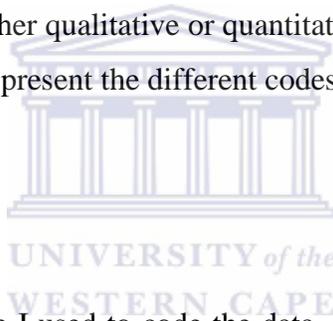
During the second phase of data collection, since there was only one school where the contextual questionnaire was to be administered to Grade 12 students, I used the same ethical procedures as those that I had used with students to whom I administered the test. After I had introduced myself to the students and explained to them the study objective, I requested their free consent through their signing of the consent form. I then distributed the questionnaire and explained them the instructions. Then I asked them to fill it out and insisted that they should explain their choices by writing their comments in the free space provided after each question.

As some participants in the questionnaire to be administered to English language teachers were selected from the schools where the tests were administered to students, I instructed the teachers to complete the questionnaire while their students were writing the test. For those participants who were selected from other schools, I took prior contact with them and either sent them the questionnaire or went to meet them in their schools during class hours.

## **5.9. Data analysis**

The data I had collected from the questionnaires and the tests were analyzed both quantitatively and qualitatively. Since the study design was descriptive, the core of the analyses was the description of all variables that were identified to affect the validity of the ESE. Therefore, depending on the kind of data to be used and whether the variable could be better described through content analysis or protocol analysis, or both, I described each variable either qualitatively or quantitatively or both qualitatively and quantitatively. This is a triangulation of methods and data that is consonant with the assumption of validity as argument where the researcher is advised to gather substantial evidence from diverse sources in order to provide argument supporting the meanings and interpretations of test scores (Kane, 2002, 2004, 2011).

The use of both qualitative and quantitative analyses for certain variables aimed to provide a much insightful description that either qualitative or quantitative description on its own could not provide. In the following section, I present the different codes assigned to data.



### **5.9.1 Data coding**

This section presents the procedure I used to code the data. Data coding is the first stage in the analyzing process. It is “a systematic way in which to condense extensive data sets into smaller analyzable units through the creation of categories and concepts derived from the data” (Sharon, 2004, p. 137). Data coding is important as it facilitates the organization, retrieval, and interpretation of data and leads to conclusions on the basis of that interpretation (Sharon, 2004).

The coding of data in this study pertained to (1) the codes I had developed for the strategies used by the participants, (2) the different tests administered to the participants, and (3) the individual test questions included in the different tests. I wish to mention that I developed and finalized the first two types of codes during data collection stage, while the coding scheme of individual test questions was developed after the data collection stage. Appendix 3 presents a table of different codes I used for the 18 strategies respondents used to complete the test tasks. In that table, S1=strategy one; that means read the whole text carefully; S2=strategy 2; that means read the whole text rapidly; etc.

The codes for the different 8 tests that I administered to different groups of participants have been presented in Table 5.4 (see page 141). In this table, **T1** means the 2008 ESE edition test administered to participants from General Subjects, subject area Mathematics and Physics; and **T2** means the 2010 edition test administered to participants from General Subjects, subject area Biology-Chemistry; etc.

Finally the codes I developed for individual test questions are presented as follow:

Item code		Meaning
1	T1.1	Item number 1 of test 1
2	T3.9	Item number 9 of test 3
3	T5.5	Item number 5 of test 5
4	T8.4	Item number 4 of test 8
5	Etc.	

### 5.9.2 Data analysis procedures

Since the present study uses a descriptive design, I decided to attempt both qualitative and quantitative analyses in order to describe the variables that likely affect the validity of the ESE. Therefore, each study variable was analyzed either qualitatively or quantitatively, or both qualitatively and quantitatively; depending on the way it was operationalized and the kind of data needed to investigate it. In this perspective, the variables that were investigated from data collected through contextual questionnaires were to be analyzed qualitatively and quantitatively. Likely, the variables that were to be investigated from data obtained from the strategies questionnaire were to be analyzed using quantitative protocol analysis methodology. Those variables that were to be investigated from the analysis of the ESE papers were mainly analyzed by using qualitative content analysis methodology, although quantitative analyses were to be sometimes conducted. Finally, data obtained from participants' results to the tests administered were analyzed using Item analysis methodology.

### **5.9.2.1 Qualitative analyses**

In the scope of this study, the qualitative analyses consisted of two broad categories: the first categories consisted of the interpretations of qualitative data that were derived from written comments and explanations the participants had offered on their choices of responses from the objective contextual questionnaires. These comments aimed to provide more information on the variables to be investigated and they were very focused on the targeted issue and generally they depended on the range of choices selected by the participants. This range of choices selected by the participants provided guidance to group their responses in different categories (for example, the participants who commented on the choices “I never read at home/ I rarely read at home/ or I usually read at home etc...”). After I had grouped the data in categories, the next step was to group data within each category in different patterns. The likelihood of responses guided me in designing and labeling the patterns. For instance, ten participants may have selected the choice that “they never read at home”. If three of them commented that they never read at home because they lacked reading resources and four commented that they were not motivated in reading, while the three others referred to economic factors that prevented them from focusing on classwork at home, I therefore used these three response patterns to substantiate in this response category.

Regarding the qualitative data obtained from the content analysis of the 8 ESE papers, the first step consisted in inspecting each test paper in order to identify the presence of the variable under investigation. Then, came the step where a description of how the identified variable was actually reflected in the data. To take an example: in investigation text genres, I first read all the 8 text passages from the eight test papers and identified for each text passage a specific genre used. Then if two texts were narratives, I had to study the chronological presentation of events in each narrative and my report consisted in presenting similar points to these two narratives as well as nuances specific to each of the two narratives.

### **5.9.2.2 Quantitative analyses**

I hasten to mention that considering the descriptive nature of this study, the quantitative analyses were strictly limited to the use of descriptive measures which were presented in frequency tables

and/or cross-tabulations. These descriptive measures were of two kinds: (a) measures of central tendency and (b) measures of dispersion.

- **Measures of central tendency**

Since I had collected the quantitative data from eight different groups of students to whom eight different tests were administered, there was a need to work with the mean scores that could reflect all the eight groups rather than working with individual scores from individual groups. This necessitated the use of the *mean* (M) as measure of central tendency. Also known as average, the mean is a statistics that is calculated by adding up all the observations/scores in each group, and then dividing to the number of participants in that group. My choice of the *mean* as measure of central tendency was justified by the fact that this measure, unlike other central tendency measures (*mode, median*), takes all observation/scores into account; and this makes it a sensitive measure of central tendency (Eysenck, 2004).

- **Measures of dispersion**

By using a mean score in order to describe the frequency of a given variable as observed in the eight test groups, it is clear that there were some individual group scores that were either below or above the mean score. This is to suggest that, there is a *variance* that may be observed in individual group scores, and sometimes this variance may be large enough to necessitate some explanations. In descriptive statistics, the appropriate statistics needed to identify this variance is the *standard deviation* (SD). This statistics explains the dispersion or "variation" in test scores from different test groups. It therefore provides an opportunity to see whether the observations/scores in a group are similar to each other or whether they are spread out (Eysenck, 2004).

In descriptive statistics, the higher the SD, the more dispersed are scores across groups; hence the need to explain the reason of the dispersion. One main reason is that a high standard deviation may mean that there is/are one/some scores that skew (scores that are either very high or very low compared to the mean score). There is not a definite score range that characterizes the boundary between high SD and low SD. Nevertheless, high SD are scores that are generally above one-fourths (1/4) or the mean score, and low SD are scores that are generally below one-fourths (1/4) of the mean score. To take an example, if the mean score observed across 8 groups

is 10, and the SD is 3; this means that the SD is high and it suggests that some individual scores are far below or above 10. But if the SD is 1.7, this means that the SD is low and it suggests that most scores across groups cluster around the score mean.

- **Item analysis**

Since one of the objectives in the study is to evaluate the extent to which some factors account for the complexity of test tasks, the investigation of item difficulty (and sometimes distractors efficiency) appears to be a relevant route to discriminate between easy test tasks and difficult test tasks.

Item difficulty (ID) or sometimes called item facility (IF) expresses the proportion of examinees who answered the item correctly. It is a simple and straightforward statistics that consists in taking the number of candidates who got an item right and divide this number to the total number of candidates who answered that item. For example, if 70 of 100 candidates got an item right, the item difficulty index for that item is computed by dividing 70/100 and this makes .7, indicating that the item is easy. Item difficulty ranges from 0.0 (none of the examinees answered the item correctly) to 1.0 (all of the examinees answered the item correctly). Values that are closer to 0.0 suggest that the item is difficult; values closer to .5 suggest that the item is of moderate difficulty while values that are closer to 1.0 suggest that the item is easy.

In order to evaluate the quality of distractors, I conducted distractor efficiency statistics. This statistics consisted in computing the frequency of each of the 5 distractors (4 distractors plus alternative 6) and evaluate whether or not they were functioning. It is generally agreed that any distractor that cannot be selected by more than 5 % of examinees is considered non-functioning; suggesting that it is a distractor of poor quality.

## **5.10 Research ethics**

This study was conducted in an ethical manner, following the ethical guidelines of the university. I firmly adhered to the University of the Western Cape procedures by seeking for ethics clearance from the Senate Research Committee before starting data collection. I strictly observed ethical procedures with regard to the information I provided to the participants and the

seeking of their consents to participate in the study. This is to suggest that, students' and teachers' participation in the study was based on their free consent and I included in the study only those students and teachers with the positive signed consents. I also reminded the participants of their right to withdraw from research at any stage and this without any consequence. They also had right to anonymity in that their names did not appear on the questionnaires and were not mentioned in the writing up of the findings. I ensured that all the documents to be given to participants, where necessary, were translated into French. Further to this, I guaranteed the participants confidentiality in that their answers from the questionnaires were to be stored in a safe place. By doing so, I ensured that the principles of integrity in all aspects of the research, accountability, and fairness had been strictly observed.

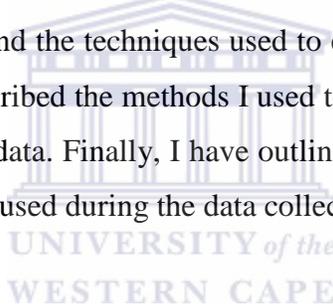
Since the test materials were to be obtained from the "*Inspection Provinciale de l'enseignement*" (Provincial Education Services) which is an educational body in charge of planning, constructing, and administering the national test at the provincial level, I asked for permission from the "*Inspecteur Provincial de l'Enseignement*" to use the test and conduct the research. I also assured the educational services that I would keep these papers in a secret place and that at any time anybody not directly associated with this research would have access to them.

Since the participants were current Grade 12 students actually engaged in the national test preparation, I asked the consent of the school principals to gain access and informed consent in order to conduct the study. During the investigation, I adhered to each school's code of conduct and I made sure that the learning was not disrupted by my administration of the test.

The different documents I used during data collection stage were the (1) Information sheet for research participants (see Appendix 8), Statement by the researcher (see Appendix 9), Informed consent Grade 12 students (see Appendix 10), Informed consent for English language teachers (see Appendix 11), Letter to the Provincial education services (see Appendix 12), and letter to school principals (see Appendix 13).

## **Summary of the chapter**

In this chapter, I have described the research design and methodology I used in this study. Since the objective of the study was to explore issues that affect the validity of the ESE validity, I have adopted a descriptive design using both qualitative and quantitative data. In order to explore and describe the different strategies examinees used to read the text and answer test questions based on it, I selected two methods: the protocol analysis and the content analysis. I have also described the research site with a particular focus on the study population, and the number of schools sampled. I have also presented the research participants who were Grade 12 high/secondary school students and English language teachers from which data were collected. Then I have described the sampling procedure and I have shown a multi-level stage of sampling characterized by a combination of stratified sampling, quota sampling and simple random sampling. I have also described the three instruments used to collect the data, the population to which these instruments targeted and the techniques used to ensure the validity and reliability of these instruments. I have also described the methods I used to analyze qualitative data as well as those used to analyze quantitative data. Finally, I have outlined the ethical aspect of the research by providing necessary documents used during the data collection stage.



## CHAPTER SIX

### FINDINGS RELATED TO THE ACTUAL CONTEXT OF THE ENGLISH STATE EXAMINATION

#### 6.0 Introduction

Reading in general and the assessment of reading in particular do not occur on a vacuum; rather, they occur in a specific context. The understanding of this context helps to understand how examinees read the text and complete the test tasks. In the third chapter where I have described this study framework, one relevant hypothesis was that examinees' use of strategies also depended on their individual characteristics, the conditions under which learning takes place, as well as the conditions under which testing occurs. Following this hypothesis, the first objective of this study was to describe the context of the English state examination as this context was expected to affect examinees' execution of test tasks. Therefore, a related research sub-question was: *“What is the actual context of the ESE and to what extent this context potentially influences examinees' performance on the test?”*

In this chapter, I present and analyze data that sought to describe the actual context of the English state examination. In light of the content of the instrument used to collect data, I have organized this chapter in the following three sections: first, the description of Grade 12 students-participants' characteristics; then the description of teachers-participants' characteristics, and last the description of the actual conditions under which the teaching/learning of reading takes place.

The data necessary to answer this research question are taken from three instruments: (1) Section 1 of the strategies questionnaire administered to 496 students; (2) the contextual questionnaire administered to twenty-six (26) Grade 12 students; and (3) the contextual questionnaire administered to twenty-seven (27) English language teachers. The data from the first instrument are analyzed quantitatively in order to provide a picture of the frequency of occurrence of

different observations. The data from the second instrument are analyzed qualitatively in order to provide deeper insights necessary for grasping the reasons of the participants' choices. Finally, the data from the third instrument are analyzed both quantitatively and qualitatively so as to provide, not only a picture of the frequency of occurrence of different observations, but also deeper insights into the reasons of the participants' choices.

I hasten to mention that, although this chapter strictly relates to the analysis and presentation of study findings, for the sake of narrative immediacy and primacy of findings, I will from time to time conduct an initial discussion of some of the findings. Such an epistemological choice aims to draw attention on the salient findings and clarify their relationships to the study aim and objectives. Nevertheless, in the Discussion Chapter, I expect to provide an extensive argument on the initial discussion provided in this chapter. In the following section, I describe the Grade 12 students' characteristics as reflected in the data.

## **6.1 Grade 12 students' characteristics**

In this section, I describe Grade 12 students' characteristics by looking at the 496 student-participants' skills in English language, their motivation for reading, their attitudes towards reading, their experience with the ESE and their attitudes towards the ESE.

### **6.1.1 Grade 12 students' skills in English**

In order to investigate Grade 12 students' skills in English, the 496 student-participants were asked, in the first part of the strategies questionnaire, to indicate how they would rate their level of proficiency in their use of English language. Table 6.1 on the next page provides information on the participants self-rating of their proficiency level in English. In this table, the proficiency levels of the 496 participants are presented in five levels ranked from higher levels to lower levels (very good, good, average, low and very low). On the other hand, the responses provided by the 496 participants are presented in eight groups that correspond to the eight tests administered (T1, T2, T3, T4, T5, T6, T7, T8). The analysis of the data takes account of the

mean score (M) while the standard deviation score (SD) is used to indicate variations of scores between individual groups.

Table 6.1: Participants self-rating of their proficiency level in English

Proficiency level N=496		T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)	T7 (%)	T8 (%)	M (%)	SD
1	Very good	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Good	2.2	3.1	2.0	3.1	5.0	4.3	1.9	3.5	3.1	1.1
3	Average	14.7	17.6	17.9	16.1	13.1	17.1	12.5	15.8	15.6	1.9
4	Low	21.2	23.1	21.6	31.9	29.5	33.2	15.6	20.0	24.5	5.9
5	Very low	62.0	56.3	58.5	49.0	52.4	45.4	70.0	60.8	56.8	7.3
<b>TOTAL</b>		<b>100.0</b>									

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; T7=Test7; T8=Test8; M=Mean; SD=Standard Deviation**

In light of the information presented in Table 6.1, I wish to state the following findings:

Across the eight groups, the majority of participants reported their proficiency level in English to be either very low (56.8%) or low (24.5%). On the contrary, no participant (0.0%) reported his/her proficiency level in English to be very good; while few participants reported that their proficiency level in English was either average (15.6 %) or good (3.1%).

In order to gain more insights on this finding, I also examined the qualitative reports from the 26 students-participants in order to understand the reasons of participants' perceptions and self-rating of their proficiency level in English. In these reports, the participants who reported their proficiency level in English to be very low or low either mentioned their poor capacity to cope with basic language skills, their inability to match the English spelling and pronunciation as it is in French; or again a certain attitude that supports that English is nothing more than an additional language that places on them extra linguistic demands that they are not capable to accomplish. The following comments can serve to illustrate the afore-mentioned points:

**Respondent 1:** I don't understand English, I can't speak it; it is a difficult course.

**Respondent 2:** Really, English is a very difficult language for me. I don't understand what my teacher says. When she asks me the question, I am unable to say something.

**Respondent 3:** This is a very difficult language, they write it in one way and you must pronounce words in another way. This is too confusing.

**Respondent 4:** I don't like English. The pronunciation is different from the French pronunciation. When you are given a text, you find two or three words with the same vowel; but when you read these three words in the same way, the teacher tells you that it is wrong. Sometimes my friends laugh at me after reading a word because I have badly pronounced it.

**Respondent 5:** It is not our language. I speak Ciluba, Lingala, Swahili and French. Do I really need to learn again this language (English)? It is very difficult. I can't be good at it.

**(See Appendix 23 for additional responses provided on a selected basis)**

However, the few participants who reported their proficiency level in English to be either average or good mostly referred to their capacity to cope with the basic language skills. These three excerpts can illustrate this view:

**Respondent 1:** I understand almost everything, and I try to speak a bit. I am comfortable with reading and writing. I think I am good at English.

**Respondent 2:** I like English because I can read a text and understand it. I can even write a letter in English. I cannot say that I am bad at English.

**Respondent 3:** I have no problem in English classes. I am a good learner.

**(See Appendix 23 for additional responses provided on a selected basis)**

The afore-mentioned findings are relevant to this study. They provide evidence to support that the majority of student-participants were aware of their poor abilities in English language. This has some implications for testing: if the ESE constructors are aware of examinees' poor skills in English language, they need to construct tests that include tasks that reflect the examinees' skills. I shall elaborate on this issue in the Discussion Chapter.

### **6.1.2 Grade 12 students' motivation for reading**

In order to investigate Grade 12 students' motivation for reading, the first section of the main questionnaire administered to the 496 participants comprised three questions that sought to investigate participants' motivation for reading. The first question requested the participants to

indicate their frequency of reading at school and at home (Question 3); the second question requested the participants to indicate their frequency of attendance to English classes (Question 5); and the last question requested the participants to indicate the number of hours per week they preferred to learn English (Question 7). Additional information on students' motivation was also provided by the qualitative reports obtained from the 26 students to whom the contextual questionnaire was administered; and from the questionnaire administered to the 27 English language teachers. In the following sections, I present results of these three investigated issues.

#### **6.1.2.1. Reading frequency at school and at home**

Students' reading at school and at home is meant to provide insights for understanding the motivation they have for reading activity. I hypothesized that students with high motivation for reading frequently read at home while those with low motivation for reading less frequently/do not read at home. Besides, I hypothesized that students who frequently read in English classes can have high motivation for reading while those who less frequently read in English classes can have less/no motivation for reading in English. Table 6.2a on the next page provides information on student-participants' reading frequency at school and at home for the participants who were actually studying English five 5 hours per week; while Table 6.2b on the next page provides information on participants' reading frequency at school and at home for the participants who were actually studying English 2 hours per week. I wish to mention that, among the eight groups of participants, six groups included the participants who were learning English five hours a week (T1, T2, T3, T4, T5, T6) while two groups (T7, T8) comprised the participants who were learning English two hours a week.

The evidence in Table 6.2a on the next page suggests that, for participants who were actually studying English five hours per week, only 20.7 per cent were regularly reading in English at school while the majority of the participants either 'sometimes' (40.3%) or 'rarely' (39.1%) read at school. On the other hand, only 5.6 per cent of participants were regularly reading in English at home while the majority of participants either 'never' (44.4%), 'rarely' (30.1%) or 'sometimes' (18.25) read at home.

Table 6.2a: Reading frequency at school and at home for participants who were actually studying English 5 hours per week

Reading frequency N=420		T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)	Mean (%)	SD
At school	Regularly	22.7	22.7	19.5	28.1	22.4	8.6	20.7	6.5
	Sometimes	45.3	45.3	39.0	40.4	40.2	31.4	40.3	5.1
	Rarely	32.0	32.0	41.5	31.6	37.4	60.0	39.1	11.0
	Never	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	<b>TOTAL</b>	<b>100.0</b>							
At home	Regularly	5.3	4.7	6.1	1.8	8.4	7.1	5.6	2.2
	Sometimes	20.0	17.2	15.9	17.5	21.5	17.1	18.2	1.4
	Rarely	28.0	43.8	20.7	36.4	29.0	22.9	30.1	9.0
	Never	46.7	34.4	57.3	44.3	41.1	42.9	44.4	7.4
	<b>TOTAL</b>	<b>100.0</b>							

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; SD=Standard Deviation

Table 6.2b: Reading frequency at school and at home for participants who were actually studying English 2 hours per week

Reading frequency N=76		T7 (%)	T8 (%)	Mean (%)	SD
At school	Regularly	0.0	0.0	0.0	0.0
	Sometimes	12.5	17.9	15.2	3.8
	Rarely	87.5	82.1	84.8	3.8
	Never	0.0	0.0	0.0	0.0
	<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	
At home	Regularly	0.0	0.0	0.0	0.0
	Sometimes	2.1	10.7	6.4	6.1
	Rarely	12.5	35.7	24.1	16.4
	Never	85.4	53.6	69.5	22.5
	<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	

T7=Test7; T8=Test8; SD=Standard Deviation

The information presented in Table 6.2b on the previous page indicates that, among the participants who were studying English two hours per week, none of them (0.0%) reported to regularly read in English at school while all the participants reported to read either 'rarely' (84.8%) or 'sometimes' (15.2%) at school. Furthermore, no participants (0.0%) reported to regularly read in English at home while all of them either 'never' (69.5%), 'rarely' (24.1%) or 'sometimes' (6.4%) read in English at home.

Based on the results from Table 6.2a and Table 6.2b, I wish to make two observations: First, students' reading at school is very restricted and that its frequency depends on the time allocated for learning English. This is to suggest that, the more students have much time for studying English, the more likely they tend to read at school. Second, reading in English at home is a neglected activity; and it is almost nonexistent for participants actually studying English only two hours per week. This is to suggest that, the less time students have for learning English at school, the less motivated they are to read in English at home.

Although I will elaborate on these findings in the Discussion Chapter, I hasten to state at this juncture that, in order to be construct-valid, the ESE need to include tasks that are sensitive to the students' actual reading amount at school and at home so as to reflect the actual test context.

The qualitative reports from the other 26 student-participants provided additional information on students' frequency of reading at school and at home. The participants who reported to rarely read at school mentioned three striking reasons that accounted for this. While some of them explained that they read rarely at school because of the reduced number of hours per week for learning English as scheduled in the curriculum, other participants attributed the lack of reading to the time they spent to copy the text in their notebooks. Still, some other participants simply attributed the fault to their teachers. The following six excerpts can act as evidence:

**Respondent 1:** We have just 2 hours per week; how can we read regularly?

**Respondent 2:** First, I would say that even in class we do not really read. We just have 2 hours per week for English. I think we can read more if we have much time.

**Respondent 3:** It takes 1 hour to copy the text that is written on the board. If it is a long text, when are we then going to read?

**Respondent 4:** Our teacher is too lazy; he takes much time to analyze one text. He takes 2 sessions for writing it on the chalkboard, 2 sessions for teaching vocabulary on the text; 2 or 3 sessions for teaching grammar on the same text; 1 or 2 sessions for text comprehension and discussion. After, he takes 1 session for dictation of a paragraph of the same text, and 1 session for essay writing. So, 1 text may take us 2 months.

**Respondent 5:** My teacher of English rarely comes to teach us. When he comes, he takes so much time to tell us long stories of his personal life.

**Respondent 6:** He [the teacher of English] has another job in a Pakistani shop where he serves as a translator. He rarely comes to class and whenever he comes to teach; he just sits down and asks us to copy the text.

**(See Appendix 23 for additional responses provided on a selected basis)**

As for the majority of participants who reported that they never read at home, the qualitative reports mainly pointed to a lack of reading materials, or they pointed to the content of reading materials not relevant to their needs, or again to their poor socio-economic conditions that do not permit them to read at home. The following four extracts can illustrate this point:

**Respondent 1:** I don't read in English at home because I don't have textbooks. I don't have anything to read.

**Respondent 2:** I cannot spend my time to read a text that I believe I have copied it with so many mistakes. I think it is important that we use printed texts. These texts can motivate us to read at home. When the teacher gives the book [teacher book] to one student and asks him to write the text on the board; the student does not pay attention to write correctly the text.

**Respondent 3:** Those texts like 'the discovery of radium', 'Uncle Bernard in London', I don't understand them. I can't spend my time to read them at home.

**Respondent 4:** I don't have time to read at home. When I come back from school, I quickly rush to the market to sell some plastics in order to get some money to pay my studies.

**(See Appendix 23 for additional responses provided on a selected basis)**

### 6.1.2.2. Attendance to English classes

In order to augment the understanding of Grade 12 students' motivation for reading, the 496 student-participants were also asked to report their frequency of attendance to English classes. I hypothesized that students with high motivation in English tend to regularly attend English classes, while those students with low motivation may not regularly attend English classes. Table 6.3 presents the frequency of attendance to English classes for the 496 participants.

Table 6.3: Students' attendance to English classes

Frequency of attendance to reading classes N=496		T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)	T7 (%)	T8 (%)	Mean (%)	SD
1	Regularly	53.3	60.9	53.7	47.4	50.5	68.6	41.3	35.7	51.4	10.4
2	Sometimes	22.7	25.0	25.6	29.8	43.9	17.1	39.4	25.0	28.6	8.9
3	Rarely	24.0	14.1	20.7	22.8	5.6	14.3	19.3	39.3	20.0	9.8
4	Never	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>TOTAL</b>		<b>100.0</b>									

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; T7=Test7; T8=Test8; SD=Standard Deviation**

From the statistics presented in Table 6.3, it appears that, only slightly more than half (51.4 %) of participants reported attending English classes regularly, while nearly the other half of participants reported attending English classes either sometimes (28.6 %) or rarely (20.0 %).

At this juncture, I hasten to mention that this finding suggests that the many participants might not be motivated in the English course as they do not regularly attend English classes. This observation is corroborated by the data provided by the questionnaire administered to 27 English language teachers (see Appendix 5). In this questionnaire, question 7 requested teacher-participants to tick from a suggested list of ten (10) statements what they believed were variables that most negatively affected their reading classes in particular and English classes in general. The information provided in Appendix 14 signals that, of the 27 teacher-participants, 17 participants (63.0 %) reported that lack of students' involvement in English classes was one of

the variables that most negatively affected their reading classes in particular and English classes in general.

The qualitative reports from the other 26 student-participants provide additional information on students' attendance to English classes. Three main reasons emerged to explain why some participants regularly attended reading classes: first, the growing importance of English in today's world life; secondly, the Congolese government's need to make English a second additional language together with French; and lastly, the different job opportunities the knowledge of English would provide them with. The following six extracts can act as evidence:

**Respondent 1:** English language is very important; the world is in rapid change and I must have at least basic knowledge of English.

**Respondent 2:** There is nothing I can do today without some knowledge of English.

**Respondent 3:** They [the government] want to make English a second international language beside French; so I must make effort to speak it.

**Respondent 4:** They are telling us that at university, English is a compulsory subject in the first year of all faculties. They say that by 2015, students must be able to study different university subjects in French as well as in English. So, I must be prepared now.

**Respondent 5:** What can you do today if you do not know English? To get a good job at all these United Nations organizations that are working here to bring peace in the country, you need to be skillful in English. So I want to be good at English to get a well-paying job in one of these organizations.

**Respondent 6:** I like English because I know that if I am good at English, I will easily get a good job.

**(See Appendix 23 for additional responses provided on a selected basis)**

As for those respondents who reported that they attended English classes either sometimes or rarely, two explanations emerged. The first explanation was that students' less attendance to English classes was a result of teacher's lack of competence and motivation for teaching; and the second explanation related to their perception of English as difficult subject as a result of lack of reading resources. Here are some illustrative reports from five participants:

**Respondent 1:** I would like to be good at English; but it is since Grade 9 that I was disappointed by the English teacher. He was not a good teacher, he was not

motivated and he was coming to teach rarely; then I lost all my motivation. Now, I feel I cannot catch up whatever I can do.

**Respondent 2:** I don't see why I can regularly attend to English classes when my teacher [of English] rarely comes to teach.

**Respondent 3:** Whenever we have English class, I just go out. This is a very difficult course. We don't have textbooks. The text is written on the chalkboard with so many mistakes. I don't feel any interest to attend the English classes regularly.

**Respondent 4:** I think we need some materials for learning English. The use of these materials can motivate us to participate. I have lost all interest in English because of the way it is taught. I don't understand anything.

**Respondent 5:** I don't know if in English the same word can be pronounced differently. Today our teacher pronounces a word in one way; tomorrow he pronounces the same word in another way; few days later another pronunciation. Then what can you learn from such confusion?

(See Appendix 23 for additional responses provided on a selected basis)

The qualitative reports from the 27 teacher-participants provide additional information on the issue of students' participation in English classes. Two main emerging patterns are that students do not regularly attend English classes because they perceive it as a difficult subject; and also because they have low motivation for English subject. Nevertheless, these teacher-participants are of the view that the students who regularly attend English classes are good at English. The following explanations from three teacher-participants can support these views:

**Respondent 1:** Some of my students don't regularly attend English classes. They repeatedly say that English is difficult; and this has as consequence poor proficiency level in English. Those who attend regularly are generally good at English. So I think the only way to be good at English is to attend English classes regularly.

**Respondent 2:** Many students don't attend my classes. When I come in, I find that some students have left the class and they are in the school yard chatting. These students really don't have interest in the English course.

**Respondent 3:** I am always concerned about my students' poor participation in my classes.

(See Appendix 24 for additional responses provided on a selected basis)

### **6.1.2.3. Participants' preferences for the time allocated for learning English**

According to the English curriculum (Programme National d'Anglais, 1988), the number of hours for teaching English is 5 hours per week in general schools, including Commercial and related subjects, and 2 hours per week in other technical schools like Agriculture, Dressmaking, Mechanics etc.

Although the number of hours per week to learn English is determined by the English curriculum, students' preferences for the time allocated to learn English can provide insights for understanding their motivation for reading. In the context of English as a foreign language where the classroom is the main setting for language learning, I hypothesize that motivated students can prefer much time to learn English while less motivated students can prefer a reduced time to learn English.

In order to determine students' preferences for the time allocated to learn English so as to augment the understanding of Grade 12 students' motivation for reading, question 8 of the main questionnaire (see Appendix 5) requested the 496 student-participants to indicate the number of hours per week they preferred to learn English. The two tables on the next page provide information on participants' preferences for the number of hours to learn English for the participants who were actually studying English 5 hours per week (Table 6.4a) and those who were actually studying English 2 hours per week (Table 6.4b). I wish to mention that, in Table 6.4a, the data presented relate to the 420 participants who were included in the six groups to whom the following six tests were administered (T1, T2, T3, T4, T5, T6). On the other hand, in Table 6.4b, the data presented relate to the 76 participants who were included in the two groups to whom test 7 (T7) and test 8 (T8) were administered. However, information on participants' preferences for the number of hours to learn English for all the eight groups together is provided in Appendix 15.

Table 6.4a: Preferences of number of hours per week for 420 respondents who were actually studying English 5 hours per week

Number of hours N=420	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)	Mean (%)	SD
1 hour per week	6.7	9.4	13.4	3.5	6.5	20.0	9.9	6.0
2 hours per week	38.7	23.4	19.5	24.6	23.4	25.7	25.9	6.6
3 hours per week	16.0	26.6	20.7	21.1	22.4	17.1	20.7	3.8
4 hours per week	12.0	17.2	18.3	12.3	15.0	8.6	13.9	3.6
5 hours per week	16.0	21.9	28.1	28.1	23.4	25.7	23.8	4.6
more than 5 hours per week	10.7	1.6	0.0	10.5	9.4	2.9	5.8	4.9
<b>Total</b>	<b>100.0</b>							

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; SD=Standard Deviation

Table 6.4b: Preferences of number of hours per week for 76 respondents who were actually studying English 2 hours per week

Number of hours N= 76	T7 (%)	T8 (%)	Mean (%)	SD
1 hour per week	58.0	54.0	56.0	2.8
2 hours per week	42.0	25.0	33.5	12.0
3 hours per week	0.0	21.0	10.5	14.8
4 hours per week	0.0	0.0	0.0	0.0
5 hours per week	0.0	0.0	0.0	0.0
more than 5 hours per week	0.0	0.0	0.0	0.0
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	

T7=Test7; T8=Test8; SD=Standard Deviation

From the information provided in Table 6.4a, I wish to state the following observations:

Among the participants who were actually studying English five hours per week, only few of them (23.8%) preferred to study English five hours a week as scheduled in the English curriculum. On the contrary, the majority of the participants (a total number of 70.4%) preferred a more reduced time of two hours a week (25.9 %), three hours a week (20.7%), four hours a week (13.9%) or simply one hour a week (9.9%). However, the data indicate that only 5.8 per cent of participants preferred to study English more than five hours a week. This finding suggests that the majority of participants who were actually studying English five hours per week had preference for a reduced time for learning English. This finding further suggests that the majority of participants may have a low motivation for learning English.

On the basis of information in Table 6.4b provided on the previous page, participants who were actually studying English two hours per week had different preferences. Only one-thirds (33.5 %) of participants were satisfied with two hours per week as planned in the English curriculum. On the contrary, more than half of participants (56 %) preferred that this actual time be reduced to one hour per week. On the other hand, only a minority of participants (10.5 %) preferred the actual two hours per week to be increased to 3 hours per week; while no participant (0.0 %) preferred an increase of up to 4 or 5 hours per week. This finding suggests that participants who were studying English two hours per week had preference for lesser time than the actual time planned for learning English. This further suggests that the participants may have low motivation to participate to English classes. Although I intend to elaborate on this issue in the Discussion Chapter, I hasten to state that, in order to be construct-valid, the ESE needs to also take account of students' preferences for the time to learn English as this can provide some indication of their degree of motivation for reading in English.

The qualitative reports from the 26 student-participants provide additional information on the students' preferences for the time allocated for learning English. The main argument that emerged from the participants who preferred to study English in the time planned in the curriculum pointed to the growing importance of English and the necessity to use English in communication and grab some job opportunities. The following excerpts provide illustrative views of three participants:

**Respondent 1:** English as an important subject. If we want to be able to speak, read and write it, we need many hours per week to learn it.

**Respondent 2:** I want more time than five hours [per week] because this enables me to become good at English and stand a chance for grabbing some job opportunities.

**Respondent 3:** I am for an increase of the time for learning English. Two hours are not enough to achieve good communication skills. Yet, if one is not able to communicate, he cannot pretend to now English.

**(See Appendix 23 for additional responses provided on a selected basis)**

However, the majority of participants who preferred a reduced time to learn English had a different view. From my scrutiny of their responses, their arguments reflect an overall negative attitude towards English subject, a belief that English is a challenging subject; and a need to allocate much learning time to other subjects believed to be more important than English subject. The following four excerpts can serve to illustrate these views:

**Respondent 1:** Five hours are too much. It is also too boring. This time must be reduced to two hours as in other technical schools.

**Respondent 2:** This is a difficult course. Five hours are too much; why not one or two hours; this will be enough.

**Respondent 3:** English subject is not as important as Math. I think we need to reduce the time for learning English to two hours so that we can have much time to learn Math subject.

**Respondent 4:** English is not an easy subject. It is good that the time to learn it be simply reduced.

**(See Appendix 23 for additional responses provided on a selected basis)**

### **6.1.3 Grade 12 students' experience with the ESE**

In order to investigate Grade 12 students' experience with the ESE, the first section of the main questionnaire administered to 496 participants included a question (Question 2) that aimed to probe participants' experience with the ESE. In this question, the participants were asked to indicate whether or not they had taken the ESE before. It is worth mentioning that students who have written the ESE before can be considered to have some experience with this test. Table 6.5 on the next page presents information on the 496 participants' experience with the ESE.

Table 6.5: Students' experience with the ESE

Have you written the ESE before? N=496		T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)	T7 (%)	T8 (%)	Mean (%)	SD
1	Yes, I have written the ESE before	34.7	35.3	29.3	31.6	25.2	19.9	24.2	27.1	28.4	5.3
2	No, I have never written the ESE before	65.3	64.7	70.7	68.4	74.8	80.1	75.8	72.9	71.6	5.0
<b>TOTAL</b>		<b>100.0</b>									

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; T7=Test7; T8=Test8; SD=Standard Deviation**

The information contained in Table 6.5 indicates that, across the eight test groups, only few participants (28.4%) reported having some experience with the ESE by having written it sometime before in the past. On the contrary, the majority of participants (71.6 %) across the eight test groups reported not having written the ESE before, suggesting a lack of some experience with the ESE.

However, having taken the ESE test before is not the only way students can acquire some experience with the ESE. Another way of training students to familiarize themselves with the ESE can be the teacher's frequent use of the multiple-choice testing method in classroom assessment. Since the multiple-choice question is the assessment method used with the ESE, I hypothesized that its use in classroom assessment can help students get some experience with the ESE; therefore, they can have some familiarity with the ESE and write it with much confidence.

In order to investigate this issue, I analyzed teacher-participants' responses from question 3 of the contextual questionnaire (see Appendix 6). In this question, participants were asked to indicate their frequency of use (Usually-Sometimes-Rarely-Never) of seven suggested methods used in classroom assessment. These seven methods were: (1) Open questions, (2) Multiple-choice questions, (3) Oral questions, (4) Essay questions, (5) Yes-No questions, (6) True-False questions, and (7) Debates.

Table 6.6 on the next page provides information on participants' frequency of use of different classroom assessment methods. According to the interpretation of information provided in this table, the open questions were the classroom assessment method *usually* used by nearly all participants (96.3 %). However, the MCQ was *sometimes* used by only one in four (25.9 %)

participants, while it was *rarely* used by nearly half of the participants (48.1 %). On the other hand, one in four (25.9%) participants *never* used the MCQ in classroom assessment. Although I will elaborate on this finding in the Discussion Chapter, I hasten to state at this juncture that, this finding suggests that the multiple-choice question, which is the ESE method, is not a central method of classroom assessment. Therefore, this finding prompts attention to the issue of students' familiarity with the ESE.

Table 6.6: Frequency of classroom assessment methods

Testing methods used in classroom		Usually		Sometimes		Rarely		Never		Total	
		N	%	N	%	N	%	N	%	N	%
1	Open questions	26	96.3	0	0.0	1	3.7	0	0.0	27	100.0
2	Multiple-choice questions	0	0.0	7	25.9	13	48.2	7	25.9	27	100.0
3	Oral questions	0	0.0	0	0.0	9	33.3	18	66.7	27	100.0
4	Essay questions	0	0.0	3	11.1	13	48.2	11	40.7	27	100.0
5	Yes-No questions	3	11.1	3	11.1	19	70.4	2	7.4	27	100.0
6	True-False questions	0	0.0	3	11.1	12	44.4	12	44.4	27	100.0
7	Oral presentations	0	0.0	0	0.0	6	22.2	21	77.8	27	100.0

In order to gain a deeper understanding of teacher-participants' use of different classroom assessment methods, I also scrutinized their comments provided in the qualitative reports. In these reports, the majority of the participants provided the reasons of their preference for the open questions in classroom assessment. Among the main reasons mentioned, some participants believed that the open question is easier to construct while other participants mentioned that the open question best provides an indication of students' language proficiency. The following extracts from five respondents illustrate these two views:

**Respondent 1:** It is easy for me to construct open questions. I do not have time to construct multiple-choice questions.

**Respondent 2:** It just takes me few minutes to design a test when I use open questions.

**Respondent 3:** I use open questions because they give me the opportunity to know which students have understood the course materials and which ones have not

understood. With the MCQ, you cannot tell that. Many students easily pass answers to their colleagues.”

**Respondent 4:** Open questions are good because they enable to check which students have understood the course materials and which ones have problems.

**Respondent 5:** With open questions I am able to identify areas where my students have problems so that I can plan corrective tasks.

**(See Appendix 24 for additional responses provided on a selected basis).**

On the other hand, the few participants who used the multiple-choice questions justified their use by the need to prepare their students to write the ESE. The following are reasons formulated by three participants:

**Respondent 1:** I know that in classroom assessment, we must initiate our students to the MCQ because the ESE is in an MCQ format.

**Respondent 2:** I must familiarize my students with multiple-choice questions. Otherwise, they will look confused when they write the ESE,

**Respondent 3:** My students should be able to identify all tricks used in the multiple-choice questions. This is the only way they can build confidence and develop positive attitude towards the ESE.

**(See Appendix 24 for additional responses provided on a selected basis)**

Finally, the participants who never used the MCQ mentioned various reasons that include the difficulty to construct them, the lack of samples of ESE papers, the belief that MCQ do not provide a good indication of students’ abilities, and/or that they are easy to answer. The following comments can serve as evidence:

**Respondent 1:** I never use the MCQ. They are very difficult to construct; and they need so much time. I think I am not prepared to do that. Also, they give an unfair advantage to weak students.

**Respondent 2:** I never use the MCQ because I don’t have samples of ESE test papers. The time I get some of these ESE papers, I will use them.

**Respondent 3:** I don’t like them [MCQ]. They don’t provide a good indication of students’ abilities.

**Respondent 4:** I never use them. They are difficult to construct. Besides, they make students lazy.

**Respondent 5:** They are easy to answer. I don't use them.

**(See Appendix 24 for additional responses provided on a selected basis)**

The aforementioned quantitative and qualitative findings relating to the classroom assessment methods suggest that the multiple-choice technique, although it is the assessment method in the national test, is not the main classroom assessment method. This is to suggest that the use of this method in classroom context is very limited; and this can have an impact on students' performance on the ESE on the ground that the students can find the multiple-choice technique challenging as they are not familiar with this technique. In light of this suggestion, I hasten to state that, in order to be construct-valid, the ESE needs to also include tasks that also take into consideration the candidates' degree of familiarity with the multiple-choice technique.

#### **6.1.4 Grade 12 students' attitudes towards the ESE**

In order to investigate students' attitudes towards the ESE, the first section of the main questionnaire administered to 496 participants included a question (Question 6) that aimed to elicit their attitudes towards the ESE. In this question, the participants were asked to indicate their degree of agreement with each of the five statements relating to their attitudes towards the ESE. The first statement sought to determine whether or not participants agreed that all Grade 12 students must write an English test. The second statement intended to find out whether or not participants agreed that the ESE was just a way of failing the candidates in the national test. The third statement aimed at finding out whether or not participants agreed that they sometimes did not feel to write the English test. The fourth statement sought to know whether or not participants agreed that basing the ESE on reading comprehension only was fair. Finally, the fifth statement aimed to elicit information on whether or not participants thought that they did not mind failing on the English test. Additional information on students' attitudes towards the ESE was also provided by the qualitative reports obtained from the 26 student-participants to whom the contextual questionnaire was administered, and from the contextual questionnaire administered to 27 teacher-participants. Table 6.7 on the next page provides information on participants' attitudes towards reading as reflected in their responses to the questionnaire.

Table 6.7: Participants' attitudes towards the ESE

<b>1. All grade 12 students must write an English test</b>										
<b>N=496 Options</b>	<b>T1 (%)</b>	<b>T2 (%)</b>	<b>T3 (%)</b>	<b>T4 (%)</b>	<b>T5 (%)</b>	<b>T6 (%)</b>	<b>T7 (%)</b>	<b>T8 (%)</b>	<b>Mean (%)</b>	<b>SD</b>
Agree	57.3	60.9	59.8	68.4	74.8	68.6	41.7	35.7	58.4	13.5
Disagree	28.0	39.1	26.8	29.8	25.2	14.3	45.8	25.0	29.3	9.5
Don't know	14.7	0.0	13.4	1.8	0.0	17.1	12.5	39.3	12.4	12.9
<b>Total</b>	<b>100.0</b>									
<b>2. The ESE is just a way of failing the candidates at the national test</b>										
Agree	21.3	14.1	14.6	31.6	13.1	14.3	29.2	21.4	20.0	6.8
Disagree	53.3	76.6	81.7	68.4	79.4	51.4	56.3	57.1	65.5	11.7
Don't know	25.3	9.4	3.7	0.0	7.5	34.3	14.6	21.4	14.5	11.0
<b>Total</b>	<b>100.0</b>									
<b>3. I sometimes feel that I don't need to write the English test</b>										
Agree	38.7	6.3	26.8	31.6	34.6	17.1	14.6	21.4	23.9	10.3
Disagree	41.3	93.8	72.0	68.4	63.6	82.9	85.4	78.6	73.2	15.1
Don't know	20.0	0.0	1.2	0.0	1.9	0.0	0.0	0.0	2.9	6.5
<b>Total</b>	<b>100.0</b>									
<b>4. I feel that basing the ESE on reading comprehension only is not fair</b>										
Agree	54.7	56.3	63.4	71.9	66.3	48.6	64.6	46.4	59.0	8.4
Disagree	21.3	23.4	23.2	8.8	25.2	34.3	27.1	35.7	24.9	7.8
Don't know	24.0	20.3	13.4	19.3	8.5	17.1	8.3	17.9	16.1	5.3
<b>Total</b>	<b>100.0</b>									
<b>5. I don't mind failing on the English test</b>										
Agree	9.3	6.3	9.8	21.1	12.2	14.3	14.6	28.6	14.5	7.2
Disagree	90.7	93.8	90.2	79.0	87.9	85.7	85.4	71.4	85.5	7.2
Don't know	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Total</b>	<b>100.0</b>									

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; T7=Test7; T8=Test8; SD=Standard Deviation

In light of the data provided in this table, I wish to make the following observations:

(1) Concerning the first statement, the information contained in Table 6.7 suggests that, across the eight test groups, not all participants were of the view that all Grade 12 students must write

an English test as requested by the Department of Education. Rather, only slightly more than half (58.4 %) of participants were of the view that all Grade 12 students must write an English test, while a considerable number of participants (29.3 %) were of the opposite view. Also, some participants (12.4 %) could not tell whether or not it is a must for Grade 12 to write an English test.

(2) In regard to the second statement, although a minority, some participants (20.0%) believed that the ESE is a way of failing candidates on the national test. Nevertheless, the majority of participants (65.5%) did not support this view. Besides, some other participants (14.5 %) could not be able to say whether or not the ESE is a way of failing candidates at the national test.

(3) With respect to the third statement, although it is an obligation for all Grade 12 students to write an English test, nearly one in four participants (23.9 %) reported that they sometimes felt they did not need to write this test. However, the majority of respondents (73.2 %) had not this feeling; rather, they agreed that all Grade 12 students must write the ESE.

(4) Concerning the fourth statement, the majority of participants (59.0 %) were of the view that it is not fair that the ESE is based only on text passage comprehension and that other language skills are not tested. On the contrary, one in four participants (24.9 %) did not see any problem that the ESE be based on text passage comprehension, while some other participants (16.1%) could not tell whether or not it is fair to base the ESE on text passage comprehension only. This finding corroborates results from the questionnaire that was further administered to teachers where data indicate that 52 % of teachers were of the view that the ESE must not test only reading; rather it must also test other skills such as speaking, listening, and writing.

(5) Regarding the last statement, although in a minority (14.5 %), some participants reported that they did not mind failing the ESE. However, the majority of participants (85.5 %) were not of this view.

These findings provide evidence of some negative attitudes on the part of some participants towards writing the ESE as these participants sometimes feel they must not write the ESE, or that the ESE is a way of failing them to the national test, or again that it is not a fair test since it is based only on text passage comprehension and neglects other language skills. Although I intend elaborating on this issue in the Discussion Chapter, I hasten to state that, since validity is about

the appropriateness and relevance of meanings and interpretations made on the basis of test scores, I believe that, in order to be construct-valid, the ESE needs to include tasks that are also sensitive to examinees' attitudes towards the test.

The qualitative reports from the other 26 Grade 12 student-participants provide additional information on the students' attitudes towards the ESE. As to the first statement, students who were of the view that all Grade 12 students must write an English test formulated three main reasons: the first reason relates to the importance of the English subject, the second reason relates to the obligation students have to write an English test, while the third reason relates to the students' needs to be assessed on their abilities in English. These following five extracts illustrate this point:

**Respondent 1:** This is an important subject. We must write it.

**Respondent 2:** It is part of the curriculum; and we must write it. We have no choice.

**Respondent 3:** If you refuse to write this test, you lose 9 per cent; and your chance to get your certificate becomes reduced.

**Respondent 4:** It is good to be evaluated by some people other than your teachers. I feel proud to write the English test.

**Respondent 5:** An external exam is the only way to ensure that what we were taught is what is requested to be taught. I believe the English test is a good measure for that.

**(See Appendix 23 for additional responses provided on a selected basis)**

On the other hand, participants who were not of the view that all Grade 12 students must write an English test linked their views to the difficulty of English subject and the fear of not passing the test. The following comments from two participants can support this point:

**Respondent 1:** Why do we need to write an English test when they know we are going to fail? Why don't they make it an optional subject? All of us don't need English. How many of those who teach English have gone to an English country? We can live without this language, and they don't need to make it a must.

**Respondent 2:** It would be good for us not to write the English test. We can just study English as a school subject; but not write it.

**(See Appendix 23 for additional responses provided on a selected basis)**

Concerning the second statement, participants who were of the view that the ESE was a way of failing candidates in the national test provided as main explanation the fear to fail this test, as this failure might result either in not obtaining the national certificate or obtaining it with low mark that cannot enable them to have direct access to university. The following comments can support this point:

**Respondent 1:** I am sure I am not going to get any marks in this test [ESE]. I am working hard in other subjects to catch up. I'm afraid I may not get my certificate with good marks and be denied to start university by next year.

**Respondent 2:** They know that English is difficult; but they have included nine questions [9 percentages] for the English test. We are going to fail.

**Respondent 3:** They know that most students are not good at English; and that we do not like it. Still, they design exams that are very difficult. This is to make us fail.

(See Appendix 23 for additional responses provided on a selected basis)

Regarding the third statement, most participants who were of the view that they sometimes felt they did not need to write an English test did not make any sound comment (see participants' comments in Appendix 23). Nevertheless, it might have appeared that fear of failing the test might sometimes have created in these students an attitude of unwillingness to write the English test. I believe an interview might have clarified this issue better.

As far as the fourth statement is concerned, participants who were of the view that the ESE must test not only reading but also other skills based their view on two arguments: first, they believed that reading was not the only skill they needed to acquire while learning English; instead, other skills were equally important. Second, they believed that reading comprehension questions appeared to be more difficult than questions testing other language aspects. One participant who was of this view argued on the difficulty of comprehension questions as every reader could construct his/her own comprehension of the text; thus, making it difficult to expect a single correct answer from possible multiple comprehensions. The following comments from three participants clarify these views:

**Respondent 1:** Why only reading? It is as if reading is the only thing we must know.

**Respondent 2:** If the ESE includes sections of questions on grammar and vocabulary; we will be able to pass it quite easily. But reading [comprehension] questions are difficult because you cannot understand the text the way it is written. You can have your understanding and when you look at alternatives you choose one you think relates to your understanding; yet this is a wrong answer. But with a question on grammar, we all understand it in the same way. It is good to include other questions on grammar and vocabulary.

**Respondent 3:** I want the English test to include also questions on grammar and vocabulary; and not only on text comprehension.

**(See Appendix 23 for additional responses provided on a selected basis)**

Qualitative reports from the 27 teacher-participants provided additional information on the issue of basing the English test on text comprehension only. Those participants who were of the view that the English test must not only test text comprehension, but also reading presented three arguments: To some, there is a need to focus on all skills in order to promote language proficiency in the learners. To others, the actual focus on reading has resulted in a washback effect in English language teaching. Still, others believed that the teaching of English must put priority on spoken language, not written. The following are three excerpts that support each of these three views:

**Respondent 1:** Why do we need to reduce English proficiency to reading comprehension? It is as if other skills are of no use. I think other skills must be tested to ensure that students are capable to be competent in English.

**Respondent 2:** You see what happens now. Other skills are neglected and even in course planning, these other skills deserve no attention.

**Respondent 3:** Language is speech; not writing. So, reading does not reflect proficiency in English language.

**(See Appendix 23 for additional responses provided on a selected basis)**

On the other hand, three arguments were articulated by teacher-participants who were of the view that the English test must test only text comprehension. The first argument pointed to the place of comprehension in language learning. The second argument pointed to the difficulty teachers perceive their students have in other skills, while the third argument related to the

practicality of assessing oral skills. The following three excerpts illustrate each of these arguments:

**Respondent 1:** Since comprehension is basic to language, it does not matter to base the ESE on reading comprehension.

**Respondent 2:** If other skills like speaking and writing are tested, most students will fail the test.

**Respondent 3:** It will be too costly to attempt to test oral skills like listening and speaking. This cannot be feasible.

**(See Appendix 23 for additional responses provided on a selected basis)**

In light of the above findings from both quantitative and qualitative data, it appears that participants had different attitudes towards the English state examination although positive attitudes appear to prevail than negative attitudes. Therefore, it is important for test developers to be careful by constructing the ESE that includes test tasks that also profile students' attitudes.

#### **6.1.5. Additional insights on Grade 12 students' attitudes and motivation towards the ESE**

Additional insights on students' attitudes and motivation for the ESE are investigated through the description of participants' preferences for the number of hours to write the ESE, the number of questions to be included in the test, the number of test alternatives, and the number of parallel forms.

- **Participants' preferences for the number of hours to write the test**

Students' preferences for the number of hours to write the ESE can provide additional information on their motivation for the test and attitude towards the test. In regard to students' skills in English, their motivation for reading as well as their attitudes towards the ESE, it can be hypothesized that when students are less motivated for reading and they have negative attitudes towards the ESE, they may have preference for more time than needed for writing the ESE.

In my previous study (Katalayi, 2011), I mentioned that the actual number of hours allocated to write the ESE is roughly 2 hours for candidates who have to answer 9 questions and 1 hour and half for candidates who have to answer 5 questions. I therefore concluded that this time was too

much and I suggested that 1 hour was enough for all candidates to write the ESE. I wish to mention that this observation serves as a guide for understanding the extent to which participants' preferences are in accordance with the suggested time of one hour as investigated by Katalayi (2011).

Question 8 of the first section of the main questionnaire administered to the 496 student-participants aimed to elicit their preferences for the number of hours to write the ESE. In this question, the participants were asked to indicate the number of hours they preferred to write the ESE. Table 6.8a and Table 6.8b on the next page present statistics on participants' preferences for the number of hours to write the ESE for the 420 participants included in the six groups who actually have 2 hours (6.8a) and the 76 participants included in the two groups who actually have 1 hour and half to write the ESE.

From the information provided in Table 6.8a, I wish to state the following:

For the 420 participants who were in the six groups where ESE is actually written in 2 hours, the majority of these participants (57.4%) preferred to write the ESE either in 2 hours (44.8 %) as planned by test administrators or in more than two hours (12.6 %). On the contrary, slightly more than one-thirds (35.3%) of participants preferred a reduced time of 1 hour and half, and only 7.4 per cent of participants preferred to write the ESE in 1 hour.

Table 6.8a: Number of hours to write the test for groups who actually have 2 hours to write the ESE

<b>Number of hours N=420</b>	<b>T1 (%)</b>	<b>T2 (%)</b>	<b>T3 (%)</b>	<b>T4 (%)</b>	<b>T5 (%)</b>	<b>T6 (%)</b>	<b>Mean (%)</b>	<b>SD</b>
1 hour	10.7	12.5	8.5	7.0	2.8	2.9	7.4	4.0
1 hour and half	26.7	32.8	32.9	45.6	33.6	40.0	35.3	6.6
2 hours	58.7	40.6	46.3	33.3	49.5	40.0	44.8	8.8
more than 2 hours	4.0	14.1	12.2	14.0	14.0	17.1	12.6	4.5
<b>TOTAL</b>	100.0	100.0	100.0	100.0	100.0	100.0	100.0	

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; SD=Standard Deviation**

Table 6.8b: Number of hours to write the test for groups who actually have 1hour and half to write the ESE

<b>Number of hours N=76</b>	<b>T7 (%)</b>	<b>T8 (%)</b>	<b>Mean (%)</b>	<b>SD</b>
1 hour	27.1	46.4	36.8	13.7
1 hour and half	66.7	53.6	60.1	9.3
2 hours	6.3	0.0	3.1	4.4
more than 2 hours	0.0	0.0	0.0	0.0
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	

**T7=Test7; T8=Test 8; SD=Standard Deviation**

On the basis of data provided in Table 6.8b, I wish to report this finding:

The majority of participants (60.1%) who were in the two groups where the ESE is written in one and half hour preferred to write the ESE in 1 hour and half as planned by test administrators. Besides, although in a minority (3.1%), some participants preferred this time to be increased to two hours. On the contrary, slightly more than one-third (36.8%) of participants had preference for a reduced time of one hour.

However, in order to understand the reasons why student-participants tended to have preference for more time to write the ESE, I scrutinized the qualitative reports from the 26 student-participants. For participants who preferred two hours or more, some of them justified their preferences for an extended time to the possibility of assistance and collaboration. These three excerpts illustrate this view:

**Respondent 1:** Two hours are not much. This is not our language. We need much time to write the test, maybe there can be some kind of assistance and if the time is reduced; we will not be able to get this assistance and collaborate.

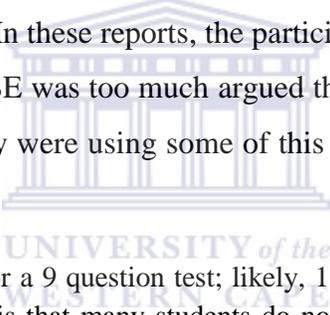
**Respondent 2:** With 4 different alternate forms, we need much time so that we can check who has got the same form with me. After each of us has answered the test questions, we can therefore try to compare our answers.

**Respondent 3:** As English is a very challenging subject, the only one way to pass it is to organize ourselves in order to get answers from outside. I am sure that if we are given much time; we can do so.

(See Appendix 23 for additional responses provided on a selected basis)

In order to validate the afore-mentioned views relating to the number of hours for writing the ESE, I also examined the data provided by the contextual questionnaire administered to the 27 teacher-participants. The information provided in the table in Appendix 16 indicates that, less than half (40.0 %) of teachers were of the view that the actual time allocated to students for writing the ESE is too much; therefore, this time needs to be reduced. However, nearly one-thirds (30.0 %) of participants did not support this view as they reported that the actual time was quite proportional to time needed to complete the ESE tasks.

In order to understand teacher-participants' differing views, I also examined their comments provided in the qualitative reports. In these reports, the participants who were of the view that the actual time foreseen to write the ESE was too much argued that students did not actually use this time for writing the test; rather they were using some of this time for collaboration. One teacher explained:



2 [two] hours are too much for a 9 question test; likely, 1 hour and half is too much for only 5 questions. The reality is that many students do not use all this time to write the test; instead, they use much of this time for struggling to communicate among each other.

In light of teacher-participants' comments in qualitative reports, the participants who preferred the actual time for writing the ESE to be maintained mentioned the need for students to have much time so as to cope with the English test. The following comments can clarify this position:

**Respondent 1:** The majority of Grade 12 students have proficiency deficit in English. I believe students need much time to read the text passage as many times as they can so as to comprehend it. So, the actual time is proportional to the test demands and students' skills.

**Respondent 2:** Our students are really struggling with reading. They need much time.

**Respondent 3:** Since English is a challenging subject, I think students need much time. This will help them to read the text and answer all the questions. Otherwise, they will leave some questions unanswered.

(See Appendix 24 for additional responses provided on a selected basis)

Although I will elaborate on this issue in the Discussion Chapter, I hasten to state that, these findings suggest that the majority of student-participants and some teacher-participants wanted the actual time for writing the test to be maintained or to be increased, a view that does not tally with Katalayi's (2011) suggestion. Therefore, I can conclude that participants' preferences for an increased time (or at least to maintain the actual time) for writing the ESE can suggest a negative attitude they have towards the ESE.

▪ **Participants' preferences for the number of questions to be included in the ESE**

Students' preferences for the number of questions to be included in the ESE provide additional information on their motivation for the test and their attitudes towards the test. I hypothesized that Grade 12 students with some negative attitudes towards the ESE may tend to prefer a reduced number of questions on the ground that few questions may not affect their final results on the national test. At this juncture, I wish to mention that the actual number of questions included in the ESE is 9 questions for candidates who study English five hours per week and 5 questions for those candidates who study English two hours per week. In my previous study (Katalayi, 2011), I reported that this number of questions was too insufficient and I suggested that the ESE be lengthened to at least 20 questions for all candidates regardless their groups or subject areas.

In order to investigate this issue, question 9 of the first section of the main questionnaire submitted to the 496 student-participants (see Appendix 4) requested them to indicate their preferences for the number of questions to be included in the ESE. Table 6.9a on the next page presents data on participants' preferences for the number of questions for the 420 participants included in the six groups to whom a nine question test was actually administered. On the other hand, Table 6.9b on the next page presents data on participants' preferences for the number of questions for the 76 participants included in the two groups to whom a five question test was actually administered.

My examination of the information in Table 6.9a reveals the following:

Slightly more than one-fourths (26.4%) of participants who belonged to the groups where the ESE includes 9 questions preferred the ESE to include the actual number of 9 questions. On the

contrary, slightly more than half of participants (50.9%) preferred the number of questions to be reduced to either 5 questions (35.9 %) or just 3 questions (15.0 %). Furthermore, although a minority, some participants preferred the actual number of questions to be increased to 12 questions (16.3 %) or to more than 12 questions (6.4 %).

Table 6.9a: Participants' preferences for the number of questions to be included in the ESE for groups who actually had 9 questions

<b>Number of questions N=420</b>	<b>T1 (%)</b>	<b>T2 (%)</b>	<b>T3 (%)</b>	<b>T4 (%)</b>	<b>T5 (%)</b>	<b>T6 (%)</b>	<b>Mean (%)</b>	<b>SD</b>
3 questions	10.7	17.2	23.2	8.8	15.9	14.3	15.0	5.1
5 questions	38.7	31.3	40.2	40.4	30.8	34.3	35.9	4.4
9 questions	20.0	39.1	20.7	21.1	31.8	25.7	26.4	7.6
12 questions	16.0	9.4	12.2	26.3	14.0	20.0	16.3	6.1
More than 12 questions	14.7	3.1	3.7	3.5	7.5	5.7	6.4	4.4
<b>TOTAL</b>	<b>100.0</b>							

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; SD=Standard Deviation**

UNIVERSITY of the  
WESTERN CAPE

Table 6.9b: Participants' preferences for the number of questions to be included in the ESE for groups who actually have 5 questions

<b>Number of questions N=76</b>	<b>T7 (%)</b>	<b>T8 (%)</b>	<b>Mean (%)</b>	<b>SD</b>
3 questions	58.3	39.3	48.8	13.5
5 questions	41.7	50.0	45.8	5.9
9 questions	0.0	10.7	5.4	7.6
12 questions	0.0	0.0	0.0	0.0
More than 12 questions	0.0	0.0	0.0	0.0
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	

In light of my scrutiny of the statistics in Table 6.9b, I wish to report the following finding:

Less than half (45.8 %) of participants who belonged to the groups where the ESE includes 5 questions preferred the ESE to include the actual number of 5 questions. On the other hand, nearly half of the participants (48.8%) preferred the number of questions to be reduced to 3 questions. Although in a minority, some participants (5%) preferred the actual number of 5 questions to be increased to 9 questions as in other groups.

Although I intend to elaborate on the findings from the preceding two tables (Tables 6.9a and 6.9b) in the Discussion Chapter, I wish to state, at this juncture, that the majority of student-participants preferred the actual number of ESE questions to be reduced, or at least to be maintained.

In order to understand the reasons of participants' preferences, I examined the comments furnished in the qualitative reports of 26 student-participants. In light of these comments, it appears that the participants who preferred the actual number of questions to be reduced explained their choice by their belief that few questions in the English test were not likely to negatively affect their general performance on the national test. The following are extracts from two respondents:

**Respondent 1:** When the ESE has 9 questions, this number of questions accounts for 9 % of the total marks. Since we are not sure to get many questions right, a reduced number of questions is advantageous to us; I can fail in English but still get my certificate with good marks.

**Respondent 2:** Imagine the ESE included only two or three questions. If I don't get these questions right, I cannot worry because my chances to get my certificate are still intact.

**(See Appendix 23 for additional responses provided on a selected basis)**

However, the student-participants who preferred the ESE to include more than the actual number of 9 or 5 questions gave two main reasons. To some, more questions could cover the course content and this nearly could reflect their actual abilities. To others, more questions increase their chance to pass the test as some of these questions may be easy. These extracts from four participants illustrate these two points of view:

- Respondent 1:** 9 [nine] questions are not enough. How can we learn English for 4 years and at the end they ask 9 questions to check our ability? This cannot provide a clear indication of our abilities. We need many questions.
- Respondent 2:** Those people cannot expect to include everything we have learned in 5 questions. This is not good. More questions are needed.
- Respondent 3:** I believe that I stand a chance to pass the ESE if it includes more questions than it does now. With few questions, you take all risks to get most of them either right or wrong. I am not for taking such a risk.
- Respondent 4:** I like the ESE to include more than 20 questions. This gives more chance to pass this test.

(See Appendix 23 for additional responses provided on a selected basis)

▪ **Students' preferences for the number of alternatives to be included in the ESE**

Students' preferences for the number of alternatives to be included in each test question also yielded additional information on their motivations for the test and their attitudes towards the test. The actual number of alternatives is six among which five alternatives are explicitly given and the sixth one is implied; that is, examinees select it in case the test question does not include the correct alternative among the five suggested. I hypothesized that, students who are not good at English and who have some negative attitudes to the ESE can prefer the test to include few alternatives as this can give them a high chance to guess the correct answer.

In order to investigate this issue, question 10 of the first section of the main questionnaire administered to 496 student-participants (see Appendix 4) requested the participants to indicate their preferences for the number of alternatives to be included in each test question. Table 6.10 on the next page provides information on participants' preferences.

In light of my inspection of the statistics in Table 6.10, I wish to mention the following considerations:

Across all test groups, no participants (0.0 %) preferred the ESE to have the actual number of 6 alternatives. On the contrary, more than half (53.8 %) of participants had preference for a 3 alternative test. However, less than half of participants preferred either the 4 alternative test (26.5

) or the 5 alternative test (19.7 %). This finding suggests that all the participants preferred the actual number of six alternatives to be reduced. Although I intend to elaborate on this finding in the Discussion Chapter, I hasten to state that this finding calls for a need to question the actual number of six alternatives.

Table 6.10: Participants’ preferences for the number of alternatives to be included in each test question

Number of alternatives N=496	T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)	T7 (%)	T8 (%)	Mean (%)	SD
3 alternatives	60.0	62.5	41.5	40.4	43.9	48.6	62.5	71.4	53.8	11.7
4 alternatives	22.7	14.1	24.4	40.4	23.4	34.3	31.3	21.4	26.5	8.3
5 alternatives	17.3	23.4	34.2	19.3	32.7	17.1	6.3	7.1	19.7	10.3
6 alternatives	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>TOTAL</b>	<b>100.0</b>									

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; T7=Test 7; T8=Test8; SD=Standard Deviation

Student-participants’ qualitative reports provided additional information on students’ preferences. The main reason why the majority of respondents preferred a three alternative test was that this reduced number of alternatives is not likely to confuse examinees compared to the actual number of six alternatives. Also, these participants were of the view that with a three alternative test, their chances of getting the correct answer by simply guessing was maximized. The following are comments from two respondents:

**Respondent 1:** 3 [three] alternatives are good because they do not confuse us. So many alternatives create much confusion.

**Respondent 2:** 3 [three] alternatives are good. We can easily guess the correct answer. With 6 alternatives, you cannot easily guess.

**Respondent 3:** With 6 alternatives, it becomes difficult for us get the correct option. But with only three alternatives, it is easy.

(See Appendix 23 for additional responses provided on a selected basis)

- **Students’ preferences for the number of parallel/alternate forms to be included in the ESE**

Students’ preferences for the number of parallel/alternate forms to be included in the ESE provide additional information on their motivations for the ESE and their attitudes towards this test. The actual number of alternate forms is 4 for candidates whose test includes 9 questions and 2 forms for those whose test includes 5 questions.

In order to investigate this issue, I have analyzed the responses provided by the 496 participants to question 11 of the first section of the main questionnaire (see Appendix 4). The aim of this question was to probe participants’ preferences for the number of alternate forms to be included in the ESE. Table 6.11a summarizes participants’ responses on the number of alternate forms for the 420 participants included in the six groups that actually have 4 alternate forms. On the other hand, Table 6.11b summarizes participants’ responses on the number of alternate forms for the 76 participants included in the two groups that actually have two forms.

Table 6.11a: Preferred number of parallel forms for groups who actually have 4 forms

<b>N parallel forms N=420</b>	<b>T1 (%)</b>	<b>T2 (%)</b>	<b>T3 (%)</b>	<b>T4 (%)</b>	<b>T5 (%)</b>	<b>T6 (%)</b>	<b>Mean (%)</b>	<b>SD</b>
1 single form	54.7	39.1	45.1	40.4	43.9	45.7	44.8	5.0
2 parallel forms	29.3	28.1	26.8	38.6	32.7	42.9	33.1	5.8
4 parallel forms	16.0	32.8	28.1	21.1	23.4	11.4	22.1	7.1
<b>TOTAL</b>	<b>100.0</b>							

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; SD=Standard Deviation**

Table 6.11b: Preferred number of parallel forms for groups who actually have 2 forms

<b>N parallel forms N=76</b>	<b>T7 (%)</b>	<b>T8 (%)</b>	<b>Mean (%)</b>	<b>SD</b>
1 single form	72.9	71.4	72.2	1.1
2 parallel forms	27.1	28.6	27.8	1.1
4 parallel forms	0.0	0.0	0.0	0.0
<b>TOTAL</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	

**T7=Test 7; T8=Test8; SD=Standard Deviation**

In light of the information provided in Tables 6.11a, I wish to report the following finding:

For the 420 participants who were in the six groups where the ESE has four parallel forms, only a minority of them (22.1 %) preferred the ESE to have the actual number of 4 parallel forms. On the contrary, across all the six groups, the majority of participants had preference for a single form test (44.8 %) first, and the two parallel form test (33.1 %) next.

On the basis of the information in Tables 6.11b I wish to report the following finding:

For the 76 participants who were in the two groups where the ESE has two parallel forms, only a minority of them (27.8 %) preferred the ESE to have the actual number of two parallel forms. On the contrary, the majority of participants (72.2 %) preferred a single form test.

These two findings suggest that the student-participants preferred a test with a reduced number of alternate forms; if not just a single form test. Although I shall elaborate on this issue in the Discussion Chapter, I hasten to state here that these findings suggest participants' low motivation for and negative attitude towards the ESE. Therefore, in order to be construct-valid, I expect the ESE to include tasks that also reflect these identified students' characteristics.

Participants' qualitative reports provided additional information on their preferences for the single form test. My scrutiny of the comments furnished by the 26 student-participants has revealed two main reasons that account for participants' preference for a single form test. First, the belief that the use of a single form test is advantageous as it can enable students to exchange answers during test writing. Secondly, the belief that the use of a single form test not only reflects the classroom testing where all students write the same test, but it also reflects fairness as everybody is subjected to the same test. The following five excerpts can illustrate these two reasons:

- Respondent 1:** I want a single form. This enables us to easily compare answers. With 4 alternate forms, you must be careful to look at your neighbor's answer sheet.
- Respondent 2:** I like a single form test because it is easy for us to collaborate.
- Respondent 3:** These people want only to make our lives difficult. With four alternate forms, it becomes difficult to know who has the same exam as you. Therefore, you cannot have any help.

**Respondent 4:** When we study English, we are in the same class. The teacher asks the same questions to everybody. Those who succeed are those who deserve to succeed. But when students who have studied under the same conditions are given different questions, this is no longer fair because some [parallel] forms might be easier than other [parallel] forms. So if all candidates are given the same form; the assessment of these students is fair.

**Respondent 5:** Why so many complications? When we write school examinations, we are all given the same questions. Why do we need to have different questions in the state examination? Do they want to say that school examinations are not good?

**(See Appendix 23 for additional responses provided on a selected basis)**

The knowledge of the students' characteristics is not sufficient to fully understand the context of the ESE. Additional information needs to be provided on the Grade 12 English language teacher, and this is the concern in the following section.

## **6.2 The Grade 12 English language teacher characteristics**

The understanding of the Grade 12 English teacher characteristics provides additional information that augments the understanding of the actual context of the ESE. Therefore, in this section, I investigate the Grade 12 English teacher characteristics in terms of their qualifications, teaching experience, training, motivations for teaching English and attitudes towards teaching profession.

### **6.2.1 Teacher qualification and experience**

Teacher qualification and experience provide information that helps to understand the context of English language teaching; hence, understand students' performance on the ESE. In order to investigate this issue, I have analyzed the responses provided by the 27 teacher-participants to question 1 and question 2 of the contextual questionnaire. Question 1 requested participants to indicate their highest degree in education while question 2 requested them to indicate how long they have been teaching English in Grade 12 (see Appendix 6). Table 6.12a presents information

on teacher-participants' qualifications while Table 6.12b presents information on their teaching experience.

Table 6.12a: Participants' qualification

Qualification		N=27	%
1	B.Ed (Hon.)	9	33.3
2	B.Ed	13	48.2
3	No degree	5	18.5

**B. Ed (Hon.)=Bachelor degree in Education, Honours; B. Ed=Bachelor degree in Education**

Table 6.12b Participants' teaching experience

Teaching experience		N=27	%
1	Less than 3 years	6	22.2
2	Between 3-5 years	10	37.0
3	More than 5 years	11	40.8

The information in Table 6.12a suggests that the majority of teacher-participants had a teaching degree in English that was either a B.Ed (48.2 %) or B.Ed/Honours (33.3 %) degree. On the contrary, participants who had no English teaching degree were only 18.5 per cent. This finding suggests that the majority of teacher-participants were formally qualified to teach English subject.

On the other hand, the information in Table 6.12b indicates that the majority of teacher-participants had some experience teaching English in Grade 12. More specifically, 40.8 per cent of participants had more than five years' teaching experience and 37.0 % had between three and five years' teaching experience. On the other hand, participants who had not any experience in teaching English in Grade 12 were only 22.2 %. This finding suggests that the majority of participants had some experience in teaching English in Grade 12.

### 6.2.2 Teachers' in-service training

In order to gain a deeper understanding on the context of English language teaching, I have also investigated the issue of teachers' in-service training. I hypothesized that when in-service

trainings are regularly organized for teachers, teachers' participation to these trainings contributes to improve their quality; therefore, this contributes to students' achievement. In order to investigate this issue, I have analysed the responses from question 8 of the contextual questionnaire administered to the 27 participants (see Appendix 6). In this question, participants were requested to tick from a list what they believed were variables that most negatively affected the reading classes in particular and English classes in general. The statistics contained the table provided in Appendix 14 suggests that almost all participants (93 %) were of the view that lack of teacher's in-service training was one of the variables that negatively affected reading classes in particular and English classes in general.

In order to augment my understanding of this issue, I have inspected the qualitative reports from the same participants. In most of these reports, participants expressed their unhappiness at the lack of training sessions organized for English teachers. They believed that such a lack was detrimental to teaching profession as teachers were not updated with new developments in teaching. One respondent made this comment:

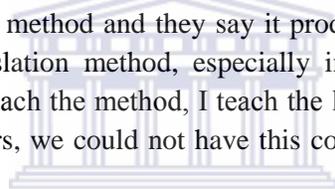
To the best of my knowledge, since 2001 when I graduated and started teaching English; there has never been any training organized for English teachers who are currently practicing. How can teachers keep on with new developments in language teaching? Everything I know of teaching English is what I had learned at university.

Furthermore, participants believed that continuous training, especially in an English speaking country, was important as it is likely to enable teachers who have been trained to teach in an environment where English is not spoken outside the classroom to familiarize themselves with how the language is actually used by people who speak it. One respondent who espoused this view made this argument:

English is not the language used in this country. It is important that those who teach it be regularly trained. This training programme may focus on sending teachers for some time to an English speaking country even those which are neighbor to us like Zambia, Tanzania or Uganda. This may give them the opportunity to learn how language is actually used. Sometimes we just teach bookish English and not actual English; and this does not benefit to learners.

Of utmost importance in teachers' in-service training is the issue of methods to be used for teaching English. Some respondents elaborated on this issue by complaining about the consequences of lack of in-service trainings. According to these respondents, one of these consequences is the actual confusion in the use of teaching methods by English teachers. The following extract is an argument developed by one respondent who was of this view:

I have been teaching English since 1980. Since then, I remember having participated in a training session only two times. The first time was in 1984 when the British Council in Kinshasa organized a training session on the communicative approach to language teaching. The second time was in 1986 when the Peace Corps came to organize a seminar on the teaching of literature. That is all. The past 20 years no seminar has been organized. This has a big consequence on the use of teaching methods. There is a big confusion in our country depending on where you completed your studies. Some teachers pretend they have been better trained because they use the communicative approach. Others repeatedly support that the audio-lingual approach is best especially in situations like our country where the communicative approach cannot fit because of lack of appropriate resources. Still others support the direct method and they say it produces good results. Some even still use the Grammar Translation method, especially in technical schools. Myself I usually support that I don't teach the method, I teach the language. If there were regular seminar organized for teachers, we could not have this confusion. We could know what we are all expected to do.

  
UNIVERSITY of the  
WESTERN CAPE

### **6.2.3 Teacher's motivation for teaching English**

Teacher's motivation for teaching English provides additional information to understand the context of English language teaching; therefore, understand students' performance on the English test. I hypothesized that students' motivation for learning and their attitudes to English in general and reading in particular can also depend on the level of teacher's motivation for teaching. Hence, teachers with high motivation for teaching English can be an asset for students' achievement.

In order to investigate this issue, Question 6 of the contextual questionnaire administered to the 27 participants requested them to indicate their views (agree - disagree - can't tell) on some suggested statements (see Appendix 6). The analysis of participants' responses in the table provided in Appendix 16 indicate that only slightly more than half of teachers-participants (56%) reported that they were motivated to teach English while nearly less than a half (44%)

reported that they were not motivated to teach English. Furthermore, the data in the same table indicate that most participants (85 %) were of the view that their poor working conditions negatively affected their motivation for teaching.

In order to understand the reasons why some teacher-participants indicated that they were not motivated for teaching English, I also examined their comments from the qualitative reports. My scrutiny of these comments suggests that, the poor working conditions were a main reason that explained their low motivation for teaching. The following comments can illustrate this point:

**Respondent 1:** My salary does not enable me to make ends meet. I have to do other jobs to make more money. I don't have time to do properly classwork. I am not motivated.

**Respondent 2:** With the current poor working conditions of Congolese teachers, there is nothing the government can expect from us. In my case, I rarely give tasks to my students because I won't be able to mark them. Yet, since English is a foreign language, learning can be enhanced through the design and use of many tasks that require the learners to use the language. Surely, what I am doing today is not teaching.

(See Appendix 24 for additional responses provided on a selected basis)

Other participants who indicated that they were not motivated for teaching English mentioned among other reasons, the lack of appropriate resources for teaching, the lack of mechanism for their empowerment, and the low motivation from the learners. The following comments can illustrate these reasons:

**Respondent 1:** I wish I were not a teacher. I don't feel I can do what I am supposed to do. With no resources, no empowerment, teaching activity becomes simply a waste of time.

**Respondent 2:** I am really disappointed. When I enrolled at university to become a teacher in the early eighties, I had many expectations. I was dreaming to be a model for my family as well as my country. But what I am today is nothing just a good-for-nothing man. I have no motivation for teaching.

**Respondent 3:** With populated classrooms, no teaching supports, no motivation from the learners, no discipline at school, I feel that my place should not be here. If I am an opportunity to do other things, I won't hesitate to grab it.

(See Appendix 24 for additional responses provided on a selected basis)

In light of the aforementioned findings relating to teacher motivation, it appears that the majority of teachers have low motivation for teaching English. Yet, I can hypothesize that when students are taught by a teacher with low motivation for teaching, their achievement can be affected.

Therefore, I am curious to find how the ESE includes tasks that also take into account teachers' low motivation for teaching.

The knowledge of Grade 12 students' characteristics as well as the knowledge of English language teachers' characteristics is necessary, but it is not sufficient for understanding the context of the ESE. Other necessary information must be provided by the understanding of the condition under which the teaching/learning of reading takes place. I intend to address this issue in the following section.

### **6.3. Conditions under which the teaching/learning of reading takes place**

Reading in a foreign language mainly takes place in a formal context that is the classroom. Therefore, the description of how reading activity is actually conducted in classroom can provide insights for understanding the conditions under which the teaching/learning of reading takes place. In this section, I investigate the actual conditions under which reading activity takes place. In order to achieve this aim, I provide a description of three issues that have emerged from my scrutiny of the data. These issues relate to the class size, the reading resources used by the teachers and the learners, and the availability and quality of the English curriculum.

#### **6.3.1. Classroom size**

In my survey of literature on the various conditions that influence reading activity in the classroom, I have stated that the number of students in a classroom is a variable that has been reported to affect the condition under which the learning of reading takes place (see Chapter 4, Section 2). In the scope of this study, I hypothesize that when reading instruction occurs in large classes, students' achievements can be negatively affected on the ground that a larger class is more challenging to handle than a smaller class.

In order to investigate this issue, I have scrutinized the responses from question 8 of the contextual questionnaire administered to the 27 teacher-participants (see table in Appendix 16). In light of the data provided in this table, it appears that the majority of teacher-participants (78

%) were of the view that a large number of students in the class was one of the factors that most negatively affected the teaching of reading.

In order to gain a deeper understanding on this issue of the effect of large classes in reading instruction, I have scrutinized the teachers-participants' qualitative reports. In light of the main issues raised in these reports, it appears that the venue capacity and the large number of students in reading classes are detrimental to efficient reading. Besides, large classes pose the problem of the limited number of textbooks; and this makes reading instruction difficult. The following comments can help illustrate these views:

**Respondent 1:** How can a small venue not equipped with enough desks accommodate 70 learners? It is very difficult to attend to each student and provide assistance. When you try to assist some, the others feel abandoned and therefore shift their focus to other things. As a result, you find out that most students have not been able to read the text and complete the reading task as requested.

**Respondent 2:** I have 51 students; but I have only 10 textbooks. It becomes very difficult to teach reading in good conditions. More than five students agglutinate around one textbook and this usually creates chaos.

**Respondent 3:** I have many students; and I can't assist each and every student. I think that few students could learn better than the actual number of students in my class.

(See Appendix 24 for additional responses provided on a selected basis)

The aforementioned quantitative finding and the participants' comments from the qualitative reports suggest that the learning of English, in general and of reading, in particular is confronted to the issue of large classes; and this is exacerbated by paucity of reading resources. In light of this conclusion, I am curious to see the extent to which the ESE tasks are also sensitive to this particular learning condition.

### 6.3.2 Reading resources

In reviewing studies that have addressed the issue of the impact of reading resources in reading achievement (see Chapter 4, Section 2), it was generally admitted that the quality and availability

of reading resources used during reading instruction also provided evidence for the understanding the test context.

In order to investigate the availability of reading resources, the 496 participants were asked, in the first part of the questionnaire, to select from a list the different sources they used for reading at school and at home. Table 6.13 provides data on the participants' use of different sources for reading.

Table 6.13: Sources of reading at school and at home

Sources of reading N=496		T1 (%)	T2 (%)	T3 (%)	T4 (%)	T5 (%)	T6 (%)	T7 (%)	T8 (%)	Mean (%)	SD
At school	Texts copied from the board	80.0	65.6	73.2	64.9	73.8	65.7	58.3	35.7	64.7	13.5
	Photocopied texts	44.0	42.1	35.4	38.6	37.4	45.7	41.7	64.3	43.7	9.0
	Textbooks	6.7	28.1	25.6	31.6	21.5	17.1	0.0	0.0	16.3	12.6
	Magazines/newspapers/journals/brochures	0.0	0.0	0.0	0.0	0.0	0.0	14.6	8.1	2.8	5.5
	Books	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Internet	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Don't read	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
At home	Texts copied from the board	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Photocopied texts	0.0	29.6	26.8	8.8	0.0	0.0	14.6	35.7	14.5	14.6
	Textbooks	2.7	20.3	15.9	8.8	11.2	5.7	0.0	0.0	8.1	7.4
	Magazines/newspapers/journals/brochures	8.0	18.8	19.5	19.3	12.2	20.0	10.4	25.0	16.6	5.8
	Books, novels	4.0	9.4	8.5	10.5	7.5	14.3	0.0	0.0	6.8	5.1
	Internet	5.3	10.9	12.2	3.5	12.2	5.7	2.1	7.1	7.4	3.9
	Don't read	65.3	65.6	62.2	59.7	69.2	65.7	79.2	60.7	65.9	6.2

Based on my scrutiny of the statistics in this table, I wish to present these findings:

Concerning the issue of reading at school, of the seven suggested sources of reading materials, two sources were frequently used by participants: first, the text passages copied from the chalk board (64.7 %); and then the photocopied texts (43.7 %). Although textbooks constitute a primary source of reading materials as articulated in the English Curriculum, the use of textbooks was very scarce as only 16.3 per cent of participants reported using them. The high variance observed ( $SD=12.6$ ) suggests a higher frequency observed among participants from the six groups who were studying English five hours per week ( $T1=6.7\%$ ;  $T2=28.1\%$ ;  $T3=25.6\%$ ;  $T4=31.6\%$ ;  $T5=21.5\%$ ;  $T6=17.1$ ) compared to a lower frequency observed among participants from the two groups who were studying English only twice a week ( $T7=0.0\%$ ;  $T8=0.0\%$ ). However, the same data suggest that magazines/newspapers/journals/brochures were the least frequently used reading resources (2.8%). On the other hand, no participants (0.0%) reported using books and internet as reading resources in classroom.

In regard to reading at home, the same data (Table 6.13) indicate that the majority of participants (65.9 %) across the 8 test groups did not read at home. On the other hand, the minority of participants who reported reading at home used one or some of these reading materials: Magazines/newspapers/journals/brochures (16.6 %), photocopied texts (14.5 %), textbooks (8.1 %), internet (7.4 %) and books/novels (6.8 %).

Although I will elaborate on these findings in the Discussion Chapter, I hasten to state, at this juncture, that these findings reflect a paucity of reading materials not only in the classroom, but also at home. Therefore, they suggest a very limited exposure of the students to reading activity, and this limited exposure needs to be considered in the development of the ESE.

In order to triangulate these findings, I inspected the frequency of use of different reading sources by English language teachers. Question 5 of the questionnaire administered to teacher-participants (see Appendix 6) requested them to indicate the frequency of use of different suggested reading sources. Table 6.14 provides information on teacher-participants' frequency of use of different reading sources.

Table 6.14: Frequency of use of different reading sources by teachers

Reading sources N=27		Usually		Sometimes		Rarely		Never		Total	
		N	%	N	%	N	%	N	%	N	%
1	Textbooks	2	7.4	4	14.8	10	37.0	11	40.7	27	100.0
2	Books	0	0.0	0	0.0	4	14.8	23	85.2	27	100.0
3	Newspapers/ Magazines/brochures	0	0.0	0	0.0	7	25.9	20	74.1	27	100.0
4	Internet sites	0	0.0	0	0.0	0	0.0	27	100.0	27	100.0

Based on the data presented in this table, I wish to make the following observations:

(1) Regarding the use of textbooks as a reading resource, the data indicate that only a minority of teacher-participants (7.4 %) reported ‘usually’ using textbooks as reading resource. On the contrary, most teachers either ‘never’ used the textbooks (40.7%), or ‘sometimes’ used them (14.8 %), or again ‘rarely’ used them (37.0 %);

(2) In regard to the use of books as reading resource, the data indicate that most teacher-participants (85.2 %) ‘never’ used books, and only very few (14.8 %) ‘rarely’ used books in reading classes;

(3) Concerning the use of Newspapers/magazines/brochures as reading resources, the data indicate that the majority of teacher-participants (74.1 %) ‘never’ used newspapers, magazines, or brochures in reading classes; and that only few of them (25.9 %) ‘rarely’ used these reading resources;

(4) Finally, in relation to the use of internet as a reading resource, the data indicate that all the teacher-participants (100.0 %) ‘never’ used internet in reading classes.

These findings on the reading resources suggest a scarcity of reading materials necessary for teachers to enhance the teaching of reading in classroom contexts. They prompt the issue of the discrepancy between the assumed teaching context that is elaborated in educational policies as articulated in the English curriculum and the actual context as translated in real classrooms. I will elaborate on this issue in the Discussion Chapter.

The qualitative reports from both student-participants and teacher-participants provided additional information necessary to augment our understanding of the issue of reading resources. The student-participants who reported not reading either at home or at school mentioned four reasons that explained this reading behavior. Among these reasons, I wish to mention a lack of reading materials, poor quality of reading materials, an abject paucity of classroom reading, and a lack of motivation for reading. The following four excerpts illustrate each of these views:

**Respondent 1:** What am I going to read? I don't have textbooks or books. What can I read? I don't know why the educational authorities do not provide the school with textbooks for English like what happens for French where we use a textbook. It is as if English is not important.

**Respondent 2:** One of our classmates writes the text on the board; and this is with many spelling mistakes. So I cannot spend my time to read a text that I know is written with many spelling mistakes.

**Respondent 3:** We do not read frequently at school; why do I need to read at home? I am not interested in reading at home.

**Respondent 4:** I prefer to use my time to read the accounting course; I don't have time to read English. This is not my language, and I don't need to know it. Where shall I go with English?

(See Appendix 23 for additional responses provided on a selected basis)

Furthermore, the qualitative reports from the teacher-participants revealed that the main reason why teachers were not using necessary reading resources was that these resources were nonexistent; or in case they existed, they were not enough to be used by all the learners. The following excerpts from can illustrate these two reasons:

**Respondent 1:** Since I have been teaching English, the school does not have textbooks. I have managed to photocopy some texts from 2 [two] textbooks and I have to write them on the chalkboard for the students to copy in their notebooks.

**Respondent 2:** We do not have a school library; we do not have internet connection at school; where can I get the text? I have my textbook but my students do not have any copy; so I have photocopied some texts I have found good for students. Photocopied texts are the only one reading material I use in reading classes.

**Respondent 3:** There is huge problem of textbooks. What I do is to search for some relevant materials from technical brochures and magazines and I make some copies of these materials. This is not easy because the majority of students cannot afford these photocopied texts. But at least, these texts have the advantage to provide the learners with visual support for reading.

**Respondent 4:** My school has only five (5) copies of the textbook. Yet, I have 53 students in my class. It becomes difficult to use these five copies in reading classes. The only solution I have is to ask one student to write the text on the board and the other students to copy the text from the board. This is not without consequence. It takes so much time to complete writing the text and also it is too demotivating as some students just refuse to copy the text from the board.

(See Appendix 24 for additional responses provided on a selected basis)

### 6.3.3 The English curriculum

The English curriculum (Programme National d'Anglais) is a document that states the objectives of the English course, the methods to be used for teaching English and testing, the sequence of materials to be covered by the English teacher, and the time devised for teaching English in different subject areas. The English curriculum has devised five hours per week for teaching English in general schools and two hours per week for teaching English in most technical schools. The availability of the curriculum and its actual use by English teachers, its quality in terms of content and the degree of its completion by English teachers can provide useful information for understanding the context of reading in general and the context of the ESE in particular. In the present section, I propose to investigate these issues through the analysis of teacher-participants' responses and comments on the contextual questionnaire.

In order to investigate the aforementioned issues relating to the English curriculum, question 6 of the questionnaire included four statements that aimed to prompt information on the availability of the English curriculum to the teachers, its actual use by the teachers, how the teachers perceived its content to be, and the extent to which they completed topics foreseen in this

curriculum. The table in Appendix 16 provides information that relates to these issues. On the basis of information contained in this table, I propose the following observations:

(1) Less than half of teacher-participants (40.7 %) did not have the English curriculum; suggesting that the majority of the participants did not have the English curriculum as part of their teaching documents.

(2) Regarding the actual use of this document by English teachers, the data indicate that only slightly more than a half (55.6%) of participants reported using the English curriculum to plan their course outline. This is to suggest that many participants were planning their course outline without using the curriculum.

(3) As regards to the quality of the English curriculum, the majority of participants (66.7 %) were of the view that the English curriculum was old-fashioned; and therefore, it needed to be revised and updated. This finding prompts the question of the content of materials used in teaching English in general and reading in particular.

(4) Concerning the issue of completion of topics foreseen in the English curriculum, the same data indicate that all the participants (100.0 %) admitted that all the topics foreseen in the curriculum were never completed in the course of the academic year; a finding that suggests a gap between what is planned and what is actually taught.

The qualitative reports from teacher-participants provided deeper information necessary to understand the issue of the availability, use and quality of the English curriculum, as well as its completion. The participants who did not have the curriculum indicated that the school did not have this document, although they acknowledged its uttermost importance in the teaching of reading. One participant commented:

It is really a pity that, as an English teacher, I do not have the curriculum. This document is very important because it contains all the topics that teachers must use to teach their students. Yet, when I asked for it to the school principal, he told me that this document was not part of documents delivered by the provincial government and that I had to check with English teachers from other schools to get it and print a copy for myself.

Other participants complained that, even during their training time as teachers, they were not given a copy of this document by their academic institution. One participant made the following comment:

When I was having my English methodology module, the lecturer repeatedly insisted that before we graduate, we should make sure that we have all necessary documents required for use by the English teachers. Unfortunately, the Department of English was not able to provide us with these documents at that time. Upon graduation, I realized that I had none of them and it is not now that I can expect to get them in my school since what counts more for the school principal is to find teaching materials for the main subjects; not for English that they view as minor subject.

Regarding the quality of the English curriculum, participants who were of the view that the English curriculum was old-fashioned explained that it was long ago that the curriculum was implemented and that it must be updated to reflect new advances in information technology. One of these participants wrote:

Our curriculum suffers from old age. Nowadays, things are advancing, the world is progressing and changes are necessary. The curriculum needs to be adapted to reality and this cannot be expected with the curriculum that was elaborated in the eighties when computers and internet were not known by most Congolese scholars.

Other participants elaborated on the issue of updating the English curriculum in terms of its content. These teachers believed that the literary excerpts used today do no longer fit the actual context. One participant made this comment:

In teaching general English, most authors suggested are situated in a certain time and space. Take for example Chinua Achebe, Cyprian Ekwensi or Amos Tutuola. Their writings belong to a certain epoch of independent Africa. If it is a good idea that our students must be initiated to reading African literary writers in English, we must look at the content of the themes developed by these writers. I think that the list of the writers must be updated along with the themes developed in order to be closer to learners' real life. I don't know, maybe there are writers who have addressed the problems of corruption, war, insecurity or human rights that plague the African states nowadays. Why not use their writings if there are some.

Still, other participants who elaborated on the content of the English curriculum indicated an inconsistency between the curriculum content and the prescribed textbook contents. One participant who espoused this view made the following comment:

I have an English curriculum; but there is little relationship between the curriculum content and the textbook content. So it becomes difficult to teach all the required topics suggested in the curriculum because there is not any support for some topics.

As regards to the degree of completion of the curriculum, many participants explained why they could not complete the materials they had planned to teach in the course of the year. One main reason reported was attributed to the many interruptions that usually occurred during academic year. One participant gave this explanation:

I have been teaching English for the past 13 years and I do not remember a single time when I completed all the topics scheduled in the curriculum. There are always so many interruptions. Sometimes it is teachers who interrupt working because they are not paid; or other times they interrupt teaching because they ask an increase of salary which the government refuses to make. At other times students regularly interrupt courses for weeks or months because they ask for a reduction of school fees but the educational authorities refuse to attend to their demands. Negotiations take too long before school starts again.

Some other participants explained the non-completion of the materials designed in the curriculum by the socio-political instability the country is going through. The following are two excerpts that support this view:

**Respondent 1:** Although we are at the center of the country and we have never been directly affected by war, insecurity and violence that characterize the country; we still face the consequences of this instability. Many children who were displaced by war in the Kivu come here to start school sometimes only two months before the end of the school year. Since there is no special program designed to respond to such situations, you put these children in the same class with those who started at the beginning of the school year. You realize that you have almost two or three different groups in one class and you find out that you are actually teaching three groups at different speed. As a consequence, you end up by not completing the entirety of topics scheduled in the curriculum. When next year these students pass to the next grade, the English teacher does not recapitulate the previous year's materials; instead, he just starts the program for that grade. With this gap, students hardly catch up.

**Respondent 2:** With the past general elections and the chaos that has resulted from these elections, schools were closed for more than two months. When we finally resumed school, there was nothing arranged by educational authorities to catch up with the time that was lost. As a result, we just try to do what we could do, and many lessons foreseen in the curriculum could not be taught.

(See Appendix 24 for additional responses provided on a selected basis)

The aforementioned quantitative and qualitative findings related to the English curriculum suggest that English instruction, in general and reading instruction, in particular appear to occur in a context where the learning objectives, the teaching and assessment methods and the sequence of materials appear to be less clearly articulated because the curriculum is not available to the majority of teachers, and that it is of bad quality. In light of this conclusion, I am curious to find how the ESE include tasks whose degree of complexity reflects this contextual issue.

### **Summary of the chapter**

In this chapter, I have presented and analyzed the data that sought to describe the English state examination. In light of this analysis and presentation, it was found that:

- (1) The majority of student-participants acknowledged that their abilities in English language were poor;
- (2) Student-participants' reading at school was very restricted; and its frequency largely differed according to participants' main subject areas; suggesting that reading frequency at school can be correlated to the time allocated for learning English;
- (3) Reading at home was not an activity carried by most participants. This was due to a lack of reading materials, the content of reading materials not being relevant to students' needs, or students' poor socio-economic conditions;
- (4) The majority of student-participants were not motivated in the English course as they did not regularly attend English classes. This low motivation was due to teachers' lack of competence and motivation for teaching, students' perception that English is a difficult subject, and the lack of reading resources;
- (5) The majority of student-participants preferred to learn English in lesser time than the actual time planned for learning English; as a result, they were less exposed to the English language;
- (6) The majority of student-participants did not have some experience with the ESE because the MCQ is almost not used as a classroom assessment method that can help them to familiarize with the ESE;

(7) Although the majority of student-participants had a positive attitude towards the ESE, some of them had a negative attitude and this was reflected in their feelings and/or beliefs that they must not write this test, that the ESE is a way of failing them to the national test, or again that the ESE is not a fair test since it is based only on text passage comprehension by neglecting other language skills;

(8) The majority of student-participants preferred the actual time for writing the ESE to be maintained or to be increased on the ground that an extended time could offer them an opportunity for assistance and collaboration;

(9) The majority of student-participants preferred the actual number of questions (9 in general subjects and 5 in technical subjects) to be reduced, or at least to be maintained on the ground that fewer questions in the English test were not likely to negatively affect their general performance on the national test;

(10) The majority of student-participants preferred the actual number of six alternatives to be reduced to three alternatives as a reduced number of alternatives could not only confuse them, but also enable them to have high chance for getting some correct answers by guessing;

(11) The majority of student-participants had preference for a single form test on the ground that a multiple alternate form test does not reflect the classroom testing where all students write the same test; besides, it does not reflect fairness;

(12) Although the majority of teacher-participants were formally qualified to teach English and they had some experience teaching English in Grade 12; they were not involved in in-service trainings as well as continuous trainings that could enhance their professional development;

(13) Reading instruction was negatively affected by large number of students in the class, an abject paucity of reading materials, teachers lacking the English curriculum as part of their teaching documents, the English curriculum being not updated, and the non-completion of all the topics foreseen in the curriculum in the course of the academic year;

The description of the ESE context provides a basis to examine the kind of tasks included in the ESE and to evaluate the extent to which these tasks take account of the actual test context. I propose to investigate this issue in the next chapter.

## CHAPTER SEVEN

### FINDINGS RELATED TO THE VARIETY AND COMPLEXITY OF DIFFERENT READING TYPES AND PROCESSING LEVELS AND THE EXTENT TO WHICH THE ESE TASKS ARE APPROPRIATE TO THE ESE CONTEXT

#### 7.1 Introduction

The second objective of this study was to determine the variety and degree of complexity of different reading types examinees attempt while completing the ESE tasks; and to evaluate the extent to which the ESE tasks are appropriate to the ESE context. In light of this objective, this chapter is expected to provide evidence of the extent to which the ESE tasks target the different types of reading and processing levels; and the extent to which the cognitive demands required to complete the ESE tasks are appropriate and meaningful to the ESE actual context.

The process-oriented approach I have used in this study relies on the identification and description of strategies examinees use to complete the test tasks in order to examine the reading construct. One of the main assumptions of this approach was that the goals that are open to the reader and that characterize reading are either careful reading or expeditious reading; and these two reading types occur either at the global level or local level. In light of this assumption, four reading types were distinguished: (1) careful reading at global level, (2) careful reading at local level, (3) expeditious reading at global level and (4) expeditious reading at local level.

Indeed, whether someone reads carefully or expeditiously at global level or local level, she/he has to execute one or some of the following operations in order to complete test tasks: encode lexis, parse sentences, establish propositional meaning, make inferences, build text mental model, and create a text macrostructure. These operations encompass the six (6) processing levels presented in the conceptual framework. The first three levels are generally qualified to be lower level processing while the last three operations are considered to be higher level

processing. According to this framework, the task complexity is function of the level on which the examinee processes the text in order to complete test tasks. Hence, test questions that request lower level processing are expected to be easier than those test questions that target high level processing.

In this chapter, I use this taxonomy in order to explore the variety and complexity of different reading types and processing levels examinees engage while completing the ESE tasks. All data presented in this chapter are taken from the eight tests administered to the 496 student-participants. In order to achieve the objective of this chapter, I propose to present three main points. First, I present the item frequency per test for the four reading types. Such a presentation aims to picture the items frequency across the eight tests as well as across the four reading types. I have hypothesized that, in order to be construct-valid, the ESE needs to include test items that request participants to conduct *all* the four reading types. Secondly, I present the frequency of strategies the participants had used for completing the test tasks in terms of the four reading types. Such a presentation aims to provide a view of strategies participants deployed in answering test questions. In light of this aim, I have hypothesized that, in order to be construct-valid, the ESE needs to include tasks that require participants to use relevant strategies to answer individual test items. Thirdly, I present the item difficulty for each of the four reading types. This presentation aims to describe the complexity of the ESE tasks in terms of the four reading types. In light of this aim, I have hypothesized that, in order to ensure its construct validity, the ESE must include tasks that whose complexity level is proportional to examinees' characteristics. In order to achieve this aim, I present the mean and item difficulty for careful reading versus expeditious reading; and the mean and item difficulty for reading at global level versus reading at local level. Lastly, I evaluate the extent to which the ESE tasks target the six processing levels in order to cover the reading construct.

However, although this chapter strictly relates to the analysis and presentation of study findings, I wish to state that, for the sake of narrative immediacy and primacy of findings, I will from time to time conduct an initial discussion of some of the critical findings. Such an epistemological choice aims to draw attention on the salient findings and their relationships to the study aim and objectives. Nevertheless, in the Discussion Chapter, I expect to elaborate on the initial discussion

provided in the present chapter. In the following section I present the frequency of strategies used for each of the four reading type.

## 7.2 Frequency of items per test for different reading types and reading skills

At the outset, I hasten to mention that the number of reading items extracted from the study sample was 56. However, while I proceeded from the identification of different strategies the 496 student-participants had used to answer these test items, I came up with 69 items. This number suggests that some items could be classified in more than one category depending on the strategies used to answer these individual items (I will provide a detailed explanation on these types of items in the Sub-section 7.3). Table 7.1 presents information on the frequency of items per test for the four reading types and different reading skills.

Table 7.1: Item frequency per test and for the four reading types

Skills tested		T1 N=7	T2 N=9	T3 N=9	T4 N=5	T5 N=7	T6 N=9	T7 N=5	T8 N=5	Total items
<b>Careful reading at global level</b>										
1	Establishing text/paragraph global comprehension	T1.1	T2.1	-	T4.1	T5.1	-	T7.2	T8.1	<b>6</b>
2	Establishing accurate comprehension of explicitly stated main ideas and supporting details across sentences	-	T2.3	-	T4.3	-	T6.1	-	-	<b>3</b>
3	Making text/paragraph inferences	T1.3	T2.8	T3.1	-	T5.2	T6.2 T6.3	-	T8.3	<b>7</b>
<b>Total items for careful reading at global level</b>		<b>2</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>16</b>
<b>Careful reading at local level</b>										
1	Identifying/understanding lexis	T1.4	-	T3.6	-	T5.7	T6.4 T6.6 T6.7 T6.9	-	T8.4 T8.5	<b>9</b>
2	Understanding syntax	T1.7	T2.5 T2.7	T3.2 T3.3	T4.4 T4.5	T5.3 T5.4	T6.5 T6.6	T7.5	-	<b>15</b>

				T3.4 T3.8 T3.9						
<b>Total items for careful reading at local level</b>		<b>2</b>	<b>2</b>	<b>6</b>	<b>2</b>	<b>3</b>	<b>6</b>	<b>1</b>	<b>2</b>	<b>24</b>
<b>Expeditious reading at global level</b>										
1	Skimming to find the title that best fits the text/paragraph	T1.1	T2.1	-	T4.1	T5.1	-	T7.2	T8.1	<b>6</b>
2	Skimming the text/paragraph to locate needed information	T1.2	-	-	T4.2	-	-	-	T8.2	<b>3</b>
3	Search reading to quickly locate implicit information	T1.3	-	T3.1	-	-	-	-	-	<b>2</b>
4	Search reading to quickly locate explicit pieces of information scattered through the whole text/paragraph	T1.5	T2.2 T2.4 T2.6	T3.4 T3.7	T4.3	-	T6.2 T6.3	T7.1	-	<b>10</b>
<b>Total items for expeditious reading at global level</b>		<b>4</b>	<b>4</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>21</b>
<b>Expeditious reading at local level</b>										
1	Scanning to locate explicit information located in a specific part of the sentence/paragraph	T1.4 T1.7	T2.5	T3.2	T4.5	T5.3	-	T7.3 T7.4	-	
<b>Total N items for expeditious reading at local level</b>		<b>2</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>8</b>
<b>Total N items for the four reading types</b>		<b>10</b>	<b>10</b>	<b>11</b>	<b>8</b>	<b>7</b>	<b>11</b>	<b>6</b>	<b>6</b>	<b>69</b>

T1=Test 1; T2=Test; T3=Test3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8; T1.1=Test 1, Item number 1; T2.1= Test 2, Item number 1; etc.

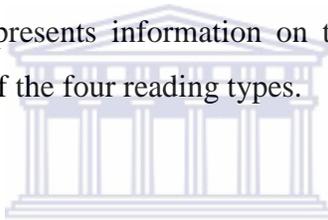
### 7.2.1 The number of reading items included in each reading type

In the Methodology Chapter, I stated that the study sample consisted of 56 reading items collected from eight (8) tests that were administered to 496 participants. In Appendix 7, I presented these 56 reading items. Indeed, in order to determine the number of reading items included in each reading type, I have examined different strategies participants had used to answer the individual test items. Then, on the basis of the study framework, I have classified the

56 test items in their corresponding categories. Table 7.1 presents the item frequency per test for the four reading types. The information provided in this table indicates that the study sample included test items that covered all the four reading types. The reading type that included the highest number of items was careful reading at local level (24 items), followed by expeditious reading at global level (21 items), then careful reading at global level (16 items) and finally expeditious reading at local level (8 items).

## **7.2.2 Reading skills necessary for answering different test item types**

In reading a text in order to comprehend it and answer the questions based on it, the readers deploy specific skills necessary for answering individual types of test items. In this section, I wish to provide, in light of the study data, the different reading skills targeted by the 56 reading items. Table 7.1 (see page 215) presents information on the different reading skills and the number of items included in each of the four reading types.



### **7.2.2.1 Careful reading at global level: Number of skills and items**

The information presented in Table 7.1 (see page 215) suggests that, the 496 participants used three main skills in order to answer the test items that targeted careful reading at global level. The first skill required the participants to establish the text/paragraph global comprehension. This skill was needed to answer the following six (6) items [T1.1; T2.1; T4.1; T5.1; T7.2; and T8.1]. The second reading skill required examinees to establish accurate comprehension of explicitly stated main ideas and supporting details across sentences. This skill was necessary to answer these three (3) items [T2.3; T4.3; and T6.1]. Finally, the third skill called for examinees to make text/paragraph inferences. This skill was needed to answer these seven (7) items (T1.3; T2.8; T3.1; T5.2; T6.2; T6.3; and T8.3). I wish to mention that Appendix 7 provides a full presentation of all these test questions.

### **7.2.2.2 Careful reading at local level: Number of skills and items**

The data presented in Table 7.1 (see page 215) indicate that, in order to answer test items that required the participants to carefully read the text at local level, two (2) skills were deemed necessary. The first skill required the participants to identify/understand lexis. This skill was needed to answer nine (9) test items (T1.4; T3.6; T5.7; T6.4; T6.6; T6.7; T6.9; T8.4 and T8.5). The second skill required examinees to understand syntax. This skill was needed to provide answer to fifteen (15) items (T1.7; T2.5; T2.7; T3.2; T3.3; T3.4; T3.8; T3.9; T4.4; T4.5; T5.3; T5.4; T6.5; T6.6; and T7.5). A full presentation of these test questions is provided in Appendix 7.

### **7.2.2.3 Expeditious reading at global level: Number of skills and items**

The information presented in Table 7.1 (see page 215) indicates that, the test items that required the participants to read the text expeditiously at global level required four skills. The first skill was skimming to find the title that best suits the text/paragraph. This skill was needed to answer the six (6) item questions (T1.1; T2.1; T4.1; T5.1; T7.2; and T8.1 [see Appendix 7]) that requested the participants to select the title that best suited the text or a specific paragraph. At this juncture, I wish to mention that these aforementioned six items have previously been classified in careful reading at global level (see Section 7.2.2.1) as they also required participants to conduct this type of reading. I will provide a detailed explanation on these types of test items further. The second reading skill demanded participants to skim the text in order to locate needed pieces of information. The participants used this skill in order to answer the following three (3) test items (T1.2; T4.2; and T8.2 [see Appendix 7]). The third reading skill requested the participants to speedily read the entire text in order to figure out implicit information. This skill was necessary to answer these two (2) test items (T1.3; T3.1 [see Appendix 7]). The last reading skill required participants to quickly read the entire text in order to locate explicit pieces of information scattered in different parts of text. This skill was needed to answer the following ten (10) item questions (T1.5; T2.2; T2.4; T2.6; T3.4; T3.7; T4.3; T6.2; T6.3 and T7.1 [see Appendix 7]).

### 7.2.2.4 Expeditious reading at local level

The information offered in Table 7.1 (see page 215) suggests that one reading skill was needed in order to answer test questions that targeted expeditious reading at local level. This skill required participants to scan the text in order to locate explicit information located in a specific part of the sentence or paragraph. This skill was used in answering these eight (8) test items (T1.4; T1.7; T2.5; T3.2; T4.5; T5.3; T7.3; and T7.4 [see Appendix 7]).

### 7.3 Frequency of strategies used for each of the four reading types

This section aims to feature the occurrence of different strategies the participants had used for completing the different test tasks. As I have indicated previously (see section 7.0), I have hypothesized that, in order to be construct-valid, the ESE needs to include tasks that require examinees to use the relevant strategies to answer individual test items. Table 7.2 presents information on the frequency of strategies the 496 student-participants had used for answering the 56 test items included in the four reading types. In this table, the strategies are classified in two categories: reading strategies [from S1 to S11] and test taking strategies [from S12 to S18] (see Appendix 3 for a presentation of these strategies). The strategies frequency is indicated in percentage to reflect their relative occurrence. When a strategy is scored 100 %, this indicates that all participants used it in order to answer a particular test item.

Table 7.2: Frequency of strategies the 496 participants used for answering the 56 test items

Reading skills		Frequency of Strategies used	
<b>1. Careful reading at global level</b>			
1	Establishing text/paragraph global comprehension	Reading strategies	S1 [79%]; S3 [58%]; S2 [42%]; S4[38%]
		Test taking strategies	S12 [42%]; S17 [33%]; S13 [21%]; S18 [8%]
2	Establishing accurate comprehension of explicitly stated main ideas and supporting details across sentences	Reading strategies	S3 [100%]; S2 [40%]; S10 [10]
		Test taking strategies	S12 [80%]; S13[10%]; S17 [10%]; S18 [10%]
3	Making text/paragraph inferences	Reading	S2 [43%]; S1 [29%]; S3 [29%];

		strategies	S10 [24%]
		Test taking strategies	S12 [33%]; S13 [33%]; S17 [24%]; S16 [14%]; S15 [9%]
<b>2. Careful reading at local level</b>			
1	Identifying/understanding lexis	Reading strategies	S8 [43%]; S1 [29%]; S5 [24%]; S9 [24%]; S11 [24%]; S7 [10]
		Test taking strategies	S17 [38%]; S12 [33%]; S13 [24]; S18 [19]
2	Understanding syntax	Reading strategies	S7 [79%]; S11 [33%]; S3 [21]; S8 [6%]; S2 [3%]
		Test taking strategies	S13 [61%]; S14 [18%]; S17 [7%]; S12 [2%]; S15 [2%]
<b>3. Expeditious reading at global level</b>			
1	Skimming to find the title that best fits the text/paragraph	Reading strategies	S1 [91%]; S2 [87%]; S3 [87%]; S4 [70%]
		Test taking strategies	S12 [42%]; S17 [33%]; S13 [21%]; S18 [8%]
2	Skimming the text/paragraph to locate needed information	Reading strategies	S2 [91%]; S4 [36%]; S3 [18%]; S1 [9%]; S10 [9%]
		Test taking strategies	S13 [81%]; S16 [72%]; S17 [7%]
3	Search reading to quickly locate implicit information	Reading strategies	S1 [60%]; S4 [60%]; S2 [40%]; S3 [40%]
		Test taking strategies	S14 [40%]; S17 [40%]; S12 [20%]; S15 [20%]
4	Search reading to quickly locate explicit pieces of information scattered through the whole text/paragraph	Reading strategies	S2 [72%]; S3 [28%]; S7 [20%]; S1 [20%]; S8 [4%]
		Test taking strategies	S17 [48%]; S13 [28%]; S16 [24%]; S18 [24%]; S12 [20%]; S15 [8%]
<b>4. Expeditious reading at local level</b>			
1	Scanning to locate explicit information located in a specific part of the sentence/paragraph	Reading strategies	S11 [63%]; S7 [42%]; S3 [32%]; S4 [21%]; S8 [11%]
		Test taking strategies	S17 [47%]; S14 [42%]; S12 [26%]; S13 [26%]; S18 [21%]; S15 [11%]

**S1=Strategy number 1; S2=Strategy number 2; S3=Strategy number 3; S4=Strategy number 4; ... S18=Strategy number 18 (See Appendix 3: Strategies codes)**

### **7.3.1 Strategies used for answering careful reading at global level items**

In this section, I present the different strategies (reading and test taking strategies) the participants used for answering careful reading at global level items. This presentation focuses on three points. In the first point, I present the strategies participants used to answer test questions that required their skill to establish the text/paragraph global comprehension. In the second point, I present those strategies participants deployed to answer test questions that required their skills to establish accurate comprehension of explicitly stated main idea and supporting details across sentences. In the last point, I present the strategies participants used to answer test questions that targeted their skill to make text or paragraph inference.

#### **7.3.1.1 Frequency of strategies the participants used for answering test questions that required their skills to establish text/paragraph global comprehension**

The information presented in Table 7.2 (see page 219) indicates that, the participants used four (4) reading strategies to answer the 6 test questions that required their skill to establish the text or paragraph global comprehension. All these six test questions requested the participants to select the alternative that contained a title that best suited the text or a specific paragraph of the text.

In order to answer these test questions, most participants (79%) had first to read the whole text carefully (S1) and some of them (42%) could reread it rapidly (S2) in order to build its gist. If the question requested them to find the title that best suited a specific paragraph, the majority of participants (58%) had to read a specific paragraph of the text carefully (S3) first and then rapidly (S4=38%) in order to build the paragraph gist. The detailed information on strategies patterns for individual test items is provided in the table in Appendix 17.

In regard to the test taking strategies used by participants to answer these items, the information in Table 7.2 (see page 219) substantiates that four (4) main strategies were used. Slightly one in five participants (21%) had to consider the five alternatives and immediately select the alternative that appeared familiar to them (S13). However, less than half of participants (42%) first considered the five alternatives and then postponed consideration of what could appear to them as the best alternative (S12) before selecting their answers through elimination of other

alternatives (S17=33%). For the minority of participants (8%), they only selected the “correct” alternative through guessing (S18). The detailed information on strategies used for answering the aforementioned individual test items is provided in Appendix 17.

### **7.3.1.2 Frequency of strategies the participants used to answer test questions that required their skills to establish accurate comprehension of explicitly stated main idea and supporting details across sentences**

The information presented in Table 7.2 (see page 219) substantiates that the participants used three (3) reading strategies and four test taking strategies in order to answer the three (3) test items that requested their skill to establish accurate comprehension of explicitly stated main idea and supporting details across sentences. Two of these three items (T2.3 and T4.3 [see Appendix 7; Section 2 and Section 4]) requested participants to indicate according to the text, what they think may be the reason of a particular incident or decision. The other item (T6.1 [see Appendix 7; Section 6]) requested participants to indicate the sentence that best summarized a specific paragraph.

In order to answer these questions, strategies use suggests that all the participants (100%) read the target paragraph carefully (S3) with the aim to comprehend all supporting paragraph details as elaborated across different sentences. However, some participants (40%) had to read first the whole text rapidly (S2) in order to construct its representation before they focused on the target paragraph and read it carefully. However, few participants (10%) augmented their careful reading of the target paragraph by trying to draw some conclusions on what the paragraph implied (S10=10%). The detailed information on strategies used for answering the aforementioned individual test items is provided in Appendix 17.

Concerning the test taking strategies participants used to answer these types of questions, the information provided in the same table reveals that, the participants used four (4) strategies. Most participants (80%) generally considered the five alternatives first and then postponed consideration of the “correct” alternative (S12) before they could finally make their choices either through elimination of other alternatives (S17=10%) or through guessing (S18=10%). However, only the minority of participants (10%) could consider the five alternatives and then immediately focus on the alternative they considered to be the “correct” one (S13). The detailed

information on strategies used for answering the aforementioned individual test items is provided in Appendix 17.

### **7.3.1.3 Frequency of strategies used by participants in answering test questions that required their skill to make text/paragraph inference**

The information presented in Table 7.2 (see page 219) suggests that the participants used four (4) reading strategies and five (5) test taking strategies in order to provide answers to the seven (7) test questions (T1.3; T2.8; T3.1; T5.2; T6.2; T6.3; and T8.3 [see Appendix 7; Sections 1, 2, 3, 5, 6 and 8 respectively]) that required their skill to make inferences. All these test questions requested participants to find the statement that reflected (or did not reflect) the information contained in the text. In order to answer these textually implicit questions, the participants had to show their skill to combine various pieces of information across the different sentences in the text and make inferences on the basis of these pieces of information.

For some test items whose pieces of information necessary for answering them were scattered across different parts of the text, the main reading strategy participants deployed was a careful reading of the whole text (S1=43%). However, some examinees either preceded or followed this careful reading with its rapid reading/rereading (S2=29%). On the other hand, for some other items whose pieces of information were to be found across sentences within a single specific paragraph, participants had to carefully read that specific paragraph (S3=29%). In both cases, some participants (24%) indicated that after this careful reading of the text/ paragraph, they had to draw some conclusions based on what the text/paragraph implied (S10). Additional information on strategies used for answering the aforementioned individual test items is provided in Appendix 17.

In regard to the test taking strategies, the participants used two main strategies in order to answer these item questions. They considered the five alternatives and then postponed to choose the “correct” alternative (S12=33%), or they considered the five alternatives and immediately selected the alternative they found was the “correct” answer (S12=33%). Others participants (24%) selected their “correct” alternative through the process of falsification and elimination of other alternatives they could find not plausible (S17). Last, but not least, few participants (9%)

made their choices of the “correct” answer by using some clues provided in the way the item question was constructed (S15). Appendix 17 presents detailed information on strategies patterns for the aforementioned items.

### **7.3.2 Frequency of strategies used for answering careful reading at local level items**

In this sub-section, I present the frequency of different strategies participants had used for answering careful reading at local level items. This presentation focuses on the strategies used by participants to answer test questions that required their skills to (1) identify and/or understand lexis and (2) understand grammar.

#### **7.3.2.1 Frequency of strategies used by participants to answer items that required their skill to identify and/or understand lexis**

The information displayed in Table 7.2 (see page 219) indicates that, the participants used six (6) reading strategies and four (4) test taking strategies in order to approach the nine (9) test questions that required their skills to identify or/ and understand some words used in the text. The nine (9) questions included in this category tested different vocabulary skills. Of these nine questions, three (3) required the participants to identify the object/thing that appeared/did not appear in the picture/diagram (T3.6; T5.7 and T6.4 [see Appendix 7; Sections 3, 5, and 6 respectively]). Such test questions required their capacity to not only identify the object or thing in the picture, but also to provide a proper name to that object/thing. One test question required examinees to identify the correct sentence about the picture (T6.6 [see Appendix 7; Section 6]). At this juncture, I wish to mention that, this test question is one of the test questions that I have classified in more than one category as it requires the knowledge of vocabulary and grammar. In order to answer this test question, participants had to demonstrate their ability to identify and name various objects contained in the picture as well as their ability to identify the correctly parsed sentence where the objects were included. One other test question required the participants to identify the name of the specific part of the letter (T6.7 [see Appendix 7; Section 6]). This item question, specific to participants from Secretary-Computing subject area, required

participants' knowledge of letter writing parts. The remaining test questions requested participants' knowledge of abbreviations used for different titles (T6.9 [see Appendix 7; Section 6]), or their knowledge to match words contained in one list to actions performed by these words in another list (T8.5 [see Appendix 7; Section 8]), or again their knowledge of word antonyms (T1.4; T8.4 [see Appendix 7; Sections 1 and 8, respectively]).

In order to answer the aforementioned questions, the information contained in Table 7.2 (see page 219) indicates that, nearly half of participants (43%) indicated that they resorted to their knowledge of vocabulary (S8). However, some participants (29%) indicated that they had first to carefully read the whole text before making their choice (S1). This careful reading of the whole text may have aimed to provide examinees the opportunity to grasp the text context and then enable them to efficiently use context clues to figure out the word meanings. For test items that targeted the technical words used in letter writing, some participants resorted to their knowledge of letter writing parts (S5=24%). In some cases, participants had first to look for parts of the text that the writer had recommended (S11=24%) in order to locate a particular word before using clues to figure out its meaning (S9=24%). Data also signal that, for other vocabulary items, some participants (10%) had to resort to their knowledge of grammar (S7). This is the case of the question T6.6 (see Appendix 7; Section 6) that equally called for examinees' knowledge of grammar, as I have explained earlier. The detailed information on strategies used for answering the aforementioned individual test items is provided in Appendix 17.

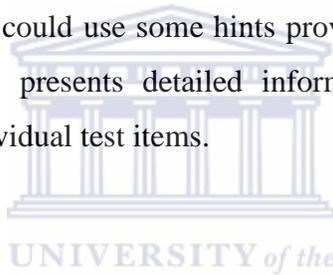
Concerning the test taking strategies participants used to answer these types of questions, the information in Table 7.2 (see page 219) signals that participants used four (4) main strategies. One of these strategies was the examinees' selection of their "correct" option through logical elimination of other alternatives (S17=38%). For some other participants (33%), they could show uncertainty by considering the five alternatives and then delayed to decide which alternative was the correct answer (S12). However, only few participants (24 %) were able to consider the five options and immediately focus on the "correct" answer (S13). Last but not least, data provide evidence that suggests that some participants (19%) selected their answers to some vocabulary items through simple guessing (S18). The detailed information on strategies used for answering the aforementioned individual test items is provided in Appendix 17.

### 7.3.2.2 Frequency of strategies used by participants in answering test questions that required their skill to understand syntax

The information presented in Table 7.2 (see page 219) indicates that participants used five (5) reading strategies and five (5) test taking strategies in order to answer the fifteen (15) test questions that required their skill to understand syntax. The fifteen questions included in this category tested different grammar structures (see Table 7.1 on page 215). Among these questions, five (5) questions required examinees' knowledge of the use of tag-questions (T3.8; T4.4; T5.4; T6.5; T7.5 [See Appendix 7, Sections 4, 5, 6 and 7]), while four (4) questions required their understanding of the use of pronouns and their antecedents (T1.7; T2.5; T3.2; T5.3 [See Appendix 7; Sections 1, 2, 3 and 5 respectively]). The other six remaining questions tested examinees' understanding of distinct syntactic structures that are the different parts of speech (T4.5 [See Appendix 7; Section 4]), use of tenses (T3.3 [See Appendix 7; Section 3]), knowledge of comparison of adjectives (T3.4 [See Appendix 7; Section 3]), use of reported speech (T3.9 [See Appendix 7; Section 3]), knowledge of the asking of *wh*- question (T2.7 [See Appendix 7; Section 2]) and use of prepositions (T6.6 [See Appendix 7; Section 6]).

In regard to the use of reading strategies, the information displayed in Table 7.2 (see page 219) confirms that, in order to answer these items, most participants (79%) used as main reading strategy their knowledge of grammar (S7) as this strategy aided them to identify the correctly parsed sentence. Although this strategy was necessary for answering grammar-based questions, the data suggest that it was not sufficient to approach all the item questions. Other strategies were used as well; for instance, for answering some questions, the participants had first to look for parts of the text that the writer had recommended in order to locate a particular grammar structure (S11=33%). After the structure had been located, the examinees could sometimes carefully read that text part (S3=21%) in order to understand the actual use of that structure. In some cases, examinees had to resort to their knowledge of vocabulary (S8=6%) for item questions that assessed both grammar and vocabulary (see test question T6.6 discussed previously). In rare situations, examinees could read the whole text rapidly (S2=3%) in order to answer an individual grammar item. Appendix 17 presents detailed information on the strategies used for answering the aforementioned individual test items.

Concerning the use of test taking strategies, the evidence offered in Table 7.2 (see page 219) suggests that, in order to answer these grammar-based questions, the majority of participants (61%), through their knowledge of grammar, considered the five question options and immediately focused on the option they perceived was the “correct” answer (S13). In some instances, some participants (18%) produced their own answers immediately after they had read the item question, and then could finally look at the five alternatives to confirm their answers (S14). Detailed data presented in Appendix 17 signal that, the participants used this strategy in answering tag-questions (See test questions T3.8; T4.4; T5.4; T6.5; T7.5 in Appendix 7; Sections 3, 4, 5, 6, and 7 respectively). In some other instances, few participants (7%) selected their “correct” answer though simple eliminations of other options they perceived were not reasonable in the context of use of the structure (S17). Still in other circumstances, very few participants (2%) could ponder the five options but postpone the choice of the correct answer until further consideration (S12); or again they could use some hints provided in some options (S15=2%) to make their choice. Appendix 17 presents detailed information on the strategies used for answering the aforementioned individual test items.



### **7.3.3 Frequency of strategies used for answering expeditious reading at global level items**

In this sub-section, I present the frequency of different strategies participants used for answering expeditious reading at global level items. This presentation focuses on the strategies they had used to answer test questions that required their skills to (1) skim the text/paragraph in order to find the title that best suited the text/paragraph, (2) skim the text/paragraph in order to locate the needed information, (3) search read the text in order to locate implicit information, or again (4) search read the text/paragraph to quickly locate explicit pieces of information scattered through the whole text/paragraph.

### **7.3.3.1 Frequency of strategies the participants used for answering the test questions that required their skill to skim the text/paragraph in order to find the title that best suited the text/paragraph**

The information presented in Table 7.2 (see page 219) states that examinees used seven (4) reading strategies and four (4) test taking strategies in order to answer the six (6) test questions (T1.1; T2.1; T4.1; T5.1; T7.2; and T8.1 [See Appendix 7; Sections 1, 2, 4, 5, 7 and 8 respectively]) that required their skill to skim the text/paragraph in order to find the title that best suited the text/paragraph.

At this juncture, I hasten to mention that these six questions have already been included in the other category of questions that required examinees' skills to establish paragraph/text global comprehension discussed in the first point of Section 7.2.1. This is to suggest that, in order to answer this type of test questions, examinees had to conduct both careful reading and expeditious reading. Having said that, I feel it is not necessary to present again the frequency of reading and test strategies used by participants to answer these test questions because I have already presented this in the first point of sub-section 7.2.1. Nevertheless, I hasten to state at this juncture, that this type of test questions can offer evidence that reading activity may not be linear; rather, it may be a fluid activity that reflects the dynamic nature of language. Besides, this type of items offers the opportunity to critically interrogate many existing taxonomies (see Cohen & Upton, 2006, 2007; Weir & Khalifa, 2008a; for example) that so far seem to categorize the different reading types as discrete and opposing entities where an item categorized in A cannot be also categorized in B. Yet, the data in this study confirm that different reading types can sometime be integrating and overlapping entities and individual items can be categorized in both A and B groups. I will elaborate on this finding in the Discussion Chapter.

### **7.3.3.2 Frequency of strategies the participants used to answer test questions that required their skill to skim the text/paragraph in order to locate needed information**

The evidence offered in Table 7.2 (see page 219) states that, the participants used five (5) reading strategies and three (3) test taking strategies in order to provide answers to the three (3) test questions (T1.2; T4.2 and T8.2 [see Appendix 7; Sections 1, 4 and 8, respectively]) that

required their skill to skim the text/paragraph in order to locate needed information. All these three questions included in this item category requested participants to identify the statement that reflected the implicit information in the text.

In regard to the use of reading strategies, the information presented in Table 7.2 signals that, almost all the participants (91%) read the whole passage rapidly (S2) in order to quickly locate needed information. If the needed information was located in a specific paragraph, examinees could then read that individual paragraph rapidly (S4=36%) and sometimes also carefully (S3=18%). Since all these test questions presented a stem with information not explicitly stated in the text (for example “*In which part of the text can we find the idea that ‘one person can decide on the place where to construct a city?’*” [T4.2; see Appendix 7; Section 4]). However, strategies use suggests that, some examinees (9%) had to draw conclusion on the basis of what the text/paragraph implied (S10). Appendix 17 presents detailed information on the strategies used for answering the aforementioned three test items.

Regarding the frequency of test taking strategies, the statistics in Table 7.2 (see page 219) substantiate that most participants (81%) pondered the five alternatives and immediately focused on the one they perceived was the correct answer (S13). This straight selection of the “correct” option may be justified by the fact that once the examinee was capable to locate the specific part of the text where information needed to answer the question was situated, it might have appeared clear for him/her to easily find the correct option immediately after reading all the five options. In rare cases, participants could either select the “correct” alternative either through paragraph or passage overall comprehension (S16=7%) or through simple elimination of some alternatives they perceived not to be plausible (S17=7%). Appendix 17 presents detailed information on the strategies used for answering the aforementioned three test items.

### **7.3.3.3 Frequency of strategies the participants used to answer test questions that required their skill to search read the text/passage in order to quickly locate implicit information**

The information presented in Table 7.2 (see page 219) suggests that, the participants used four (4) reading strategies and five (4) test taking strategies in order to answer the two (2) test

questions (T1.3 and T3.1 [see Appendix 7; Sections 1 and 3]) that required their skills to search for information not explicitly stated in the text. These two items required the participants to find the sentence that best related to the information implicit in the text. Therefore, these items required the examinees to struggle in order to locate information requested by the test question by quickly reading the entire text as they were not certain of the precise place and form in which the needed information would appear.

I hasten to state that the particularity of this type of questions is that participants had to conduct both careful reading and expeditious reading of the entire text in order to locate the necessary pieces of information. And since the needed information was implicit in the text, examinees had to make text inferences in order to approach this type of items. This is the reason why I have also included these two items in the category of items that test careful reading at global level where participants' skill to infer from implicit information in the text was needed to answer these questions. Such a classification of individual items in two different reading types substantiates the overlapping nature of reading skills as evidenced by research. I shall elaborate on this in the Discussion Chapter.

Concerning the use of reading strategies, the information contained in Table 7.2 (see page 219) confirms that, in order to answer these items, the majority of participants had to read the whole text both carefully (S1=91%) and speedily (S2=87%) in order to locate the part of the text where the needed information was localized. Once the necessary information was located in a specific paragraph, the participants could read that specific paragraph both carefully (S3=87%) and expeditiously (S4=70%). Appendix 17 presents detailed information on the strategies used for answering the aforementioned two test items.

Regarding the frequency of test taking strategies, the statistics in Table 7.2 (see page 219) suggest that for some participants, once the information needed was located in a specific part of the text and inferences made, they either produced their own answers immediately after reading the question and then after could look at the five suggested choices for confirmation (S14=40%), or they simply proceeded by the elimination of all choices that did not appear to be plausible so as to remain with few choices to choose from (S17%=40%). In view of the complex nature of this type of questions, some participants chose their "correct" answer either by considering the five choices and then postponed the selection of the "correct" choice (S12=20%), or by using

some clues provided in the items (S15=20 %). Appendix 17 presents detailed information on the strategies used for answering the aforementioned two test items.

#### **7.3.3.4 Frequency of strategies used by participants to answer test questions that required their skill to search read the text in order to quickly locate different pieces of information scattered through the whole text**

The data presented in Table 7.2 (see page 219) indicate that the participants used five (4) reading strategies and six (4) test taking strategies in order to answer the ten (10) questions that required their skill to search read the text in order to quickly locate different pieces of information scattered through the whole text. All the ten items (T1.5; T2.2; T2.4; T2.6; T3.4; T3.7; T4.3; T6.2; T6.3 and T7.1 (see Appendix 7; Sections 1, 2, 3, 4, 6 and 7, respectively) requested the participants to indicate the sentence that agreed/did not agree with text information. In order to find out this sentence, the participants had to rapidly read the entire text, trying to identify any piece of information necessary to answer the test question. Of the ten (10) test questions included in this category, three (3) had also already been categorized in other previous categories. This is the case of test items T4.3 (classified in careful global reading and targeting the skill of establishing accurate comprehension of explicitly stated main ideas and supporting details), and test items T6.2 and T6.3 (classified in careful global reading and targeting the skill of making text/paragraph inference).

Regarding the use of reading strategies, the information contained in Table 7.2 (see page 219) confirms that, in order to answer these items, the majority of participants (72%) had to rapidly read the entire text (S2) in order to locate any pieces of necessary information to the completion of the item task. Since these pieces of information were scattered in the different parts of the text, expeditious reading of the entire text was not sufficient in itself. Therefore, some participants used other types of reading that were conducted before or after the expeditious reading in order to augment their comprehension of the text and locate needed pieces of information. This is the case of careful reading of a portion of the text (S3=28%) that was generally conducted after expeditious reading of the whole text. This type of reading aimed to search any useful piece of information located in a specific paragraph and that could contribute to answering the test

question. Some participants attempted a careful reading of the whole text (S1=20%) either before or after they had conducted the main search reading activity. However, since the participants could not predict the location of information necessary for answering this type of items, and since they could not anticipate the form the needed information could take, some participants found it important to augment their understanding of the text by using their knowledge of grammar (S7=20%). This strategy could aid them in parsing sentences so as to construct an appropriate comprehension of specific piece of information located in a specific part of the text. However, the data indicate that, in some situations, the participants' knowledge of grammar was not enough as some of them resorted to their knowledge of vocabulary in order to understand the actual meaning of some words through their context (S8=4%).

I wish to state that, the use of multiple strategies to answer this type of questions explains why I have classified item questions T4.3, T6.2 and T6.3 into more than one category. To be more specific, item question T4.3 (see Appendix 7; Section 4) required the participants' skill to establish accurate comprehension of explicitly stated main ideas and supporting details across sentences. In order to answer this test question, most participants read the whole text carefully (S2) first, and once the information needed could be located, they then read part of the text carefully (S3). Likewise, test questions T6.2 and T6.3 (see Appendix 7; Section 6), since they required the participants to indicate the sentence that agrees/does not agree with the text, demanded their skills of inferring from implicit information contained in the text as all the five alternatives stated information that was implicit in the text. Therefore, both a careful reading of the entire text (S2) as well as a careful reading of a specific part of the text (S3) assumed to be relevant strategies to answer these test questions. Appendix 17 presents detailed information on the strategies used for answering the aforementioned individual test items.

As for the use of test taking strategies, the information contained in Table 7.2 (see page 219) confirms that, the participants used different strategies in order to answer test questions that required their skills to search read the text in order to quickly locate different pieces of information scattered through the whole text. For some test questions, participants just selected the "correct" option through an elimination of the other options (S17=48%). For other test questions, they made their choices by selecting the correct option just after inspecting all the five options (S13=28%) or by simply guessing (S18=24%), or again through overall text

comprehension (S16=24%). In rare cases, some participants made their choices by using clues provided in the five options (S15=8%).

#### **7.3.4 Frequency of strategies used for answering expeditious reading at local level items**

In this last sub-section, I present the frequency of different strategies the participants had used for answering expeditious reading at global local items. The main skill that characterizes these items is scanning. This is to suggest that, in order to provide answer to these test questions, participants had to look for specific words and phrases in the specific parts of the text, usually indicated by the writer.

The information presented in Table 7.2 (see page 219) indicates that the participants used five (5) reading strategies and six (6) test taking strategies in order to answer the eight (8) test questions (T1.4; T17; T2.5; T3.2; T4.5; T5.3; T7.3; and T7.4 [see Appendix 7; Sections 1, 2, 3, 4, 5 and 7]) that required their skill to scan a specific part of the text in order to locate explicit piece of information.

I hasten to state that five (5) of these eight test questions have already been classified in other categories. More specifically, T1.4 has been categorized in careful reading at local level as it involved the examinees' ability to carefully read a text portion and understand lexis. This generally involved examinees' knowledge of vocabulary. Likewise, T1.7; T2.5; T3.2 and T4.5 have been classified in careful reading at local level as these questions required the participants to carefully read a specific part of the text and understand syntax by parsing sentences. Such understanding generally involved examinees' knowledge of grammar. Like in the other cases discussed so far, these test questions provide evidence that suggests that careful reading and expeditious reading can sometime be interacting activities and the boundary between these two reading types can sometimes be blurred; suggesting that a specific reading task can equally well requires examinees to process the text both carefully and expeditiously.

Regarding the use of reading strategies, the information contained in Table 7.2 (see page 219) confirms that, in order to answer these items, the majority of participants (63%) had to look for parts of the text indicated by the writer (S11 [for example T1.4: "*The contrary of 'hardly' underlined in the last paragraph is:*"]). Once the target word located, the participants had first to carefully read that text portion (S3=32%) in order to either use clues provided by the context to

understand the word, or parse the sentence in order to understand the actual use of the grammar structure. In most cases, participants were not able to identify/understand lexis or understand grammar until they could also read expeditiously the same text portion (S4=21%); thus combining both careful reading and expeditious reading. I wish to mention that, the use of careful reading in this context seems to be conducted in order to confirm the understanding of the target lexis or grammar structure. In both types of reading, participants had to resort to their knowledge of grammar (S7=42%) as well as their knowledge of vocabulary (S8=11%). In Appendix 17, I have presented detailed information on the strategies used for answering the aforementioned individual test items.

Concerning the use of test taking strategies, the information provided in Table 7.2 (see page 219) confirms that two strategies were dominant: For some test questions, almost nearly a half of participants selected the “correct” option after the process of elimination of all implausible options (S17=47%); or they produced the answer immediately after reading the question, and then after looked at the five options only to confirm the answer (S14=42%). This latter strategy was most common with grammar items that required the participants’ skill to ask tag questions (T3.8; T4.4; T5.4; T6.5 and T7.5) or to report what somebody had said (T3.9). In Appendix 17, I have provided extensive information on the patterns of strategies used for answering these individual test items. Depending on the item’s specific task and the examinee’s skill, other participants made their choice either by considering the five choices and then postpone the selection of the “correct” choice until further confirmation (S12=26%) or by inspecting the five choices and immediately focusing on the “correct” choice (S13=26%), or again by simply guessing the answer (S18=21%). Furthermore, the data provide evidence of the use of some clues provided in the choices to guide examinees’ selection of the “correct” answer (S15=11%).

#### **7.4 Reading types and task complexity**

In this section, I present the task complexity of the four reading types as reflected in the relative difficulty of different test questions that targeted each of the four reading types. In order to achieve this objective, I first computed the item difficulty index of each individual item (see Appendix 17). This is a straightforward statistics that consisted in taking the number of

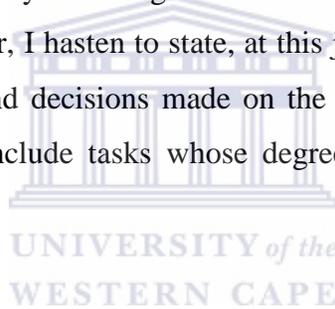
candidates who got an item right and dividing this number by the total number of candidates who answered that item. From the results of item difficulty per item, I computed the mean item difficulty for each different reading skill included in each of the four reading types. In order to show any significant variance, I computed the *standard deviation*. This is a statistics that aimed to reflect the dispersion (variance) observed among item difficulty index of items included in the same group. Table 7.3 presents information on the mean item difficulty for each reading type. In this table, values closer to .0 suggest that the item is difficult; values closer to 1 suggest that the item is easy, whereas values that cluster around .5 suggest that the item is of moderate difficulty.

Table 7.3: Mean item difficulty for different reading skills included in each reading type

Reading type		N items	Mean item difficulty	SD
<b>1. Careful reading at global level: Item difficulty for different skills</b>				
1	Establishing text/paragraph global comprehension	6	.4	0.20
2	Establishing accurate comprehension of explicitly stated main ideas and supporting details across sentences	3	.3	0.17
3	Making text/paragraph inferences	7	.5	0.18
<b>Total careful reading at global level</b>		<b>16</b>	<b>.4</b>	<b>0.18</b>
<b>2. Careful reading at local level: Item difficulty for different skills</b>				
1	Identifying/understanding lexis	9	.5	0.18
2	Understanding syntax	15	.6	0.16
<b>Total careful reading at local level</b>		<b>24</b>	<b>.6</b>	<b>0.17</b>
<b>3. Expeditious reading at global level</b>				
1	Skimming to find the title that best fits the text/paragraph	6	.4	0.20
2	Skimming the text/paragraph to locate needed information	3	.7	0.06
3	Search reading to quickly locate implicit information	2	.6	0.07
4	Search reading to quickly locate explicit pieces of information scattered through the whole text/paragraph	10	.4	0.15
<b>1. Total expeditious reading at global level</b>		<b>21</b>	<b>.5</b>	<b>0.12</b>
<b>Expeditious reading at local level</b>				
1	Scanning to locate explicit information located in a specific part of the sentence/paragraph	8	.5	0.20
<b>Total expeditious reading at local level</b>		<b>8</b>	<b>.5</b>	<b>0.20</b>

#### **7.4.1 Item difficulty for careful reading at global level test questions**

The data presented in Table 7.3 on the previous page indicate that the test questions that targeted careful reading at global level were difficult (Mean item difficulty=.4). However, within this category of items, if the test questions that required participants to establish text/paragraph global comprehension and those that aimed to establish accurate comprehension of explicitly stated main ideas and supporting details across sentences appeared to be difficult for participants (ID=.4 and .3 respectively), those test questions that aimed to make text/paragraph inference appeared to be of moderate difficulty (ID=.5). These results suggest that careful reading at global level may be a challenging task participants might have been able to complete. In light of this suggestion, a relevant question is to know whether or not, in light of the actual ESE context described in Chapter 6, test developers need to include many test questions that require examinees to read the text carefully and at global level. Although I propose to answer this question in the Conclusion Chapter, I hasten to state, at this juncture, that since validity is about the relevance of interpretations and decisions made on the basis of test scores, in order to be construct-valid, a test needs to include tasks whose degree of complexity is proportional to examinees' characteristics.



#### **7.4.2 Item difficulty for careful reading at local level test questions**

The evidence provided in Table 7.3 on the previous page suggests that, the test questions that targeted careful reading at local level were of moderate difficulty (Mean item difficulty=.6) as confirmed in the item difficulty indices of test questions that required examinees' skill to identify/understand lexis (Mean ID=.5) and those test questions that required participants to show their understanding of syntax (Mean ID=.6). Nevertheless, it is worth mentioning that these last two indices suggest that the participants found the test questions that required their knowledge of grammar easier than those that required their knowledge of vocabulary. In Appendix 17, I have provided detailed information on item difficulty for individual test questions included in this category.

### **7.4.3 Item difficulty for expeditious reading at global level test items**

The information presented in Table 7.3 (see page 235) indicates that, the test questions that targeted expeditious reading at global level were of moderate difficulty level (ID=.5). However, the relatively high SD (0.12) suggests some variations across test questions that targeted different reading skills. More specifically, test questions that required examinees' skill to skim the text in order to find the title that best fits the text/paragraph and those test questions that demanded the examinees' capacity to quickly read the text in order to locate explicit pieces of information scattered through the whole text/paragraph appeared to be difficulty (ID=.4) for participants. On the contrary, test questions that required the participants to skim the text/paragraph in order to locate needed information appeared to be easy for participants (ID=.7), while those questions that demanded participants' ability to quickly read the text/paragraph in order to locate implicit information appeared to be of average difficulty (.6). In Appendix 17, I have provided further details on the item difficulty for individual items included in this category.

### **7.4.4 Item difficulty of expeditious reading at local level test questions**

The statistics offered in Table 7.3 (see page 235) signal that, participants found the test questions that targeted expeditious reading at local level to be of moderate difficulty level (ID=.5). All these test questions required participants' capability to scan in order to locate explicit information located in a specific part of the sentence/paragraph. Detailed information on the difficulty level of individual test questions is provided in Appendix 17.

### **7.4.5 Comparing task complexity for different reading types**

This section summarizes results on task complexity by providing evidence on the relative focus between for different reading types, and by comparing task complexity for test questions targeting careful reading and test questions targeting expeditious reading on the one hand, and test questions targeting reading at global level and test questions targeting reading at local level, on the other hand.

- **Careful reading versus expeditious reading: item frequency and mean item difficulty**

This sub-section aims to provide evidence on the relative focus between careful reading and expeditious reading as well as the relative difficulty of tasks that targeted these two reading types. Table 7.4 presents information on the item frequency and mean item difficulty for the test questions that targeted careful reading and expeditious reading.

Table 7.4: Comparing test focus and task difficulty between careful reading and expeditious reading

Careful reading versus expeditious reading: Item frequency and mean item difficulty		Item frequency		Mean item difficulty	SD
		N=69	%		
1	Careful reading	40	58	.5	0.18
2	Expeditious reading	29	42	.5	0.16

In light of the statistics presented in Table 7.4, it can be concluded that, the ESE constructors put more focus on careful reading tasks (58%) than on expeditious reading tasks (42%). This is to suggest that, there are more test questions that required examinees to read the text carefully than those that required them to read the text rapidly. Furthermore, the same statistics suggest the equality of complexity of tasks that target these two types of reading as reflected in their equal mean item difficulty (.5), indicating that test tasks that target these two reading types were of average difficulty. Since the present study conceptualizes construct-validity in terms of the appropriateness of test tasks to the test context, these findings prompt the issue of the ESE construct validity. I will elaborate on this issue in the Discussion Chapter and the Conclusion Chapter.

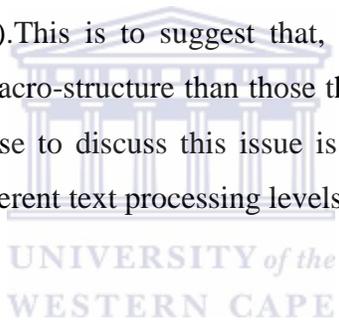
- **Reading at global level versus reading at local level: item frequency and ESE focus**

This sub-section aims to provide evidence on the relative focus between reading at global level and reading at local levels as well as the relative difficulty of tasks that targeted these two reading types. Table 7.5 on the next page provides information on the item frequency and mean item difficulty of test questions that target reading at global level and reading at local level.

Table 7.5: Comparing test focus and task difficulty between reading at global level and reading at local level

Reading at global level versus reading at local level: Item frequency and mean item difficulty		Item frequency		Mean item difficulty	SD
		N=69	%		
1	Reading at global level	37	54	.5	0.15
2	Reading at local level	32	46	.6	0.19

The information presented in Table 7.5 suggests that the ESE constructors put more focus on reading at global level than reading at local level as the data indicate that the ESE includes more test questions that require examinees to read at global level (54%) than questions that require them to read at local level (46%). This is to suggest that, there are more test questions that required examinees to build test macro-structure than those that require them to process the text at the propositional level. I propose to discuss this issue in details in the next section that relates to the description of the different text processing levels required by the ESE tasks.



### **7.5 The reading construct as investigated through processing levels targeted by the ESE tasks**

Although the description of the different reading types examinees conduct while writing the ESE provide evidence on the complexity and relative importance of ESE tasks, this information is not sufficient to clearly grasp what specific tasks might appear easy or problematic to examinees. This is the reason why a description of task complexity in terms of processing levels targeted by the test tasks may augment the understanding of ESE task complexity.

The process-oriented model of reading I have used in this study hypothesizes that the intensity and quality of reading comprehension processes examinees engage in can vary considerably with different levels of reading comprehension assessed by individual test items. In light of this hypothesis, the model suggests that in order to fully cover the reading construct, the test needs to include tasks that require examinees to process the text at the following six levels: lexical

encoding, syntactic parsing, establishing propositional meaning, inferencing, building mental model and creating a text level representation.

In order to investigate the processing levels targeted by the ESE test questions, I have classified, on the basis of strategies use perspective, the 56 test questions in the different six levels. Level 1 relates to the encoding of words from the text. In this level, I have included all test questions that required participants' skill to understand vocabulary. Level 2 relates to the parsing of individual text sentences. In this level, I have included all test questions that required participants' skill to understand grammar. Level 3 relates to establishing propositional meaning at clause or sentence level. In this level, I have included all test questions that required participants to add any external knowledge to the text in order to understand the actual meaning of individual propositions. Level 4 relates to inferencing, and it comprises all test questions that required participants to link information necessary to answer the test question to information not explicitly stated in the text. Level 5 relates to building the text mental model. In this level, I have included all test items that required the participants' ability to establish accurate comprehension of explicitly stated main ideas and supporting details across sentences. Finally, level 6 relates to creating text level structure. In this level, I have included all test questions that require examinees to carefully read the text at global level in order to construct its macro-structure.

In text processing literature, Levels 1, 2 and 3 are lower levels as they require reading at local level while levels 4, 5 and 6 are higher levels since they require the reader to engage with the text at global level. One main assumption in text processing theory is that lower level tasks are generally easier to complete than higher level tasks.

In Table 7.6 (see page 242), I have classified the 56 items from the corpus according to this taxonomy. This classification not only enables to explore the degree of complexity of the ESE tasks in relation to examinees' characteristics, but it also provides a picture of the extent to which the ESE tasks appropriately cover the reading construct.

Before I present the data on the different levels at which ESE tasks are processed, I hasten to voice some crucial considerations. First, I have classified some test items in more than one processing level. This is the case of test question T6.6 (see Appendix 7; Section 6) that required both lexical encoding (Level 1) and syntactic parsing (Level 2); test question T3.4 (see Appendix

7; Section 3) that required the participants' capacity to construct text macro-structure (Level 6) as well as their capacity to parse sentence (Level 2); test question T4.3 (see Appendix 7; Section 4) that required the participants to both build a text mental model (Level 5) and create its macro-structure (Level 6). Other test questions that I classified in two levels are T6.2 (Levels 4 and 6), T6.3 (levels 4 and 6), T1.3 (levels 4 and 6) and T3.1 (levels 4 and 6). In Appendix 7; Sections 1, 3, and 6, I have provided the content of these questions while in Appendix 17, I have provided details on the strategies used to answer these test items. Secondly, some other test questions from the corpus could not be classified in any of the six levels. This is typical of some test questions on letter writing where I found that some questions based on the letter did not actually require examinees to process the letter at any level. These test items are T5.6, T5.7 and T5.8 (see Appendix 7; Section 5). For example, T5.7 required examinees to indicate the part of the letter that is not included. In order to answer this question, examinees did not process the letter at any of these six levels, but they only used their knowledge of letter writing parts to identify the missing part (see Appendix 17 for details on strategies used to answer these questions).

These two considerations justify why the total number of items in Table 7.6 (see next page) does not tally with the total number of items in the corpus. These considerations can also provide evidence that the six processing levels as articulated in the literature may not strictly be hierarchical levels where readers linearly process the text from lower level to higher level. Rather, they suggest that processing levels may be just labels and that actual text processing may be conducted in a forth and back process. This suggestion appears to reflect the fluid nature of reading as well as the dynamic nature of language. I will elaborate on this finding in the Discussion Chapter.

### **7.5.1 The processing levels targeted by the test items**

Table 7.6 (see next page) presents information on the different processing levels targeted by the 56 sampled test questions. For each processing level, the test questions that required the participants to process the text at this level are computed as well as their relative frequency in percentage and their item difficulty.

Table 7.6 Processing levels required by the ESE tasks

Processing level		Test Items	N=57	%	Mean ID
Level 1	Lexical encoding	T1.4; T3.6; T5.6; T5.7; T6.4; T6.6; T6.7; T6.9; T7.4; T8.4;	10	17.5	.5
Level 2	Syntactic parsing	T1.7; T2.5; T2.7; T3.2; T3.3; T3.4; T3.8; T3.9; T4.4; T4.5; T5.3; T5.4; T6.5; T6.6; T6.7; T7.5	16	28.1	.6
Level 3	Propositional meaning	No items found	0	0.0	-
Level 4	Inferencing	T1.3; T2.8; T3.1; T5.2; T6.2; T6.3; T7.3; T8.3	8	14.0	.5
Level 5	Building a mental model	T2.3; T4.3; T6.1	3	5.3	.3
Level 6	Creating a text level representation	T1.1; T1.2; T1.3; T1.5; T2.1; T2.2; T2.4; T2.6; T3.1; T3.4; T3.7; T4.1; T4.2; T4.3; T5.1; T6.2; T6.3; T7.2; T8.1; T8.2;	20	35.1	.4

A scrutiny of the information presented in this table indicates that, the 56 test items included in the study sample targeted five of the six processing levels. This finding suggests that the test items included in the study sample did not cover the entire reading construct. Furthermore, the data reveal that two levels were most represented in the corpus: level 6 (creating text level structure) and level 2 (syntactic parsing) that included 20 items (35.1%) and 16 items (28.1%) respectively. Two other levels were fairly represented: level 1 (lexical encoding) and level 4 (inferencing) that included 10 items (17.5) and 8 items (14.8%) respectively. However, one level was rarely represented in the corpus. This is level 5 (building the text mental model) comprised only 3 items (5.3%). On the contrary, no test questions targeted level 3 (propositional meaning).

### 7.5.2 ESE focus: Lower level processing or higher level processing?

In order to find out whether the focus of ESE is on lower level processing or higher level processing, I have computed all test items that requested lower level processing and those that requested higher level processing.

The evidence provided in Table 7.7 suggests that the 57 test items from the study sample included more test items that required examinees to process the text at the higher level (31 items=54.4%) than test items that required examinees to process the text at lower level (26 items=45.6%). This finding prompts the issue of the appropriateness of ESE tasks to the ESE context as it can be argued whether constructing a test that includes more tasks that require examinees to process the text at higher level is appropriate to the ESE specific context.

Table 7.7 Lower level processing versus higher level processing

Processing levels				N	%	Mean ID
1	Lower level	1	Lexical encoding	10	17.5	.5
		2	Syntactic parsing	16	28.1	.6
		3	Establishing propositional meaning	0	0.0	-
	<b>Total lower level processing item questions</b>			<b>26</b>	<b>45.6</b>	<b>.6</b>
2	Higher level	1	Inferencing	8	14.0	.5
		2	Building a text mental model	3	5.3	.3
		3	Creating a text level representation	20	35.1	.4
	<b>Total higher level processing items</b>			<b>31</b>	<b>54.4</b>	<b>.4</b>

Although the answer to this question forms a part of the answers to the study research question that I propose to provide in the Conclusion Chapter, I hasten to mention at this juncture that, by considering the Grade 12 students' characteristics and the teaching and learning environment of English course in general and reading in particular, it sounds inappropriate to design an English test where examinees must complete the tasks that mainly request them to process the text at higher level. This finding raises the issue of construct- underrepresentation variance and construct-irrelevant variance. This issue is relevant to this study as the understanding of these two concepts may throw some light for evaluating the ESE validity in terms of content coverage as well as appropriateness of test tasks to the test context. This is to suggest that, the relevant approach to evaluating the ESE validity appears to be the one that examines test tasks not only in terms of the extent to which they target the six processing levels that encompass the whole range of reading construct, but also in terms of the extent to which the complexity of these tasks tap into the ESE context. This conceptualization of construct validity appears to be a development

that this study proposes to consider. I will elaborate on this issue in the Discussion Chapter; and in the Conclusion Chapter, I will propose a reconceptualization of the notions of construct-underrepresentation and construct-irrelevant variance. This, I believe can qualify as one of my contributions to the body of testing literature.

### **Summary of the chapter**

In this chapter, I have presented and analyzed the data that sought to determine the variety and degree of complexity of the different reading types examinees conduct while completing the ESE along with the cognitive demands required by different text processing levels to complete these tasks. The following is the summary of findings:

- (1) The 56 test items included in the study sample covered all the four reading types. However, the reading type that included the highest number of items was careful reading at local level, followed by expeditious reading at global level, then careful reading at global level and finally expeditious reading at local level;
- (2) Test items that targeted careful reading at global level required the skills to (a) establish the text/paragraph global comprehension, (b) establish accurate comprehension of explicitly stated main ideas and supporting details across sentences, and (c) make text/paragraph inferences;
- (3) Test items that targeted careful reading at local level required the skill to (a) identify and/or understand lexis and (b) to understand grammar;
- (4) Test items that targeted expeditious reading at global level requested the skills to (a) skim the text in order to find the title that best suits the text/paragraph, (b) skim the text in order to locate needed information, (c) speedily read the entire text in order to figure out implicit information, and (d) quickly read the entire text in order to locate explicit pieces of information scattered in different parts of text;
- (5) Test items that targeted expeditious reading at local level called for the skill to scan the text in order to locate explicit information located in a specific part of the sentence or paragraph;

(6) In order to answer test questions that targeted careful reading at global level, the participants had to read the whole text carefully (S1) and sometimes reread it rapidly (S2) in order to build its gist. Then, they had to read a specific paragraph of the text carefully (S3) first and then rapidly (S4) in order to build the gist of that paragraph;

(7) In order to answer test items that targeted careful reading at local level, participants used their knowledge of vocabulary (S8) and/or grammar (S7);

(8) In order to answer test items that targeted expeditious reading at global level, participants had to read the whole passage rapidly (S2) in order to quickly locate needed information. If the needed information was located in a specific paragraph, they had to read that individual paragraph rapidly (S4) and sometimes also carefully (S3);

(9) In order to answer test questions that targeted expeditious reading at local items, participants had to look for parts of the text indicated by the writer (S11) before they could read that text portion carefully (S3); and sometimes expeditiously (S4);

(10) In order to answer the 56 test items, participants mainly used the following test taking strategies: consider the five alternatives and immediately select the alternative that appeared familiar to them (S13), or consider first the five alternatives and then postpone consideration of what could appear to them as the best alternative (S12) before selecting the answer either through elimination of other alternatives (S17), or through guessing (S18). In some cases, they had to produce their own answers immediately after reading the question and then could inspect the five alternatives in order to confirm their choices (S14), or again they could use some clues provided in some alternatives (S15);

(11) Test questions that targeted careful reading at global level were difficult for the majority of participants. Within this category, test questions that required examinees to establish text/paragraph global comprehension and those that aimed to establish accurate comprehension of explicitly stated main ideas and supporting details across sentences appeared to be more difficult than test questions that aimed to make text/paragraph inferences;

(12) Test questions that targeted careful reading at local level were of moderate difficulty for the participants. Within this category, test questions that required examinees' skill to identify and/or

understand lexis appeared to be more difficult than those test questions that required participants to show their understanding of syntax;

(13) Test questions that targeted expeditious reading at global level were of moderate difficulty. Within this category, test questions that required examinees' skill to skim the text in order to find the title that best fits the text/paragraph and those test questions that demanded examinees' capacity to quickly read the text in order to locate explicit pieces of information scattered through the whole text/paragraph appeared to be difficulty for participants. On the contrary, test questions that required participants to skim the text/paragraph in order to locate needed information appeared to be easy for participants, while those questions that demanded participants' ability to quickly read the text/paragraph in order to locate implicit information appeared to be of average difficulty;

(14) Test questions that targeted expeditious reading at local level were of moderate difficulty level;

(15) There were more test questions that targeted careful reading than those that targeted expeditious reading. Nonetheless, both careful reading test items and expeditious reading test items were of moderate difficulty level;

(16) There were more questions that targeted higher level processing than those targeting lower level processing. Yet, test questions that targeted higher level processing were more cognitively demanding than those that targeted lower level processing;

(17) The 56 test questions from the corpus did not target all the six processing levels in equal proportions; besides, it was found that no test questions targeted level 3 (propositional meaning);

Having summarized the main chapter findings in this section, I propose to elaborate on them in the Discussion Chapter. However, at this juncture, I wish to present and analyze, in the following chapter, data that investigate the textual features that likely impact on the difficulty of ESE tasks.

## **CHAPTER EIGHT**

### **FINDINGS RELATED TO THE TEXTUAL FEATURES THAT POTENTIALLY AFFECT THE VALIDITY OF THE ENGLISH STATE EXAMINATION**

#### **8.0 Introduction**

The process-oriented model I have situated in this study acknowledges the impact of textual features on text comprehension and test performance. In this chapter, I propose to investigate different textual features that likely impact on the difficulty of ESE tasks; and evaluate the extent to which the ESE includes texts whose features are appropriate to the ESE context. The textual features that are investigated in this chapter include: 8.1 the content of ESE texts, 8.2 the different genres included in the ESE texts and 8.3 the density of ESE texts. The main data used in this chapter are the eight text passages and one letter included in the eight ESE papers. Nevertheless, these are sometimes supplemented by the reports of the 496 student-participants from the strategies questionnaire.

However, although this chapter strictly relates to the analysis and presentation of study results, I wish to state that, for the sake of narrative immediacy and primacy of findings, I will from time to time conduct an initial discussion of some of the study findings. My desire to adopt such a methodological presentation is justified by my need to highlight the relevance of the salient findings to the study and foreshadow the relationships of these findings to the study aim and objectives. Nevertheless, in the Discussion Chapter, I expect to elaborate on the initial discussion provided in the present chapter. In the following section, I investigate the content of the ESE texts and the extent to which these ESE text contents are relevant to ESE context.

## 8.1. The content of ESE texts

One theory that informs this study is the construction-integration theory. This theory stresses the role of the reader's knowledge in text processing and comprehension. The main assumption of the theory is that text comprehension combines a construction phase in which a textbase that contains the propositional meaning is constructed from the linguistic input, and an integration phase in which the textbase is integrated into the reader's knowledge in order to form a coherent mental representation of what the text is about (Kintsch, 1988, 1998, and 2004, van Dijk & Kintsch, 1983). In light of this assumption, it is expected that, in order to comprehend a text, the reader must use his/her prior knowledge to integrate meanings of individual sentences into a coherent representation of events contained in the text. Therefore, readers with high level of background knowledge on the text content may comprehend the text better than those readers with low level of background knowledge on the text content.

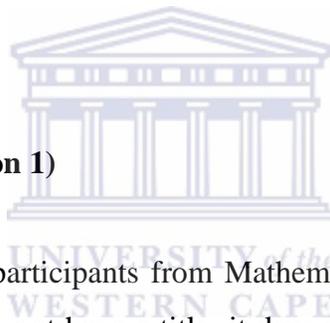
In order to investigate the content of ESE texts and evaluate the extent to which these texts are appropriate to the ESE actual context described in chapter 6, I have investigated two variables: first, the participants' degree of familiarity with the text content, and then their perceptions of text content difficulty.

At this juncture, I wish to mention that I have investigated text content familiarity in relation to the text theme and the extent to which this theme relates to examinee's background knowledge, and/or in relation to the kind of themes suggested in the English curriculum that students are recommended to explore in classroom reading. I have hypothesized that when an examinee reads a text whose content relates to his/her background knowledge or is similar to the kind of texts he/she has been reading in classroom context, this examinee can perceive the text content to be familiar.

On the other hand, I have examined text content difficulty in relation to three variables: first, the extent to which the vocabulary used in the text may be encoded by examinees through their knowledge of vocabulary or through some clues provided in the context. Secondly, the extent to which the text includes grammar structures that are suggested in the English curriculum as materials to be covered in classroom learning situations. In relation to these two variables, I hypothesized that, when readers process texts with vocabulary that they can easily encode and

when they can easily parse text sentences, they are likely to comprehend these texts. Thirdly and lastly, the presence/absence of a text title and subtitles. I hypothesized that when examinees process a text with a title and subtitles, they better understand them as the presence of a text title/subtitles can guide them in figuring out what the text is going to be about.

However, my own perception, although useful, is nevertheless not sufficient enough to establish text content difficulty and familiarity. This is the reason why I have also presented the 496 participants' perceptions of text content difficulty and familiarity from the reports obtained from Section 3 of the Strategies questionnaire (see Appendix 3). In this section, two (2) questions requested the participants to indicate their perceptions of text content difficulty (Question1) and their perceptions of their degree of familiarity with text content (Question3). In the following subsection, I present T1 (test one) that was that was included in the test materials (see Appendix 7).



#### **8.1.1. T1 (see Appendix 7, Section 1)**

This test was administered to 75 participants from Mathematics-Physics subject area. The text passage used in the test paper does not have a title; it does not have subtitles either. The text is about the various consequences people experience with AIDS. The text tells a story of a young happy couple that soon discovers that one partner (the husband) is HIV-infected. Despite their joy at the birth of their first daughter, the child cannot survive as she dies from AIDS; and soon after, the husband also dies from AIDS-related complications. As a consequence, the wife finds that the only way for her to survive is not to get married again.

In regard to text familiarity, this text appears to be familiar to students as it deals with a current issue that affects their immediate environment. Also, since the issue of HIV-AIDS is one of the most talked about and publicized issue in international, national, and local media, students are expected to be familiar with this text content.

In regard to content difficulty, this text is written in a style that a good reader can be able to approach without much difficulty. Many words and expressions that are used can be understood through reader's knowledge of vocabulary or context clues (For example: *AIDS*, *HIV prevention*,

*cheerful smile, tragic story, ... were overjoyed at the birth of their daughter, to be in good health, she stopped gaining weight, etc.*). Also, most grammar structures used are those planned in the English curriculum and therefore, they can be approached by an average reader. This is the case of the use of simple past tense plus time markers to narrate a past event (*At the outset, Asha seemed to be in good health. Yet, she stopped gaining weight...; Kirongozi ... also began to get sick; One day he collapsed...*); or the use of time adverbials to indicate time sequence (*In 1997, Yasekuru...; At the age of three, Asha died...; A few years later, Kirongozi...; Yasekuru now lives...*). Therefore, since the text content appears to be familiar to the reader, the latter can easily activate his/her schemata to comprehend the text. Thus, the content of this text may be easy or of average difficulty for the majority of participants.

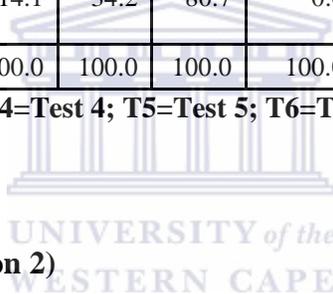
However, the 75 participants to whom the test was administered had different perceptions of text difficulty and content familiarity. Although this text discussed a theme that was believed to be familiar to them, the information in Table 8.1 on the next page indicates that only few participants (38.7%) perceived the text content to be familiar to them. On the contrary, the majority of participants (61.3%) perceived the text content to be somewhat familiar to them. Nevertheless, none of the participants perceived the text content to be unfamiliar to them. This finding suggests that all participants perceived, to some degree, some familiarity with the text. This finding appears to be consistent with my evaluation of text content familiarity.

Concerning the content difficulty, the data in Table 8.1 on the next page suggest that only one in three participants (33.3%) perceived the text content to be easy, while the majority of participants reported that the text was either of average difficulty (44.0%) or difficult (22.7%). If the frequency of participants who found the text content easy is summed with the frequency of participants who found it of average difficulty, the total number of participants who could comprehend the text to some degree were 77.33 %; suggesting that the majority of participants could read the text and comprehend it. This finding appears to be also in agreement with my previously stated perception of text content difficulty.

Table 8.1: Students' perceptions of text content difficulty and familiarity

Options		T1	T2	T3	T4	T5	T6	T7	T8
<b>Text content difficulty</b>									
1	Very easy	0.0	3.1	0.0	0.0	0.0	0.0	0.0	0.0
2	Easy	33.3	9.4	6.1	0.0	4.7	8.6	58.3	7.2
3	Average difficulty	44.0	9.4	46.3	0.0	64.5	31.4	27.3	82.1
4	Difficult	22.7	56.3	41.5	50.9	24.3	60.0	14.6	10.7
5	Very difficult	0.0	21.9	6.1	49.1	6.5	0.0	0.0	0.0
<b>Total</b>		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
<b>Text content familiarity</b>									
1	Familiar	38.7	6.3	0.0	0.0	50.5	20.0	56.3	32.1
2	Somewhat familiar	61.3	79.7	65.9	19.3	49.5	80.0	43.8	67.9
3	Not familiar	0.0	14.1	34.2	80.7	0.0	0.0	0.0	0.0
<b>Total</b>		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8



### 8.1.2. T2 (see Appendix 7, Section 2)

This test was administered to 64 participants from the Bio-Chemistry subject area. The text passage used in the test paper does not have a title and subtitles. The text is about trouble between students and the police as a result of a conflict between students from two neighbouring schools. The students hardly accept the police presence close to them although the police have been deployed to help them feel secure. As a result of the trouble, robbery of students' valuable objects is observed, sexual abuses are committed by the police, and some students are caught to be jailed.

In regard to text familiarity, the main text theme seems to be familiar since it depicts the common situation in most Congolese towns where the police is usually perceived as the potential enemy to the citizens. However, the text details describe a setting that is specific to one town, Kinshasa; and many readers may find it difficult to visualize this setting; therefore, this may render text comprehension difficult.

Concerning the text content familiarity, there are many words and phrases that can be classified as part of students' active vocabulary (For example: *drivers, taxi buses, trip, police, crazy students, sexual abuses, residential flats, phone cells, golden jewelry, girl-friend*, etc.). However, the text contains some words that can be problematic to readers who are not familiar with the text setting (For example: *launching tears' gaz from their guns; [...] along Mabanga road over the NATIONAL Commercial School's fence*). Nevertheless, many of these problematic words can be encoded through context (For example: *Men in dark blue uniform; [...] the protection of the school from the revenge of [...]; [...] to invade that school; Some demonstrators were caught to be jailed.*).

Furthermore, most the grammatical structures used in the text are those kinds of structures that are planned in the English curriculum. This is the case of the simple past tense that is used to report events that took place sometime in the past (For example: *That Wednesday morning a larger number of drivers **drove** their car...; Passengers **wondered** whether they could continue their trip; Worldwide radio and TV reporters **announced** that the situation ...; One student **died**; etc.*), the descriptive adjectives that are used to describe people or events (For example: *men in **dark blue** uniforms, flocks of **furious young** men; **harmful** objects, **crazy** students; etc.*); or the time adverbials that are used to indicate a specific time when the event occurred (For example: *That Wednesday morning...*).

However, the information presented in Table 8.1 on the preceding page suggests that only a slight minority of participants (6.3 %) perceived the content of this text to be familiar to them, while the majority of participants (79.7%) perceived it to be somewhat familiar. On the other hand, few participants (14.1 %) reported that the text was not familiar to them. These results seem to be in accordance with my perception that indicated that the text can be perceived familiar to some examinees, but because of the many specific details around the text setting, some examinees may find the story setting less familiar to them; resulting in some limited degree of familiarity with the text content. This finding suggests that, since narratives have a specific setting, the reader's familiarity with this setting likely impacts on his/her comprehension of the text. I will elaborate on this issue in the Discussion Chapter.

In relation to the content difficulty, the data presented in Table 8.1 (see page 251) indicate that the majority of participants found the text either difficult (56.3%) or very difficult (21.9%). On

the contrary, only few participants perceived the text content to be either very easy (3.1%), easy (9.4%) or of average difficulty (9.4%). This finding is not consistent with my evaluation of text content difficulty. According to my evaluation, since the text included many vocabulary words that examinees could encode, and that most grammar structures used in the text were those kinds of structures recommended in the English curriculum, examinees were expected to comprehend it. The reason may be that, participants' perceptions of text content difficulty might have been in accordance with how they perceived the text to be familiar to them; suggesting that the less familiar the text is, the more difficult its content will be.

### **8.1.3. T3 (see Appendix 7, Section 3)**

Like the preceding texts, the text included in T3 does not have a title or subtitles. The text narrates a car crash that happens at an intersection in Kinshasa during the busy traffic hour in the morning. A harsh discussion ensues between a man and a woman, the two people who are involved in the car crash. Each accuses the other of careless driving. One witness gives her account of the accident but she seems to sympathize with the other woman involved in the accident. Since no one is injured, people who are around calm everyone down and they continue their way.

In relation to text familiarity, although the content of this text is not similar to the content of texts that are planned in the English curriculum, it nevertheless depicts a social problem that is common in most Congolese big towns where car accidents are common. Thus, for students who are familiar with urban life, the text may be somewhat familiar to them. However, the story setting is specific to Kinshasa and the story details that contribute to the construction of the story line and text macrostructure relate to specific avenues, roads and crossroads. Therefore, some students who are not familiar with this specific setting may find difficult to visualize the story scene and construct their own text meaning; suggesting that this text may appear difficult for some readers.

The text vocabulary is appropriate to the town context (*loud crash, not signalling, turn left at the traffic lights, driving to work, gave her own account of the accident, walking along the*

*pavement, nobody was injured*) and vocabulary words and phrases can be encoded through context from reader's background. Grammar structures are those structures students are supposed to have practiced in classroom learning (*wh-* clauses [*When I arrived on the scene...*], tenses [*On April 18, I was standing on the corner of Mbala Avenue...*], reported speech [*Ms. Mujinga said she had signalled...*], etc.)

However, the eighty-two (82) participants who wrote this test had different views concerning text content familiarity and difficulty. The data in Table 8.1 (see page 251) suggest that no participant (0.0%) found the text content to be familiar. Nevertheless, the majority of participants (65.8%) perceived the text content to be somewhat familiar to them. On the contrary, one-thirds (34.2 %) of participants perceived the text content not to be familiar to them.

In regard to text content difficulty, only a minority of participants (6.1 %) perceived the text content to be easy while nearly half of participants (46.3%) perceived the text content to be of average difficulty. On the contrary, almost the other half of participants perceived the text content to be either difficult (41.5%) or very difficult (6.1%).

These statistics seem to corroborate my evaluation of when I suggested that, although the text content depicts a social reality common in Congolese towns, the content of this text is not the kind of texts that are planned in the English curriculum; and that the text may be perceived to be difficult by students who cannot visualise the specific spatial details necessary to construct text macrostructure. This finding suggests that narratives with a specific spatial setting might be difficult for the majority of examinees who are not familiar with that setting. This finding is valuable as it helps interrogate the validity of reading tests that include texts that depict town setting as there may be a serious problem for students from the rural areas to make a mental model of such texts as these students may not be able to visualize scene descriptions and details. Therefore, such reading tests may be biased as they may advantage candidates living or who have lived in specific town (Kinshasa in this case) over those who have never been at this town; or again, those living in rural areas. I will provide extensive argument on this issue in the Discussion Chapter.

#### 8.1.4. T4 (see Appendix 7, Section 4)

The text passage in this test paper does not have a title or subtitles. The text content relates to urban planning, the need to have organized plans and well-conceived designs at all stages of city's birth and development. The text also explains how great cities were originally planned to make them beautiful by ensuring that the traffic moves freely, and that communications and commercial centers operate effectively. The author gives examples of how engineers took much care to design some large cities such as Alexandria (Egypt), or Washington, DC (USA).

Regarding text familiarity, this text content is not the kind of texts that are planned in the English curriculum for learners. Besides, it depicts a reality that is not relevant to the Congolese setting where the design and construction of all actual big cities are traced back to the colonial period situated between 1885 and 1960, a period not familiar to all students. Furthermore, examples of big cities design and planning (Alexandria, Washington, DC) that are given in the text are not familiar to students. These three reasons can make the content of this text to be not familiar to the majority of examinees.

Since the text content can appear less familiar to the majority of students, many students can find it difficult to encode some words and phrases as they may fail to activate relevant schemata necessary to use clues provided by the context to understand these words and phrases. Words, names and phrases like “*urban planning, well-conceived designs, archaeologists, Mesopotamia, Alexander the Great, Potomac River, Pierre Charles l’Enfant, harbour, open-air ceremonial space, domed Capitol building*” may be problematic to many examinees to encode since they lack appropriate background on the text content; therefore, the examinees may find it difficult to construct an appropriate text macrostructure and this may result in poor performance on the test.

The statistics in Table 8.1 (see page 251) suggest that none of the participants (0.0%) perceived the text content to be familiar to them. On the contrary, most participants (80.7%) perceived the text content to be not familiar to them. Furthermore, only nearly one in five (19.3%) participants perceived the text content to be somewhat familiar to them. These statistics seem to support my evaluation of text content familiarity.

Concerning the text content difficulty, the data in Table 8.1 (see page 252) suggest that, all the participants perceived the text content to be either difficult (50.9%) or very difficult (49.1%); suggesting that no participants (0.0%) perceived the text to be either easy or of average difficulty. These statistics also confirm my evaluation of text content difficulty when I have argued that the text includes many words and phrases that readers may find difficult to encode since these words are not appropriate to participants' background knowledge. From this finding, it appears that content familiarity may interact with content difficulty as suggested previously.

#### **8.1.5. T5 (see Appendix 7, Section 5)**

This test was designed for students from the Commercial and Administration subject area. That is why this test paper, unlike the preceding test papers, includes a text passage and a letter. Both the text passage and the letter do not have titles or subtitles. The text narrates a story of a young man who goes to the bank in order to open an account. He is excited when he sees everything in the bank. He asks to see the bank manager and when he meets him, the latter refers him to the clerks who assist him to open an account and deposit some money. The letter, on the other hand, is written by the manager of a company to apologize to a client about delay in payment due to cash-flow.

Regarding content familiarity, both the text and letter contents address themes that are related to the students' subject knowledge and that are recommended by the English curriculum for these types of students. Therefore, any average student from this subject area can find the text and the letter easy to read. Moreover, both the text and the letter contain many technical words and phrases that can be easily encoded either through students' familiarity with these words or their ability to use clues provided by the context (*booking-offices, clerks, open an account, safe place to keep the money, consult the manager, accountant, fairly large amount, deposit, slip, to draw a cheque, to draw out money, cheque book, Ekunde Sales Services (ESS), we drew your attention to ..., customers, cash-flow problem, payment, outstanding accounts, to delay, settle the debts very soon, liquidation, company, Yours sincerely*). Furthermore, most grammatical structures included in both the text and the letter are those kinds of structures that are planned in the English curriculum; therefore, examinees may easily parse text sentences. This is the case of the

use of simple past tense to narrate events that took place at a certain moment in the past (*One day I entered into a bank, I was happy to see nice booking-offices, I got excited, the clerks looked at me. I saw much money which attracted me. Everything enjoyed me in the bank. I attempted to open an account ...*); or the use of adjectives to provide a description of some text characters and specific settings (*The accountant was a tall, cool man; the manager was a grave, calm man; He led the way to a private room; etc.*) that also profile the kind of structures Grade 12 students are supposed to be familiar with.

The data contained in Table 8.1 (see page 251) provide evidence that half of the participants (50.5%) who wrote this test reported that both the text content and the letter were familiar to them. The other half of participants (49.5%) reported that the text and letter contents were somewhat familiar to them. This finding suggests that in general, participants found the reading materials relevant and appropriate to them. It is therefore consistent with my evaluation of the text content familiarity.

In relation to text and letter content difficulty, the data contained in Table 8.1 (see page 251) indicate that, despite their familiarity with text and letter contents, only a negligible minority of participants (4.7 %) perceived the text and the letter contents to be easy, while the majority of participants (64.5%) found the text and letter contents to be of average difficulty level. On the contrary, nearly one in three participants (30.8%) still perceived the text and letter contents to be either difficult (24.30%) or very difficult (6.5%). This finding does not support my evaluation of content difficulty of the reading materials as I previously assumed the content of this text to be easy because of students' familiarity with the theme as well as with their familiarity with the vocabulary words and grammatical structures included in both the text and the letter. This finding suggests that the content of reading materials can be familiar to readers; but still, readers can perceive these materials to be difficult. This really raises the issue of the contribution of reader's topical knowledge of the text, his/her knowledge of grammatical structures and that of vocabulary breadth in text comprehension. The finding provides evidence of a possible interaction among these three variables. Although I believe that text comprehension is a fluid and complex activity that can be hardly measured, this finding suggests a need to conduct predictive studies based on multiple correlation methodologies in order to determine the relative

contribution of each of these variables in text comprehension. Nevertheless, this issue overlaps the scope of the present study and therefore will not be discussed in this study.

#### **8.1.6. T6 (see Appendix 7, Section 6)**

This test was designed for students from Secretary- Computing subject area. The test paper contains a narrative that tells the story of the fox and a group of geese from a farm-house. The fox, which is very hungry, comes upon a group of geese and wants to eat them; but one clever goose suggests to him to do them a last favour by letting them say their prayers. As they are saying their prayers very loudly, the farmer quickly comes out with a gun and the fox quickly runs back to the wood.

Concerning text content familiarity, this text discusses a theme that is not part of kind of themes requested in the English curriculum for students from Secretary-Computing subject area. According to this curriculum, students should read different texts and letters that contain information on how to run an office, the different hierarchies in a company, the issues of company management, the use and importance of computers in business, etc. (see Programme National d'Anglais, 1988). Therefore, since this text content does not relate to any of these themes that reflect the kind of reading materials students are requested to read in classroom context, many examinees can find this text unfamiliar to them. Nevertheless, since this narrative tells a story that is part of students' cultural knowledge, it is not surprising that many students can still find its content familiar.

Regarding the text content difficulty, the text contains many words and expressions that can be difficult to the students since some of them require students' knowledge of these words and not their capacity to use clues to encode them. For example, words and expressions like *fox; geese; ducks; came upon; you've got the better of us this time; crackled/cracklings, etc.* might really be difficult to be encoded through context clues; hence any reader who does not know them might find it difficult to construct the text meaning. Nevertheless, the text content difficulty appears to be compensated for by the use of grammatical structures that are part of the English curriculum. This is the case of adjectives that are used to describe (*beautiful fat white gees; reasonable*

*request*); the use of both simple past, past perfect, and past continuous tenses to narrate past events (*Once upon a time, there **was** a fox. And the fox **was** very hungry. He **hadn't eaten** for a long time, and he **was** just **dying** for a good meal*), the use of direct speech (*"I am going to eat you!" said the fox; The goose said, "erm,...I'd like to say my prayers"*), pronouns (*He hadn't eaten for a long time;* ), articles (*a fox; the goose; some geese*), modals (*Can we please ask one last favour?/ The noise of the crackling could be heard*), etc.

The information presented in Table 8.1 (see page 251) indicates that, although the text content does not relate to the kind of texts suggested in the curriculum, no participant (0.0%) found the text content not familiar as I have suggested. On the contrary, the majority of participants (80.0%) found the text to be somewhat familiar to them. The data provides evidence that confirms that 1 participant in 5 (that is, 20.0% of participants) perceived the text content to be familiar to them. This finding suggests that the participants might have activated their cultural schemata and this might have resulted in their perceptions of their familiarity or some degree of familiarity with text content. Nevertheless, the main question is to know whether or not texts with such contents are appropriate for examining the reading construct with candidates from this subject area. I propose to answer this question in the Discussion Chapter.

Concerning text content difficulty, the data in Table 8.1 (see page 251) suggest that, the majority of participants (60.0%) perceived the text content to be difficult; a finding that conforms with my evaluation as I stated that the text contained many words and expressions that can appear difficult to students since some of these words and expressions demanded only the students' knowledge of these words and not their capacity to use clues to encode them. Nevertheless, this finding does not reflect the participants' perception of text content familiarity. This is to suggest that, although many participants perceived the text content to be familiar to them, they nevertheless perceived its content to be difficult to understand, confirming that a text may be familiar, but still be difficult to comprehend. This finding contributes to our understanding of text difficulty as a complex construct. I propose to elaborate on this finding in the Discussion Chapter.

### 8.1.7. T7 (see Appendix 7, Section 7)

This test was administered to 48 participants from the Agriculture subject area. The text included in this test paper does not have a title or subtitles. The theme developed in the text is about the agricultural potentialities of DR Congo due to its diversity of climate and soils. The author gives a description of various produces that can be produced in different parts of the country, as well as the main challenges that so far prevent the agricultural development of the country.

Concerning the text content familiarity, the text theme relates to the kinds of themes that are suggested in the English curriculum, such as soil composition and fertility, soil erosion, sowing, weeding and harvesting, etc. Therefore, participants can find the text content to be familiar to them.

Regarding the text content difficulty, the majority of words and phrases (for example, *agricultural capacity, arable land, farming, coffee, beans, potatoes, high plains, cool temperature, fertile soil, cultivation of tomatoes, vegetables, subsistence farming, crops, corn, cassava, freshwater species, timber industry, etc.*) are technical words that are relevant to the participants' subject knowledge and many of these words are also cognates and their encoding may be straightforward (For example: *agricultural capacity* [French: *capacite agricole*], *arable land* [French: *terre arable*], *coffee* [French: *cafeirs*], *potatoes* [French: *patates*], *cool temperature* [French: *temperature basse*], *fertile soil* [French: *sol fertile*], *cultivation of tomatoes* [French: *culture des tomates*]). Besides, most grammatical structures included in the text relate to the structures that are planned in the curriculum. This is the case, for instance, of the descriptive adjectives (for example: *agricultural capacity, high plains of the east, cool temperature, huge potential fish, etc.*), plurals (for example: *potatoes, tomatoes, freshwater species, etc.*), simple present tense to describe (*The Democratic Republic of Congo **has** vast unrealized agricultural capacity; subsistence farming **produces** food crops in tropical areas...*, etc.), and prepositions to locate place (for example [...] *palm oil and sugar cane are grown **in** plantations* ).

The information provided in Table 8.1 (see page 251) states that all the participants who wrote this test found the text content to be either familiar (56.3%) or somewhat familiar (43.7%), a finding that seems to be consistent with my evaluation. As to text content difficulty, the

information in the same table suggests that the majority of participants perceived the text content to be either easy (58.3%) or of average difficulty (27.1%). This finding also appears to correspond with my evaluation.

These two findings suggest that when the text relates to readers' subject knowledge and when it includes many words that are part of their technical vocabulary and/or are similar in form to the French equivalents, readers may perceive these kinds of texts to be familiar and they may perceive their contents to be relatively easy.

#### **8.1.8. T8 (see Appendix 7, Section 8)**

This test was administered to 28 participants from Dressmaking subject. The text passage included in the test paper is a description of a dressmaker's house and the different items in it. Like the other texts in the sample, this text does not have a title or subtitles.

As regards text content familiarity, the text theme relates to the kinds of themes suggested in the English curriculum for Dressmaking students: dressmaking, cutting, dressing, weaving, housing, etc. Therefore, since the text content relates to the participants' subject knowledge, they can find the text content to be familiar to them.

Concerning text content difficulty, the majority of words and phrases included in the text are technical words that are relevant to participants' subject knowledge (For example, *cutting-table, treadle sewing-machine, wardrobe, up-to-date dresses, skirts, blouses, clothes' patterns, curtains*). Examinees do not need to use context clues to encode these words; rather, they need only to use their knowledge of technical vocabulary. Besides, the grammatical structures included in the text are those structures that are recommended in the English curriculum. Since the text is a description of a dressmaker's room, prepositions/prepositional phrases of place are the most frequently used grammatical structure (For example: *Gbamogigi's house stands **near the lake**; **Next to the corner** a cupboard is full of ...; **At the entrance of her room** stands a cutting table; **At the corner**, a large wardrobe...; **Near the window**, the perfumed curtains..., etc.). Therefore, a good reader can comprehend the content of this text with relatively some ease.*

The information presented in Table 8.1 (see page 251) suggests that, all the participants perceived the text content to be either familiar (32.1%) or somewhat familiar (67.9%); a finding that appears to be consistent with my evaluation of text content familiarity. As to text content difficulty, the data in the same table confirm that, although only very few participants (7.2%) perceived the text content to be easy, the majority of participants perceived the text content to be of average difficulty (82.1%); suggesting that most participants could comprehend the text with some ease. This finding also seems to correspond with my evaluation of text content difficulty.

These two findings suggest that when a text relates to readers' subject knowledge, readers may perceive it to be familiar and they may perceive their content to be relatively easy. However, the use of specialized texts that relate to students' subject knowledge must be considered with much caution because the test questions based on such texts might tend to measure subject knowledge rather than reading comprehension; hence, yield construct-irrelevant variance. I propose to return to this issue in the Discussion Chapter.

The reader's familiarity with text content is not sufficient to understand many textual variables that impact on text comprehension. We also need to understand the text genre and the kind of organizational structures typical to each genre. I wish to address this issue in the following section.

## **8.2 Identifying different genres included in ESE texts and the extent to which the ESE tasks tap into these different genres**

In processing a text for comprehension, constructing text meaning is likely to be affected by the reader's knowledge of text genre. This is because text genre can function as textual schemata, and therefore limits the meaning-potential of a particular text (Alidib, 2004). In light of this statement, when readers are familiar with different text genres and they are able to recognize their specific features, this familiarity can provide an important frame of reference that helps them to comprehend the text. In the scope of this study, it is hypothesized that the validity of a reading test also depends on the extent to which the test tasks tap into the structure of the text. In light of this hypothesis, I explore, in this section, the different genres used in the eight text

included in the study sample and I evaluate the extent to which the 56 ESE test items included in the study sample request information that taps into the different genres.

Cohen and Upton (2007), and Manxia (2008) classify texts into three genres: narration, argumentation, and exposition; and each of these genres has at least one or more major structural organization, such as time sequence, description, classification, comparison/contrast, cause/effect, and problem/solution, with the information presented from more than one perspective or point of view. Nevertheless, the specific context of the ESE as reflected in the English curriculum (Programme National d'Anglais, 1988) requires the inclusion of one additional "genre" called Practical writing. This genre relates to students' capability to read and understand some practical simple materials such as letters, memos, notices, invitations, especially in technical subjects.

In order to identify the different genres used in the sampled texts, I first inspect all the texts and classify them in argumentation, narration, exposition, or practical writing. Then I conduct a qualitative content analysis of texts included in each identified genre; and in the report I include only common features of all texts included in each genre category. Furthermore, in order to evaluate the extent to which the 56 test questions request information relevant to the specific genres, I examine the relationship between the text genres used and skills being assessed by individual test items and the degree of cognitive demands required by these skills. In the following section, I present the taxonomy of text genre from the sampled texts.

The information presented in Table 8.2 on the next page suggests that the eight texts and one letter from the study sample are classified in three genres. These genres are narration (T1, T2, T3, T5, and T6), exposition (T4, T7, and T8) and practical writing (T5). Besides, each of these three genres is organized in one or different text structures. The five texts that use narrative genre are classified in terms of the following structural organizations: time sequence, description, cause/effect, and problem/solution. On the other hand, the three texts that use exposition genre are classified in terms of the following structural organizations: description, comparison, and cause/effect. Finally, the letter is organized around problem/solution and cause/effect structures.

Table 8.2 ESE texts classified in genres and related structures

Genre types		Test papers		Discourse modes	
1	Narration	1	T1	1	Time sequence
				2	Description
				3	Cause/effect
		2	T2	1	Description
				2	Cause/effect
		3	T3	1	Description
				2	Time sequence
		4	T5	1	Description
				2	Problem/solution
		5	T6	1	Description
2	Problem/solution				
3	Cause/effect				
2	Exposition	1	T4	1	Description
				2	Comparison
		2	T7	1	Description
				2	Cause/effect
		3	T8	1	Description
				2	Cause/effect
3	Argumentation		No text found		
4	Practical writing	1	T5	1	Problem/solution
				2	Cause/effect

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8

### 8.2.1 ESE based on narrative texts

In the five narrative texts included in the study sample (see these texts in T1, T2, T3, T5, and T6 in Appendix 7), the story starts with the opening sentence that sets the space (place) and the time. The spatial setting is either real or fictitious. The following extracts provide an illustration. Note that the information relating to the spatial setting is in italics:

**T1:** *At the rural market of Burungangu*, Yasekuru sells groundnuts, maize, rice cakes and soya beans. She appears happy and healthy. Her cheerful smile but hides a tragic story.

- T2:** That Wednesday morning a large number of drivers drove their car and taxi buses through that irritating black smoke *along MABANGA ROAD over the National Commercial School's fence.*
- T3:** On April 18<sup>th</sup> at 9.15 hrs, *I was standing on the corner of Mbala Avenue and Kitoko road.* I heard a loud crash and turning round I saw two cars stopped in the middle of the road.
- T5:** One day *I entered into a bank.* I was happy to see nice booking-offices. I got excited. The clerks looked at me. I saw much money which attracted me. Everything enjoyed me in the bank.

On the other hand, the time setting is also either real or fictitious. The following extracts can provide an illustration. Note that the information relating to the time setting is in italics:

- T2:** *That Wednesday morning* a large number of drivers drove their car and taxi buses through that irritating black smoke along MABANGA ROAD over the National Commercial School's fence.
- T3:** *On April 18<sup>th</sup> at 9.15 hrs.,* I was standing on the corner of Mbala Avenue and Kitoko road. I heard a loud crash and turning round I saw two cars stopped in the middle of the road.
- T5:** *One day* I entered into a bank. I was happy to see nice booking-offices. I got excited. The clerks looked at me. I saw much money which attracted me. Everything enjoyed me in the bank.
- T6:** *Once upon a time,* there was a fox. And the fox was very hungry. He hadn't eaten for a long time, and he was just dying for a good meal. So, he went out of his wood and walked towards a farm where he knew there were some geese, chickens and ducks.

The story characters are generally people (For example: *Yasekuru, Kirongozi, Asha* [T1]; *Apolosa, the taxi drivers, the students, the police* [T2]; *Mr Mbela, Mrs Mafuta, Ms. Mujinga* [T3]; *Mr Montgomery* [T5]); but sometimes they are animals that are personified (For example: *the fox, geese* [T6]).

Since these are narratives with a reporting feature, the writer is anonymous and generally is himself/herself the central character. The following two excerpts provide an example:

- T3:** *I heard a loud crash and turning round I saw two cars stopped in the middle of the road*

T5: One day *I* entered a bank. *I* was happy to see nice booking-offices. *I* got excited.

Sometimes, the writer speaks at the place of the characters by using the third person pronouns (he/she, his/her, or himself/herself); or he/she sometimes lets them speak for themselves using the first person I as in these three extracts:

T1: *Her* cheerful smile hides a tragic story

T1: “*I* have got HIV. *I* will not get married. *I* will not have any more children”

T6: “*I* am going to eat you”, said the fox.

In all the five narratives included in the study sample, episode 1 is the initiating event, or the problem. This problem is generally formulated in the first sentence or paragraph, and it sets up the story main content. The following excerpts provide an example:

T1: At the rural market of Burungangu, Yasekuru sells groundnuts, maize, rice cakes and soya beans. *She appears happy and healthy. But her cheerful smile hides a tragic story.*

T2: [...], *a large number of drivers drove their cars and taxi buses through that irritating black smoke [...]*

T3: I heard *a loud crash* and turning round I saw two cars stopped in the middle of the road;

T6: Once upon a time, there was a fox, and *the fox was hungry. He hadn't eaten for a long time, and he was just dying for a good meal.*

Once the problem is set in the first episode, episode 2 is the response to the problem. For example, in T2, the second sentence (*Some [drivers] stopped abruptly*) and the third sentence (*Passengers were puzzled and wondered whether they could continue their trip*) are a logical response to an abnormal situation the drivers face. This is also the case of T6 where the response from the fox's hunger is presented in the next sentence immediately after the problem is set (*So, he went out of his wood and walked towards a farm where he knew there were some geese, chickens and ducks*).

In episode 3, the narrator describes the event/incident proper. For instance, in T1, the narrator describes how one of his/her character who was apparently in good health suddenly started showing signs of HIV-AIDS infection. This excerpt from the text provides an illustration:

At the outset, Asha seemed to be in good health. Yet, she stopped gaining weight and contracted one infection to another. At the age of three, Asha died from AIDS.

In T2, the main incident is also described in these terms:

Meanwhile, facing the police, flocks of furious young men were feverishly throwing to their opponents any kind of harmful objects: pieces of stones, bricks, iron sticks... Whereas men in dark blue uniform were launching tears' gaz from their guns in order to disperse those crazy students.

The incident in T6 is also described by the narrator in these terms:

In a field near the farm the fox came upon a flock of beautiful fat white geese. They didn't hear him coming and when they were trapped, they couldn't get away, the fox was going to eat them. 'I am going to eat you!' said the fox. The geese were terrified and didn't know what to do. One of the geese who was a little cleverer than the others turned to the fox and said 'well, Mr Fox, you've got the better of us this time, you are going to eat us, we have no way of escape. Can we ask you one last favour?'[...] The goose said 'erm ... I'd like to say my prayers.'

In episode 4, the narrator may describe the consequence of the incident. In some narratives, the consequence can culminate in the denouement of the story. In T2, for example, the consequence of the confrontation between the students and the police results in damages, abuses, robbery and arrests. This excerpt from the text illustrates that:

A lorry bumped a tree and was badly damaged. Its driver mortally wounded. Apolosa, a well-known policeman, was seriously injured. One student died; sexual abuses were committed in ladies' residential flats. Phone cells, golden jewelry and money robbery were observed. Some demonstrators were caught to be jailed.

Also, in T6, after the fox had allowed the geese to say their prayers before he could eat them, the geese started praying so loudly that their cracklings could alert the farmer. As a consequence, *"The farmer of course heard the cracklings, knew something was wrong, took his gun, rushed out of the farmhouse, rushed down to the field and there he saw the fox."* And this consequence culminates in this denouement:

At the same time, the fox saw the farmer and of course had no time to eat any of the geese, turned round and ran away as fast as he could back to his wood, as hungry as he had left that morning.

Concerning the discourse modes used in the five narratives included in the study sample, the information presented in Table 8.2 (see page 264) indicates that a total of four discourse modes are used in order to communicate directly with the reader. These modes are: description, time sequence, cause-effect, and problem-solution.

In a narrative structure literature (Mandler & Johnson, 1977; Kintsch, Mandel, & Kozminsky, 1977; Dymock, 2007; Thorndyke, 1977), the descriptions of different events are made in such a way that the readers can anticipate the outcome of the story. In this perspective, readers are expected to deduce motives of different story characters and connect the different relations among events. In light of the preceding statement, the five narratives included in the sample make use of adjectives, adverbials, and prepositions to describe story events. However, in order to make the description vivid and maintain suspense so as to enable the reader visualize the event and/or story scene, the narrator uses adjectives either to describe characters, scenes, or emotions. The following two extracts can illustrate how the narrator uses adjectives to describe characters. Note that the descriptors are italicized.

T2: Apolosa, a *well-known* policeman; was seriously injured.

T5: The accountant was a *tall, cool* man. . . The manager was a *grave, calm* man.

On the other hand, the following three extracts illustrate how the narrator uses adjectives to visualize the story scene. I wish to mention that the information relating to scene description is in italics.

T1: At the *rural* market of Burungangu, Yasekuru sells groundnuts, maize, rice cakes and soya beans. She appears happy and healthy. Her cheerful smile but hides a tragic story.

T2: Yet, their security equipment could not protect them from the anger of their *close* neighbours.

T3: Mr. Mbela, 56, an engineer had crashed *into the side door* of the white Mazda belonging to Ms. Mujinga, 25, a headmistress.

Finally, the following two extracts illustrate how the narrator uses adjectives to describe characters' emotions:

T2: Meanwhile, facing the police, flocks of *furiously young men* were feverishly throwing to their opponents any kind of harmful objects: pieces of stones, bricks, iron sticks...

T6: “What a *terrible* noise they make saying their prayers!” The fox exclaimed.

Furthermore, in the five narratives, adverbials and place prepositions are sometimes used to describe the intensity of the incident, scene or emotion. Consider the following four extracts that illustrate this point:

T2: A lorry bumped a tree and was *badly* damaged. Apolosa [...] was *seriously* injured.

T1: Yasekuru now lives *in one room’s hut in the suburbs of Burungangu ...*

T3: I heard a loud crash and turning round I saw two cars stopped *in the middle of the road ...*

T6: When the geese were saying their prayers [...], they opened their mouths and cackled *as loudly as* they could [....].

Since narratives involve reporting a story, event or incident, time sequence is a discourse mode that is generally used. Time sequence is generally translated through the use of the past tenses (simple past tense, past perfect, past progressive) together with adverbials that introduce the first sentence of the narrative. This use of past tense is found in the opening sentence of all the five narrative and it goes on through the whole story. This use of past tense throughout the whole text is likely to enable readers figure out that they are reading a narrative text, and eventually they may deploy appropriate strategies necessary to construct the text story line that may enable them to comprehend the text. These three excerpts can illustrate this point:

T3: *On April 18<sup>th</sup> at 9.15 hrs, I was standing* on the corner of Mbala Avenue and Kitoko road. I *heard* a loud crash and turning round I *saw* two cars stopped in the middle of the road. I *did not see* the crash itself. When I *arrived* on the scene the two drivers *had got out* of their vehicles and *were arguing* noisily.

T5: *One day I entered* into a bank. I *was* happy to see nice booking-offices. I *got* excited. The clerks *looked* at me. I *saw* much money which *attracted* me. Everything *enjoyed* me in the bank.

T6: *Once upon a time, there was* a fox. And the fox *was* very hungry. He *hadn’t eaten* for a long time, and he *was* just *dying* for a good meal. So, he *went out* of his wood and *walked* towards a farm where he *knew* there were some geese, chickens and ducks.

In narrative structure literature (Mandler & Johnson, 1977; Kintsch, Mandel, & Kozminsky, 1977; Dymock, 2007; Thorndyke, 1977), time sequence is also expressed through the use of transitional words and expressions such as first/firstly, second/secondly, then, next, later, soon,

finally, earlier, afterwards, meanwhile, eventually, etc. However, there is evidence that the five narratives included in the study sample do not use transitional words and expressions to indicate the chronology of events. In all these narratives, time sequence is simply expressed through juxtaposition of clauses, sentences and paragraphs and transitional words and expressions are almost not used. More specifically, narratives T1 and T2 do not have any time sequence transitional word or expression. On the other hand, narratives T3, T5 and T6 include only one time sequence transition each. More specifically, T3 includes the transitional phrase “*After several minutes*” to introduce the last paragraph which is also the denouement of the story; the narrative T5 includes one transitional word “*Then*” that starts the concluding sentence of the narrative; while narrative T6 includes one transitional phrase “*At the same time*” that also starts the last sentence of the narrative.

Among the main features of narratives, Alidib (2004) argues that most narratives involve a problem or goal that is faced by the main character; and it is this problem or goal that guides the main episodes of the story. Following this argument, problem-solution appears to be a mode narrators sometimes use to tell the story. Related to the problem-solution mode is the cause-effect mode. This suggests that in some narratives, a problem is sometimes articulated in terms of its cause(s) and the effects that result from this problem. In the five narratives from the study sample, there is evidence of the use of these two discourse modes.

For example, the story in narrative T5 is about a young man who has a problem of opening a bank account in order to keep his money. But since he does not know what to do, he goes to the bank and instead of telling his problem to the clerks so that they can help him, he asks for meeting the bank manager. When he meets the bank manager, the latter is surprised and finally finds as a solution to refer the young man to the clerk so that they can assist him open an account. At this juncture, I wish to mention that, although the text in T5 does not include any rhetorical device to show the problem-solution relationship, the following excerpt illustrates how the main character, who is also the narrator, expresses his problem and how he finds solution to it. The coordinating conjunction “so” that is used indicates problem-solution mode:

*So* I went in [the bank] and looked timidly round at the clerks. I had an idea that a person who wants to open an account must consult with the manager. [...] I asked him [the accountant] to see the manager alone. [...] The manager felt that I had terrible secret to

reveal. He led me the way to private room. I told him that I have come to open an account and I intend to keep all my money in this bank. [...] The manager got up and opened the door. He called on the accountant, Mr. Montgomery, and he said that this gentleman is opening an account; [...] I rose and went to the accountant's window and pushed the money at him.

Likely, the story in narrative T6 is about the fox, the central character who has a problem of hunger and who must find solution to this problem by finding something to eat. As a solution, he decides to go to a farmhouse where he expects to find some preys like geese, chickens and ducks. The narrator explicitly uses the coordinating conjunction “so” as a rhetorical device to indicate problem-solution mode. The following excerpt from T6 substantiates this claim:

Once upon a time there was a fox. And the fox was very hungry. He hadn't eaten for a long time, and he was just dying for a good meal. *So*, he went out of his wood and walked towards a farm where he knew there were some geese, chickens and ducks. [...] when they were trapped, they couldn't get away, the fox was going to eat them. 'I am going to eat you!' said the fox.

As to cause-effect relationship, the narrative T1 illustrates the main cause of women's infection of HIV and the consequences that result from this infection. The following excerpt can act as evidence:

Women can be infected through normal sexual intercourse and are at risk of contracting HIV. In one hand, they mustn't discuss sexuality. In the other, African men commonly abuse many sexual partners. This is why women risk sex abuse if they refuse. To exclude women from AIDS matters does not favour HIV prevention.

Having conducted a qualitative content analysis of the five narratives included in the study sample, I now wish to evaluate the extent to which the test questions relating to these narratives request information relevant to narrative. I propose to discuss this issue in the following section.

- **Evaluation of the validity of the ESE tests based on the five narrative texts**

In Chapter One, Section 1 relating to the validity theory, I stated that validity is about the relevance of inferences and decisions made on the basis of test scores (Messick, 1989; Kane, 2001, 2002, 2004). In this section, I evaluate the extent to which the five ESE tests using based

on narrative texts include test items that target information that is appropriate to narrative structure.

From the content analysis I have conducted on the five sampled narratives, I hasten to state that the five narratives generally follow the story line of most narratives as described by Enright *et al.* (2000) and Alidib (2004). Viewed from the structure perspective, the five narratives can be seen as an effective sample of narratives that can be used to assess examinees' skill to construct the narrative story line. However, since validity is about the appropriateness and relevance of inferences and decisions made on the basis of test scores, the main issue is not whether or not the narratives included in a test session are well structured, but it is about the extent to which test questions based on these narratives reflect the narrative structure. In order to answer this question, I have hypothesized that, in order to be construct-valid, examinees' scores on a reading test based on narrative texts must provide an indication of their abilities to read the narrative and reconstruct its story line. Consistent with this hypothesis, the test questions must target information that requires examinees' skills to reconstruct the narrative story line with a focus on the chronological sequencing as well as the spatial organization of narrative events. More specifically, the test questions must require examinees to identify/describe the narrative setting, to identify the main character and auxiliary characters, to analyze the overall plot through the identification of the story problem, the response, the action and the outcome or resolution. In light of this scope, a valid reading test based on narrative is expected to include test items that generally focus on the identification and description of time and place descriptors, the description and comparison of characters' appearances, personality and emotions, their motives and social relationships, the identification and location of the story problem, the identification and location of the response to the story problem, the descriptions and judgment of characters' actions to solve the problem, and the identification and personal judgment of the outcome of the story.

In light of the preceding details, I now turn to evaluate the validity of the ESE based on narrative texts. However, since it is neither feasible nor necessary to examine the test questions included in all the five narratives, I have purposely selected narrative T2 (see Appendix 7, Section 2) in order to examine the extent to which the 8 test questions not only tap into information relevant to this narrative, but also are ordered in such a way that this ordering reflects the story line of the

text. Although the findings expected from this analysis cannot be generalized to the other four narratives, I believe that they nevertheless provide some information on the extent to which ESE tests based on narratives are sensitive to narrative structure and information ordering.

The main content and storyline of narrative T2 have been presented in the previous section. At this stage, I only recall that this text is a narrative that uses description and cause-effect as narrative modes. Therefore, two important things to consider are the narrative story line and its picture. In order to achieve a good understanding of this text, examinees must read for details through scanning and understanding explicitly stated information. Therefore, I expect comprehension questions to be mainly based on these two skills. Concretely, scanning may call for items that focus on the narrative setting (where the story takes place, when it takes place, and who are the main characters involved in the story). These might be operationalized in question **1, 2** and **3**. Likely, scanning may also call for items that relate to each story episode.

**Question 4** may logically focus on the initiating event of the story. The test writer can either focus on adjectives that are used to amplify the incident (*a large number of drivers, that irritating black smoke*), or he/she can focus on the identification of the narrative problem/conflict. By focusing on adjectives that amplify the incident, the test writer can request examinees to indicate the option that shows that there is something unusual that both the passengers and the drivers are experiencing. Therefore, the test question can request examinees' capacity to identify the adjective(s) they perceive describe the unusual situation. If the test developer decides to focus on the identification of the problem, the question can request examinees to select the alternative that best states the problem in the first paragraph. Such a question will require examinees' capacity to understand the main points and supporting details that are necessary for understanding the main idea of the first paragraph.

**Question 5** can logically relate to the second story episode, the response to the problem. The test writer can request examinees to identify, through reading for explicitly stated information, the emotional reaction of the "*drivers*" and "*passengers*". Useful hints to this question can be these two phrases: "*Some [drivers] stopped abruptly*" and "*Passengers were puzzled and wondered whether they could continue their trip*".

**Question 6** can relate to the incident itself as described in the third narrative episode. The test writer can choose between focusing on the identification of the two groups involved in the conflict, and focussing on the identification of different objects used by each group to challenge each other. If the test writer decides to focus on the identification of the two opponent groups, he/she can focus on the two metaphors used in the narrative: “*flocks of furious young men*” and “*men in dark blue uniform*”; therefore, the question can request the examinees to use context clues and their background knowledge to find which words these two metaphor actually refer to (in this case they refer to “*students*” and “*police*”). On the contrary, if the test constructor decides to focus on the identification of different objects used by each group to challenge each other, he/she can request examinees to select from the alternatives the one that include/does not include objects used by the two groups to challenge each other. Such question will request examinees’ capacity to scan a specific part of the text in order to locate precise pieces of information.

**Question 7** can relate to episode 4 and the test writer has a choice between asking examinees to identify, through scanning, the rhetorical devices that are used to amplify the incident and asking them to identify through scanning the various consequences of the incident. In case the test writer chooses the first option, the test question can focus on such devices, and it will require examinees’ capability to recognize and encode words needed for comprehending the text. The following rhetorical devices in bold in this excerpt can serve the question focus:

*Worldwide* radio and TV reporters announced that the situation was getting worse [...] A lorry bumped a tree and was **badly** damaged. Its driver was **mortally** wounded. Apolosa, a well-known policeman was **seriously** injured. One student died; **sexual abuses** were committed ...

On the contrary, if the test writer decides to ask examinees to identify through scanning the various consequences of the incident. The same paragraph can provide hints for examining test takers’ capacity to understand details necessary to identify the different consequences of the incident.

Finally, **question 8** can relate to the last paragraph where the narrator uses the flashback technique to describe the cause of the incident. The test writer can ask examinees to identify, through search reading, the main cause of the conflict between the students and the police. The opening two sentences of the last paragraph (*The incident broke up because students hardly*

*stand the police presence close to them. Among other reasons, the [...] students' leader was deadly beaten in [...])* can provide hint for this test question. Or again, the test writer can ask examinees to select the alternative that states the reason why the police was sent to the school. In this latter case, the following sentence from the last paragraph (*The Chief District of the police then decided the protection of the school from the revenge of the Congo Technical Institute's students*) can serve as hint to item construction.

However, since validity is about the relevance and appropriateness of inferences and decisions made on the basis of test score, the main issue at this stage is to know the extent to which the 8 questions based on this text not only relate to those questions I have suggested, but also follow the linear and incremental order that I have established from question 1 to question 8. I hypothesized that, in order to ensure the valid ESE test based on a narrative test, the test must include tasks that request information that tap into narrative structure and these tasks must be ordered in order to reflect the order in which the information is presented in the narrative. In order to investigate this issue, I have examined the information requested by the eight test questions as well as the ordering of these test items included in this test paper.

**Question 1** (T2.1) requires examinees to establish global comprehension of the text in order to identify the text title. Insights from strategies use (see Appendix 17) suggests that, since the text title is not provided, examinees must read the whole text both carefully and expeditiously. While reading the text carefully, examinees struggle to establish accurate comprehension of explicitly stated main ideas and supporting details throughout the entire text. On the other hand, while reading the text expeditiously, examinees skim the whole text in order to fix their comprehension and decide what appears to be the best title that suits the text. As my evaluation of this test question, it appears that, since this question requires careful reading and expeditious reading at global level, its scope overlaps the narrative structure where the story macrostructure is essentially built around story details. Also, since this test question does not target the narrative setting that is the first element of narrative structure, it appears to be inappropriate; therefore, this test questions does not need to be included in the test.

**Question 2** (T2.2) relates to the consequences of the incident as described by the narrator in the second paragraph. Insights from strategies use (see Appendix 17) suggest that this question requires examinees to search read in order to quickly locate explicit pieces of information

scattered through the whole paragraph. The information needed to answer this question relates to the fourth episode of the narrative, which, according to the previously suggested story element, needs to be addressed in the seventh position (Question 7). Nevertheless, since the question requires examinees to search for details within a single paragraph, it appears to be a good item, but it needs to be shifted to the seventh position.

**Question 3** (T2.3) relates to the last paragraph that gives the narrative flashback relating to the causes of the incident. If one takes into account the story line in the text, this question should be in the last position (question 8). Insights from strategies use (see Appendix 17) indicate that this question required the examinees to read at paragraph level in order to establish accurate comprehension of explicitly stated pieces of information and related supporting details across all the sentences included in the paragraph. As my evaluation of this test item, I state that this item taps into information related to narrative structure, but it needs to be shifted to the eighth position in order to be more appropriate.

**Question 4** (T2.4) relates to the understanding of vocabulary (objects) that are used by the characters in the incident; as indicated in episode 3 of the story line (see paragraph 2). This question corresponds to the sixth question in my suggestion of ordering. As for cognitive operations necessary for answering it, this question requests examinees' capacity to search read the first text paragraph in order to identify the specific words necessary for answering the question (see strategies use in Appendix 17). As my evaluation, this item is a good item as it targets information relevant to narrative structure; but it needs to be relocated to the sixth position.

**Question 5** (T2.5) relates to the last paragraph as the test developer asks examinees to recognize an antecedent to a pronoun by identifying the relationship between words in different sentences. This is a relevant operation that aims to assess examinees' linguistic ability, more specifically their ability to parse sentences. However, I believe that the order of this test item is problematic as it needs to be shifted to a later position.

**Question 6** (T2.6) requires the examinees to process the text both carefully and expeditiously at global level in order to identify pieces of information scattered throughout the whole text. Since this question requires searching for information not located in a specific part of the text, but in

the entire text, the cognitive demands needed to answer this question may be very high; thus making the item to be difficult for examinees. Furthermore, from the narrative structure viewpoint, this question is not appropriate as it targets information that overlaps the scope of narrative structure whose main aim is to reconstruct the story line through searching for specific information. If we can agree that a narrative aims to present the story line of events, and that it is built around text details, we can admit that test questions that require the examinees to read the whole text overlap the scope of appropriate test questions. By appropriate test questions I mean those questions that consider the extent to which the cognitive demands needed to complete the test tasks are consistent with the structural organization of the text.

**Question 7** (T2.7) calls for examinees' metacognition skill of prediction as they must make inferences on the basis of their global comprehension of the text. Besides, this question requests examinees' knowledge of grammar in order to identify the correctly parsed sentence in the suggested five alternatives. Although such cognitive demands might be critical to examinees, this question has a communicative function that relates to the kind of materials requested in teaching and learning situations. Therefore, this item appears to be an appropriate item as examinees may find the item question familiar to them and may get it right.

**Question 8** (T2.8) points to the narrative purpose: to narrate with the objective of giving advice. Insight from strategies use (see Appendix 17) suggest that, in order to answer this question, the participants had to process the whole text both carefully and expeditiously and make inferences on the basis of textual information. Therefore, although the cognitive demands are high (both careful reading and expeditious reading at global level), this question appears to be a good question as it requests information that targets the narrative structure.

- **Summary of the evaluation of validity of tests based on narrative texts**

In order to sum up my evaluation of narrative texts, I wish to remind my readership that the main issue was twofold: firstly, to evaluate the extent to which the test questions based on the five narratives targeted information that taps into the narrative structure; and secondly, to evaluate the extent to which the information targeted by test question is appropriate to the cognitive demands relevant to narratives.

In order to provide the final answer to the first issue, the analysis and discussion conducted so far on the 8 test questions reveal that, although some test questions targeted information that tap into the narrative structure, the majority of test questions were not ordered in such a way as to reflect the narrative story line.

In order to provide the final answer to the second issue, the analysis and discussion conducted have revealed that some test questions did not call for operations that are relevant to narratives. This is to suggest that, since it is assumed that narratives are mainly based on details and descriptions, appropriate questions are those that require examinees to read at local level (both carefully and expeditiously) in order to locate needed pieces of information. Yet, evidence from the analysis of the eight questions has suggested that the majority of test questions required examinees to process the text at global level. In light of this finding, I can conclude that the skills required by the majority of test questions in order to complete test tasks appear to be more cognitively demanding than expected; therefore, the test tasks appear to be inappropriate to the test expectations.

Although I will elaborate on these two findings in the Discussion Chapter, I hasten to state at this juncture, that there is evidence to suggest that ESE based on narrative texts appears to be less construct-valid; therefore, the ESE constructors need to design tests that include test questions that take account of the narrative structure and the underlined cognitive demands appropriate to narratives. Having analysed the ESE tests based on narrative texts, I now turn to analyse the ESE tests based on expository texts. This is the concern of the following section.

### **8.2.2 ESE based on expository texts**

Expository texts are those that primarily serve to inform the readers. In my previous study (Katalayi, 2011) I mentioned that expository texts are the most often used type in the teaching textbooks in DR Congo although they are not frequently used in the ESE. This less frequent use of expository texts in testing session is reflected in Table 8.2 (see page 265) where it is indicated that, of the nine texts included in the eight test papers, only three texts (T4, T7 and T8) were expository texts. The data in the same table indicate that three modes of exposition were used by the writers: description, comparison and cause-effect modes. Description was the most used

mode of writing while comparison and cause-effect were two modes that supported the writer's effort to describe the information.

In addressing the issue of the construct validity of a reading test based on expository text, the main assumption is that, since expository texts primarily aim to present information, they are expected to be structured around main ideas and supporting details; suggesting that readers must process the text at both global and local levels. Also, because expository texts mainly use description as mode of writing, the use of descriptors and present tenses characterise most expository texts.

### 8.2.2.1 Structure of the three expository texts included in the study sample

In all the three expository texts included in the study sample, the use of present tenses (simple present, present perfect, present progressive) is well indicated straight in the opening sentence of the text as illustrated in the following extracts:

- T4:** Look at urban centres. They *are* great cities of the world. We often *do not think* about how they were originally planned. [...] However, urban planning *requires* a city to be beautiful, the traffic to move freely...
- T7:** The Democratic Republic of Congo (DRC) *has* a vast unrealized agricultural capacity. The majority of its lands *is* arable and suitable for farming.
- T8:** In Boyasegeze City, Joyce Gbamogigi's house *stands* near the lake. It *has* seven rooms, among them a living room, [...]. Although her room *is* not very large, it *is* really comfortable.

In order to make description vivid, the author uses descriptive adjectives as well as prepositions of place. The use of adjectives to make description vivid is illustrated in the following extract from T7:

With *huge* potential fish could be a *valuable* source of food, but it is not exploited. The lakes in eastern and southern regions are a *massive* reserve of a variety of *freshwater* species such as tilapia. The river Congo is another *important* source with *major fishing* ports in Kisangani and Bandaka.

Consider this second paragraph of T8 and see how prepositional phrases (in bold and italics) serve to make the description vivid:

*At the entrance* of her room stands a cutting-table. *Close to it* she puts a treadle sewing-machine which provides a convenient place to work. *At the corner*, a large wardrobe is full of up-to-date dresses, skirts, blouses and many clothes' patterns. *Through the window* the rising sun fills her room every morning until sunset. *Near the window* the perfumed curtains emit the smell of roses whose scent makes her relax. The soft and warm linens make her sleep peacefully.

### 8.2.2.2 Evaluation of the validity of tests based on expository texts

If we can agree that expository texts primarily aim to inform the readers, and that description is the basic mode of exposition, in order to be construct-valid, we expect reading tests based on expository texts to mainly include those questions that request the examinees' ability to process the text at both global and local level in order to get the gist of the information presented in the text as well as to locate specific pieces of information located at global and local levels. Viewed from this text processing perspective, tests questions are expected to require examinees' capacity to demonstrate how well they have been able to visualize text details through the identification and understanding of descriptors and place prepositions. At this juncture, I hasten to inspect the test questions included in T4, T7 and T8 in order to evaluate the extent to which these questions are appropriate for examinees.

Questions T4.1 (see Appendix 7, Section 4), T7.2 (see Appendix 7, Section 7) and T8.1 (see Appendix 7, Section 8) required the examinees to find the title that best suited the text/paragraph. Insights from participants' strategies use (see Appendix 17) indicate that these three questions requested examinees' capacity to carefully read the text at global level in order to construct its macrostructure. Hence, since text gist construction may provide a ground for the identification of text details located at global and local levels, I can qualify these three test questions as appropriate questions. However, since these three questions required higher level text processing, they may appear challenging to the majority of examinees as results from the analysis participants' individual characteristics have revealed that the majority of examinees' skills in English were very low (see Chapter 6). Therefore, these three test questions may appear less appropriate for examinees in light of their individual characteristics.

In order to validate this evaluation, the item difficulty index for each of these three test questions was computed. The data presented in Appendix 17 suggest that the item difficulty index for the three questions was .1 for T4.1, .1 for T7.2 and .5 for T8.1. This is to suggest that, as I have predicted, participants found these three test questions very difficult (T4.1 and T7.2) and of average difficulty (T8.1). This finding also supports the conclusion stated in chapter 7 when it was found that test questions that required high level text processing were more difficult than those test questions that required lower level processing. This finding has an implication in the scope of this study that I wish to formulate in form of a question: “Considering examinees’ characteristics, which types of questions need to be included in the ESE based on expository texts: global questions or local questions, or both?” Although I will provide the answer to this question in the Discussion Chapter, I hasten to state that, since reading construct cannot be examined outside reading context, a valid reading test must not only take account of the extent to which test tasks are appropriate to the kind of cognitive demands necessary for constructing text macrostructure, but it must also take account of the degree of complexity of test tasks in relation to the contextual parameters such as examinees’ reading skills, motivation for and attitudes towards reading activity.

Questions T8.4 (see Appendix 7, Section 8), T8.5 (see Appendix 7, Section 8), T4.4 (see Appendix 7, Section 4) and T7.5 (see Appendix 7, Section 7) required examinees to carefully read the text at local level in order to identify or understand particular words (T8.4 and T8.5) or understand syntax (T4.4 and T7.5). Since these four questions targeted examinees’ capacity to encode words and parse sentences, they required lower level text processing; therefore, they are appropriate questions.

Questions T4.2 (see Appendix 7, Section 4), T4.3 (see Appendix 7, Section 4) and T7.1 (see Appendix 7, Section 7) required the examinees to read expeditiously the text by skimming it in order to identify the paragraph where a specific piece of information was located (T4.2) or to search read in order to quickly locate pieces of information scattered through the whole text (T4.3 and T7.1). Since these questions targeted examinees’ capacity to process the text expeditiously at global level, they are good items although they required higher level processing that may be challenging to some examinees.

### **8.2.2.3 Summary of the evaluation of validity of tests based on expository texts**

In order to sum up the evaluation of expository texts, I wish to recall that expository texts primarily aim to inform the readers, and that description is the basic mode of exposition. Therefore, viewing their informative nature, reading tests based on expository texts are expected to include mainly those questions that call for the examinees' ability to process the text at both global and local level in order to get the gist of the information presented in the text as well as to locate specific pieces of information located at both global and local levels. Moreover, since description is the basic mode of exposition, test questions are also expected to require examinees' capacity to demonstrate how well they have been able to visualize text details through the identification and understanding of descriptors and place prepositions.

In light of the preceding scope, the analysis and discussion conducted on the test questions from the three tests substantiate that these test questions actually targeted examinees' ability to process the text at both global and local level in order to get the gist of the information presented in the text as well as to locate specific pieces of information located at both global and local levels. However, there were no questions that targeted examinees' capacity to identify and understand descriptors and place prepositions necessary to visualize text details.

This finding suggests that ESE constructors seem to be aware of the need to base ESE test based on expository texts around gist construction through reading at both global and local levels. However, they seem to be unaware of the need to include test tasks that feature the description of text details. This suggestion prompts the issue of under-representation variance I have discussed previously (see Chapter 2). I will elaborate on this issue in the Discussion Chapter.

### **8.2.3 ESE based on practical writing texts**

In the context of the present study, practical writing consists of reading materials that are used in technical subjects and that include letters, memos, notices, and invitations. Since some of these materials are used to test students' reading abilities (for example letters), they are considered as a distinct genre.

The information I have presented in Table 8.2 (see page 265) indicates that, of the nine texts included in the eight test papers, only one (T5) is a letter. The same source indicates that two modes of letter writing are used by the writers: problem-solution and cause-effect modes.

At this juncture, I wish to remind my readership that, letters have conventional presentations and they include different parts. They also have different contents. They are written to fulfil different functions such as to apply for a position, to order goods, to apologize for delivery delay, etc. Besides, there are conventional expressions and phrases that are used to introduce different letter parts (salutation, complimentary clause, etc.) or express different emotions and moods (apology, regrets, etc.).

In light of the preceding description of letter form and content, the validity of a multiple-choice reading test based on a letter can be established from a letter form and content perspective. This is to suggest that, a reading test based on a letter is construct-valid if the test includes questions that request examinees' ability to identify and understand the letter's parts and contents as well as their ability to identify, understand and write in different ways the different conventional expressions included in the letter.

In the letter that is used in T5, the different parts of the letter are indicated (sender's address, inside address, date, salutation, body of the letter, conclusion, complimentary close, and signature). Formally, this is a well-written letter that reflects the kind of letters used by students in learning situations and that can also be used in normal life situations.

### **8.2.3.1 Evaluation of the validity of the test that is based on letter writing**

In this subsection I propose to examine the validity of the three questions T5.5, T5.6 and T5.7 (see Appendix 7, Section 5) that were based on the letter. **Question T5.5** required examinees' ability to identify the main idea of the third paragraph of the letter. This question targeted the content of the third paragraph that is about the expression of the firm's regret. In order to answer this question, evidence from strategies use (see Appendix 17) indicates that the participants had to read the whole letter carefully in order to construct its gist before reading the target paragraph both carefully and expeditiously in order to infer the paragraph content from main ideas and

supporting details. Since this item targets the construction of the letter gist, it is an appropriate item that taps into the target construct. However, since this test item requires high level processing that may be challenging to the majority of examinees, its validity must be taken with much caution on the ground of its appropriateness to examinees' individual characteristics.

In order to gain a better understanding on the preceding evaluation, the item difficulty index of this test question was computed. The main hypothesis was that, in the context of the DR Congo ESE, test questions that request higher level processing are cognitively demanding; therefore the majority of examinees most likely get them wrong. The data presented in Appendix 17 indicate that the item difficulty index of T5.5 was .4, indicating that the test question was difficult for examinees. This finding, like some other previous findings, suggests that a test item can be valid when the information required to answer it taps into the target construct, but still be inappropriate to the specific context of the test. As pointed out earlier, this finding suggests that reading construct need not be examined outside reading construct. I propose to elaborate on this finding in the Discussion Chapter.

**Question T5.6** required examinees' knowledge of letter writing parts in order to identify the item option that included the proper signature in reply to the letter. This is an appropriate test question, and this evaluation is corroborated by insights from participants strategies use (see Appendix 17) as well as the item difficulty index (.5).

Finally, **question T5.7** targeted examinees' skill to identify the part of the letter that does not appear in the letter. In order to answer this question, examinees used their knowledge of letter writing parts (see Appendix 17 for strategies used). This is also a good question; its item difficulty is .7 (see Appendix 17), suggesting that participants found this question to be easy.

### **8.2.3.2 Summary of the evaluation of validity of the test based on letter writing**

In light of the results from the preceding analysis, the questions based on letter writing generally appear to be appropriate as they targeted examinees' skill to read the letter and understand its parts, form and content. Nevertheless, the few number of test questions suggest that the test may

not cover the entire range of letter writing skills. This prompts the issue of construct-underrepresentation. I intend to elaborate on this issue in the Discussion Chapter.

The evaluation of the content of ESE texts is necessary for evaluating the quality of the ESE. Likely, the evaluation of the extent to which the ESE include test questions that are appropriate to the different genres is crucial in validating the ESE. However, the evaluation of validity of ESE texts will not be sufficient if we do not investigate their density. I address this issue in the following section.

### **8.3 Examining the density of ESE texts and their appropriateness to the ESE context**

One of the theories that inform this study is the Construction-Integration theory. This theory describes text comprehension as a process where a reader first constructs a textbase from the text propositions and then integrates this textbase into his/her background knowledge to form a coherent mental representation of the text (Kintsch, 1988, 1998, 2004; van Dijk & Kintsch, 1983).

In light of this description, the difficulty in processing a text is also influenced by its propositional density. This is the ratio of the number of propositions to the total length of the passage. Therefore, it is generally hypothesized that propositionally dense texts appear to be more difficult to process and comprehend than less propositionally dense texts.

In text comprehension literature (Kintsch, 1988, 1998; Embretson & Wetzel, 1987; for example), propositional density is generally investigated through an analysis of the number of paragraphs included in a text, the number of sentences included in each paragraph, and the average number of words included in the sentence. I have also adopted this methodology in order to examine propositional density of the eight texts included in the study sample.

However, since text density is associated with text difficulty and task complexity, addressing the issue of ESE text density from this angle is likely to produce results that hardly enable to assess the construct validity of ESE texts. Since my epistemological stance supports that the construct validity of a reading test needs to be addressed in the broader context of the actual reading

activity, I hypothesize that the validity of ESE also depends on the extent to which the density of ESE texts is proportional to the density of classroom texts examinees have been reading in classroom.

In light of this hypothesis, in order to evaluate the validity of ESE texts in relation to their density, I propose to first describe the propositional density of the eight texts and one letter included in the sample, and then I compare their density to the density of texts and letters used in English classes. At this juncture, I hasten to mention that the English curriculum (Programme National d'Anglais, 1988, 2007) is mute as to the length of texts to be used in reading classes. However, my scrutiny of the three textbooks widely in use in Grade 12 will serve as a basis for determining the main patterns of classroom text density.

In order to provide a description of the text density, I have considered four variables. The first variable is the number of words included in the whole text. This statistic was obtained by simply counting the words included in each text. This statistics provides the picture of text length, and enables to decide whether texts used in ESE are long, short or of average length. The second variable is the number of paragraphs included in the whole text. This statistic was also obtained by simply counting the number of paragraphs in each text. Since a paragraph is a group of sentences around a single idea, it can be hypothesized that if there are more paragraphs included in the text, there will be more ideas; as a result, the text becomes dense and susceptible to comprehension difficulty. The third variable is the average number of sentences in a paragraph. This statistics was computed by counting the total number of sentences included in the text and dividing it to the number of paragraphs in the same text. Since longer paragraphs may be more difficult to process and comprehend than shorter paragraphs, this statistics is useful to picture the length of paragraphs, and this may help to understand text density and task complexity. Finally, the fourth variable is the average number of words included in a sentence. This variable was investigated to picture sentence length on the assumption that longer sentences may be difficult to parse than shorter sentences. This statistics was calculated by taking the number of words included in the whole text and divide it to the number of sentences included in the same text. Table 8.3 on the next page provides information on density of texts included in the study sample.

Table 8.3: Density of texts included in the study sample

Text density variables		T1	T2	T3	T4	T5 <sup>1</sup>	T6	T7	T8	Mean	SD
1	Number of words included in the whole text	232	266	272	298	492	291	181	172	276	99.5
2	Number of paragraphs included in the whole text	4	3	5	4	7	3	5	3	4	1.4
3	Average number of sentences included in a paragraph	5	6	3	4	5	6	2	4	4	1.4
4	Average number of words included in a sentence	12	15	18	19	14	16	18	14	16	2.4

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8; SD=Standard deviation**

Table 8.4: Average density of classroom texts

Text density variables		Cartledge Book 4			English for Africa			Today's English			GM	SD
		S	L	M	S	L	M	S	L	M		
1	Number of words included in the text	587	1216	902	398	857	628	758	1659	1209	913	290.66
2	Number of paragraphs included in the text	5	9	7	6	9	8	6	10	10	8	0.58
3	Average number of sentences included in a paragraph	5	8	7	6	8	7	6	9	9	7	0.58
4	Average number of words included in a sentence	23	17	20	11	13	12	21	19	20	17	4.61

**S=Short text; L=Long text; M=Mean; GM=General mean; SD=Standard deviation**

From my scrutiny of information contained in Table 8.3 and Table 8.4, I wish to make the following observations:

<sup>1</sup> The description takes account of both the text and the letter

Concerning the number of words included in the text, the statistics in Table 8.3 and Table 8.4 suggest that classroom texts had an average number of 913 words per text while ESE texts had an average number of 273 words. This finding suggests that classroom texts were three times longer than ESE texts. Therefore, one question worth asking is to know if, for the sake of appropriateness, it is better that the density of texts included in the ESE relates to the density of classroom texts, or for task complexity purpose, the texts used in reading assessments must be less dense than those used in classroom reading. This issue appears to be more difficult than expected, and I will elaborate on it in the Discussion Chapter.

Regarding the number of paragraphs in the text, the information in Table 8.3 and Table 8.4 indicates that classroom texts had an average number of eight (8) paragraphs while ESE texts had an average number of four (4) paragraphs. This finding suggests that classrooms texts had twice the number of paragraphs compared to ESE texts. This finding is relevant to this study in that, if a paragraph is a group of sentences constructed around a single main idea, it is reasonable to conclude that a text that is constructed around more main ideas may be more challenging to readers than a text constructed around few main ideas. However, evidence needed to supports this claim can be gathered through correlational studies that, unfortunately, overlap the scope of this study. Nevertheless, I wish to state at this juncture that, this finding clarifies the real need for ESE developers to be thoughtful about the density of the ESE text and the main ideas around which these texts need to be constructed. I will elaborate on this issue in the Discussion Chapter.

In regard to the average number of sentences included in a paragraph, the data presented in Table 8.3 and Table 8.4 (see previous page) indicate that classroom texts have an average number of seven (7) sentences in a paragraph while ESE texts have an average number of four (4) sentences in a paragraph. This finding suggests that classroom texts have almost twice the number of sentences included in a paragraph compared with ESE texts. This finding is also relevant in this study. It prompts the following question: can we assume that it is more difficult for readers to comprehend a paragraph when it is elaborated through more sentences than when it is elaborated through few sentences? In other words, if we agree that a paragraph is elaborated around one main idea and if each sentence included in a paragraph states a detail that supports that main idea, can we conclude that readers may understand a main idea better when it is elaborated through more sentences than when it is elaborated through few sentences, or vice-versa? The

answer to this question can be provided through correlation studies, and this overlaps the scope of the descriptive nature of the present study. However, this finding prompts the necessity for test developer to be sensitive to the density of sentences included in text paragraphs in order to ensure the quality of ESE texts.

Regarding the average number of words included in a sentence, the evidence offered in Table 8.3 and Table 8.4 (see page 287) confirms that classroom texts have an average number of seventeen (17) words included in a sentence while ESE have an average number of sixteen words (16) included in each sentence. This finding indicates that the length of the sentence is almost the same in both classroom texts as well as in ESE texts; suggesting that examinees might find some familiarity with ESE sentence length; and this may positively contribute to their attempts to parse ESE text sentences.

In light of the preceding argument on the validity of the ESE texts in relation to their density and on the basis of the aforementioned results, I wish to state the following conclusion: Since the texts included in the ESE are far less dense than those examinees use in classroom reading, the ESE texts appear to be less appropriate to the ESE context. Implicit in this conclusion, ESE scores may hardly reflect the examinees' actual text comprehension in non-testing situations. I will elaborate on this issue in the Discussion Chapter.

### **Summary of the chapter**

This chapter has presented and analyzed data that sought to investigate the different textual features that potentially impact on the difficulty of ESE tasks, and evaluate the extent to which the ESE includes texts whose features are appropriate to the ESE context. Therefore, the following findings have been reported:

(1) Although there has been evidence of some texts that were not familiar to participants, I have classified most texts to be either familiar or somewhat familiar. On the contrary, only few participants perceived the text content to be familiar whereas the majority of them perceived the texts to be either somewhat familiar or not familiar;

(2) Text passages participants perceived to be familiar were those whose content addressed current issues that affected their immediate environment; or those texts that addressed themes that were related to their subject knowledge;

(3) Text passages participants perceived to be somewhat familiar were essentially texts that depicted a social situation relevant to their backgrounds but with details that described a setting that was not familiar to them; therefore, making it difficult to visualize the text setting;

(4) Text passages participants perceived not familiar were essentially those texts that were not appropriate to the kind of text contents they were using in classroom reading, and whose text details did not relate to their social backgrounds;

(5) According to my evaluation, most texts included in the study sample were either easy or of moderate difficulty. However, the majority of participants perceived the text content to be either difficult or of moderate difficulty;

(6) Text passages participants perceived to be difficult were those texts whose contents were not appropriate to the kind of text contents participants were familiar with in learning situation, and whose details did not relate to their social backgrounds;

(7) The eight passages and one letter included in the eight test papers related to three genres: narration, exposition and practical writing;

(8) Narrative texts were structured in four types: time sequence, description, cause/effect, and problem/solution structures; expository texts were organised in three types: description, comparison, and cause/effect structures; while the practical writing text was organized in problem/solution and cause/effect structures;

(9) The five narratives used three main descriptors: adjectives, adverbials, and prepositions;

(10) In the five narratives, time sequence was generally translated through the use of the past tenses together with adverbials that introduced the first sentence of the narratives. Also, the past tenses were used to narrate past events; and this was found in the opening sentence and could go on through the whole story;

(11) The five narratives were truncated as they did not include time sequence devices that could provide the reader with the evolution of the narrative story line necessary to help examinees to comprehend the text;

(12) Although the five narratives followed the story line appropriate to most narratives, the questions based on these narratives did not necessarily follow the linear and incremental order appropriate to narratives;

(13) Some test questions based on the five narratives did not call for operations that were appropriate for narratives;

(14) The three expository texts were structured around main ideas and supporting details; suggesting that examinees had to process the text at both global and local levels. Furthermore, it was found that the use of descriptors and present tenses characterised the three expository texts;

(15) Test questions included in the three expository tests actually called for examinees' ability to process the text at both global and local level in order to get the gist of the information presented in the text as well as to locate specific pieces of information located at both global and local levels. However, there were no test questions that requested examinees' capacity to identify and understand descriptors and place prepositions necessary to visualize text details;

(16) Test questions included in the test based on letter writing were generally appropriate to the kind of tasks that required the examinees' skill to read the letter and understand its parts, form and content. Nevertheless, the few number of test questions suggest that the test could not cover the entire range of letter writing skills;

(17) Classroom texts were three times longer than ESE texts; suggesting that ESE texts might be less appropriate to the ESE context;

In the Discussion Chapter, I will elaborate on these findings by presenting their relevance and implications to the present study. At this juncture, I present, in the following chapter, findings that pertain to item variables that impact on task complexity.

## **CHAPTER NINE**

### **FINDINGS RELATED TO ITEM FEATURES THAT POTENTIALLY AFFECT THE VALIDITY OF THE ENGLISH STATE EXAMINATION**

#### **9.1 Introduction**

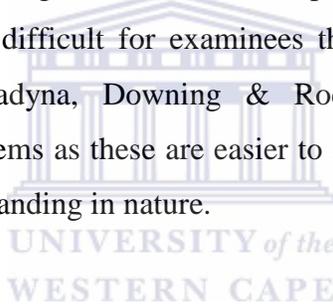
The multi-componential processing model I have used in the present study hypothesizes that the features of test items impact on the examinees' selection of strategies to construct text meaning and complete test tasks. Following this hypothesis, the model suggests that the different properties of an item can influence examinees' scores by enabling some examinees to get an item right or wrong simply by the way it is constructed. However, if examinees' scores can be inflated by factors irrelevant to their abilities to process the text, comprehend it and complete test tasks, we can state that such scores cannot reflect the examinees' actual reading abilities. Therefore, we can conclude that such scores cannot serve as a basis for meaningful and appropriate decisions. On the basis of this consideration, in the present chapter, I propose to present and analyse data that sought to determine item features that can affect the validity of the ESE. In order to achieve this objective, I propose to first look at the stem structure before I inspect the phrasing of the key. Next, I investigate the quality of alternatives, and finally I examine the quality of the sixth option (option 6). I have used two main types of data in this chapter: the quantitative analysis of participants' responses to the eight tests and the qualitative examination of the eight (8) ESE papers. In the following section, I examine the ESE stem.

#### **9.2 ESE stem quality and its appropriateness to the ESE context**

The stem is the first part of a multiple-choice question that states the question. Different item writing guidelines (see Haladyna, Downing & Rodriguez, 2002) advise test writers to construct tests with good stems since it is agreed that quality stems contribute to the quality of the test

question. In order to examine the quality of stems used in ESE, I have used three variables: (1) stem focus, (2) stem completeness, and (3) stem orientation. In examining stem focus, I intend to find out the extent to which the ESE stems clearly present the central idea of the test question. As evidenced by literature (Downing *et al.*, 1999; Haladyna, Downing & Rodriguez, 2002), when examinees read test questions where the stem presents enough information that allow them to know what is being asked, they better perform on the test than when they read test questions with stems that require them to first review the different item alternatives in order to understand what is being asked.

In examining stem completeness, my objective is to establish whether the stem is presented in a form of a complete statement or an incomplete statement, and also whether it is positively or negatively worded. There is evidence (Haladyna, 1999; Haladyna, Downing & Rodriguez, 2002) that suggests some preference for using test items with complete stems because incomplete stems are sometimes found to be more difficult for examinees than complete stems. Also, there is evidence (Haladyna, 1999; Haladyna, Downing & Rodriguez, 2002) that advises test constructors to develop positive stems as these are easier to comprehend than negative stems as the latter are more cognitively demanding in nature.



### **9.2.1 ESE stem focus**

In order to examine the ESE stem focus, I inspected the formulation of the stems in the 56 test questions and I conducted a qualitative evaluation of those stems that I perceived to be focused and those that I perceived to be unfocused. I used Downing *et al.*'s (1991) operational definition of a focused stem where they suggest that a stem is focused when the examinee can pose an answer based only on the stem. In light of this definition, I classified the 56 test questions in two groups: those that had focused stems and those that had unfocused stems.

The information presented in Table 9.1 on the next page suggests that, of the 56 test questions included in the study sample, 34 test questions (61.0%) had focused stems while 22 test questions (39.0%) had unfocused stems. This finding suggests that the ESE uses stems that are both focused and unfocused although focused stems are more frequent than unfocused ones.

Table 9.1: ESE stem focus, completeness and orientation for the eight (8) tests

Variables			T1	T2	T3	T4	T5	T6	T7	T8	Total	
			N=7	N=9	N=9	N=5	N=7	N=9	N=5	N=5	N=56	%
1	Stem focus	Focused	5	3	7	4	6	3	2	4	34	61
		Unfocused	2	6	2	1	1	6	3	1	22	39
		<b>Total</b>	<b>7</b>	<b>9</b>	<b>9</b>	<b>5</b>	<b>7</b>	<b>9</b>	<b>5</b>	<b>5</b>	<b>56</b>	<b>100</b>
2	Stem completeness	Complete	7	9	7	3	1	8	4	4	43	77
		Incomplete	0	0	2	2	6	1	1	1	13	23
		<b>Total</b>	<b>7</b>	<b>9</b>	<b>9</b>	<b>5</b>	<b>7</b>	<b>9</b>	<b>5</b>	<b>5</b>	<b>56</b>	<b>100</b>
3	Stem orientation	Positive	6	6	8	5	7	8	3	5	48	86
		Negative	1	3	1	0	0	1	2	0	8	14
		<b>Total</b>	<b>7</b>	<b>9</b>	<b>9</b>	<b>5</b>	<b>7</b>	<b>9</b>	<b>5</b>	<b>5</b>	<b>56</b>	<b>100</b>

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8

However, a question worth asking is to know how the focused and unfocused stems were reflected in the data. A qualitative content analysis of the eight test papers and an investigation of participants test taking strategies reveal that the following types of stems were focused:

(a) The six (6) test questions that asked examinees to select the option that suits as the title of the text/paragraph (T1.1, T2.1, T4.1, T5.1, T7.2 and T8.1 [see Appendix 7]).

Consider this test question: T1.1

The title that best suits the fourth paragraph of the above text is:

1. Yasekuru's sickness
2. Kirongozi's death
3. A tragic experience
4. African women and AIDS
5. Widow Yasekuru's situation

In order to answer this test question and other test questions listed previously, examinees had to, after they had carefully and/or expeditiously read the text/paragraph, read the question stem and immediately produce their own answer and then look at the five alternatives to confirm the answer produced (S14). I wish to mention that Appendix 17 presents extensive details of the

frequency of strategies used by participants to answer the six test questions listed in this category.

(b) The three (3) test questions that asked examinees to skim the part of the text where a specific idea was located (T1.2, T4.2, T8.2 [see Appendix 7]).

Consider this test question: T1.2

The event that affected deeply Yasekuru's life is found in the ... paragraph:

1. Fifth paragraph
2. Second paragraph
3. First and fourth paragraph
4. Second paragraph
5. First paragraph

Like in the preceding example, evidence of strategies use in Appendix 17 indicates that some examinees were able to read the question and produce their own answers before they could inspect the five alternatives to confirm their choice (S14).

(c) The fifteen (15) test questions that required the examinee to understand syntax (T1.7, T2.5, T2.7, T3.2, T3.3, T3.4, T3.8, T3.9, etc. [see Appendix 7]). Here is an example (T3.9):

“I finally calmed everyone down.” The author said he finally ... everyone down

1. calms
2. had calmed
3. calmed
4. would calm
5. has calmed

In order to answer this test question, some examinees who were aware of how to use tense shift when we turn a sentence from direct to reported speech immediately produced their answer (had calmed) before going through the suggested five options (S14). Data in Appendix 17 provide extensive details on the participants' use of S14 to answer this test question and other test questions classified in this category.

On the other hand, **unfocused stems** are reflected in the following types of items:

(a) The test questions that requested examinees to search read the text in order to locate explicit pieces of information scattered through the whole text (T1.5, T2.2, T2.6, T3.4, T3.7, T4.3, T6.3, T7.1 [see Appendix 7]). An example of this type of question is test question T1.5:

All the following sentences agree with the text, except:

1. Yasekuru will not get married because she is aware of her sickness
2. Kirongozi suffered from a well-known illness
3. Yasekuru's husband had a lot of sexual partners
4. Yasekuru's husband was a car driver
5. Kirongozi died at the eight year of marriage

In order to answer this question and many other questions classified in this group, examinees could not read the stem and immediately pose an answer based only on the stem. On the contrary, they had to skim and scan the text in order to find pieces of needed information. When one piece of information was found, the examinees inspected the five options and eliminated that option until few options remained before they could finally select their answers. Therefore, (logical) elimination of alternatives (S17) appeared to be the main test taking strategy for answering this type of test questions. Appendix 17 provides evidence on the participants' frequent use of S17 to answer most of the test questions included in this category.

(b) Test questions that required examinees to scan the text in order to locate explicit information located in a specific part of the sentence or paragraph (T2.3, T2.4, T6.2, T7.3, and T7.4 [see Appendix 7]). For example, test question T2.4:

The following objects were used by students in the incident, except:

1. pieces of stones
2. any harmful objects
3. iron sticks
4. tears' gaz
5. pieces of bricks

In order to answer this question and other similar questions listed previously, the majority of examinees had to scan the specific sentence that contained the list of objects used by students in the incident. Therefore, the answer to the question was provided only after the examinees had scanned the sentence and then after they could read the five options and finally focus on the

option that they found to be the “correct” option, suggesting that the main test taking strategy for this type of questions is S13 that reads “consider the option and focus on the familiar option”.

### **9.2.2 Stem completeness**

In order to investigate the stem completeness in the eight test papers, I examined the formulation of each test question stem and therefore decided whether it was a complete stem or an incomplete stem. Any stem written without suspension marks (...) was assumed to be complete whereas any stem written with suspension marks was classified as incomplete. Although there is still inconclusive results on which of these two kinds of stems is appropriate for examinees, most item writing guidelines (see for example, Haladyna, Downing & Rodriguez, 2002) urge test constructors to design MC test with complete stems. I also adopt this suggestion on the ground that a complete sentence may be easier to parse than an incomplete one. Nevertheless, this is not to imply that I undervalue the validity of the incomplete stems.

The data presented in Table 9.1 (see page 294) indicates that the majority of test items (N=43; 77%) included in the study sample were complete stems, while only few stems (N=13; 23%) were incomplete stems.

### **9.2.3. Stem orientation**

In order to investigate the stem orientation in the eight test papers, I examined the formulation of each test question stem and therefore decided whether it was positively oriented or negatively oriented. In item writing literature (see Haladyna, Downing & Rodriguez, 2002), a positively oriented stem is a stem that presents information in an assertive statement. Such a stem does not include any negation whereas a negatively oriented stem generally includes at least one word with negative/contrast meaning (for example: none of, neither, except...).

Since there is evidence (see Ascalon *et al.*, 2007; Schaefer, 2009; Tamir, 1993) that supports that negatively-oriented stems are more difficult to understand than positively-oriented stems, I have also adopted this view to consider test questions that use negative stems as questions that might

be problematic for examinees; therefore, the use of these types of questions needs to be restricted.

The statistics presented in Table 9.1 (see page 294) suggest that, of the 56 test questions included in the study sample, most questions (N=48; 86%) had positive stems while only few questions (N=8; 14%) had negative stems.

However, all the eight (8) test questions that used negative stems (T1.5, T2.2, T2.4, T2.6, T3.6, T6.4, T7.1, T7.4 [see Appendix 7]) used one negative word: “*except*”, as in the following item question T2.6:

All the following sentences agree with the text, except:

1. iron sticks were launched to policemen
2. the driver died as the lorry bumped at the tree
3. Apolosa was injured by a harmful object
4. the Chief District of the police ordered to protect the N.C.S.
5. sexual abuses were committed by policemen

In order to provide the answer to this question as well as other questions included in this category, participants’ strategies use suggests that the majority of examinees had to reread the question in order to understand that the good answer was the option that *did not* include the information contained in the text.

However, if we can hypothesize that test items that use negatively oriented stems may be difficult for examinees as they are more cognitively demanding, one way to test this hypothesis would be to look at the item difficulty of such test questions as well as examinees’ perceptions of item difficulty level. On the basis of this information provided in Appendix 17, the item difficulty index for the eight items classified as having negative stems as presented in Table 9.2 on the next page confirms that, of the eight questions that had a negative stem, six questions (thus 75%) had an item difficulty situated in the range between .2 and .4, suggesting that these test questions were difficult for participants.

Table 9.2: Item difficulty index for test questions with negative stems

Test item	T1.5	T2.2	T2.4	T2.6	T3.6	T6.4	T7.1	T7.4
Item difficulty	.3	.2	.3	.3	.7	.8	.4	.3

However, one important point worth noting with the data in Table 9.2 is that two test questions have their item difficulty of .7 (T3.6) and .8 (T6.4), suggesting that these two questions were easy. This result, since it is in contradiction with the main trend and existing literature, needs some explanation. The following is an example of test item T6.4:

The following items are birds, except:

1. duck
2. goose
3. cow
4. chicken
5. pigeon

My inspection of this test item as well as test item T3.6 (see Appendix 7, Section 3) reveals two important things: first, these two items required reading at local level (lexical encoding) contrary to the other six questions that required reading the text at global level. Since examinees had to process the text at sentence level in order to encode the lexis, this may have partly contributed to low difficulty as evidence suggests that test questions that require local level reading are less cognitively demanding than those that require global (text) level reading (van Steensel, Oostdam & van Gelderen, 2012; OECD, 2003; Song, 2008; Rouet *et al.*, 2001). Second, these two test items have focused stems while the other six test questions have unfocused stems. Since evidence supports that test questions with focused stems are generally easier than those with unfocused stems (Downing, 1991; Haladyna, Downing & Rodriguez, 2002), this might also have contributed to low difficulty.

Although I will elaborate on this issue in the Discussion Chapter, I hasten to state at this juncture, that stem orientation is not a variable that solely impacts on item difficulty; rather, it may interact with many other variables (such as stem focus) to account for examinee's performance on the test. This confirms my motivation to conceptualize reading in general and reading performance in particular as a multi- componential activity, and that not only any effort to explain it by single and isolated factors may be futile, but also any attempt to rank through

sophisticated factor analysis and regression statistical analyses the “exact” factors that “actually” affect reading performance may be just a scholarly enterprise that has less educational benefits.

### **9.3 The ESE key**

The key is the correct option in a multiple-choice question. Different item writing guidelines (see Haladyna, Downing & Rodriguez, 2002, for example) advise test writers to be careful not to provide clues to the right answer, and to vary the location of the key.

Balancing the key is a procedure in which each test option should be correct in approximately equal proportion of time (Taylor, 2005). Conventional wisdom suggests that when the correct answers to a multiple-choice test are distributed in different options, this distribution minimizes the effect of guessing.

If varying the location of the key is a technical requirement that aims to ensure the psychometric properties of the items, the provision of some clues to the correct answer appears to be a serious issue as it affects examinees’ scores. This is to suggest that when the correct answer in a multiple-choice test contains some clues, these clues may enable some examinees to get the right answer without necessarily knowing the answer to the question.

In this section, I propose to use tests results from the 496 student-participants as well as the content of the eight test papers in order to investigate these two issues.

#### **9.3.1 Investigating the structure of the ESE key: evidence of clues to getting the item right**

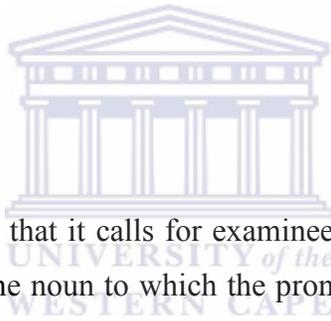
If the correct alternative to a multiple-question contains some hints that may enable some examinees to get the item right, this means that scores obtained by these examinees are not their true scores since these scores include extra scores (variance) due to factors that are not related to their comprehension of the text and their understanding of the test questions and alternatives. Technically, such examinees’ scores are believed to include construct-irrelevance variance, a threat to test construct validity.

In order to investigate this issue, I have examined the structure of 56 test questions and identified those questions that contained the correct answer with clues that potentially could enable some wise examinees to get the question right. Since such qualitative appreciation was not enough to provide evidence of the presence of these clues, I have also triangulated this analysis with an investigation of strategies used by participants to answer these specific test questions.

My inspection of the structure of the key for the 56 test items reveals that six (6) items had their keys that potentially provided some clues to wise examinees. These test items are: T3.7, T5.3, T5.6, T6.4, T6.8 and T8.5. I wish to provide an example of two of these six test items and explain how the quality of their keys may be questionable. These two test questions are T5.3 and T6.8. Test question T5.3 reads:

The underlined word “**They**” in the third paragraph of the text refers to:

1. gentleman
2. accountant
3. manager
4. people
5. person



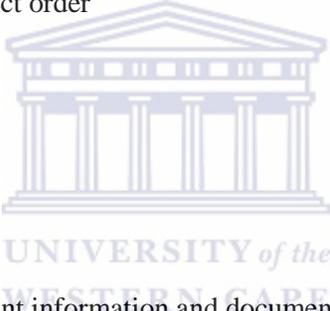
A scrutiny of this test item reveals that it calls for examinees’ skill to parse a specific sentence from the text in order to identify the noun to which the pronoun “*they*” refers. Since the part of the text where the pronoun “*they*” appears is indicated in the question stem (“*the third paragraph of the text*”), the examinees must go to that part, locate the sentence (and neighbouring sentences), parse it/them and therefore produce their own answer before going back to the question and confirm the answer from the suggested list of alternatives.

However, some examinees who are knowledgeable of the syntactic rules of personal pronouns may figure out that “*they*” must refer to a plural noun, and that only item alternatives with plural nouns are competing alternatives to inspect while any other alternative must be discarded. Therefore, by scrutinizing the five alternatives, these examinees can find that only one alternative contains a plural noun (“*people*”), suggesting that this alternative **MUST** be the correct answer. Hence, in selecting alternative **4** (“*people*”), which is in fact the correct answer, without reading the part of the text where the pronoun *they* is located, these examinees have got a mark that does not reflect their understanding of the part of recommended text as well as their understanding of the question’s expectation (to parse the sentence/s).

The table in Appendix 17 contains a column that indicates the strategies used for responding to individual test items. In this table, evidence indicates that some participants used clues provided in the alternative (S15) in order to answer this test question. Also, the item difficulty for this test question (.8) as indicated in the same table confirms that most participants got this question right. This finding is confirmed by participants' perceptions of item difficulty. The table in Appendix 21 reveals that, of the 107 participants who answered this test question, 59 (55%) reported perceiving this question as easy, 28 participants (26%) perceived it as a question of moderate difficulty while only 20 participants (19%) perceived it as a difficult test question. These two findings appear to confirm my evaluation of this test question.

The second test item is T6.8, it reads:

If you were asked to write a reply to the above letter, here are the steps you should go through. They are given in a jumbled way. Give the correct order

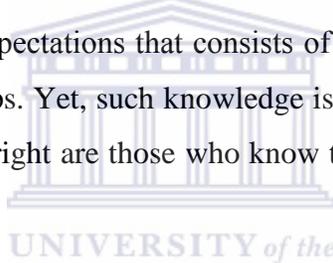
- 
1. write your first draft
  2. revise your first draft
  3. write an outline
  4. assemble all the relevant information and documents
  5. write your aim
  6. arrange the points in the order of importance
  7. write and type your final version

1. 1 – 3 – 4 – 2 – 5 – 7 – 6
2. 1 – 6 – 4 – 2 – 3 – 7 – 5
3. 4 – 1 – 2 – 6 – 7 – 3 – 5
4. 3 – 7 – 6 – 1 – 2 – 5 – 4
5. 5 – 4 – 6 – 3 – 1 – 2 – 7

This question is an example of matching multiple-choice questions which are predicted to be challenging for examinees (see Haladyna, Downing & Rodriguez, 2002). In order to answer this question, examinees have to know the different letter writing steps. This may be a challenging activity. Some examinees may know the steps but may find it difficult to order them. Others may

know the first step (“*write your aim [5]*”) and the last step (“*write and type the final version [7]*”), but may not be able to provide the correct order of other steps that are situated between the first and the last steps. Still, other examinees may not even know any letter writing step, but by reading from numbers 1 to number 7, may sense that any letter writing normally ends when someone writes and types the final version (7). In both cases, some wise examinees may inspect the five question alternatives and search for the alternative that starts with 5 and ends with 7, ignoring the other numbers situated between these two ends. Since there is only one alternative (alternative 5) that starts with 5 and ends with 7, they may select this alternative which is effectively the correct answer. Or again, some others may just search for the alternatives that end with number 7, ignoring the other alternatives. Since there is only one alternative that ends in 7 (alternative 5), these examinees may select this alternative, which is the correct answer.

Like in the preceding example, examinees’ selection of alternative 5 is far to reflect their understanding of the question’s expectations that consists of the knowledge of the letter writing steps and the ordering of these steps. Yet, such knowledge is a way to provide evidence that test takers who have got this question right are those who know the different letter writing steps and their proper ordering.



The table in Appendix 17 provides evidence that substantiate that some participants used clues provided in the alternative (**S15**) in order to answer this test question. Although apparently this kind of matching MC question is among questions predicted to be difficult, the item difficulty for this test question (.5) as indicated in the same table confirms that this question was of average difficulty, suggesting that half of participants got it right. This finding appears to confirm my judgement. However, this finding is not confirmed by reports from examinees’ perceptions of item difficulty. The information in the table in Appendix 21 suggests that 71 % of the 35 participants who answered this question perceived it as a difficult question while 20 % perceived it of moderate difficulty and only 9 % perceived it as an easy question. Nevertheless, this finding suggests that examinees’ perceptions of item difficulty may not necessarily be translated in actual item difficulty. Examinees may perceive an item as easy but still miss it; or they may perceive an item as difficult but still get is right.

Although the number of test items that provide evidence of clues provided in the key to get the item right is very few (6 items= 11%), the findings from these two examples have relevance to

this study. They suggest that scores obtained from tests that include these test items may fail to reflect examinees' reading abilities since these scores are inflated with irrelevant variance that makes these test items to be easier than they really are.

### **9.3.2 Investigating the structure of the ESE stem: evidence of the distribution of the key**

Since multiple-choice tests are, by their essence, susceptible to guessing, some examinees can receive credit on test items simply by guessing from the response patterns of the test. Although empirical evidence does not show any impact of the response pattern on reading test performance (see Taylor, 2005, for example), most theoretical investigations (Haladyna, Downing & Rodriguez, 2002; Sechrest, Kihlstrom & Bootzin, 1993; Ory & Ryan, 1993; Ellsworth, Dannel & Duell, 1990) advocate balancing the key on the assumption that test wiseness examinees can use a response set for guessing.

In light of the preceding argument in favour of balancing the key, I have computed the distribution of the correct answer among the six possible answers in order to find out whether or not ESE constructors are aware of the necessity to locate the key in such a way that it is “equally” distributed among all the six possible choices. Table 9.3 on the next page presents the frequency of items per each choice for each test.

The information displayed in this table confirms that the responses of the 56 test items were distributed among all the 6 alternatives/choices. However, this distribution was not balanced. If we consider that with 56 items and 6 possible choices, each choice should be the correct answer 9 times (17%), it clearly appears that there response pattern is not “equally distributed” since some choices are more included as correct answer than other choices. This is the case of choice 3 (25%), choice 1 (21%), and choice 4 (18%) that are more frequently used (their percentage is above the mean of 17 %) than choice 2 (16%), choice 5 (14%) and choice 6 (5%).

Although there is no evidence that the response pattern can impact on test performance, this finding provides evidence that test developers may not sometimes be thoughtful about the issue of “equal distribution” of the correct option in a test. Yet, most theoretical studies recommend such a distribution.

Another important conclusion from this finding is that the implied sixth choice (*“If there is no good answer, write 6”*) is a good answer for only 3 of the 56 items (5%); suggesting that this choice is almost not included by test constructors as the correct answer. This finding interrogates the validity of choice 6. I will investigate this issue in Section 9.5.

Table 9.3: Frequency of items per choice for the eight (8) tests

Key	T1 N=7	T2 N=9	T3 N=9	T4 N=5	T5 N=7	T6 N=9	T7 N=5	T8 N=5	Total N=56	%
Choice 1	1	3	1	1	2	1	1	2	12	21
Choice 2	1	1	3	2	1	1	0	0	9	16
Choice 3	2	2	3	1	2	3	1	0	14	25
Choice 4	2	2	1	0	1	2	1	1	10	18
Choice 5	1	1	1	0	1	2	1	1	8	14
Choice 6	0	0	0	1	0	0	1	1	3	5
<b>Total items</b>	<b>7</b>	<b>9</b>	<b>9</b>	<b>5</b>	<b>7</b>	<b>9</b>	<b>5</b>	<b>5</b>	<b>56</b>	<b>100</b>

T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8

#### 9.4 The ESE distractors

The distractors are incorrect alternatives/choices/options of the multiple-choice item. This issue of distractors is a highly sensitive issue because the quality of multiple-choice items depends largely on the quality of the distractors. Since the ESE includes six alternatives among which one is the correct alternative and the five others are distractors, the examination of the distractors quality can provide some understanding on how examinees use different test taking strategies to select the correct answer and discard other options. In the present section, I propose to evaluate the quality of ESE distractors by examining their plausibility.

In item analysis literature (Haladyna, Downing & Rodriguez, 2002; Downing *et al.*, 1999; Ascalon *et al.*, 2007; Schaefer, 2009), it generally suggested that the most important quality of a distractor is plausibility. This is to suggest that, a plausible (or functioning) distractor must be selected by some examinees as correct answer since it is expected to be similar to the correct answer.

In the scope of this thesis, I have investigated the distractor's plausibility through two procedures: the first procedure is achieved by computing the distractor weights. This is a straightforward statistics that consists in computing, for each distractor included in each test item, the number (and percentage) of examinees who have selected this distractor. From the results obtained from the first method, I have conducted, in the second procedure, a qualitative content analysis of the ESE papers to explain both functioning and non-functioning distractors. Based on item analysis guidelines (see Haladyna, Downing & Rodriguez, 2002), I have considered a functioning distractor as any distractor selected by *at least* 10 per cent of examinees; and non-functioning distractors as any distractor not selected or selected by less than 10 per cent of examinees. In order to validate my judgments, I have also examined the reports of strategies used by the participants to answer test items.

Table 9.4 on the next page summarizes information on the functioning distractors for the 56 test questions. However, the extensive information relating to distractors' weight for the 56 test items across the eight tests is provided in Appendix 19. In this Appendix 19, I have classified the 56 items in six categories: In the first category, I have included all the test questions that contained five (5) functioning distractors; the second category includes all the test questions that had four (4) functioning distractors; the third category includes the test questions with three (3) functioning distractors; the fourth category comprises the test questions with two (2) functioning distractors; the fifth category comprises the test questions with one (1) functioning distractor; and finally the last category included the test questions that had no functioning distractors.

In light of the information summarized in Table 9.4, there is evidence to suggest that, of the 56 test questions in the sample, no test question was found to have five functioning distractors as expected. On the contrary, most test questions had either three functioning distractors (N=19; 34%), two functioning distractors (N=19; 34%) or only one functioning distractor (N=14; 25%). Furthermore, the data in the same table suggest that only 2 items (4%) had four functioning distractors; while 2 items (4%) had not any functioning distractors. This finding suggests that most test questions included in the study sample had functioning distractors in the range of 1 and 3. But at this juncture, I wish to examine the test questions included in each category in order to provide more insights on the quality of distractors used in the ESE.

Table 9.4: Classification of the 56 test questions in categories of functioning distractors.

Number of functioning distractors	N items	%	Test items
5 functioning distractors	0	0	No test item found
4 functioning distractors	2	4	T3.5; T6.1
3 functioning distractors	19	34	T1.1; T1.4; T2.2; T2.3; T2.4; T2.5; T2.6; T2.8; T4.3; T5.5; T5.6; T5.7; T6.2; T6.3; T6.5; T6.9; T7.1; T7.4; T8.5
2 functioning distractors	19	34	T1.3; T1.5; T2.7; T3.1; T3.3; T3.4; T4.1; T4.2; T5.1; T5.2; T5.4; T6.6; T6.7; T6.8; T7.2; T7.3; T8.1; T8.2; T8.4
1 functioning distractor	14	25	T1.2; T1.6; T1.7; T2.1; T3.2; T3.6; T3.7; T3.8; T3.9; T4.4; T4.5; T5.3; T7.5; T8.3
No functioning distractor	2	4	T2.9; T6.4
<b>Total</b>	56	100	

T3.5=Test 3, item number 5; T6.1=Test 6, item number 1; T1.1=Test 1, item number 1; etc.

#### 9.4.1 Test items with no functioning distractors

Of the 56 test questions included in the sample, two test questions were found not to have any functioning distractors. These items are T2.9 and T6.4. The detailed information on the number and percentage of participants who selected each of the six alternatives for these two items is found in the table provided in Appendix 19. At this point, I only examine the distractors' plausibility for one item (T2.9). This item reads:

“Tears’ gaz, fire-arms, weapons, guns and soldiers”: The title which best suits the above groups of words is:

1. army
2. communication
3. transportation
4. hospital
5. school

<b>Distractor weight (see Appendix 19): Options 1=98% (key); 2=2%; 3=0%; 4=0%; 5=0%; 6=0%</b>
---

From my inspection of this test item results, I wish to outline the following two observations:

(1) Since this item tests vocabulary knowledge, it is expected that the distractors semantically relate not only to the stem where the question is presented (“*Tears’ gaz, fire-arms, guns and soldiers*”), but also to the key (“*army*”). Yet, this is not the case where the four distractors are totally semantically distant from both the stem and the key. This may be one reason why most participants (98%) selected alternative 1 and rejected the other 5 distractors.

(2) This item is based on the text that is about a clash between students and police. All the words used in the question stem (“*Tears’ gaz, fire-arms, guns and soldiers*”) are words included in the text. After reading the text whose content is specific, examinees may have perceived the information in the other four options as unrelated to the text they have read; and this may also have justified their decisions to discard all the other options.

In order to validate these two judgements, we need to look at test testing strategies that the participants used to respond this question. The information on strategies used provided in Appendix 17 suggests that the 64 participants who wrote this test used three different test taking strategies: For some participants, after reading the test question, they immediately produced their own answer and then looked at the five options only to confirm their choices (S14). For some others, they examined the five options and immediately focused on the option/alternative they considered to be appropriate (S13). Still, other participants made their choices after they had ‘logically’ eliminated the four other options (S17). This finding suggests that participants used relevant strategies to respond this test item and the non-selection of other alternatives may indicate that these alternatives were simply bad quality distractors.

In light of the preceding conclusion, one might expect test items that have no functioning distractors to be easy for examinees. In order to validate this hypothesis, I have looked at both the actual item difficulty and the perceived item difficulty of these questions. The actual item difficulty was investigated through participants’ scores on these test items while the perceived item difficulty was investigated through participants’ reports of their perceptions of item difficulty.

The information in the table provided in Appendix 21 indicates that the item difficulty of T2.9 and T6.4 were .9 and .8 respectively; suggesting that these two test questions were actually easy. This finding confirms the preceding judgement when I concluded that test questions with no

functioning distractors may be easy for examinees. These results have been triangulated with participants' perceptions of item difficulty. The information in Appendix 21 indicates that, of the 64 participants who answered test questions T2.9 and T6.4, the majority of participants (71%) perceived T2.9 to be either easy (45%) or of moderate difficulty (26%). Likely, the majority of participants (83%) perceived test item T6.4 to be either easy (66%) or of moderate difficulty (17%). This finding confirms that the actual item difficulty of test items with no functioning distractors may relate to examinees' perceptions of item difficulty of such items.

In light of the preceding similar findings from triangulated sources, one can interrogate the validity of a test that includes some test questions with no functioning distractors. I will examine this issue later.

#### 9.4.2 Test items with only one functioning distractor

The information presented in Table 9.4 (see page 309) suggests that, of the 56 test items included in the sample, 14 test questions had only one functioning distractor. The detailed information on the number and percentage of participants who selected each of the six alternatives for these 14 test items is provided in Appendix 19. However, I have selected one test item to examine its content. This item is T7.5, and it reads:

The cool temperature and fertile soil favour the cultivation of potatoes, ... ?

1. do they
2. are they
3. don't they
4. haven't they
5. have they

<b>Distractors weight (cf. Appendix 19): Options 1=21%; 2=0%; 3=75%; 4=4%; 5=0%; 6=0%</b>
---

The following observations are based on my review of this test item:

(1) This item requested examinees' knowledge of grammar to know how to construct tag-questions. Since this structure (tag-question) is one of the grammar structures planned in the English curriculum, examinees are expected to be aware of how to construct tag-questions. Therefore, examinees may have easily, after reading the test question, produced their own answers before looking at the five alternatives in order to select the one that confirms their answer.

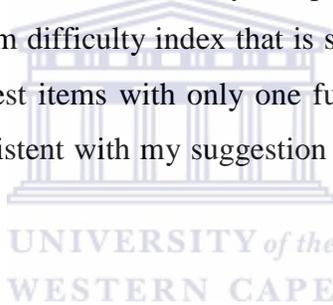
(2) A look at the explanations provided in the teaching manuals indicates that the rule for constructing the tag-question is to consider the main verb tense and then turn it in a form of a negative question. In light of this rule, if the verb in the main utterance contains an ordinary verb, the tag question will use the "do" auxiliary with its corresponding tense and person by using the opposite form (if the tense form in the verb is affirmative, the negative must be used in the tag question, and vice-versa). When this rule is internalised by the learners, any option that contains a tag question formulation that does not conform to this rule may be simply eliminated. Therefore, since the main verb in the key is "favour" [defined as <+plural, +present, +ordinary verb>], only options with "do -structure" need to be examined; hence any other options not including "do" are automatically discarded as irrelevant. This explanation can justify why only options 1 and 3 were selected and that since option 3 was the correct answer, the only option that could distract examinees was option 1. This finding suggests that, for grammar-based items, when examinees are knowledgeable of the grammar rules, their selection of the correct answer may be a straightforward activity since any option with an inappropriate structure may be simply eliminated.

(3) Following the second observation, since there are only the two options (options 1 and 3) that are plausible after the three other options have been eliminated; the possibility of getting the correct answer through guessing becomes high. Some examinees who were not capable of choosing between "do they" and "don't they" may have simply made their choice through guessing.

Since these observations are based on my judgement, there is a need to validate them through the examination of strategies used by participants to answer this test item. The table provided in Appendix 17 includes a column that presents information on strategies used to answer the 56 test items. According to the information in this table, there is evidence that the participants used four

test taking strategies in order to answer this question. Some produced their own answer after they had read the question and they could then look at the five alternatives to confirm their answer (S14). Other participants selected their “correct” answer after reading both the question and the five alternatives (S13). Some others made their selection after the process of elimination of some distractors (S17); and still, some other examinees made their selection through guessing (S18). These results appear to partially confirm my judgment enunciated earlier. They suggest that when a test item includes some distractors that do not function, examinees can sometimes recourse to strategies (such as guessing) that do not show their understanding of the construct being tested.

Like the items with no functioning distractors, items with just one functioning distractor are predicted to be easier for examinees than items with many functioning distractors on the ground that examinees have the possibility to select from only two plausible alternatives. This prediction appears to be confirmed by the item difficulty index that is situated in the range of .6 and .8 for twelve (12) of the fourteen (14) test items with only one functioning distractor (see Appendix 21). This result appears to be consistent with my suggestion that the less plausible distractors in an item, the easier the item may be.



### **9.4.3 Test items with four functioning distractors**

In a six-option multiple-choice test like the ESE, test items with four functioning distractors are the type of test items that can be desired since examinees are distracted by five of the six item alternatives; therefore, their selection of the best option potentially depends on their comprehension of the question and the meaning of each alternative. In the context of this study, I have hypothesized that, if the majority of ESE test questions have four functioning distractors, the ESE test questions will have high psychometric qualities; therefore, its quality will be enhanced. However, the data in Table 9.4 on page 317 indicates that, of the fifty-six (56) sampled test questions, only two (2) test items (T3.5 and T6.1) were found to include four functioning distractors. Details on the number and percentage of participants who selected each of the six alternatives for these 14 test items appear in the table provided in Appendix 19. However, I have selected item T6.1 in order to examine its content. This item reads:

One of the following sentences best summarizes the second paragraph

1. Geese cackled prayers loudly
2. A hungry fox found some geese in the field and wanted to eat them
3. The farmer heard the noise, came and chased the fox away
4. A goose begged for a last favour and the fox agreed
5. Geese were eaten by the fox

<b>Distractor weight (see Appendix 19): 1=26%; 2=9%; 3=14%; 4=17% (key); 5=17%; 6=17%</b>
---

From my closer analysis of this test item, I wish to make the following observations:

(1) This item is based on text comprehension; and in order to answer it, the examinees must establish accurate comprehension of explicitly stated main idea and supporting details across sentences of the second paragraph. Therefore, they must read carefully at global level.

(2) A look at the five explicit alternatives reveals that all these alternatives present information stated in the text; but to some degree. Alternative 1 presents information stated in the second paragraph; but this information is not the main idea of this paragraph; it is simply a supporting detail to the main idea. This is the most selected alternative (26%), but it is not the correct answer. Examinees who selected this alternative may have failed to distinguish the paragraph main idea from supporting details. As a conclusion, this is a very good distractor.

Alternative 3 presents information that is stated in the text. However, this information is implicitly stated in the second paragraph as details, but explicitly stated in the third paragraph as main idea. Since this is not the main idea of the second paragraph, this alternative is not a good answer. The examinees (14%) who selected this option may have read the whole text carefully and/or expeditiously before they focused on the second paragraph; therefore, they may have been confounded by the information overlap between the two consecutive paragraphs; and therefore may have failed to locate which paragraph the same information is main idea and which one it is a detail. As a conclusion, this is a very good distractor.

Alternative 4 is the correct answer. Examinees who selected it may have understood that this alternative presents the information that is central to the second paragraph.

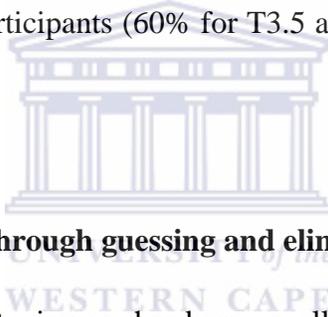
Alternative 5 presents the information that can be inferred from the second paragraph. Examinees (17%) who selected this option may have made some inferences on the second paragraph details without reading the last paragraph that presents the story denouement. Since they have answered the question by making inferences from paragraph details, this alternative is also a good distractor.

Finally, the implicit alternative 6 was selected by 17% of examinees. Since there is a section (9.4) that investigates the use of this alternative in this chapter, I intend to examine the content of this distractor in that section.

In order to validate these observations, I have examined the different strategies used by participants to answer this test item. The table in Appendix 17 presents strategies used by participants to answer the 56 test questions. In this table, the data substantiate that all participants had to read the whole text carefully (S1) before they read the second paragraph carefully (S3). Then, they could read the question, read the five options but postpone considering the “correct” option (S12); but reread the five options and then select the “correct” option either through elimination of other options (S17), by focusing on the option they found the “correct” answer (S14), or through simple guessing (S18).

This finding appears to support my judgment. The participants’ careful reading of both the whole text and the paragraph appeared to be the best activity that could enable them to first build the text macrostructure so that they could then construct the gist of the second paragraph; suggesting that a paragraph may be meaningless if it is detached from the text. On the other hand, the examinees’ reading and rereading of both the question and the five alternatives and their hesitations to make their final choice substantiate Embretson and Wetzel’s (1987) decision/falsification process in which the question stem and the options/alternatives are compared for accuracy of the text. Therefore, the three strategies used to make the decision (focus on the “correct” option [S13], select the option through elimination of other options [S17], and select option through guessing [S18]) appear to be a logical result of the examinees’ individual degree of success of the decision/falsification process. Those who successfully encoded both the paragraph and the question, and successfully evaluated the truth status of response alternatives were those who might have got the item right.

Although this test question is an example of a well-developed test item with four functioning distractors, two observations are worth making: first, since only two (2) test items (4%) are included in this category; this suggests that the number and quality of distractors included in the ESE needs to be questioned. Second, considered from the strategies use stance, these kinds of test items with four functioning distractors appear to be cognitively demanding in light of examinees' individual characteristics and the ESE context. Therefore, although I have hypothesized that test questions with four functioning distractors have better psychometric properties than those with few functioning distractors, their use in the context of this study must be taken with caution. This argument finds support when we consider the item difficulty (see Appendix 21) of these two items ( $T3.5=.3$ ;  $T6.1=.2$ ) that indicates that the two items were very difficult for examinees. Not only these two items were actually difficult for the participants, but also the participants perceived them difficult. The statistics in the table provided in Appendix 21 substantiate that the majority of participants (60% for T3.5 and 57% for T6.1) reported that they perceived these two items difficult.



## **9.5 Investigating test wiseness through guessing and elimination of distractors strategies**

In testing literature, the issue of test wiseness has been usually investigated through guessing and elimination of distractors strategies. Test wiseness is the examinee's capacity to utilize the characteristics and formats of the test and the way test items are constructed to receive a high score (Sarnaci, 1979). Viewed from this test wiseness perspective, guessing and elimination of distractors have been always considered as a threat to test validity because they lead to the overestimation of the examinee's test scores, resulting in a construct-irrelevance variance. In this section, I propose to investigate the frequency of use of guessing and elimination of distractors strategies with a focus on how these two strategies have been actually employed by participants to answer test items. Since my aim is to evaluate the extent to what these two strategies threaten the validity of the ESE, I have first identified and computed the frequency of all test items where at least one participant reported using these two strategies; then I have inspected a sample of these test items to examine how these two strategies were actually employed by participants so as to determine whether or not the use of these two strategies was construct-irrelevant as

supported in the literature. Table 9.5 presents information on the test questions where guessing and elimination of distractors were used as strategies to complete the test tasks.

Table 9.5: Frequency of guessing and elimination of distractors items for the 8 tests

TEST	Guessing items			Elimination of distractors items		
	Test items	N	%	Test items	N	%
T1 N items=7	T1.4; T1.5	2	29	T1.1; T1.2; T1.3; T1.4; T1.5; T1.6; T1.7	7	100
T2 N items=9	T2.2; T2.5; T2.6; T2.7	4	44	T2.2; T2.5; T2.6; T2.9	4	44
T3 N items=9	T3.2; T3.5; T3.7; T3.8; T3.9	5	51	T3.2; T3.3; T3.4; T3.5; T3.6.; T3.7; T3.9	7	78
T4 N items=5	T4.1; T4.4	2	40	T4.1; T4.4; T4.5	3	60
T5 N items=7	T5.4; T5.5; T5.7	3	43	T5.1; T5.2; T5.3; T5.4; T5.5; T5.6; T5.7	7	100
T6 N items=9	T6.1; T6.2; T6.8	3	33	T6.1; T6.2; T6.4; T6.5; T6.6; T6.7; T6.8	7	78
T7 N items= 5	T7.2; T7.4	2	40	T7.1; T7.4; T7.5	3	60
T8 N items=5	T8.1; T8.5	2	40	T8.1; T8.2; T8.3; T8.4; T8.5	5	100
Total= 56 items		23	41	Total= 56 items	43	77

**T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8  
T1.4=Test 1, item number 1; T2.2= Test 2, item number 2; T7.4=Test 7, item number 4; etc.**

On the basis of data in this table, there is evidence that:

(1) Both guessing and elimination of distractors were two test taking strategies some participants used to answer some test questions. However, of the 56 test questions included in the sample, guessing was used with 23 (41%) test questions while elimination of distractors was used with 43 (77%) test questions. This finding suggests that although examinees used these two strategies, they more frequently used elimination of distractors than guessing.

(2) Both guessing and elimination of distractors were used across the eight tests. This suggests that the participants used these two test wiseness strategies as part of basic activities in completing test tasks.

(3) Some participants used both guessing and elimination of distractors in answering the same question (see T1.4, T1.5; etc.), suggesting that these two test wiseness strategies can be complementary. The data in Appendix 17 indicate that, of these two strategies, participants first eliminated some distractors, and then made their final choices through guessing among the remaining options.

(4) In six tests (T2, T3, T4, T5, T7, T8), participants used guessing in answering relatively large number of test questions while in two tests (T1, T6), participants used guessing in answering few questions.

(5) In three tests (T1, T5, T8), some participants proceeded through elimination of distractors in answering all the tests questions. In other four tests (T3, T6, T7, T8), some participants used the strategy of elimination of distractors in answering the majority of test questions, while in one test (T2) this strategy was used in answering few questions.

The previous findings on guessing and elimination of distractors are relevant to the scope of this study. They suggest that some strategies used by the participants to complete test tasks may not have been construct-valid; therefore scores reported on the test might not reflect examinees' reading abilities. Nevertheless, I hasten to mention that this finding must be taken with much caution until I examine how these two strategies were actually used by study participants through concrete examples. That is the reason why I have purposely selected two test items (one for guessing and one for elimination of distractors) that I wish to examine their content.

### **9.5.1 Actual meaning of guessing strategy as investigated through a selected test item**

In order to determine the actual meaning of guessing strategy as used by the study participants, I have selected, for illustration, test question T1.4 that reads:

The contrary of “hardly” underlined in the last paragraph of the text can be:

1. happily

2. 2. commonly
3. 3. rarely
4. 4. joyfully
5. 5. unhappily

An examination of this test item reveals the following:

(1) This item requests examinees' knowledge of vocabulary through their capacity to encode the word "*hardly*". There is evidence (see Table 9.5 on page 315) that the participants who used guessing strategy to answer this test item first proceeded by eliminating other distractors. For these participants, they may have an idea of the contextual meaning of the word "*hardly*" and may have eliminated alternatives 1 (*happily*), 4 (*joyfully*) and 5 (*unhappily*) because they are not plausible to the context of sentence. Therefore, some of these participants may have found it difficult to decide between alternatives 2 (*commonly*) and 3 (*rarely*) since alternative 2 (*commonly*) is the antonym of "*hardly*" while alternative 3 (*rarely*) is the synonym of "*hardly*"; hence, they may have proceeded to guessing as the final strategy to make the decision.

(2) However, for those participants who could not figure out the contextual meaning of "*hardly*", they may have found it difficult to understand the meaning of the five options and may have only blindly guessed their answer.

These two findings suggest that guessing strategy may be either "blind" or "educated". Blind guessing occurs when examinees are not capable to understand the expectation of the test question as well as the meaning of the different alternatives. In this case, the examinees may not be able to evaluate the truth value of each alternative and therefore, may try their luck by blindly guessing the answer. In this type of guessing, the probability to get the item right is much reduced since each of the six alternatives has equal chance ( $1/6=17\%$ ). On the contrary, "educated" guessing occurs when the examinees are capable to understand the expectation of the question and the meanings of *some* item alternatives. In this case, the examinees may proceed by eliminating any non-plausible distractors and they may make their last decision by guessing between the outstanding alternatives. The participants who used educated guessing had high chance to get the item right, depending on the number of outstanding alternatives. For example in T1.4, the probability to get this item right is of  $1/2$  (that is 50%). Although I will elaborate on the implication of these findings in the Discussion Chapter, I hasten to state here, that guessing is

a test taking strategy that jeopardizes the validity of the test since it brings variance in test scores, preventing to take appropriate decisions based on examinees' scores on a test.

### 9.5.2 Actual meaning of elimination of distractors strategy as investigated through a selected test item

I selected the test item T2.6 that reads:

All the following sentences agree with the text, except:

1. Iron sticks were launched to policemen\*
2. The driver died as the lorry bumped a tree\*
3. Apolosa was injured by a harmful object
4. The Chief District of police ordered to protect the N.C.S.
5. Sexual abuses were committed by policeman

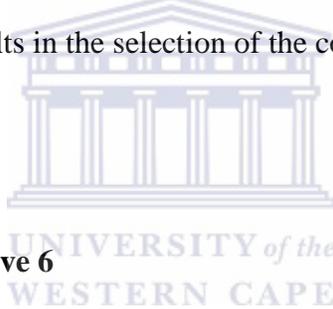
[\* I am constrained to point out that these two structures are grammatically incorrect: Iron sticks were launched *at* policemen / The driver died as the lorry bumped *into* a tree. However, for the sake of upholding the authenticity of the T2.6, I have presented the item as it appears in the test paper]

My inspection of this test item suggests the following:

(1) This item requires the examinees to read expeditiously the entire text in order to locate explicit pieces of information scattered through the whole text. Therefore, successful completion of test task depended on the participants' ability to understand the expectation of the test question as well as the ability to locate (in the text) information contained in each alternative. Since the understanding of the test question's expectation depended on the participants' ability to understand that they were requested to select the alternative that contained information that *is not* stated in the text, some participants may, every time they read the text and found the information that was not stated in one alternative, eliminate that alternative; suggesting that this *was not* the correct answer. This elimination process could be conducted until the examinee could find an alternative that contained information that *was not* stated in the text (alternative 2: "The driver **died** as the lorry bumped into a tree" instead of "the driver **was wounded** as the lorry bumped a tree").

(2) If some participants who reported having proceeded through the elimination of distractors used this logical process; it can be concluded that this strategy is construct-valid because it exemplifies examinee's genuine understanding of the text, his/her understanding of test question expectation and his/her capacity to evaluate the truth status of each item alternative on the basis of information each alternative contains.

This finding is relevant to this study. Unlike the existing conceptualization of elimination of distractors as a test wiseness strategy that threatens test validity, this strategy can be sometimes highly construct-valid; therefore, any argument that views this strategy as construct-irrelevant must be considered with much caution. I will elaborate on this finding in the Discussion Chapter. But at this juncture, I hasten to state that when elimination of distractors results in guessing between the few outstanding alternatives (see previous example on guessing subsection), this strategy is construct-irrelevant; but when elimination of distractors consists of evaluation of the truth status of alternatives and results in the selection of the correct answer, this strategy is highly construct-valid.



## **9.6 Actual meanings of alternative 6**

The ESE uses a six- alternative format of which five alternatives are explicitly given and the sixth one is implied. The meaning of alternative 6 is stated in the test rubrics as follows: "*If there is no good answer, write 6*". However, the main issue is that, by giving the opportunity to examinees to write alternative 6 in case there is no good answer, the test constructors assume that for some test items, there is not a correct alternative among the five suggested. Implicit in this assumption is this question: when do examinees actually write alternative 6: is it when they fail to identify the correct answer from the five alternatives or is it when they find that there is no good answer suggested in the five alternatives?

In this section, I wish to provide empirical evidence on the actual meanings of alternative 6 and evaluate to which extent this implicit alternative is construct-valid. In order to reach this objective, I have first computed the frequency of alternative 6 across the 56 test questions. Then after, in order to identify the actual meanings of alternative 6, I have used data from section 3 of strategies questionnaire where the 496 participants were asked to indicate, after whenever they

selected option 6, the reasons why they had selected this option (see Appendix 4). Table 9.6 presents the details on the frequency of alternative 6 as investigated in the 56 test items.

Table 9.6 Frequency of participants' selection of alternative 6 in the 56 test items

TEST	Test items where alternative was used	Total items
Test 1 [N=75]	T1.1 [1%]; T1.2 [3%]; T1.3 [3%]; T1.4 [13%]; T1.5 [20%]	5
Test 2 [N=64]	T2.2 [28%]; T2.4 [25%]; T2.5 [16%]; T2.6 [30%]; T2.8 [25%]	5
Test 3 [N=82]	T3.4 [7%]; T3.5 [13%]; T3.6 [16%]	3
Test 4 [N=57]	T4.1 [9%] ( <b>Key</b> ); T4.3 [18%]	2
Test 5 [N=107]	T5.1 [7%]; T5.5 [12]; T5.7 [20]	3
Test 6 [N=35]	T6.1 [17%]; T6.2 [34%]; T6.6 [14%]; T6.7 [9%]; T6.8 [9%]	5
Test 7 [N=48]	T7.1 [2%]; T7.2 [15%]; T7.4 [38%]	3
Test 8 [N=28]	T8.5 [25%]	1
<b>TOTAL [N=496]</b>		<b>27</b>

**T1.1=Test 1, item number 1; T2.2=Test 2, item number 2; T3.4=Test 3, item number; T8.5=Test 8, item number 5**

The information presented in Table 9.6 indicates that some participants selected alternative 6 in answering 27 of the 56 test items. Yet, this alternative was a good answer for only one test item (T4.1). The data indicate that alternative 6 was most selected (functioning distractor) in some test items (T7.4 [38%], T6.2 [34%], T2.6 [30%], T2.2 [28%], T2.4 [25%]; T2.8 [25%], T8.5 [25%], etc. ); but least selected (i.e. not functioning distractor) in some other test items [T1.1 [1%], T7.1 [2%], T1.2 [3%], T5.1 [7%], etc.). In T4.1 where the alternative 6 was a good answer, it was selected by only 9% of examinees. These findings suggest that although option 6 is an implicit option, examinees also selected it as correct answer.

However, the main issue with option 6 is to know whether or not the meaning expected by test developers when examinees select this option is the same meaning that the examinees actually have when they select this option. This issue is relevant in the scope of this thesis as it helps evaluate the relevance of this option in the context of the ESE. In the view of test constructors, option 6 should be selected when examinees find that there is **no** best option among the five suggested option. In view of this conceptualization of option 6, any other meanings provided to this option is construct-irrelevant, therefore, threatens the test validity.

In order to investigate this issue, I have used data obtained from participants' answers to the strategies questionnaire (see Appendix 4). In this questionnaire, section 3 included one question (Question 4) that requested participants to indicate, whenever they had selected option 6, the reason/s for their choice of this option. The question reads:

If you have selected option 6, tick all options that correspond to your choice of this option

- 1 I selected 6 because there was no best option among the suggested 5 options
- 2 I selected 6 because I could not find the correct answer
- 3 I selected 6 because there were more than one correct answer options
- 4 I selected 6 because I was confused with the other 5 options
- 5 I selected 6 because I sensed that it could also be part of the correct options
- 6 I selected 6 just by guessing

Table 9.7 presents information on the participants' reasons for selecting option 6 as investigated in the data.

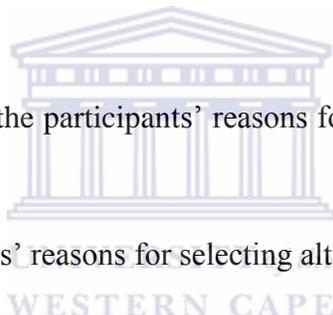


Table 9.7: Frequency of participants' reasons for selecting alternative 6

Reasons for selecting option 6		T1	T2	T3	T4	T5	T6	T7	T8	Tot.	%
1	I selected 6 because there was no best option among the suggested 5 options	0	7	4	3	7	0	11	2	34	16
2	I selected 6 because I could not find the correct answer	6	15	1	3	7	3	2	2	39	18
3	I selected 6 because there were more than one correct answer options	10	21	3	5	3	6	3	0	51	24
4	I selected 6 because I was confused with the other 5 options	4	13	10	0	13	3	2	4	49	23
5	I selected 6 because I sensed that it could also be part of the correct options	0	4	3	0	1	0	1	1	10	5
6	I selected 6 just by guessing	4	3	11	2	3	0	5	3	31	14

From the statistics in Table 9.7, the following observations are worth mentioning:

(1) There are six actual reasons why participants selected option 6; suggesting that participants had given six actual meanings to option 6.

(2) The primary meaning of option 6 “I selected 6 because there was no best option among the suggested 5 options” comes in the fourth position as the reason why participants selected this option. This is to suggest that the conceptual meaning of option 6 was overlooked by participants.

(3) The first frequently reported reason why participants selected option 6 was that they could find more than one correct answer. This reason accounts for 24% of all option 6 choices. This finding suggests that some participants were not capable of falsifying the five explicit options and come up with the “correct” option. Their falsification process might have compelled them to make the last choice; and this incapacity to make the last choice among remaining options may have given them the impression that all outstanding options were correct; therefore, the correct answer would be option 6.

(4) The second frequently reported reason why participants selected option 6 was that they were confused by the five explicit options. This reason accounts for 23 % of all option 6 choices. This finding suggests that some participants may have been unable to understand the expectation of the test question and the meaning of the different question alternatives.

(5) The third frequently reported reason why some participants chose alternative 6 as the correct answer was that they could not find the correct answer. This reason accounts for 18 percent of all option 6 choices. This finding suggests that some participants were not capable to find the correct answer; although there was one. This failure to identify the correct option may have given them the impression that the only choice they have is to choose option 6. This is some confusion in the understanding of option 6 which is used when **there is no** correct answer and not when an examinee **cannot find a correct answer**.

(6) Two less frequently reported reasons that account for the choice of alternative 6 are that some participants selected this option through simple guessing (14%) or they chose option 6 because they sensed that it could also be part of the correct options (5%).

The previous findings are relevant to this study. They suggest that the meaning of option 6 as articulated by test constructors is not the meaning option 6 actually has. Therefore, in the context of the ESE, this option appears to be construct-irrelevant since its meaning is not understood by some examinees.

### **Summary of the chapter**

In this chapter, I have presented and analyzed the data that sought to investigate the different item features that potentially impact on the difficulty of ESE tasks. The following findings have been reported:

- (1) The ESE uses stems that are both focused and unfocused although focused stems are more frequent than unfocused ones;
- (2) Focused stems were stems that included test questions that: (a) asked examinees to select the option that suits as the title of the text/paragraph; (b) asked examinees to skim the part of the text where a specific idea is located; (c) required the examinee to understand syntax;
- (3) Unfocused stems were those stems that included test questions that: (a) requested the examinees to search read the text in order to locate explicit pieces of information scattered through the whole text; and (b) required the examinees to scan the text in order to locate explicit information located in a specific part of the sentence or paragraph;
- (4) The majority of ESE stems from the sample were complete stems, while only few ESE stems were incomplete stems; moreover, most test questions had positive stems;
- (5) Although most test questions with negative stems were difficult for examinees, it was found that test questions with negative stems but that required reading at local level were easy for examinees;
- (6) Although in a minority, some items had their keys that potentially provided some clues to wise examinees;
- (7) The distribution of the ESE responses among the 6 alternatives/choices was not balanced. More specifically, choices 3, 1, and 4 were more frequently used than choices 2, 5 and 6.

Furthermore, the implied sixth choice (“If there is no good answer, write 6”) was a good answer for only 5% of all test questions;

(8) No item was found to have five functioning distractors as expected. On the contrary, most test items had either three, two or just one functioning distractors. Also, only 4% of items had four functioning distractors and 4 % of items had not any functioning distractors;

(9) Both guessing and elimination of distractors were two test taking strategies some examinees had used to answer some test questions, and these two test wiseness strategies were employed across the eight tests. However, participants more frequently used elimination of distractors than guessing;

(10) Some examinees used both guessing and elimination of distractors in answering the same test question, with elimination of distractors occurring before guessing;

(11) Guessing strategy was either “blind” or “educated”. Blind guessing occurred when examinees were not capable of understanding the expectation of the test question as well as the meaning of the different alternatives. On the contrary, “educated” guessing occurred when examinees could understand the expectation of the test question and the meaning of some item alternatives; but proceeded by eliminating any non-plausible distractors before making their last decision through guessing between the outstanding alternatives;

(12) Elimination of distractors was sometimes a highly construct-valid strategy as some examinees who used it showed their genuine understanding of the question expectation and the meanings of items alternatives;

(13) There were six actual reasons why examinees selected option 6; however, the primary meaning of option 6 “I selected 6 because there was no best option among the suggested 5 options” came only in the fourth position in terms of occurrence;

Having presented and analyzed the data in the previous four chapters, I wish to discuss the findings of my study in the next chapter.

## CHAPTER TEN

### DISCUSSION OF THE FINDINGS

#### 10.1 INTRODUCTION

The study aimed to explore the fundamental validity issues that can affect the DR Congo English state examination. The study aspired to generate an understanding of the potential issues that affect the construct validity of a test within the epistemological stance that supports a strong relationship between test construct and test context. In light of this aim, the study sought to answer the following main question: *“What are the main issues that potentially threaten the validity of the DR Congo English state examination?”* In my attempt to answer this main question, I addressed the following four sub-questions: (1) What is the actual context of the ESE and to what extent this context potentially influences examinees’ performance on the ESE tasks? (2) What is the variety and degree of complexity of the different reading types examinees conduct while completing the ESE tasks; and to what extent the ESE tasks are appropriate the ESE context? (3) What are the different textual features that affect the complexity of ESE tasks, and to what extent these textual features are appropriate to ESE actual context? (4) What are the item features that affect the validity of ESE?

In order to provide the relevant theoretical orientations to the study, I have discussed three theories: first, the validity theory that has provided a theoretical ground necessary for understanding the quality of tests needed for assessing students’ reading abilities. Secondly, the construction-integration theory that has provided an understanding of how texts that are used in reading assessments are processed and understood by readers; and finally the strategic competence theory that was meant to address how examinees deploy strategies to complete test tasks and the extent to which these strategies tap into the reading construct.

On the basis of insights drawn from the two processing models by Weir and Khalifa (2008a) and Embretson and Wetzel (1987), I have developed a multi-componential model that signposts the

social context of testing. In this model, I have hypothesized that, since cognitive processing always occurs within and is significantly affected by the test specific context, a valid reading test needs to include tasks that not only tap into the reading construct, but also those tasks that are appropriate to the specific context of reading activity.

In regard to the research design, I have used an exploratory descriptive design using both qualitative and quantitative data. Concerning the research methods, I have used protocol analysis and content analysis. While the former aimed at providing an understanding of the cognitive processes that mediate the reading construct and test performance so as to explore the different strategies examinees use to answer the test questions, the latter helped me to examine the content of the different English state examination papers and identify the different textual and item features that potentially impact on examinees' performance on ESE tasks.

The study main participants were 496 Grade 12 secondary school students who were in the final year of secondary school and who were preparing for the national test. As instruments, a strategies questionnaire and eight tests were used to elicit data from participants while a socio-contextual questionnaire was used to prompt additional information with teacher-participants. Depending on the kind of data to be used and whether the individual validity issue could be better described through content analysis or protocol analysis, or both, I have described each validity issue either qualitatively or quantitatively or both qualitatively and quantitatively. This triangulation of methods and data was meant to gather substantial evidence from diverse sources.

The presentation and analysis of the results have generated findings that are related to four domains of validity issues. The first domain has related to a description of the actual context of the ESE, and it has highlighted the kind of issues that can affect the validity of the ESE. The second domain has encompassed issues that pertain to the variety and degree of complexity of different reading types and processing levels targeted by the ESE tasks; and the extent to which the ESE tasks are appropriate to the test context. The third domain has included different issues that pertain to the textual features that impact on the complexity of ESE tasks; and the extent to which these textual features are appropriate to ESE context. Finally, the fourth domain has included issues that pertain to the different item features that impact on the complexity of ESE tasks. In the following section, I propose to discuss findings that relate to impact of the specific context of the ESE on examinees' performance on the ESE tasks.

## 10.2 The DR Congo ESE context and its expected impact on the validity of the ESE

The first objective of the study was to describe the actual context of the ESE as this context was expected to affect examinees' execution of test tasks. In light of this objective, I propose to summarize and discuss the findings relating to the ESE context and explain how this context can influence examinees' performance on the ESE tasks.

The analysis of data relating to the description of the ESE context (see Chapter 6) reveals seven main contextual issues that pertain to the English state examination and that are expected to affect examinees' performances on ESE tasks. These contextual issues relate to the (1) students' skills in English language; (2) students' motivation for reading at school and at home, (3) students' attitudes towards the ESE, (4) students' experience with the ESE, (5) paucity of reading materials, (6) availability of the curriculum and its actual use by English teachers, and (7) teachers' in-service training.

The first contextual issue that can influence examinees' performance on the ESE tasks relates to students' skills in English. The study findings reveal that the majority of participants reported that their abilities in English language in general, and in reading in English in particular, were poor. This was reflected in their poor capacity to cope with basic language skills (listening, speaking, reading and writing); the confusion they had to match the English spelling and pronunciation as compared to French language, and some attitudes that supports that English is a difficult language one cannot easily learn. This finding is relevant in the scope of this study. If we agree that validity is about the inferences and decisions that are made on the basis of test scores (Messick, 1989), in order to be construct-valid, we must ensure that a reading test includes test tasks that take account of examinees' skills in English in general, and in reading in particular. This is to suggest that, the ESE needs to include tasks whose complexity level is proportional to the examinees' abilities. This suggestion is consonant with the concept of validity as "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (AERA, APA, & NCME, 1985, p. 9). It also reflects the context of English as a foreign language that is learnt *only* at school; and that is characterized by students' limited exposure to English language. In such a limited language learning environment, students are generally expected to be poor readers who encounter different language problems. This view is supported by evidence that suggests that poor readers have problems to comprehend texts because of their

inability to recognize and encode words and to parse sentences (Nassaji, 2003; Barnett, 1986). They also have problems to make inferences and identifying causal structures in the text (Hammadou, 1991; Lu, 1999). In the third and four sections of this chapter (see Sections 10.3 and 10.4) where I propose to discuss the validity issues that pertain to reading types and processing levels required by test tasks, as well as the validity issues that pertain to ESE texts, I will evaluate the extent to which the ESE tasks included in the study sample were relevant to the participants' skills in English language.

The second contextual variable that appears to influence examinees' execution of ESE tasks is the students' motivation for reading. The study findings reveal a low motivation for reading that was reflected in a low students' reading amount at school and at home, and students' low attendance to English classes as well as their preference to learn English in lesser time than the time planned for learning English. In light of this finding, I wish to highlight the following three observations: First, students with low motivation may have problems to successfully read test materials, especially if these materials are less appropriate. This observation appears to chime in Mucherah and Yoder (2008)'s finding when they reported that students who had high motivation in their reading successfully read challenging materials, while those students who were not highly motivated had difficulty approaching the same materials. Secondly, students' low motivation for reading may result (and sometimes is the result of) in low reading amount at school and at home. This relationship was established by Lau (2009) who found that students' motivation was mostly strongly related to their reading amount; suggesting that the more one reads the more motivated one can become. Lastly, students' participation in reading classes appears to be associated with their reading motivation; an association suggested in a study by Lau (2009) who reported that, those students with negative motivation for reading were found to be reluctant to participate to reading classes.

In light of the preceding findings on reading motivation and building on the main assumption of the strategic competence theory that has provided theoretical orientations to this study, I wish to state that, since reading is a strategic activity where a reader uses a variety of strategies in order to construct meaning, readers' motivation for reading can influence their selection of reading and test taking strategies and therefore, impact on reading performance. In light of this statement, I believe that, in order to be construct-valid, the ESE needs to include tasks that also profile

students' low reading amount, less frequent participation in reading classes and low reading motivation. In the third and fourth sections of this chapter (see Sections 10.3 and 10.4), I will evaluate the extent to which the ESE tasks included in the study sample were also relevant to participants' low reading amount, less frequent participation in reading classes and low reading motivation.

The third contextual issue that can affect examinees' performances on the ESE tasks relates to the students' attitudes towards the ESE. The study results reveal that, although the majority of participants had positive attitudes towards the ESE, some of them had negative attitudes towards the ESE. Insights from qualitative reports of the participants have indicated that some participants felt and/or believed that they must not write the ESE as this test is perceived to be a way of failing them on the national test. Some other participants reported that the ESE is not a fair test since it is based only on text passage comprehension by neglecting other language skills. Furthermore, information gathered from the contextual questionnaire suggested some negative attitudes as reflected in the participants' preferences for an extended time to write the ESE, their preferences for a reduced number of questions, number of alternatives and number of alternate forms to be included in the ESE.

More specifically, the study findings indicate that many participants preferred the actual time for writing the test to be increased, a view that does not concur with a suggestion by Katalayi (2011). One main reason for participants' preference for an extended time was that an extended time might offer them an opportunity for assistance and collaboration as this could increase their chance to pass the ESE. The study findings also indicate that the majority of participants preferred the actual number of questions (9 in general subjects and 5 in technical subjects) to be reduced, or at least to be maintained; but not to be increased. One main reason for the participants' preference was that fewer questions in the English test were not likely to negatively affect their general performance on the national test. Furthermore, the majority of participants preferred the actual number of six alternatives to be reduced to three alternatives as this could enable them to have a better chance of getting some test questions right by simply guessing. The study findings also confirm that the majority of the participants preferred the actual number of four alternate forms to be reduced to a single form test. Among the reasons for this preference, some participants believed that the use of multiple alternate forms neither reflects the classroom

assessment where all the students write the same test, nor does it reflect fairness. These findings relating to the students' attitudes towards the ESE are relevant in the scope of this study. As evidenced by research, positive attitudes towards the test can produce beneficial effects while negative attitudes can erode confidence and potentially impact negatively on performance (Murray, Riazi & Cross, 2012; Rasti, 2009, Han, Dai & Yang, 2004). These findings suggest that, in developing the ESE, test developer need to be sensitive to students' attitudes towards the ESE. Since the study findings have signposted, as a consequence of negative attitudes towards the ESE, students' intention to rely on test wiseness strategies (guessing) and some unethical practices (collaboration, cheating) while writing the ESE, it is important that the ESE includes tasks that also profile students' attitudes towards the test so as to reflect their actual reading abilities.

The fourth contextual issue that can affect examinees' performances on ESE tasks relates to their experience with the ESE. One way to familiarize the students with the ESE is the teacher's frequent use of multiple-choice testing method in classroom assessments as such a familiarity potentially provides them with some experience with the ESE. Evidence from the study suggests that, although the multiple-choice method is the ESE method, the majority of teachers used the open questions as the main classroom assessment method on the ground that the multiple-choice method was difficult to construct or it could not properly evaluate the students' language ability. Yet, evidence suggests that, test familiarity is a factor that can impact on examinees' scores (Reeve *et al.*, 2009) as Peña and Quinn (1997) found that learners performed significantly better on the familiar test tasks, and that the familiar tasks were more sensitive in differentiating learners who were developing from those who were struggling.

The fifth contextual issue that potentially affects examinees' performances on the ESE pertains to the paucity of reading materials. The study findings indicate that reading instruction is severely hampered by an abject paucity of reading materials for reading at school and at home. Besides, it is also hampered by the use reading materials of poor quality. The study findings also suggest that the majority of English teachers do not have the English curriculum as part of their teaching documents, and/or they do not use the curriculum in the planning of the English course. Also, the English curriculum was found to be old-fashioned; and the teachers urged for it to be updated. This issue of abject paucity of reading materials in the context of reading in a foreign

language appears to be a critical issue in the ESE context. Since reading involves a reader and a text, the presence of the text and its quality are critical to reading in general and reading performance in particular. This suggestion appears to concur with Greaney' (1996) and Lonsdale' (1997) findings when they reported a positive relationship between textbooks availability and student achievement in reading and student performance on reading tasks.

The poor quality of reading materials as revealed by the study findings suggests that students are exposed to reading materials that appear to be less relevant to their needs. Evidence from the data suggests that students' lack of motivation for reading also comes from the content of texts they read in classroom reading instruction. The socio-economic crisis of DR Congo, the budgetary restraints by lending agencies (such as the World Bank, the International Monetary Funds), the possible lack of skills at local level and the absence of outstanding forms of independent publishing infrastructure have resulted in relying on Belgium editors and publishers for the design and supply of English textbooks. Yet, evidence seems to have established the danger of cultural imperialism and continued economic dependency generated by such reliance (Altbach & Kelly, 1988; Scrase, 1992, Crossley & Murby, 1994). Furthermore, evidence also suggests that reading resources of good quality can impact on students' reading, their motivation for reading and their performance on reading tests. This suggestion appears to chime with Lewin and Stuart's (2003) and Bukenya' s (2005) findings when they argue that textbooks of good quality can counter-balance the problems of poorly trained teachers and lack of basic facilities in schools.

The sixth contextual issue that appears to affect examinees' performances on the ESE tasks relates to the availability of the curriculum and its actual use by the English teachers. Since the curriculum is a teaching resource that indicates what should be learned and how the learning should be organized for teaching (Petrina, 2004), its availability and quality need to be considered while developing the ESE. The study findings indicate that, some teachers do not have the English curriculum as part of their teaching documents; therefore, they do not use this document in the design and planning of reading course. Furthermore, the study findings reveal that the English curriculum is old-fashioned and therefore needs to be updated. These two findings raise the issue of the availability and quality of the curriculum; therefore, they prompt the question of the content of materials used in teaching English in general and reading in

particular. This is to suggest that the English language instruction, in general and reading instruction, in particular may be provided without little/no planning. Furthermore, the study findings reveal that all the topics foreseen in the curriculum are never completed in the course of the academic year. This finding suggests that there seems to be a gap between what is planned and what is actually taught. In light of these findings relating to the English curriculum, I can state that, if the ESE constructors are not mindful of the aforementioned issues while developing the ESE, the tests that they develop can be construct-irrelevant as these tests will hardly reflect students' actual reading abilities because they appear not to tap into the actual ESE context.

The last contextual variable that appears to affect the examinees' performances on the ESE tasks relates to teachers' in-service training. There is consensus that in-service training programs for the teachers tend to increase the qualities possessed by good teachers; hence these programs positively affect teachers' performance (Harris & Sass, 2001). However, the study findings reveal a total absence of in-service trainings as well as continuous trainings for English teachers. Consequently, teachers are not able to keep abreast of new developments in teaching; besides, there is a substantial degree of confusion in their use of teaching methods. This finding is relevant in the scope of this study as evidence suggests a positive relationship between teachers' in-service training and teacher productivity (Harris & Sass, 2001), a positive relationship between in-service training and change in behaviors in both classroom and administrative work (Samupwa, 2008), a positive relationship between in-service training and teacher perceptions of the teaching profession, or again a positive relationship between in-service training and positive behavior (personality, knowledge, and commitment skills) on the part of the teachers (Jahangir, Saheen, & Kazmi, 2012; Conco, 2004).

To conclude this section, I wish to state that all the aforementioned contextual issues are expected to exert some influence on the examinees' completion of ESE test tasks. Ignoring them while developing the ESE appears to be a dangerous assessment practice that can be detrimental to the whole educational system. In light of this position, the validity of the ESE depends on the extent to which the test includes test tasks that, not only target the reading construct, but also tap into the reading context. Therefore, I am curious to find, in the next section, the extent to which the ESE tasks are sensitive to the ESE context.

### **10.3 Validity issues that pertain to the variety and degree of complexity of different reading types and processing levels targeted by the ESE tasks**

The second objective of this study was to investigate the variety and degree of complexity of different reading types and processing levels targeted by the ESE tasks; and evaluate the extent to which the ESE tasks are appropriate to the ESE context. In this section, I propose to summarize the fundamental validity issues that relate to this objective.

The study findings reveal three main validity issues: (1) the issue of preponderance of test questions that target careful reading over test questions that target expeditious reading; (2) the issue of preponderance of test questions that target higher level processing over test questions that target lower level processing; and (3), the issue of content coverage.

Concerning the first issue relating to a predominance of test questions that target careful reading over test question that target expeditious reading, the study findings reveal that, although the 56 test items included in the study sample covered all the four reading types, there were more test items that targeted careful reading than test questions that targeted expeditious reading. This finding suggests that ESE test constructors put much emphasis on careful reading than on expeditious reading. This finding appears to concur with previous findings revealed in two studies by Weir *et al.* (2008a) when they found that 77 % of strategies used by participants to complete IELTS test tasks tapped into careful reading while only 23 % of strategies tapped into expeditious reading; and Weir *et al.* (2008b) when they found that the major focus of the IELTS test papers appeared to be on careful reading although most participants used expeditious reading skills and strategies in answering test questions. This preponderance of careful reading tasks over expeditious reading tasks was also reported by Sarojani and Krishnan (2011)'s replicative study on IELTS when they found that the majority of IELTS test items tested careful reading although the majority of participants perceived expeditious reading skills as more relevant and appropriate to their academic reading activity.

In light of the above study finding and the findings from existing literature, one main question that deserves to be asked is whether a reading test must include more tasks that request examinees to conduct careful reading than tasks that require examinees to conduct expeditious reading. In order to answer this question, I hasten to state that since many existing reading

models (for instance Bernhardt, 1991; Hoover & Tunmer, 1993; Rayner & Pollatsek, 1989) are nearly premised on a careful reading model and do not take sufficient account of the different purposes of reading, test developers appear to be influenced by these conceptual models as they emphasize on test tasks that target careful reading. However, recent developments in reading theory (Khalifa & Weir, 2009) suggests that careful reading alone may be an inadequate construct for students as expeditious reading may tell us how readers can cope with other common life reading behaviours such as skimming or search reading.

In light of my conceptualization of validity as appropriateness of test tasks, I am inclined to hypothesize that the validity of a reading test depends on the extent to which the test includes tasks that elicit skills and strategies similar to those readers deploy in non-testing conditions. Considering this argument, the study findings reveal that careful reading test items appear to require a frequent combination of strategies as the examinee attempts to extract complete meanings within or beyond sentences – right up to the level of the entire text so as to construct the text macrostructure. On the contrary, expeditious reading test items appear to require less frequent combination of strategies as they involve a quick, selective and efficient reading to access needed text information. Viewed from this angle of validity as appropriateness of test tasks, in the ESE context, a valid reading test would include more tasks that target expeditious reading than those tasks that target careful reading as the former tasks appear to reflect the way readers read in non-testing conditions.

Since validity is an argument that consists in gathering, synthesizing, and summarizing various sources of evidence while validating a test (Borsboom, Mellenbergh, & Van Heerden, 2004), I wish to validate my previous claim by triangulating it with the results from the task complexity as reflected in the item difficulty. The study findings in Table 7.4 (see page 238) indicate that test questions that targeted careful reading were more difficult (.5) than those that targeted expeditious reading (.6). This is to suggest that careful reading tasks appeared to be more cognitively demanding for examinees than expeditious reading tasks. Therefore, considering the ESE actual context as described in the preceding section, I am inclined to conclude that, by including more tasks that target careful reading over tasks that target expeditious reading, the ESE appears to be less appropriate as it does not closely reflect the test actual context.

As regards the second issue, the study findings reveal a predominance of tasks that require higher level processing over those tasks that require lower level processing. Yet, the study findings reveal that higher level questions were more cognitively demanding for participants than lower level questions. More specifically, while lower level questions mainly required participants to process the text at local level with recourse to their knowledge of grammar and/or vocabulary, higher level questions mainly required participants to process the text at global level by attempting to generate more connections between their knowledge and text information. Furthermore, in order to answer lower level questions, participants generally considered the five options and focused on a familiar option, or they produced their own answers after reading the questions and then looked at the five options to confirm their choices. On the contrary, in order to answer higher level questions, the participants usually had to resort to elimination of options and guessing strategies. Yet, evidence suggests that elimination of options and guessing are two test wiseness strategies that the examinees use as a short-cut to arriving at answers (Cohen & Upton, 2006, 2007; Hirano, 2007; Rupp, Ferne & Choi, 2006). Therefore, the use of these two strategies in the context of this study may *sometimes* suggest that the process of answering higher level questions may have *sometimes* been a problem-solving process instead of being a reading process.

In light of these findings, I am inclined to state that the focus on higher level questions does not appear to reflect the actual ESE context described previously (see Section 10.2). Therefore, the ESE fails to take account of the actual characteristics of the students and the conditions under which reading instruction occurs. Consistent with this statement, I can conclude that the ESE appears to be construct-irrelevant as it includes tasks that are less appropriate to the test context. This conclusion is relevant to this study as the understanding of the concept of construct-irrelevant validity may provide some hints for evaluating the kinds of tasks that can be included in the ESE in terms of their relevance, appropriateness and usefulness to the test context.

The third issue relates to the content coverage of the ESE items. The process-oriented model of reading I have proposed in this study hypothesizes that the intensity and quality of reading comprehension processes examinees engage in can considerably vary with different levels of reading comprehension assessed by individual test items. In light of this hypothesis, the model suggests that, in order to fully cover the reading construct, the test needs to include tasks that

require examinees to process the text at the following six hierarchical levels: lexical encoding, syntactic parsing, establishing propositional meaning, inferencing, building a text mental model, and creating a text level representation. This classification not only enables to explore the degree of complexity of the ESE tasks in relation to examinees' characteristics, but it also provides a picture of the extent to which the ESE tasks appropriately cover the reading construct.

Regarding this issue of content coverage, the study findings indicate that the ESE test items targeted five of the six processing levels. Besides, some levels included more test items (level 6: creating text level structure and level 2: syntactic parsing) than other levels (level 1: lexical encoding, level 4: inferencing, and level 5: building the text mental model). This finding suggests that the test items included in the study sample partially covered the reading construct as no test questions targeted level 3 (propositional meaning). Yet, this processing level is as critical as the other five levels as it can help to assess examinees' abilities to establish propositional meaning by adding external knowledge to text propositions in order to turn them into a message that relates to the context in which these propositions occur.

The issue of the ESE not covering the entire reading construct is one of the most crucial validity threats as investigated in this study (see Chapter 7, point 7.4.2). This issue has been conceptualized in validity theory as "construct-underrepresentation", and it occurs when the test is too narrow and fails to include important dimensions or aspects of the construct (Messick, 1989; Young, 2008). However, since validity is viewed as an argument that relates theory, predicted relationships and empirical evidence to suggested score meanings and interpretations (Kane, 2002, 2004, 2006, 2011; Downing, 2003), a reading test that does not cover the whole range of reading construct may fail to capture critical evidence related to the construct being assessed. This observation is particularly relevant in the scope of this study as a 9 or 5 item test cannot cover the whole range of reading construct so as to reflect examinees' comprehension of the text. This is to suggest that, the reading construct is expected to be reflected in the use of tasks that cover the test domain. As Bakker *et al.* (2008) argue, the selection of just a part of the relevant evidence and arguments hardly reflects examinees' true abilities as many other critical aspects of the construct are missed. In light of this observation, the study findings reveal that the majority of test questions included in the study sample focused on the general aspects of reading construct while relevant detailed aspects were simply missed. This is the case of most test

questions based on the five narratives where it was found that the questions that could target the specific aspects of the narrative structure were simply missed. More specifically, it was found that the ESE based on narrative texts did not include any questions relating to the identification and description of time and spatial settings, the description and comparison of characters' appearances, personality and emotions, their motives and social relationships, the description and judgment of characters' actions to solve the problem, and the identification and personal judgment of the outcome of the story. If the test could have more questions than the actual number of 9 or 5 questions, these critical aspects would probably have been included so as to appropriately measure the whole range of reading construct.

#### **10.4 Validity issues that pertain to the ESE text structure**

The third objective of the study was to explore different textual features that potentially impact on the difficulty of ESE tasks, and to evaluate the extent to which the ESE includes texts whose features are appropriate to the ESE context. In light of this objective, the present section presents the validity issues that pertain to the structure of the ESE texts.

The study findings reveal four broad categories of validity issues that pertain to the structure of ESE texts: (1) the issues that pertain to text familiarity and difficulty; (2) the issues that pertain to the use of narrative texts; (3) the issues that pertain to the use of expository texts; and (4) the issues that relate to the density of ESE texts.

As regards the first issue relating to text content familiarity, the study findings reveal that, my evaluation of text familiarity did not necessarily match with the participants' perceptions of text familiarity. Although there was evidence of some texts whose contents were not familiar to participants, I classified most ESE texts as either familiar or somewhat familiar. On the contrary, only few participants perceived the ESE text contents to be familiar whereas the majority of participants perceived the texts to be either somewhat familiar or not familiar. Likewise, my evaluation of text content difficulty did not necessarily match with the participants' perceptions of text content difficulty. According to my evaluation, most texts included in the study sample were either easy or of moderate difficulty. On the contrary, only the minority of participants

perceived the text content to be easy, while the majority of participants perceived the text content to be either difficult or of moderate difficulty.

This finding is relevant to this study. It provides evidence for the benefits of triangulation of results in validation studies. The triangulation of my evaluation of text content familiarity and difficulty from content analysis methodology and the participants' perceptions of text content familiarity and difficulty aligns with one epistemological orientation of validity theory that suggests that test validation must combine the logical argument and empirical evidence needed to support inferences (Landy, 1986 cited by Shepard, 1993). This is to suggest that, in validating a test, various sources of evidence must be gathered, synthesized, and summarized (Borsboom, Mellenbergh & Van Heerden, 2004). In light of this epistemological orientation, my evaluation is based on a logical argument developed from my extensive readings on validity theory coupled with my long experience as an English teacher in the setting where the study is conducted. Therefore, my evaluation embodies some elements of subjectivity that need to be cross-validated with participants' perceptions that, to some extent, constitute empirical evidence that reflects how they actually perceived text content familiarity and difficulty. Consistent with this view, the validity of the conclusions that are formulated on the issue of text content familiarity and difficulty depend on the extent to which the relationship between my personal evaluation and participants' perceptions is plausible and reasonable. This task is the concern of my readership.

In light of the preceding methodological scope, the study findings reveal that text passages participants perceived to be familiar were those texts whose content addressed current issues that affected their immediate environment; or those texts that addressed themes that related to their subject knowledge. On the other hand, text passages participants perceived to be somewhat familiar where essentially those texts that depicted a social situation relevant to their background but with details that described a setting that was not familiar to them; therefore, making it difficult to visualize the setting. On the contrary, texts that participants perceived not familiar were essentially those texts that were not appropriate to the kind of text contents that the participants were familiar with in learning situation, and whose text details did not relate to their social backgrounds. This finding suggests that when readers read materials that relate to situations that are familiar to them or to their subject knowledge, or subject discipline, they more easily comprehend these materials than when they read materials that are not related to their

background/subject knowledge. This finding appears to replicate similar findings reported by Hock (1990), Yousif and Shumaimeri (2006) and Chang (2006).

Concerning the text content difficulty, the study findings reveal that ESE texts participants perceived to be difficult were those texts whose contents were not appropriate to the kind of text participants were familiar with in learning situation, and that text details did not relate to participants' social background. Therefore, the majority of participants found it difficult to encode some words and phrases as they failed to activate relevant schemata necessary to use clues provided by the context to understand key words and phrases. This finding suggests that the vocabulary of a text exerts influence on reader's capacity to comprehend the text as evidence suggests that L2 reading requires much lexical-conceptual processing (Ulijn, 1984; Brisbois, 1995; Yamachita, 1999); a finding that suggests that texts with vocabulary that is not familiar to readers are potentially difficult to be comprehended. This finding also aligns with one of the assumptions of the present study that suggests that text comprehension also depends on the reader's capacity to encode words.

Concerning the second validity issue that pertains to the use of narrative texts in the ESE, I wish to state that, since validity is about the appropriateness and relevance of inferences and decisions made on the basis of test scores, the validity of a reading test based on narratives depends on the extent to which the test questions tap into the structural organization of the narrative. In light of this statement, the study findings reveal two validity issues: first, the issue of the nature of information targeted by the test items, and then the issue of the order of test items. Concerning the issue of information targeted by test items, the study findings reveal that, although the majority of ESE test questions tapped into the narrative structure, some test questions targeted text information that was inappropriate to the narrative structure. As to the issue of the ordering of test questions, the study findings indicate that the ordering of information in the text did not relate to the ordering of information in the test questions. These two findings suggest that the ESE based on narrative texts may have low validity. These findings have particular importance in the scope of this study. The literature on reading comprehension suggests that, in processing a narrative, a reader attempts to create a micro-world of the story by constructing some representations of the story characters, events, states, goals, and actions (Zwaan, Langston, & Graesser, 1995). Therefore, in order to be appropriate, scores on a reading test based on

narratives need to provide an indication of examinees' abilities to read the narrative and reconstruct its story line. Consistent with this claim, the test needs to include tasks that require examinees' skills to reconstruct the narrative story line with a focus on the chronological sequencing as well as the spatial organization of story events (identify/describe the narrative setting, to identify the main character and auxiliary characters, to analyze the overall plot through the identification of the story problem, the response, the action and the outcome or resolution).

Another related issue concerning the nature of information targeted by test items relates to the reading types requested by ESE test items based on narratives as well as the processing levels required by these test items. Since literature (Mandler & Johnson, 1977; Kintsch, Mandel, & Kozminsky, 1977; Dymock, 2007; Thorndyke, 1977) suggests that narratives are mainly based on details and descriptions, appropriate questions on the narratives will then be those that require examinees to read at local level (both carefully and expeditiously) in order to locate needed pieces of information. Yet, the present study findings reveal that the majority of test questions based on narratives requested participants' skills to process the text at global level by trying to build the text macro-structure. In light of this latter finding, the ESE based on narrative texts may be less appropriate to the test expectations as the skills required by test questions to complete test tasks are more demanding than expected.

The second validity issue pertaining to narrative texts relates to the order of test questions. Research indicates that narratives have a hierarchical structure, and that readers are sensitive to this structure so that when it is used to guide comprehension and recall, both are facilitated and comprehension is enhanced (Shin, 2002; Glenn, 1978; Mandler, 1978; Carroll, 1985). Consistent with this insight, the validity of reading tests using narrative texts will also depend on the extent to which the test questions reflect the story line used by the writer to construct the story. At this juncture, I hypothesize that, since the examinee's knowledge of narrative story line can guide him/her to easily comprehend the text, and that this comprehension can facilitate his/her performance on the test, test questions that are ordered in such a way that they follow the narrative story line are more appropriate than those that are not ordered in that way. In light of this hypothesis, the study findings reveal that the majority of test questions from the five sampled narratives were not ordered in such a way as to reflect the narrative story line. Also, the item difficulty statistics indicated that the majority of test questions that did not follow the narrative

story line were difficult for the examinees. This finding appears to confirm previous findings by Sheehan and Ginther (2001), Freedle and Kostin (1992), Kintsch (1994, 1998), Shin (2002), and Gorin (2005). Sheehan and Ginther (2001) found that the location of relevant information within a passage was associated with the difficulty of MC questions on the TOEFL- 2000 reading test. On their part, Freedle and Kostin (1992) found that an individual's expectations regarding the location of relevant information in the text affected the activation of information previously stored in the knowledge representation. On his part, Kintsch (1998) reported that the order in which information was encoded related to the manner in which it was built into the knowledge representation. Further to this, Shin (2002) found that examinees poorly performed on most items that could not follow the story line of the text as such questions did not call for operations examinees were likely to use to construct text meaning. On their part, Kintsch (1994) and Gorin (2005) found that changing the order in which the information in the text is presented impacted on the construction of coherent text representation, and this was found to increase the difficulty of test questions pertaining to the newly ordered information.

These findings are relevant to this study. They suggest an urge for ESE constructors to design tests that include tasks that are adjusted to narrative structure and the underlined cognitive demands appropriate to narrative texts. By failing to design tests that are sensitive to narrative structure, scores on such tests hardly provide an indication of examinees' abilities to read the narrative and reconstruct its story line. Therefore, this finding prompts the issue of construct-irrelevant variance; suggesting that examinees' scores may be inflated with some variance due to factors that bring irrelevant difficulty/easiness to examinees to complete test tasks. This conceptualization of construct-irrelevant variance goes beyond the actual conceptualization of construct-irrelevant variance. This suggested conceptualization of construct-irrelevant variance stresses the dimension of the appropriateness of cognitive operations necessary for completing test tasks, a view that current scholars in reading assessment research seem to have so far missed. Although I will elaborate on this insight in the Conclusion Chapter, I wish to state, at this juncture, that a reading test is construct-irrelevant when its scores (also) include variance due to examinees' use of cognitive operations that are not relevant to the kind of operations necessary to construct the text macrostructure.

Another validity issue that pertains to the narrative text relates to the use of narrative texts with specific spatial settings. In the context of the ESE as a national test administered to students in the whole country, the use of narrative texts with specific spatial details might be difficult for the majority of examinees who are not familiar with that specific setting. This finding is valuable as it can help interrogate the validity of reading tests that include texts that depict a specific town setting as there may be a serious problem for students living in other towns or rural areas to make a mental model of such texts. Therefore, such reading tests may be biased as they may advantage candidates living or who have lived in this specific town over those who have never lived at this place. This observation finds support in comprehension theories that suggest that, when a reader processes a narrative, he/she attempts to create a micro-world of the story through some representations of the story setting, characters, events, etc. (Mandler & Johnson, 1977; Kintsch, Mandel, & Kozminsky, 1977; Dymock, 2007; Thorndyke, 1977; Zwaan, Langston, & Graesser, 1995).

The last validity issue that pertains to the use of narrative texts in the ESE relates to the use of transitions and linguistic signaling devices in order to achieve coherence. Since the five narratives included in the study sample used time sequence, cause-consequence, problem-solution and description as discourse modes (see Table 8.2 on page 265), the narratives must be written in such a way that coherence relations are made explicit by linguistic markers in order to connect different text segments, such as cause-consequence, problem-solution, and chronological order. Therefore, the use of transitions and other linguistic signaling devices can help to construct these coherent representations. Regarding this issue, the study findings reveal that, in all the five narratives, interclausal relationships were expressed simply through a juxtaposition of clauses, and no explicit marking was used. Also, the linguistic devices and conjunctions were not used. Yet, as Crosson (2008) argues, these devices act as guiding cues that can assist the readers to understand how ideas in one clause relate to the ideas in adjacent clauses. Crosson's argument appear to be consistent with existing literature on text comprehension that indicates that an understanding of interclausal relationships in the text is central to reading comprehension (Degand & Sanders, 2002; Graesser *et al.*, 2004; Van Dijk and Kintsch, 1983).

In view of the previous finding, I wish to state that by using narratives that are truncated, the ESE developers appear to confound linguistic complexity and text readability. Although the use

of simple and/or simplified linguistic structures aims to make the text readable so as to reflect the ESE context, syntactic simplicity may decrease text coherence and cohesion; hence, examinees can have difficulty to comprehend the text. Furthermore, the non-use of time sequence devices that can provide the reader with the evolution of the story can be detrimental to some readers as they can fail to grasp the story line; therefore, they can fail to comprehend the text. However, since the English Curriculum has planned the teaching of some of these transitions (see Programme National d'Anglais, 1988), it appears relevant to design texts that include time sequence devices to create coherence and guide the examinees in their attempts to construct the text macrostructure. There is no doubt that these devices act as guiding cues that can assist readers to construct the hierarchical structure of the story necessary for comprehending the text.

Regarding the second validity issue that pertains to the use of expository texts in ESE, I wish to state that, expository texts primarily aim to inform the readers, and they use description as the basic mode of exposition. Consistent with this statement, in order to be construct-valid, reading tests based on expository texts need to include mainly those questions that request the examinees' ability to process the text at both global level and local level in order to get the text gist as well as to identify specific pieces of information located at global and local levels. Viewed from this text processing perspective, tests questions must require examinees' capacity to demonstrate how well they have been able to visualize text details through the identification and understanding of descriptors.

However, the study findings reveal that, although the ESE test questions based on expository texts actually requested examinees' ability to process the text at both global level and local level in order to get the text gist as well as to locate specific pieces of information located at both global and local levels, there were no test questions that requested examinees' capacity to identify and understand the different descriptors necessary to visualize text details. This finding suggests that ESE constructors seem to be aware of the need to base the ESE based on expository texts around gist construction through reading at both global and local levels. However, they seem to be unaware of the need to include test tasks that feature the description of text details. Yet, since literature suggests that the content and linguistic features of expository text are considered more challenging for readers and that the structural details of expository text can guide readers to understand the text (Akhondi, Malayeri, & Samad, 2011; Snyder, 2010), I

believe that basing a reading test on those structural elements can enable the test developer to ensure that examinees' scores on the test reflect their comprehension of the text.

The third and last validity issue pertaining to the structure of the ESE texts relates to the density of ESE texts. In text processing literature (Kintsch, 1988, 1994, 1998; van Dijk & Kintsch, 1978), text density is usually associated to text difficulty and task complexity. This is to suggest that, readers are expected to have more challenges to comprehend dense texts than less dense texts. In light of this observation, the study findings reveal that the ESE texts were less dense than classroom texts as classroom texts were three times longer than ESE texts.

I wish to remind my readership that my epistemological stance suggests to address the issue of construct validity in the broader context of reading activity. In light of this epistemological stance, the quality of ESE texts depends on the extent to which their density is adjusted to examinees' characteristics. This observation aligns with my view of validity as appropriateness of test tasks to the actual test context. Therefore, since ESE texts are much less dense than texts students actually read in classroom instruction, I conclude that ESE texts are less appropriate to the ESE context. Implicit in this conclusion, ESE scores may hardly reflect the examinees' actual text comprehension in non-testing situations.

### **10.5 Validity issues that pertain to the ESE item structure**

The fourth study objective was to explore the item features that can affect the validity of ESE tasks. In light of this objective, the present section presents the validity issues that pertain to the structure of the ESE test questions.

The study findings reveal that five main issues appear to affect the validity of the ESE. These issues pertain to (1) the use of negative stems, (2) the use of some keys that potentially provide some clues to the correct answer, (3) the use of non-functioning distractors, (4) the examinees' use of guessing and elimination of distractors as strategies to answer test questions, and (5) the multiple meanings of the implicit alternative 6.

Concerning the first issue relating to the use of negative stems, the study results reveal that, although most test questions included in the study sample had positive stems; few test questions were worded negatively. This finding suggests that the ESE constructors may be aware of the benefit of using positive stems as they avoid as much as possible negative stems and this can contribute to writing test items with some quality. However, in the context of the ESE as described previously (see Section 10.2), it would be better to use only positive stems given that negatively worded stems are more difficult to comprehend than positively worded stems (Haladyna, Downing, & Rodriguez, 2002; Ascalon *et al.*, 2007; Tamir, 1993).

However, unlike findings from existing literature (Haladyna, Downing, & Rodriguez, 2002; Ascalon *et al.*, 2007), the present study findings reveal that the majority of test questions with negatively worded stems were easy as reflected by their item difficulty. This inconsistency of findings can be explained by the fact that all these test questions with negative stems (see Section 9.1.3) required reading at local level; suggesting that they were less cognitively demanding. This explanation appears to be consistent with previous findings by van Steensel, Oostdam and van Gelderen (2012), Song (2008) and Rouet *et al.* (2001) when they reported that, test questions requiring examinees to process the text at a lower level can be easy as these questions require examinees to read the text at local level by focusing on text micro-propositions and by quickly searching for text information through a browsing of some parts of the text. Therefore, in light of the finding reported in this study, and from research insights, I can suggest that stem orientation appears to interact with the processing level required by test tasks.

As regards the second validity issue that pertains to the use of some keys that potentially provide some clues to the correct answer, the study findings reveal that, although in a minority, some test items had their keys that potentially provided some clues to test wise examinees. The study findings also reveal that, although participants generally perceived these questions as difficult questions, test results indicated that these test questions were easy. This finding suggests that participants may have used clues provided in the key to get the items right. Therefore, they may not have understood the questions' expectations and the meanings of all the six alternatives.

Concerning the third issue relating to the plausibility of distractors, the study findings reveal that, no test item was found to have five functioning distractors as expected. On the contrary, most

test items had either three, two or just one functioning distractors. Also, only a negligible number of test items had four functioning distractors. The study findings indicate that, although very few, some test items had not any functioning distractors. This finding raises the issue of the number of alternatives to be included in the ESE. It suggests that the ESE constructors may not be capable of constructing a test with five functioning distractors. Yet, evidence suggests that the different options to a multiple-choice item are good if they are equally plausible to examinees who have not mastered the content tested (Lin, Chu, & Meng, 2010). This issue of plausibility of distractors is critical to many high-stake multiple-choice tests. In a study of a large scale test of reading comprehension with 200 multiple-choice items, Haladyna and Downing (1988) found that only 5.5 percent of test items had four functioning distractors; a finding that appear to be consistent with Haladyna and Downing's (1993) finding when they reported that the number of functioning distractors per item was only about one and that only 1 to 8 per cent of items had two or three functioning distractors.

The strategic competence theory that has provided some theoretical insights to this study is mainly based on the examinee's conscious and deliberate ability to use any strategies (cognitive and metacognitive strategies) to appropriately complete test tasks (Phakiti, 2003). Furthermore, the protocol analysis methodology used in this study is based in the identification of reading and test testing strategies examinees use to answer individual test items. These two theoretical and methodological perspectives have helped me to gain an understanding of participants' behaviours in the selections of their answers. The study findings reveal that, participants used relevant strategies to answer many test questions that included no/few functioning distractors. Furthermore, these test questions with no/few functioning distractors were generally easy for the participants. This finding suggests that the non-selection of other distractors may indicate that these alternatives were simply bad quality distractors. It also suggests that when a test question has no/few functioning distractors, it can be easy as examinees easily dismiss non-functioning distractors.

The previous study findings on the plausibility of distractors raise the critical question of the optimal number of alternatives to be included in the ESE. Although I propose to address this issue in the Conclusion Chapter in the section relating to my proposal for ESE development (see

Section 11.4), I hasten to state, at this juncture, that there is no reason for the ESE to have six alternatives if only one or two can be functioning.

The fourth validity issue that pertains to the item structure relates to guessing and elimination of distractors as test wiseness strategies. The study findings reveal that both guessing and elimination of distractors were two test taking strategies some participants used in answering some test questions. Furthermore, it was found that participants used both elimination of distractors and guessing in answering the same test questions. This finding suggests that these two strategies are sometimes complementary. The complementarity of these two strategies has also been highlighted in the literature when Sarnaci (1979) considered the process of elimination of distractors as an element of guessing strategy. In light of these two findings, I conclude that the ESE validity is also threatened by examinees' use of elimination of distractors and guessing strategies. This conclusion aligns with the theoretical conceptualization of test wiseness that supports that guessing and elimination of distractors in testing context are a threat to test validity because they result in a construct irrelevant variance as examinees' scores tend to be overestimated (Sarnaci, 1979).

Although I propose to present the relevant research insights this study has generated in the Conclusion Chapter (see Section 11.3), I wish to state, at this juncture, that one insight from the use of elimination of distractors strategy in the context of this study is that some participants who used this strategy demonstrated their genuine understanding of the text, their understanding of test question expectation and their capacity to evaluate the truth status of each item alternative on the basis of information each alternative contains. This finding suggests that elimination of distractors can sometimes be a highly-construct relevant strategy. This finding is not consistent with existing literature that suggests that elimination of distractors is a construct-irrelevant strategy (Sarnaci, 1979; Rodriguez, 2005). I will elaborate on this insight in the Conclusion Chapter.

The last validity issue that pertains to the item structure relates to the implicit sixth option. The meaning of this option is: "If there is no good answer, choose 6". The first issue that pertains to the inclusion of option 6 relates to its frequency of use as a correct answer. In regard to item writing guidelines, it is advised that the correct answer be balanced in such a way that all the

options be the correct answer (Haladyna, Downing & Rodriguez, 2002; Haladyna, 1999). Since the ESE has six options, it is expected that each of these six options is a good answer in a balanced proportion; that is about 17 per cent. However, the study findings reveal that option 6 was a good answer for only 5 per cent. This finding suggests that the ESE constructors may not be aware of including option 6 as a correct answer in a balanced proportion. This finding has one implication: if students in their test preparation stage realize that the implicit option 6 is rarely included as good answer, they may be reluctant to direct their attention to this option; therefore, they may be deceived whenever option 6 is used as a correct answer. Therefore, the main question is to know why the ESE constructors have included option 6 if this cannot be a good answer in a fairly balanced proportion.

The second issue that relates to the inclusion of option 6 relates to its meaning. The primary meaning of this option is that examinees select it in case they find that *there is no good answer* among the suggested five options. The analysis of the data reveals six actual reasons why examinees selected option 6; a finding that suggests that alternative 6 has got six actual meanings. The primary meaning of option 6 “I selected 6 because there was no best option among the suggested 5 options” came only in the fourth position as reasons why examinees selected this option; suggesting that the conceptual meaning of option 6 may be overlooked by examinees. This finding questions the validity of option 6 in terms of the actual meaning and interpretations assigned to this option.

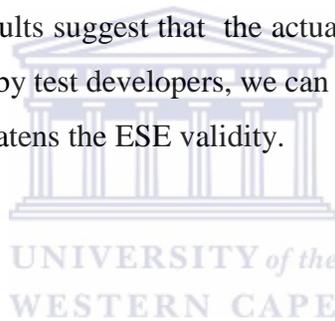
However, the most frequently reported reason why participants selected option 6 was that they could find more than one correct answer; a finding that suggests that some examinees could not falsify the five explicit options and come up with the “correct” option. Since their falsification process might have been confronted to making the last choice, this incapacity might have given them the impression that all outstanding options were correct; therefore, the correct answer would be option 6.

The second frequently reported reason why participants selected option 6 was that they were confused by the five explicit options; a finding that suggests that some examinees may have been unable to understand the expectations of the test question and the meaning of the different question options.

The third frequently reported reason why some participants chose alternative 6 is that they could not find the correct answer; a finding that suggests that some examinees were not only capable to find the correct answer; but they had also some confusion in the understanding of option 6 which is used when *there is no* correct answer and not when an examinee *cannot* find a correct answer.

Two less frequently reported reasons that accounted for participants' choice of alternative 6 were that some examinees selected this option through simple guessing or they chose it because they sensed that it could also be part of the correct options. This finding suggests that examinees who selected option 6 for these two may be those examinees who did not simply understand the question expectation and the alternative meanings.

These findings suggest that option 6 has more actual meanings than its primary meaning as articulated by test constructors. Therefore, if we agree that validity is about score meanings and interpretations, and if the study results suggest that the actual meanings of option 6 overlaps the meaning of option 6 as articulated by test developers, we can conclude that the use of the implicit option 6 in the context of ESE threatens the ESE validity.



### **Summary of the chapter**

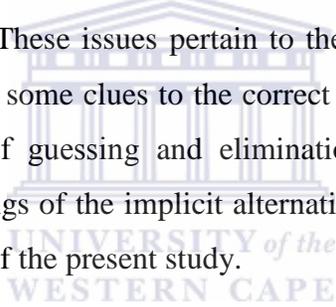
In the present chapter, I have discussed the study findings in light of the aim, objectives and main questions of the study. Regarding the first study objective that was to describe the actual context of the ESE, the discussion has focused on seven validity issues that appear to affect the context of the English state examination. These issues have related to students' skills in English language, their motivation for reading at school and at home, their attitudes towards the ESE, their experience with the ESE, the paucity of reading materials, the availability of the curriculum and its actual use by English teachers, and teachers' in-service training.

Concerning the second study objective that aimed to investigate the variety and degree of complexity of different reading types and processing levels targeted by the ESE tasks and evaluate the extent to which ESE tasks are appropriate to the ESE context; the discussion has focused on three main validity issues that potentially threaten the ESE validity. These three issues have related to the preponderance of test questions that target careful reading over test

questions that target expeditious reading; the preponderance of test questions that target higher level processing over test questions that target lower level processing; and finally, the issue of content coverage.

Concerning the third study objective that aimed to explore different textual features that potentially impact on the difficulty of ESE tasks, and to evaluate the extent to which the ESE includes texts whose features are appropriate to the ESE context, the discussion has focused on four broad categories of validity issues that appear to affect the ESE validity. These four broad issues pertain to text familiarity and difficulty, the use narrative texts, the use of expository texts, and the density of ESE texts.

Finally, concerning the fourth study objective that aimed to explore item features that potentially affect the validity of ESE tasks, the discussion conducted focused on five main issues that appear to affect the validity of the ESE. These issues pertain to the use of negative stems, the use of some keys that potentially provide some clues to the correct answer, the use of non-functioning distractors, the examinees' use of guessing and elimination of distractors as test wiseness strategies, and the multiple meanings of the implicit alternative 6. In the next and last chapter, I propose to present the conclusion of the present study.



## CHAPTER ELEVEN

### CONCLUSION

#### 11.1 A Reappraisal

In this section, I propose to recapitulate the principal parts of the study, explain what the study did and what ensued as results.

The process of investigation discussed so far is a descriptive study that I have initiated at the Department of Language Education, in the Faculty of Education, University of the Western Cape, South Africa. The investigation was mainly aimed to generate an understanding of the issues that can threaten the construct validity of the English state examination, a national exit test administered to Grade 12 school students for certification. In order to achieve this aim, I directed focus to these specific objectives: (a) to describe the actual context of the ESE and determine the extent to which this context potentially influences examinees' performance on ESE tasks; (2) to determine the degree of variety and complexity of different reading types and processing levels targeted by the ESE tasks and evaluate the extent to which the ESE tasks are appropriate to the ESE context; (3) to determine the textual features that can impact on the difficulty of ESE tasks and evaluate the extent to which the ESE texts are appropriate to the ESE context; (4) to determine item features that can affect the validity of ESE tasks, and (5) to design a proposal for the development of a more relevant and appropriate ESE.

The theoretical orientations of the study were based on three theories. The discussion conducted on the validity theory has enabled me to understand the kinds of test questions are appropriate for inclusion in the ESE. The description of the basic assumptions of the construction-integration theory has enabled to decide the kinds of text passages are relevant for inclusion in the ESE. The insights from the strategic competence theory have provided necessary information to evaluate different strategies used by study participants so as to decide which strategies are construct-relevant and which ones are construct-irrelevant. Furthermore, the design of a specific multi-

componential model of reading that fits the specific context of this study has provided evidence to maintain that the reading activity is a complex construct that can be better examined when we take into account the specific context of reading activity.

My choice of the descriptive design has enabled me to gather the evidence that I believe is necessary to answer the research questions. This design has enabled me to identify and describe the different variables that are a potential threat to the ESE validity. Further to this, since no validation studies have been conducted so far on the ESE, the use of descriptive design in this study has produced findings that have generated new insights that I believe can advance the large body of testing research. Among these insights, I hasten to mention the following: (1) construct validity is an evolving and context-dependent concept; (2) reading construct cannot be examined outside reading context; (3) elimination of distractors can sometimes be a highly construct-relevant strategy; (4) the concept of construct-irrelevance variance needs to be reconceptualized; (5) construct underrepresentation is a context-dependent concept. I propose to provide detailed explanations of these insights in the section 11.3 of the present chapter.

One important methodological aspect of the present study relates to triangulation. The concurrent use of protocol analysis and content analysis has produced findings that none of these methods alone could produce. The use of protocol analysis methodology helped to identify and describe the way the study participants searched for text information, evaluated item alternatives, and chose the *best* option. The use of a concurrent strategies questionnaire as research instrument helped to collect research data during the task completion process so as to ensure their validity. My choice of concurrent strategies questionnaire, moored to Ericsson and Simon (1993)'s epistemology supports the need to ensure the closest connection between thinking and verbalization. This was advantageous as it ensured that the study participants' reports were not distorted as their thoughts were verbalized during task completion. The use of protocol analysis in this study has culminated in the description of strategies participants used to complete test tasks so as to understand the reading construct. Therefore, the findings generated by protocol analysis methodology have generated different issues that appear to threaten the ESE validity. Furthermore, the use of content analysis methodology was motivated by the need to thoroughly describe the English state examination papers that were used as a research instrument. This

method has helped to identify and describe the different textual and item features that likely impacted on participants' abilities to process and understand the text and complete test tasks.

## **11.2 Relating the study outcomes to the research questions**

In this section, I propose to relate the sub-research questions of this study to the outcomes discussed in Chapters 6, 7, 8 and 9. I believe that the following explanations relating to these sub-research questions will usefully contribute to this conclusion as they summarize the study findings and attempt to answer the sub-research questions in an informed way.

### **11.2.1 Sub-research question 1: What is the actual context of the ESE and to what extent this context potentially influences examinees' performance on ESE tasks?**

In light of the study findings presented in Chapter 6, the actual context of the ESE can be described as follow:

(1) Grade 12 students' skills in English, in general and in reading, in particular are very low as their reading at school is very restricted due to insufficient time for learning English, paucity of reading materials, and lack of commitment of their English teachers. Further to this, students' reading at home is almost non-existent due to the lack of reading materials, their contents not being relevant to students' needs, as well as the students' poor socio-economic conditions.

(2) Students have less experience with the ESE as the multiple-choice technique is rarely used as a classroom assessment method. Furthermore, they have some negative attitudes towards the ESE and these negative attitudes are reflected in their desire to maintain or increase the actual time for writing the test, their preference to reduce the actual number of test questions, item options and alternate forms.

(3) Although the majority of teachers of English are formally qualified to teach the English subject and they have some experience teaching English in Grade 12, there is an abject lack of in-service training as well as continuous trainings. Consequently, the teachers of the English subject are not updated with new developments in teaching. Furthermore, due to poor socio-

economic conditions, many teachers have low motivation for teaching English; and they lack appropriate materials to deal with large classes. They do not have some teaching documents, and they never complete all the topics foreseen in the curriculum in the course of the academic year.

The preceding context is without consequences on the validity of ESE. I believe that if the ESE tasks do not take account of this context, students' performances on the ESE will hardly reflect their actual abilities that they have in English in general and reading in particular. Briefly stated, the closer ESE tasks are adjusted to the ESE context, the more valid the ESE.

### **11.2.2 Sub-research question 2: What is the variety and degree of complexity of the different reading types and processing levels examinees conduct while completing the ESE tasks, and to what extent the ESE tasks are appropriate to the ESE context?**

In light of the study findings presented in Chapter 7, I am inclined to state that:

(1) The majority of ESE test items target careful reading instead of expeditious reading. Yet, test tasks that target careful reading appear to be more difficult than those that target expeditious reading as careful reading tasks require a frequent combination of strategies while expeditious reading tasks appear to require less frequent combination of strategies.

(2) The majority of ESE test items target reading at global level (text level) instead of reading at local level (sentence and paragraph levels). Yet, test tasks that require global level reading appear to be more cognitively demanding than those that target local level reading.

Consistent with these two main findings, and considering the actual ESE context described at point 11.2.1, I conclude that the ESE appears to be less appropriate to the ESE context as its main focus is on tasks that are cognitively demanding to examinees.

### **11.2.3 Sub-research question 3: What are the different textual features that likely impact on the complexity of ESE tasks, and to what extent these textual features are appropriate to ESE actual context?**

In light of the study findings I have presented in Chapter 8, I wish to state that:

(1) In connection with the ESE based on narrative texts, I have found that the majority of test questions target text information that is not appropriate to the narrative structure. Furthermore, the ordering of information in the text does not relate to the ordering of information in the test questions. Besides, the majority of test questions request examinees' skills to process the text at global level rather than local level. In light of these findings, I conclude that the ESE based on narratives appear to be less appropriate to the ESE context as the test tasks hardly take account of the narrative structure and the underlying cognitive demands appropriate to narrative texts.

(2) With regard to the ESE based on expository texts, I have found that, although the ESE test questions actually request examinees' ability to process the text at both global and local levels in order to get the text gist as well as to locate specific pieces of information located at both global and local levels, there are no test questions that target examinees' capacity to identify and understand descriptors and place prepositions necessary to visualize text details. Picking upon this finding, I conclude that the ESE based on expository texts appears to be partly appropriate to the ESE context as the test fails to include other critical aspects of the reading construct.

(3) Concerning text density, the study findings have indicated that the ESE texts are less dense than the classroom texts as the latter are three times longer than ESE texts. In light of this finding, I conclude that ESE texts appear to be less appropriate to the ESE context. Implicit in this conclusion, ESE scores can hardly reflect the examinees' actual text comprehension in non-testing situations.

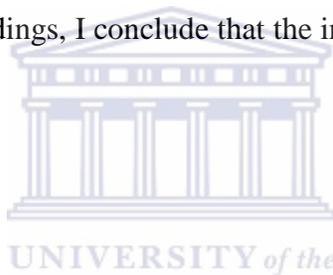
### **11.2.4 Sub-research question 4: What are the item features that potentially affect the ESE validity?**

The study has provided evidence that suggests that:

(1) Some ESE test items have their keys that potentially provide some clues to test wise examinees as the latter can get these test items right without necessarily understanding the expectations of the questions as well as the meanings of all the six alternatives.

(2) There are no test items that have five functioning distractors as expected. On the contrary, most test items have three, two or just one functioning distractors. Further to this, some ESE test questions have no single functioning distractors. This finding suggests that ESE constructors may not be capable to construct an ESE with five functioning distractors.

(3) Option 6 is not included as a correct answer in a balanced proportion as evidence indicates a good answer for only 5 per cent. This finding suggests that ESE constructors may not be aware of including option 6 as a correct answer in a balanced proportion. Further to this, the primary meaning of option 6 “if there is no answer, select option 6” is overlooked by examinees as the study has provided evidence that supports that students have six actual meanings they give to this implicit option. Based on these findings, I conclude that the inclusion of the implicit option 6 is a threat to the ESE validity.



### **11.3 New insights**

The research process conducted so far aimed to explore the issues that potentially affect the validity of the ESE. The theoretical and conceptual perspectives adopted determined the route to be taken in order to achieve this aim. The choice of the descriptive design tallied with the study aim and it contributed to identify and describe the main issues that can threaten the ESE construct validity. From the study findings obtained and the discussion presented, it appears that the present study has not only answered the study questions, but it has also generated insights that are useful beyond the local context of this study. Therefore, in this section, I propose to elaborate on these insights that I believe can advance research on construct validity in general and the assessment of reading in particular. I hasten to state that these insights are not context-independent, atemporal affirmations. But these are context-bound confirmations that offer value-loaded points for informed approaches and actions in the rather complex domain of testing. Therefore, these insights need to be enriched through a critical discursive approach with investigations made on diverse contexts. In doing so, I believe that we can advance research and we can therefore usefully clarify some construct validity issues.

The entire process of the present study has generated seven insights that are: (1) Validity: an evolving and context-dependent concept, (2) Reading context and reading construct: Two interrelated and inseparable concepts, (3) Elimination of distractors: A construct-relevant strategy, (4) Construct-irrelevance variance: A need for reconceptualization, (5) Construct underrepresentation: A context-dependent concept, (6) Reading activity: A complex process, (6) A rationale approach for reading test development: From the understanding of reading activity to the design/selection of an appropriate measurement model and not applying existing measurement models to reading activity, (7) Can a reading test be valid in all contexts? A brief appraisal of some touted international university admission tests. In the following point, I propose to explain the first insight that is “Validity: an evolving and context-dependent concept”. In the following section, I discuss the first insight.

### **11.3.1 Validity: an evolving and context-dependent concept**

The belief that the concept ‘validity’ has evolved over time is not an insight. In the Chapter two relating to the theoretical framework; specifically in Section 2.1 related to the construct validity theory, I have offered the historical developments of the concept ‘validity’ as presented by measurement scholars such as Goodwin (1997, 2002), Goodwin and Leech (2003) and Kane (2002, 2004, 2011). In light of this historical development, I have mentioned that an early definition of validity (in the 1940s) emphasized the test itself, suggesting that validity was considered as a static property of a measure (Goodwin & Leech, 2003). This is to suggest that, a test was considered valid if it measured what it was supposed to measure. However, in the 1960s, the meaning of validity shifted its focus to test use. At that time, validity was defined as the extent to which a test could produce information that was useful for a specific purpose (Goodwin, 1997, 2002). One main innovation during that time was that the concept validity was fragmented into three types: criterion-referenced validity, content validity and construct validity (Kane, 2002, 2004, 2011). However, the revolution in the conceptualization of validity emerged during the 1980s and 1990s with Cronbach’s (1980) and Messick’s (1989) works. At that time, the concept of validity was associated with the inferences and decisions that are made on the basis of test scores. Consistent with this association, the concept validity was defined in terms of the appropriateness, meaningfulness, and usefulness of the inferences made from test scores

(Goodwin & Leech, 2003; Kane, 2002, 2004, 2011). In light of this conceptualization, validity became to be viewed as a unitary concept and the concept ‘construct validity’ became an overarching term that encompassed the three traditional validity types. This conceptualization of validity is still predominant today in most testing circles. The present study has also adopted it.

However, the process of the present investigation has generated an insight that is worth mentioning here because I believe that this insight can bring our understanding of the concept ‘(construct) validity’ a step further. The study findings and the discussion I have presented suggest that the concept ‘(construct) validity’ can be better understood when it is defined from a context-dependent stance. This is to suggest that a general and contextless definition of validity does not appear to solve the harsh and current debate and criticism on the validity theory. In Chapter 2, Section 2.1.4, I have mentioned that much of the dissatisfaction with the actual conceptualization of validity is based on the view that the theory is impractical in educational contexts, it does not clearly articulate, and its basic concepts are sometimes wrongly used (see Brennan, 1998; Fremer, 2000, Borsboom, Mellenbergh, and van Herdeen, 2004; Lissitz & Samuelson, 2007, Embretson, 2011).

In light of the preceding argument and considering my suggestion to conceptualize validity from a context-dependent stance, and by looking at the scattered grasps generated by this study, I wish to mention that (1) there are different types of (construct) validity depending on the types of construct being investigated. This is to suggest that the ingredients involved in the conceptualization of the (construct) validity of a mathematic test are different from those involved in the conceptualization of the (construct) validity of a language test. By the same token, what we need to understand for defining the validity of a reading test is different from what we need to understand for defining the validity of a writing (speaking/listening) test. Furthermore, within the scope of reading test, the ingredients involved in the definition of the validity of a multiple-choice reading test are different from those involved in the definition of the validity of a constructed-response reading test. Further to this, the ingredients involved in the definition of the validity of a multiple-choice reading test based on narrative tests are completely different from those involved in the definition of the validity of a multiple-choice reading test based on expository texts. In light of this suggestion, the present study results and the discussion conducted have incorporated as ingredients of validity: the specific context of reading activity,

readers' use of strategies to comprehend the text, readers' skills to understand the expectations of the test questions, their skills to understand the meaning and implications of the different item options, and their capacity to select or discard item options from their understanding of the text. I believe that such ingredients would not be involved in a conceptualization of the validity of a constructed-response reading test. Further to this, the present study results and the discussion presented have clarified that the elements necessary for defining a multiple-choice reading test based on narratives include variables such as (a) the extent to which the test questions target information that relates to the narrative structure, (b) the extent to which the ordering of information in the text relates to the ordering of information in the test questions, (c) the extent to which the narrative details that relate to the story setting are not biased by advantaging some examinees or groups of examinees over other examinees or groups of examinees. I believe that these elements are not surely those that are involved in the definition of the validity of a multiple-choice reading test based on expository texts as the study findings and the discussion presented have indicated that the validity of a multiple-choice reading test based on expository texts relates to the extent to which : (a) the test tasks request examinees' ability to process the text at both global and local levels in order to get the text gist as well as to locate specific pieces of information located at both global and local levels; (b) the test tasks request examinees' capacity to identify and understand descriptors and place prepositions necessary to visualize text details.

The preceding argument I have made appears to support my suggestion that construct validity is a context-dependent concept. Therefore, in order to advance our understanding on this concept, we need to distance ourselves from conceptualizing construct validity from a universal and contextless stance. Rather, we need to turn our attention to the specific aspects involved in the operationalization of the construct being investigated. By doing so, we will be better served in our validation studies as this new conceptualization will appear more practical than theoretical, clearer than blurred, and more feasible than complicated; therefore, results from such studies will be beneficial to the educational enterprise. I believe that in considering this suggestion, current harsh and endless debates on the semantics of the concept construct validity will appear to be a scholarly enterprise with little educational benefits. In the next section, I propose to discuss the second research insight generated by this study.

### **11.3.2 Reading context and reading construct: two interrelated and inseparable concepts**

In order to understand this insight, I wish to mention that in my of the current processing models to examining the reading construct (see Chapter 3; Section 2), I mentioned that most of these models appear to be strongly influenced by the union of cognitive psychology and assessment (Embretson & Wetzel, 1987; Embretson, 1999; Cohen & Upton, 2006, 2007; Weir & Khalifa, 2008; Gao & Rodgers, 2007; Carr, 2003; Kostin, 2004). Consistent with this union, these models mainly rely on psychometric measures (Item analysis), the use of multiple-faceted Rasch measurement, or the structural equation modelling or again factorial analyses and linear regression methodologies. By taking such a theoretical and methodological stance, these models appear to be encapsulated in the cognitive view of language and therefore ignore the social dimension of language proficiency. Although validity studies using these models have incorporated some of the most current cognitive theories in examining the reading construct, thus providing useful information on examinees' performance on the reading test, they nevertheless appear to ignore to link the critical features of cognitive theories to the specific testing context so as to reveal the meaningful cognitive processes in a particular test domain.

However, my design of a process-oriented approach on multiple relations among different components in order to examine the issues that affect the validity of the ESE and the selection of a protocol analysis methodology based on the investigation of examinees' strategies used during the completion of test tasks have yielded another insight that I believe can advance research. This insight suggests that reading context and reading construct are two interrelated and inseparable concepts. This is to suggest that reading construct becomes a concept difficult to understand if it is explained and/or investigated outside the specific context of reading activity. In light of this suggestion, the study findings and the discussion presented have revealed that examinees' performance on test tasks appeared to be influenced by the context of reading activity. This is to suggest that the way participants used strategies to process the text in order to comprehend it, and their use of strategies to complete test tasks largely depended on the context in which reading activity was conducted in both classroom contexts and normal life. Revisiting the context of participants' reading in classroom context, the study findings have indicated that participants had low skills in English in general and reading in particular as their reading at school was an activity heavily restricted due to insufficient time for learning English, paucity of reading

materials, and lack of commitment of their English teachers. Regarding reading at home, the findings have indicated that reading activity was almost non-existent due to the lack of reading materials and students' poor socio-economic conditions. In such a context, it was found that, the preponderance of test questions that targeted careful reading over test questions that targeted expeditious reading as well as the preponderance of test questions that targeted higher level processing over test questions that targeted lower level processing did not relate to the actual ESE context. This is to suggest that, since validity is about the relevance of inferences made on the basis of test scores (Messick, 1989, Kane, 2011), in order to be construct-valid, the ESE test scores need to closely reflect the actual reading context by including more tasks that target expeditious reading and that require reading at lower level as such tasks appear to be less cognitively demanding than those that target careful reading and that require reading at higher level.

This insight is worth considering in conducting validation studies. It suggests that a test item can be valid when the information required to answer it taps into the target construct, but still be inappropriate to the specific context of the test. In order to advance our knowledge on the construct validity of reading tests, we need to agree that the development of valid reading tests can be possible if we first examine as extensively as possible the nature of reading activities and different response processes examinees engage during test taking. This entails eliciting the cognitive processing involved in not only the text and test tasks, but also in contexts beyond the test itself, that is, in those contexts involved in performing reading tasks in classroom and real life. By ignoring this route, reading appears to be an activity that occurs in a vacuum; therefore, the results generated by studies that ignore the actual reading context tend to miss the salient aspects of the reading construct as we hardly understand the actual reading behaviours. Further to this, examinees' completion of test tasks can become more of a problem-solving activity rather than a reading activity. In the following section, I propose to discuss the third insight generated by this study.

### **11.3.3 Elimination of distractors: a construct-relevant strategy**

My review of existing studies that have addressed the issue of test wiseness (see Chapter 4; Section 6.3) has revealed that elimination of distractors is a test wiseness strategy that is considered to be a threat to test validity (Sarnaci, 1979; Lin, Chu, & Meng, 2010; Haladyna & Downing, 2004; Rodriguez, 2002). These studies have also highlighted that elimination of distractors is a strategy that is generally used in conjunction with guessing strategy as the examinee tends to eliminate some item options before he/she finally guesses between outstanding options. Therefore, these studies have supported that elimination of distractors is a construct-irrelevant strategy because it tends to lead to an overestimation of the examinee's test scores, resulting in the construct- irrelevant variance.

However, the present study findings and the discussion presented have revealed that elimination of distractors may sometimes be a highly construct-relevant strategy. The study results have indicated that some participants who used elimination of distractors strategy before getting to guessing the 'correct' answer demonstrated their genuine understanding of the text, their understanding of test question expectations and their capacity to evaluate the truth status of each item alternative on the basis of information each alternative contained. Furthermore, the present study results have revealed that some participants used elimination of distractors as a response decision process in which the question stem and the options/alternatives were compared for accuracy of the text, a response procedure that appears to indicate genuine control over the item and that appears to align with Embretson and Wetzel's (1987) model of reading comprehension used in this study to provide conceptual orientations.

In light of the preceding finding, there is a need to re-conceptualize the concept of elimination of distractors strategy. One way to do this is to contextualize examinees' use of this strategy before deciding whether it is construct-irrelevant or construct-relevant. In the following section, I wish to discuss the fourth insight relating to the need to reconceptualize the concept 'construct-irrelevance variance'.

#### **11.3.4 Construct-irrelevant variance: a need for reconceptualization**

In describing the validity theory (see Chapter 2, Section 1), I have mentioned one serious threat to validity that is called construct-irrelevant variance. Validity scholars (Messick, 1989; Young, 2008; Haladyna & Downing, 2004; Downing, 2002) argue that when a test includes factors (or variables) that influence students' test scores but are not directly related to the test construct, this test is said to be construct-irrelevant. This is to suggest that the scores obtained by an examinee or a group of examinees on this test cannot serve a basis for meaningful test score interpretations as such scores are partly due to factors that are different from the construct that the test is designed to measure. In light of this conceptualization of construct-irrelevant variance, most validity studies using processing models to examine the reading construct have reported that poorly crafted test questions lead some examinees to use test wiseness strategies, and this generally results in tests that are construct-irrelevant on the ground that test scores hardly provide an indication of examinees' engagement with the reading tasks. Such a conceptualization of construct-irrelevant variance appears to be narrow; therefore, it does not offer the opportunity to capture the appropriate meaning of validity. However, the present study findings and the discussion that has ensued have generated one conceptual illumination on construct-irrelevant variance. In order to understand this illumination, I propose that we look at validity from two perspectives: first, validity as appropriateness of test tasks, and then validity as the appropriateness of cognitive operations necessary to complete test tasks.

- **Validity as appropriateness of test tasks**

By conceptualising validity in terms of the appropriateness and meaningful of decisions made on test scores (Messick, 1989; Kane, 2001, 2004, 2011), we need to acknowledge that appropriate and meaningful decisions can be made when the test includes tasks that are adjusted to examinees' characteristics. This suggestion is consistent with my epistemology that privileges test context in validating the reading construct. In light of this epistemological stance, the present study results have revealed that, although the majority of test questions were situated at one of the six processing levels that make up the reading construct, there was a predominance of test questions that required participants to process the text at higher level over those test questions that required them to process the text at lower level. Therefore, this imbalance of test tasks does not appear to reflect the actual context of the ESE. Consistent with this argument, the ESE

appears to be construct- irrelevant as the test tasks appear to be less appropriate as they do not feature examinees' actual reading abilities.

Another example of this issue of inappropriateness of test tasks in the context of the present study relates to the inclusion of short and truncated texts in ESE. The study findings have revealed that, since classroom texts were found to be longer than ESE texts, and that ESE texts did not include transitions as well as cohesive devices, I assumed that ESE texts appear to be less appropriate to the ESE context. Implicit in this conclusion is that the ESE scores may hardly reflect the examinees' actual text comprehension in non-testing situations.

- **Validity as the appropriateness of cognitive operations necessary to complete test tasks**

In processing a text for comprehending it and completing test tasks, the reader has recourse to a variety of cognitive processes and sub-processes depending on the text structure as well as the questions based on the text. For a reading test based on narrative, for example, appropriate test questions are those that require examinees to use the underlying cognitive operations readers engage while constructing the narrative story line. In light of this stance, the present study findings have revealed that the majority of test questions based on narratives did not require examinees to deploy the underlined cognitive skills appropriate to narrative texts. This finding suggests that the ESE may be construct-irrelevant as the test tasks hardly require the cognitive operations appropriate to narrative texts.

In light of the preceding argument, it is worth suggesting that validity is not to be viewed *only* in terms of the appropriateness and meaningfulness of decisions made on the basis of test scores; nor should it *only* be viewed in terms of appropriateness of test tasks to the test context. Instead, it *also* needs to be conceptualized in terms of the appropriateness of cognitive operations necessary to complete test tasks. This insight accrued from the present study findings appears to be missed by current scholars in reading assessment research. According to this suggested conceptualization, a reading test is construct-irrelevant when its scores (*also*) include variance due to examinees' use of cognitive operations that are not relevant to the kind of operations necessary to construct the text macrostructure.

### **11.3.5 Construct underrepresentation: A context-dependent concept**

In describing the construct validity theory (see Chapter 2, Section 1), I have pointed out that construct underrepresentation is a threat to test validity as it relates to a variance that occurs when the test is too narrow and fails to include important dimensions or aspects of the construct (Messick, 1989; Young, 2008). Research indicates that, in order to be construct-valid, educational tests need to be a representative sample from the test large domain of knowledge and skills (Haladyna & Downing, 2004). In reading assessment research, studies that have addressed this issue (see Young, 2008; Haladyna & Downing, 2004; for example) have generally adopted a content approach mainly based on the specification of course content and objectives and then mapping it with the test tasks so as to decide the extent to which the test includes tasks that are a representative sample of the tasks from the test domain. Therefore, these studies have reported that construct underrepresentation generally arises from poorly conceptualized or inadequately operationalized constructs. Other studies have adopted a processing approach to construct representativeness (Weir & Khalifa, 2008b; Weir, Hawkey, Green & Devi, 2008; Weir, Hawkey, Green, Unaldi & Devi, 2008; Sarojani & Krishnan, 2011) based on the evaluation of the extent to which the test items require examinees to process the text at the different six levels that make up the entirety of the reading construct. Although these studies have generated useful findings that help to understand the issue of construct underrepresentation and the extent to which it threatens test validity, they, nevertheless, appear to have overlooked some of the critical dimensions of this concept. The main reason is that these studies view the concept ‘construct underrepresentation’ as a general concept only. Therefore, they appear to ignore its context-dependent aspect. This is to suggest that the concept ‘construct-underrepresentation’ can be a function of the specific characteristics of the test and the test tasks.

In light of the preceding observation, the present study findings and the ensued discussion have revealed that, the ESE test questions based on expository texts actually requested examinees’ ability to process the text at both global and local levels in order to get the text gist as well as to locate specific pieces of information located at both global and local levels. This finding suggests that the test tasks actually tapped into the test construct as the ESE constructors seem to be aware of the need to base ESE test based on expository texts around gist construction through reading at both global and local levels. However, the same study findings have indicated that there were

no test questions that requested examinees' capacity to identify and understand descriptors and place prepositions necessary to visualize text details; suggesting that these critical features of the test construct have been simply missed. Yet, as literature suggests (Akhondi, Malayeri, & Samad, 2011; Snyder, 2010), the structural details of expository text help guide readers to understand the text. Therefore, I believe that in order to be construct-representative, the ESE based on expository texts needs to include not only test questions that target gist construction through reading at both global and local levels, but also those test questions that target descriptors and prepositions as these contribute to the construction of the text macrostructure.

The previous argument also applies to reading tests based on narratives. The study findings have revealed that the ESE test items did not include questions that targeted certain critical aspects relevant to narratives. This is to suggest that, in order to be construct-valid, scores on a reading test based on narrative texts must also provide an indication of examinees' abilities to read the narrative and reconstruct its story line. Consistent with this claim, test questions must also require examinees' skills to reconstruct the narrative story line with a focus on the chronological sequencing as well as the spatial organization of narrative events. More specifically, the test needs to include tasks that require examinees to identify/describe the narrative setting, to identify the main character and auxiliary characters, to analyze the overall plot through the identification of the story problem, the response, the action and the story outcome. However, the present study findings have revealed that the ESE test items did not target most of these critical aspects specific to narrative structure; suggesting that the specific reading construct was not fully examined as many relevant aspects of the construct were simply overlooked.

The previous two arguments based on the study findings suggest that the concept construct underrepresentation is a context-based concept that needs to be reconceptualized so as to encompass not only the two existing conceptualizations, but also to include another conceptualization that takes account of the different elements specific to different text genres. In doing so, we can take the concept of construct underrepresentation a step further as we can better understand it and our validation studies will tend to sound more clear, precise and practical; therefore, they will better serve for educational purpose. If we ignore this route, we will keep on producing reading tests that include tasks we believe cover the entire reading construct when in reality such tests are still not construct-valid as some of the critical aspects of the reading

construct in specific contexts have been simply missed. In the following section, I discuss the next insight relating to the complex nature of reading activity.

### **11.3.6 Reading activity: a complex and fluid activity**

This insight can be helpful in validating reading tests. If we are aware that reading in general and reading performance in particular is a complex and fluid activity, we will agree that any effort to explain reading activity by singling and isolating the multiple factors that are believed to interfere in reading process can be futile. This observation alludes to some existing models that have adopted a subskills approach to testing reading premised on the possibility to link particular test items or test tasks to specific subskills that they are said to tap into (see Guthrie & Kirsch, 1987; Rosenshine, 1980; Schedl *et al.*, 1996; Young, 2008). Therefore, these models heavily rely on quantitative approaches mainly based on the use of factorial analyses and linear regression methodologies. However, I wish to mention that my intention is not to undervalue the studies that have used these methodologies. They have produced useful findings regarding the validity of reading tests. Nevertheless, I wish to highlight that factor analysis methodology is limited by its essence as it focuses on “the separability of the capabilities that a reader is assumed to need in order to tackle certain test items” (Weir & Khalifa, 2008, p. 2). Besides, it is a psychometrically driven approach that deals with the factors that statistically show their contribution to successful reading performance. Therefore, factor analysis appears to ignore the actual finetuning processes the reader engages in reading the text. Further to this, linear regression methodology is limited as it attempts to model the relationship between variables that account for reading performance. This methodology heavily relies on the use of correlations (mainly multiple correlations) by attempting to determine the specific variables that account for a substantial percentage of the variance in item difficulty (see Kasai, 1997). In the scope of the present study, it would be imprudent to categorize the identified factors in those factors that most impact, those that moderately impact and those that least impact on ESE performance. I believe that such a categorization would not make a substantial contribution to the improvement of ESE validity. Instead, it will overshadow the salient problems that need to be considered for the ESE improvement.

At this juncture, I wish to highlight a relevant aspect of this insight because I believe that this aspect can bring some contribution to the large body of reading assessment research. Weir and Khalifa (2008a)'s model of reading assessment that is used in this study to provide the conceptual orientations conceptualizes reading construct on multiple dimensions with careful reading and expeditious reading that can be conducted at both global level and local level. In light of this conceptualization, subsequent research using the model have shown differences in the use of strategies between test items that require examinees to read at global level and those that require them to read at local level (Weir & Khalifa, 2008b; Weir, Hawkey, Green & Devi, 2008; Weir, Hawkey, Green, Unaldi & Devi, 2008; and Sarojani & Krishnan, 2011). Implicit in this finding is that reading and test taking strategies become discrete entities that can be labeled in those strategies that fit for answering global items and those that fit for answering local questions on the one hand, and strategies that fit for answering careful reading items and those that fit for answering expeditious reading test questions on the other hand. However, the study findings have revealed that in order to answer some test questions, participants had to conduct both careful reading and expeditious reading in almost equal proportions; and/or process the text at more than one level. This finding provides evidence to suggest that reading activity may not be linear; rather, it may be a fluid activity that reflects the dynamic nature of language. This finding offers the opportunity to critically interrogate many existing taxonomies (see Cohen & Upton, 2006, 2007; Weir & Khalifa, 2008a; for example) that so far categorize the different reading types as discrete and opposing entities where an item categorized in A cannot be also categorized in B. Yet, the data in this study have confirmed that different reading types may sometimes be integrating and overlapping entities and individual test items can be categorized in both A and B groups. Moreover, this finding suggests that the six processing levels as articulated in the literature are not strictly hierarchical levels where readers linearly process the text from lower level to higher level. But, it suggests that processing levels may sometimes just be labels and that actual text processing may be conducted in a forth and back process.

Furthermore, the study findings and the discussion presented have revealed that, in many cases careful reading was found to be used in conjunction with expeditious reading; especially with questions that targeted careful reading of a portion of the text as examinees had to first read the whole text (either carefully or expeditiously, or both) in order to integrate the paragraph macrostructure into the overall text gist. This combination of careful reading and expeditious

reading suggests that these two types of reading may involve some overlapping processes and actions. Finally, the study findings have also suggested that reading at global level may be seen as a general strategy used for answering even some test items that target information at sentence level. Therefore, global reading might aid struggling readers to build text gist necessary to understand local pieces of information.

The aforementioned findings provide evidence that supports that reading activity is fluid, dynamic, and complex and that the assessment of reading must be grounded in this specific nature of reading. This is to suggest that appropriate models of examining the reading construct need to be adjusted to the fluid nature of reading activity. Consistent with this suggestion, the use of purely quantitative models to examine the reading construct appears to be inconsistent with the essence of reading activity. Therefore, conducting validation studies by using these models can be an enterprise with little benefits to educators. In the following section, I propose to explain the last study insight that relates to the relevant approach to test development.

### **11.3.7 Can a reading test be valid in all contexts? A brief appraisal of some touted international university admission tests**

Although this insight overlaps with the scope of this study, I feel it is relevant to discuss it as my extrapolation from all the preceding insights.

It is well known today that university admission for students coming from the ‘Third World’ in the majority of outstanding Western universities is conditioned to success on some international tests. Some of these tests include the TOEFL, the IELTS, to mention only few. These language tests are proficiency tests that intend to predict if the candidate has the abilities, if admitted to the university, to successfully undertake his/her studies in the medium of English. This is to suggest that these tests generally aim to determine the extent to which a candidate can successfully read academic materials and actively participate in the academic life. I wish to mention that these tests have built a wide popularity in the world today. My engagement with the language testing field reveals that these tests are the most studied by scholars and their validity has been established. Recent studies using Weir and Khalifa (2008a)’s model (that is partly used in this study) have

also confirmed their high validity (Weir & Khalifa, 2008b, Weir, Hawkey, Green & Devi, 2008; Weir, Hawkey, Green, Unaldi & Devi, 2008; Sarojani & Krishnan, 2011).

At this juncture, I wish to mention that one main component of these tests is the reading component where the candidates are given a reading passage (or multiple short passages) and they are asked to answer multiple-choice questions based on the passage. Since this testing format is similar to the format of the ESE, I wish to propose, in light of the present study findings and discussion, my view on the validity of these touted tests. I wish to mention that my intention is not to deny the quality of these tests; instead, I want to raise, within the epistemological scope adopted in this study, the validity issue relating to the primacy of test context in investigating test validity. This issue is worth raising because these tests are administered to candidates coming from different countries with different linguistic, cultural, educational, social, political and economic backgrounds. My aim is strictly contributory in that I intend to draw researchers' attention to the need to reappraise the established validity of these tests.

After I have provided the motivation for raising this insight, I wish to mention that the present study results, the discussion of its findings and the description of the previous insights have suggested that reading construct cannot be examined outside the actual context of reading activity. In keeping with this position, when tests like TOEFL and IELTS are administered to students from diverse contexts, their validity can depend on the extent to which these tests include test tasks that are sensitive to these diverse contexts. This is to suggest that these tests need to recognize the candidates' contexts so as to align test tasks accordingly on the ground that the discrepancy between different contexts need to be adjusted so as to capture the salient cognitive capitals of different candidates. I believe that some scholars in language/reading assessment research may find this task out of reach as it appears to make our work more difficult than easy. However, I wish to highlight that by ignoring to connect the test tasks to the actual reading context of candidates, we come up with test scores that hardly reflect the actual potentials of candidates. This can be reflected in inappropriate decisions that deny some candidates access to university simply because their scores are situated in lower bands. Yet, we cannot say with absolute confidence that those candidates who have passed these tests are really capable of successfully engaging in academic reading. My interactions with some English

language educators with long international experience have revealed that some candidates who scored extremely high on these tests could not successfully engage in academic reading as predicted.

By discussing this insight as an extrapolation from the present study findings and discussion, I intend to raise the issue of social justice that appears to characterize the reading section of most touted international university admission tests. Although I will not elaborate on this issue as it overlaps with the present study scope, I wish to draw the attention of language/reading assessment researchers on a route that may illuminate the reading assessment research. I am of the view that we will benefit more from international university admission tests if test developers start by recognizing the unrecognized before they embark in developing their tests. This is to suggest that test developers need to consider as point of departure the possible multiples students' contexts that appear to be ignored so far. In doing so, their reading tests will produce scores that can generate empirical and conceptual illuminations that can advance research. How to actually achieve this is a task that overlaps the scope of this study. It is up to reading assessment scholars to embark on this route. After having provided in this section a description of the seven insights this study has generated, I now turn to propose what I believe are relevant suggestions for the development of a valid English state examination.

#### **11.4 Proposal for ESE development**

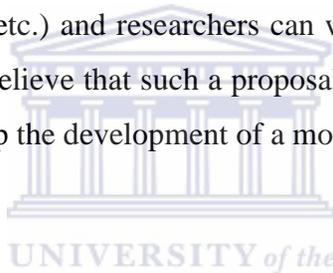
The presentation of the current study findings in Chapters 6, 7, 8 and 9 and the subsequent discussion of the study findings have provided evidence that suggests that the validity of the English state examination is threatened by many issues that are due to the way it is constructed in relation to the actual test context. However, the fifth and last objective of the present study was to design a proposal for the development of a more relevant and appropriate ESE. In light of this objective and considering the evidence produced by the study findings, I propose to present the following proposal as my contribution to the improvement of the ESE. I believe that this proposal also constitutes the main significance of the present study.

At this juncture, I wish to mention that, in order to highlight the embedded sense making that I attempt here, I have decided to use imperatives as subheadings for my proposal.

## **1. Always conduct validation studies to justify the use of the ESE**

The use of a high stake exit test such as the ESE should be justified through validity studies that are meant to provide sufficient evidence for test scores use and interpretations. This is to suggest that, the state should regularly initiate validation studies where fundamental validity issues are identified and suggestions for test improvement proposed. Therefore, since the ESE has been used as a component of the national certification test for more than forty years without being validated through empirical studies, I suggest that the Congolese government initiates and funds validation studies so as to ensure the ESE validity.

However, it is usually agreed that validating a test is a difficult enterprise that one person can hardly conduct. This thesis may serve as a point of departure from which the ESE validation agenda can be built where different stake holders (test developers, government experts, English language teachers and inspectors, etc.) and researchers can work together in order to produce a proposal for ESE development. I believe that such a proposal must take into account the various contextual parameters that handicap the development of a more valid test.



## **2. Always consider the ESE context while developing the ESE**

The description of the DR Congo English state examination context in Chapter 6 has provided evidence that suggests students' poor skills in English in general and reading in particular, little/no inclination for reading in class and at home, paucity of reading materials, less students' experience with the ESE, some negative attitudes towards the ESE, lack of commitment of the teachers of English subject, an abject lack of teachers' training and/or continuous training, and lack of teaching supports.

In light of this specific context, I propose that ESE developers include test tasks that not only target the reading construct, but also reflect the actual context of reading activity. Failure to consider this context in ESE design necessarily results in tests with low validity.

### **3. Revisit the meanings and interpretations of ESE scores**

In designing the ESE, test developers should always bear in mind that each test question included in the ESE must have a specific meaning (for instance, examinees' capacity to read a specific paragraph and find the word to which a specific pronoun relates; examinees' ability to read a paragraph and find a title that best suits that paragraph, etc.) and the entire test has also a specific meaning (for example: examinees' capacity to comprehend a text and reconstruct the text story line; examinees' capacity to read the text and visualize its details, etc.). Therefore, the ESE constructors should revise each test item to ensure that the specific meaning it targets contributes to the overall test meaning and interpretation. Consistent with this suggestion, any test item with a meaning that overlaps with the target meaning must be discarded or at least revised.

### **4. Ensure the appropriateness of ESE tasks**

Validity is about the meanings and appropriateness of test scores (Messick, 1989). In order to be valid, the test constructor must ensure that the ESE tasks are appropriate to examinees' characteristics. The study findings have revealed that the ESE includes more test items that target careful reading than test items that target expeditious reading; on the one hand, and more test items that require examinees to read at higher level than test items that require them to read at lower level; on the other hand. However, evidence from the study suggests that test questions that require careful reading are more cognitively demanding than those test questions that require expeditious reading; and test questions that require reading at higher level are more cognitively demanding than those targeting reading at lower level. In light of this evidence, I suggest that the ESE be revised so as to closely reflect the actual context of English/reading instruction by including more test questions that require expeditious reading than test questions that require careful reading; on the one hand, and more test questions that target lower processing level than those that target higher processing level; on the other hand.

## **5. Ensure the construct representativeness of ESE tasks**

One main threat that affects the ESE validity relates to construct representativeness of ESE tasks. The study findings have indicated that the majority of test questions based on narrative and expository texts targeted the general reading aspects while the critical aspects appropriate to narrative structure and exposition structure were simply ignored. In light of this finding, I suggest that the ESE be designed in such a way that it includes test questions that target the relevant detailed aspects of reading construct. In the case of ESE based on narratives for example, test questions should target most critical aspects of the narrative structure such as the identification and description of time and spatial settings, the description and comparison of characters' appearances, personality and emotions, their motives and social relationships, the descriptions and judgment of characters' actions to solve the problem, and the identification and personal judgment of the outcome of the story. Likely, test questions based on expository texts should also target the different descriptors that are used for visualizing the text.

## **6. Revisit the number of questions to be included in the ESE**

The actual number of ESE questions is 9 for general subjects and 5 for technical subjects. However, this cannot permit the inclusion of test tasks so as to ensure the construct representativeness of the ESE tasks. Therefore, I suggest that the actual number of ESE questions be increased to *at least* 20 questions so as to include tasks that target the different reading types as well the different processing levels. A look at the historical development of the ESE reveals that between 1975 and 1977, the ESE included 50 multiple-choice questions, while between 1978 and 1980, it included 20 questions (see Katalayi, 2011).

## **7. Avoid bias in the ESE texts in order to ensure fairness**

Ideally, an assessment should not discriminate between students except on grounds of the ability being assessed. This is to suggest that an assessment should not distinguish between students with different backgrounds although in some cases this is impossible to achieve. In light of this fairness stance, evidence from this study suggests that, although the majority of ESE text

passages depict a social situation relevant to the students' backgrounds, some text passages include details that describe a setting that is familiar to only some groups of students who are living or who have lived in that specific setting. Therefore, I am inclined to conclude that the ESE based on such texts can be biased as other groups of students can hardly comprehend these texts as they can hardly visualize that specific setting.

In light of this evidence, I suggest that, in order to ensure that all candidates are evaluated on the same ground, the ESE developers should ensure that the texts they use are not biased so as to advantage candidates from one geographical setting over those from other geographical settings. Therefore, the content of ESE texts should relate to common issues that relate to all Congolese regardless their towns/villages they reside. Texts with topics related to HIV-AIDS, insecurity, unemployment, political instability, political intolerance, corruption, etc. may accommodate all Congolese students.



#### **8. Diversify the genres of ESE texts**

The study findings have indicated that, although texts used in classroom reading encompass all the four genres (narration, exposition, argumentation and practical writing), the ESE texts are mainly focused on narrative genre with little attention on expository genre while argumentation genre is simply ignored. In light of this finding, I suggest that the ESE be constructed in such a way that it includes various text genres. The predominant use of narrative texts over other text types may be very detrimental to reading instruction as some teachers of English may be tempted to ignore to include the other genres in reading instruction.

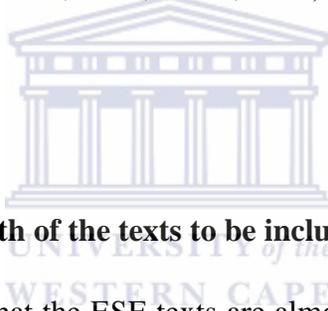
#### **9. Ensure the coherence of ESE texts through the use of transitions and cohesive devices**

The study findings have revealed that the majority of ESE narrative texts are truncated as they do not use transitional words and expressions to indicate the chronology of events. Rather, time sequence is simply expressed through juxtaposition of clauses, sentences and paragraphs. In light

of this finding, I propose that ESE texts be written in such a way that they include transitional words and signaling devices so as to ensure text coherence necessary for comprehension.

#### **10. Ensure that the order of test questions is in alignment with the order of text information**

The study findings have revealed that the majority of ESE test questions are not ordered in such a way to target the order of information in the text. Consistent with this finding, I suggest that the ESE test questions be ordered in such a way that the information required by the test question relates to the way information is ordered in the text. Not relating the order of test questions information to the order of text information can impact on examinees' capacity to comprehend the text and complete test tasks (Kintsch, 1998; Gorin, 2005).



#### **11. Revisit the structure and length of the texts to be included in the ESE**

The study findings have revealed that the ESE texts are almost three times shorter than the texts students actually use in classroom reading. Picking upon this finding, I propose that the ESE texts be longer than they actually are so as to reflect the length of classroom texts and enable examinees not only to read for gist, but also to read for details. Besides, the use of longer texts is beneficial as it can allow for constructing more questions than the actual number of nine or five questions.

#### **12. Ensure the quality of ESE test questions**

The study findings have suggested that the ESE validity is also threatened by the use of unfocused, incomplete and negatively oriented stems. Furthermore, the ESE includes some stems that provide clues to the correct answer. In light of this finding, I propose that, in order to ensure the quality of ESE test questions, the test developers construct test questions with stems that are

focused, complete and positively oriented so as to facilitate examinees' comprehension of test questions. Also, I urge test developers to check all the test questions in order to ensure that none of them includes a stem that provides some clues to the correct answer.

### **13. Revisit the number of alternatives to be included in the ESE**

There is no reason for the ESE to have six alternatives if only one or two may be functioning. Since the study findings suggest that ESE constructors may not have the capacity to construct a test with five functioning distractors, I believe that a three-alternative test with one correct answer and two functioning distractors can be more appropriate to the ESE context. This suggestion is not only in accordance with suggestions from most empirical studies (Haladyna & Downing, 1988; Owen & Froman, 1987; Rodriguez, 2005), but it also reflects examinees' preferences for the actual number of six alternatives to be reduced to three alternatives. Besides, in the Congolese context characterized by paucity of resources, the move from a six- alternative test to a three- alternative test has an additional benefit to allow for more space that can be used to include more questions. This is to suggest that we can easily move from a 9 question test to a 20 question- test without much additional space that can require additional resources.

### **14. Reconsider the use of the implicit alternative 6**

Evidence from this study reveals that the implicit alternative 6 has got multiple meanings. Yet, if we agree that validity is about score meanings and interpretations, we can conclude that the use of the implicit option 6 in the context of ESE threatens the ESE validity. Therefore, I propose that the use of implicit option 6 be strictly prohibited.

## **11.5 Limitations**

Although I believe that I have been able to address the aim and objectives of my study, I also believe that the present study is far from being perfect given its context and setting. Apart from the non-involvement of the Congolese government in the sponsoring of this study and related

financial difficulties due to this non-involvement, I wish to propose only those limitations that I believe appear to weaken the quality of this dissertation.

First, although the multi-componential model I have developed proposes five components (reading construct, reading context, reading types and processing levels, text structure and item structure), I was not able to clearly explain the multiple interactions between these five components. I believe that an attempt to show the interactions among the five components as stated in the suggested model would have produced useful illuminations on how the suggested model articulates. However, I saw such an enterprise too demanding as it could have generated a fairly large amount of data that I would have found daunting to handle and offer more definitive explanations.

Secondly, one of the main research instruments, the strategies questionnaire, requested respondents to write the strategies in the order they had used them. However, this ordering has been overlooked in my analysis. Although existing studies that have used the same instrument have not taken into account the order in which strategies occur during task completion, I believe that this study would have produced more useful insights if my analysis of participants' strategies use had taken account of the order in which participants had deployed the different strategies.

Thirdly, since I did not include as study participants test constructors, it has been difficult to understand the reason why some of the validity issues reported in this study could not have been overcome. The main reason is that the government officials did not wish to assist me in this matter. They might have privileged test secrecy on behalf of research interests and educational issues; or they might have felt uncomfortable to disclose some of the practical behaviours that characterize the ESE in general and the state examination, in particular.

Lastly, methodologically, a simple association of variables (I do not mean correlation) that are predicted to affect the ESE validity could have provided a clearer indication of how various variables can be associated to affect the ESE validity. I deliberately ignored this route as it would have made the analyses too large although I am aware of the benefit it could have provided in terms of relevant insights. In the following section, I propose what I believe can be conducted as further research based on the present study outcome.

## 11.7 Further research

In order to provide a suitable conclusion to this chapter, I propose to present my agenda for the present study results dissemination, the research agenda on the English state examination, and suggestions for further research needed from the present study insights.

### 1. Study results dissemination agenda

In order to be accessible to a wider scholarly audience, I intend to disseminate the results of the present study in different scholarly journals. Although some of the present study results have already been published (see Katalayi & Sivasubramaniam, 2013), I intend to publish the following specific study results because I believe they can illuminate reading assessment research:

- Can reading construct be examined outside the reading context? An investigation of the construct validity of an English as a foreign language test;
- The validity of a multiple-choice reading test based on narrative texts;
- The validity of a multiple-choice reading test based on expository texts;
- Elimination of distractors: A construct-irrelevant strategy? An investigation of test takers strategies use;
- The validity of the DR Congo English state examination implicit ‘option 6’;
- Test context, text density and task complexity: An investigation of the (construct) validity of a reading test used with EFL students.

### 2. Research agenda on the ESE

The validation of a language test like the ESE is a hard process that one cannot pretend to complete in a single study and within an academic time limit. Rather, it is a program that should involve all stakeholders as well as testing experts and researchers. My Master thesis (2011) and this Doctoral dissertation are just a point of departure from which more detailed studies must be conducted in order to improve the quality of the ESE. Since my epistemological stance supports a consideration of the context of reading activity in examining reading construct, there is a need to explore students’ actual reading behaviours in classroom instruction through ethnographic approaches so as to determine which kind of ESE tasks better relate to these specific behaviours.

Such studies will offer the advantage to identify actual reading problems students face in classroom instruction and then align ESE tasks on classroom tasks. Furthermore, there is a need to target the larger context of the Congolese educational setting through studies that investigate stake holders' views on the ESE. These studies will offer the opportunity to understand the actual form and content of the ESE and why some validity problems explored in this study cannot be overcome. Therefore, these studies will enlighten us on the practical aspects that make it difficult to address some issues reported in the present study. At the same time, such studies will likely propose suggestions that are more realistic than those proposed in the present study.

### **3. Further research needed from the present study insights**

In light of the insights generated by the presented study, there is a need for reading assessment researchers to investigate on the following issues in order to take reading assessment research a step further:

The multi-componential model developed in this study to investigate the validity of reading tests needs to be validated through other studies based on other contexts. By applying this model to contexts other than the Congolese context, we are likely to come up with the model limitations; and some suggestions can be made for its improvement.

Since the present study supports that reading construct needs to be examined in the actual context of reading activity, we will be served if the reading tests are aligned on students' actual reading behaviours in classroom. Therefore, one best way to explore students' actual reading behaviours in classroom may be through ethnographic methodologies where the researcher interacts with the students and the reading teacher through classroom observations. As a classroom member, the researcher will be in the position to understand how students actually read, the actual reading problems they face, the quality of reading materials that are used as well as the potentials of the students. As a micro-society, the classroom context will therefore provide the researcher with a clear understanding of reading activity; therefore, the researcher will be able to evaluate the appropriate tasks that connect with the test context. Sivasubramaniam (2004)'s study provides useful methodological insights related to the use of ethnographic methodology in exploring students' classroom reading.

Furthermore, there is a need for researchers to investigate detailed critical aspects of the reading construct in validating reading tests. Most current validation studies focus on general aspects of reading while salient features of reading construct are ignored. By shifting focus to critical aspects of reading construct, we will be able to understand the fine turnings of reading activity. By taking this route, our validation studies will generate suggestions that are clearer, more practical and more useful than they actually are.

Since this study supports that the concept ‘construct validity’ is an evolving concept, there is a need to re-launch the debate on the conceptualization of this concept. Such a debate is useful as it can provide us with the various critical meanings of validity concept; therefore, the substantial degree of confusion that is currently around this concept can be clarified. This debate will focus specifically on the conceptualization of concepts such as ‘score meanings’, ‘validity as appropriateness of test tasks’, ‘construct underrepresentation’ and ‘construct irrelevance variance’.

As a conclusion to this chapter, I wish to mention that there are many issues that can be researched in the scope of insights generated by the current study. However, the challenges are many as well. The main challenge is the significantly low level of involvement of Applied linguists in validation studies. As long as validation research in reading assessment will be dominated by measurement scholars and cognitive psychologists as it appears to be today, we will hardly achieve a relevant evaluation of existing reading tests. This is to suggest that, as long as we will not acknowledge that the construction of good reading tests is not only function of the capacity of the test to discriminate good readers from poor readers (as it seems to be today), but it is also function of the extent to which the test reflects the essential aspects of reading activity in specific contexts, we will be far from producing validation studies with relevant suggestions that can improve the tests being used. At present, reading assessment research and practice seem to ignore the social and cultural evidence of reading behaviour which is visible in the culture itself. Therefore, validation research in reading assessment appears to be a scholarly enterprise with little educational and social benefits. By the same token, the concept ‘validity’ appears to be a difficult concept to both researchers and stakeholders. This should serve as my concluding word to this study.

## REFERENCE LIST

- Adelman, J.S., & Brown, G.D. 2008. Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review*, 115(1), 214–229.
- AERA, APA, & NCME. 1985. *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- AERA, APA, & NCME. 1999. *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- AERA, APA, & NCME. 2011. *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Akhondi, M., Malayeri, F. A., & Samad, A. 2010. How to teach expository text structure to facilitate reading comprehension. *The Reading Teacher*, 64(5), 368-372.
- Albrecht, J.E., & O'Brien, E.J. 1993. Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology; Learning, Memory, and Cognition*, 19, 1061-1070.
- Alderson, C. J., & Banerjee, J. 2002. Language testing and assessment (Part 2). *Language Teaching*, 35(4), 79-113.
- Alderson, C.J. 1993. The relationship between grammar and reading in an English for academic purposes test battery. In Douglas, D. & Chapelle, C., (Eds.) *A new decade of language testing research: selected papers from the 1990 Language Testing Research Colloquium: dedicated in memory of Michael Canale*. Alexandria, VA: Teachers of English to Speakers of Other Languages.
- Alderson, C.J. 2000. *Assessing reading*. Cambridge: Cambridge University Press.
- Alidib, Z. A. 2004. The effects of text genre on foreign language reading comprehension of college elementary and intermediate readers of French. Unpublished Thesis, The Ohio State University, Ohio.

- Altbach, P. G., & Kelly, G. P. 1988. Textbooks in the Third World: policy, content. *Compare*, 17, 93-106.
- Ascalon, M.E., Meyers, L.S., Davis, B.W., & Smits, N. 2007. Distractor similarity and item-stem structure: Effect on item difficulty. *Applied Measurement in Education*, 20(2), 153-170.
- Atai, C., & Soleimany 2009. On the effect of text authenticity and genre on EFL learners' performance in C-Tests. *Pazhuhesh-e Zabanha-ye Khareji*, 49, 109-123.
- Bachman, L., & Palmer, A. 1996. *Language testing in practice*. Shanghai: Shanghai Foreign Language Education Press.
- Bachman, L. 1990. *Fundamental considerations in language testing*. Oxford, Oxford University Press.
- Bacon, S. M., & Finnemann, M. D. 1990. A study of the attitudes, motives, and strategies of university foreign language students and their disposition to authentic oral and written input. *The Modern Language Journal*, 74(4), 459-472.
- Bajtelsmit, J. W. 1975. Development and validation of an adult measure of secondary cue-using strategies on objective examinations: The test of obscure knowledge. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington.
- Bakker, M., Beijaard, D., Roelofs, E., Tigelaar, D., Sanders, P., & Verloop, N. 2008. The impact of construct-irrelevant variance and construct under-representation in assessing teachers' coaching competence. In Bakker, M. *Design and evaluation of video portfolios: Reliability, generalizability, and validity of an authentic performance assessment for teacher*. Openaccess. leidnuiv.nl.
- Barnett, M.A 1986. Syntactic and lexical/semantic skills in foreign language reading: importance and interaction. *Modern Language Journal* 70, 343-349.
- Barry, S., & Lazarte, A. 1998. Evidence for mental models: How do prior knowledge, syntactic complexity, and reading topic affect inference generation in a recall task for nonnative readers of Spanish? *Modern Language Journal*, 98, 176-193.

- Barry, S., & Lazarte, A. 1995. Embedded clause effects on recall: Does high prior knowledge of context domain overcome syntactic complexity in students of Spanish? *Modern Language Journal*, 79, 491-504.
- Beck, I.L., Perfetti, C.A., & McKeown, M.G. 1982. Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology* 74(4), 506-521.
- Becker, M., McElvany, N., & Kortenbruck, M. 2010. Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. *Journal of Educational Psychology*, 102: 773-785.
- Bernhardt, E.B. 1991. *Reading development in a second language*. Norwood, NJ: Ablex.
- Best, R., Ozuru, Y., Floyd, R., & McNamara, D.S. 2006. Children's text comprehension. Effects of genre, knowledge, and text cohesion. In S. A. Barab, K. E. Hay, D. T. Hickey (Eds.), *Proceedings of the Seventh International Conference of the Learning Sciences* (pp. 37-42). Mahwah, NJ: Erlbaum.
- Biancarosa, C., & Snow, C. E. 2006. *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.), Washington, DC: Alliance for Excellent Education.
- Bishay, A. 1996. Teacher motivation and job satisfaction: A study employing the experience sampling method. *J. Undergrad. Sci.* 3, 147-154.
- Block, E. 1986. The comprehension strategies of second language readers, *TESOL Quarterly* 20(3), 463-494.
- Bloome, D. (1985). Reading as a social process. *Language Arts*, 62(2), 134-142.
- Board, C., & Whitney, D. R. 1972. The effect of selected poor item-writing practices on test difficulty, reliability and validity. *Journal of Educational Measurement*, 9, 225-233.
- Borsboom, D., Mellenbergh, G., & van Herdeem, J., 2004. The concept of validity, *Psychological Review* 111, 1061-1071.

- Bossers, B. 1992. *Reading in two languages: a study of reading comprehension in Dutch as a second language and in Turkish as a first language*. Rotterdam, The Netherlands: Drukkerij van Driel.
- Brantmeier, C. 2005. Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. *The Modern Language Journal*, 89, 37-57.
- Brennam, R. L., 1998. Misconceptions at the intersection of measurement theory and practice. *Educational Measurement; Issues and Practice*, 17(1), 5-9.
- Breznitz, Z. 2006. *Fluency in reading: Synchronization of processes*. Mahwah, NJ: Erlbaum.
- Briggs, D.C., Alonzo, A. C., Schwab, C., & Wilson, M. 2006. Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33-63.
- Bribois, J.E. 1995. Connections between first-and second language reading. *Journal of Reading Behaviour*, 27, 565-584.
- Briton, B. K., Glynn, S. M., Mayer, B. J., & Penland, M. J. 1982. Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 74, 51-61.
- Brown, J. D. 1997. An integrated view of context: Implications of miscue analysis. Unpublished doctoral dissertation, University of Arizona, Tucson.
- Brown, J.D. 2000. What is construct validity? *Shiken: JALT Testing & Evaluation SIG Newsletter*, 4(2), 8-12.
- Brualdi, A. 1999. Traditional and modern concepts of validity. ERIC/AE Digest, <http://www.ericdigests.org/2000-3/validity>.
- Bruno, J. E., & Dirkzwager, A. 1995. Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic perspective. *Educ Psychol Meas* 55(6), 959-966.
- Bukenya, M. 2005. Findings and recommendations of the NAPE Report: Levels of competency in literacy and numeracy of primary school Pupils in Uganda. Kampala.

- Canale, M., & Swain, M., 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Carver, R.P. 1992. Effect of prediction activities, prior knowledge, and text type upon amount comprehended: Using raiding theory to critique schema theory research. *Reading Research Quarterly*, 27, 165-174.
- Chalhoub-Deville, M. 1997. Theoretical models, assessment frameworks, and test construction. *Language Testing*, 14(1), 3-22.
- Chamot, A.U., Kupper, L., and Interstate Research Associates, VA. 1989. Learning strategies in foreign language instruction, *Foreign Language Journal*, 22(1), 13-24.
- Chang, C. 2006. Effects of topic familiarity and linguistic difficulty on the reading strategies and mental representations of nonnative readers of Chinese. *Journal of Language and Learning*, 4(2), 172-198.
- Chappelle, A. 2005. *ESOL Tests and testing: A resource for teachers and program administrators offers guidance to professionals in selecting English language tests*. TESOL Publications.
- Cizek, G. J., & O'Day, D. M. 1994. Further investigations of non-functioning options in multiple-choice test items. *Educational and Psychological Measurement* 53(4), 861-872.
- Cizek, G. J. 2012. Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, 17(1), 31-43.
- Cohen, A. & Upton, T. 2007. I want to go back to the text: Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24(2), 209-250.
- Cohen, A. & Upton, T. 2006. *Strategies in responding to the new TOEFL reading tasks*. Educational Testing Services, 666 Rosedale Rd Princeton, NJ 08540-2218.
- Conco, Z. P. 2004. How effective is in-service training for teachers in rural school contexts. Unpublished Master Thesis, University of Pretoria, South Africa.

- Constitution de la République Démocratique du Congo. 2006. Kinshasa, KAS/ LINELIT.
- Crehan, K. D., & Haladyna, T. M. 1989. The validity of two item-writing rules. *Journal of Experimental Education*, 59(2), 183-192.
- Cronbach, L.J. 1980. Validity on parole: How can we go straight? New directions for testing and measurement: Measuring achievement over a decade. *Proceedings of the 1979 ETS Invitational Conference* (pp. 99-108). San Francisco: Jossey-Bass.
- Cronbach, L.J. 1988. Five perspectives on validity argument. In H. Wainer (Ed.), *Test validity*, (pp. 3-17), Hillsdale, NJ: Erlbaum.
- Cronbach, L.J. 1989. Construct validation after thirty years. In R.L. Linn (Ed.). *Intelligence Measurement, theory and public policy* (pp. 147-171). Urbana: University of Illinois Press.
- Crossley, M., & Murby, M. 1994. Textbook provision and the quality of school curriculum in developing countries: Issues and policy options. *Comparative Education* 30(2), 99-114.
- Darling-Hammond, L. 2000. Teacher quality and student achievement: A review of state policy evidence. *Journal of Education Policy Analysis*, 8(1), 145-157.
- Davies, A. 1990. *Principles of Language Testing*. Oxford: Blackwell.
- Davey, B. 1988. Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers, *Journal of Experimental Education*, 56, 67-76.
- Degand, L., & Sanders, T. 2002. The impact of relational markers on expository text comprehension in L1 and L2. *Reading and writing: An interdisciplinary Journal*, 15, 739-757.
- Devitt, S. 1997. Interacting with authentic texts: Multilayered processes. *The Modern Language Journal*, 81(4), 457-469.
- DFID, 2010. *DFID's Performance in 2008–09 and the 2009 White Paper. Fourth Report of Session 2009–10*. London: The Stationery Office Limited.

- Doehring, D. G., Hoshko, I. M. 1977. Classification of reading problems by the Q-technique of factor analysis. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 13(3), 281-294.
- Donaldson, K. 2008. Opportunities gained and lost: Perceptions and experiences of sixth grade students enrolled in a title I reading class. Doctoral dissertation. Retrieved from ERIC.
- Douglas, D. 2000. *Assessing languages for specific purposes*, Cambridge: CUP.
- Downing, S. M., & Haladyna, T. M. 2006. *The handbook of test development*. Mahwah, NJ: Lawrence Erlbaum.
- Downing, S.M., Dawson-Saunders, B., Case, S.M., & Powell, R.D. 1991. The psychometric effects of negative stems, unfocused questions, and heterogeneous options on NBME Part I and Part II characteristics. A paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Downing, S.M. 2003. Validity: On the meaningful interpretation of assessment data. *Medical Education*, 37, 830-837.
- Drum, P.A., Calfee, R.C., & Cook, L.K. 1981. The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, 16(4), 486-514.
- Dudycha, A.L., & Carpenter, J.B. 1973. Effects of item format on item discrimination and difficulty. *Journal of Applied Psychology*, 58(1), 116-121.
- Dymock, D. 2007. *Community adult language, literacy and numeracy provision in Australia: diverse approaches and outcomes*. Adelaide: NCVER.
- Eccles, J.S. 2005. Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C.S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105-121), New York: The Guilford Press.
- Eide, E. R., Goldhaber, D. D., & Showalter, M. H. 2004. Does catholic high school attendance lead to attendance at a more selective college? *Social Science Quarterly*, 85, 1335–1352.

- Elkin, J., & Lonsdale, R. 1997. *Focus on the child: Libraries, literacy and learning*. The Library Association, London.
- Ellsworth, R. A., Dannel, P., & Duell, O. K. 1990. Multiple-choice test items: What are textbook authors telling teachers? *Journal of Educational Research*, 83, 289-293.
- Elo, S. & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115.
- Embretson, S. E., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, 38, 343-368.
- Embretson, S.E. & Wetzel, D. 1987. Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement* 11, 175-93.
- Embretson, S. (1999). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, 3(3), 245-257.
- Embretson, S. 2011. Construct validity: A universal validity system or just another test evaluation procedure? *Educational Researcher*, 40, 3-15.
- Ericsson, K. A., & Simon, H. A. 1993. *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Eysenck, M. W. 2004. *Psychology: An International Perspective*. Psychology Press.
- Faerch, C., & Kasper, G. 1987. From product to process: Introspective methods in second language research. In F. Faerch and G. Kasper (Eds.), *Introspection in second language research* (pp. 5-23). Clevedon: Multilingual Matters Ltd.
- Feldt, L. S., & Brennan, R. L. 1989. Reliability. In R. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed., pp. 103-146). Phoenix, AZ: American Council on Education/Oryx Press.
- Field, J. 2010. Cognitive validity, in L. Taylor (Ed.) *Examining speaking (Studies in Language Testing)*, Cambridge: Cambridge University Press.
- Field, J. 2004. *Psycholinguistics: the Key Concepts*. London: Routledge.

- Fletcher, J.M. 2006. Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), 323-330.
- Freedle, R. & Kostin, I. 1993. The prediction of TOEFL reading item difficulty: Implications for construct validity. *Language Testing* 10(3) 133-170.
- Freedle, R., & Kostin, I. 1992. Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, 16(1), 2-32.
- Freire, P., & Macedo, D. 2013. *Literacy: Reading the word and the world*. London: Routledge.
- Fremer, J., 2000. Promoting high standards and the “problem” with construct validation. *NCME Newsletter* 8(3), 1.
- Gernsbacher, M.A. 1995. Activating knowledge of fictional characters’ emotional status. In C. Weaver, S. Mannes, and C. Fletcher (eds.), *Discourse comprehension: Essays in honor of Walker Kintsch* (pp. 141-155), Hillsdale, NJ: Erlbaum.
- Gibb, G. B. Test wiseness as secondary cue response, (Doctoral dissertation, Stanford University), Ann Arbor, Michigan.
- Gillis, J. J, DeFries, J. C, & Fulker, D. W. 1992. Confirmatory factor analysis of reading and mathematics performance: A twin study. *Acta Genet Med Gemellol* , 41(4):287-300.
- Glazerman, S., Mayer, D., & Decker, P. 2006. Alternative routes to teaching: The impacts of teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management* 25(1), 75-96.
- Glenn, C. 2005. Outcomes assessment in higher education: Using assessment instruments to improve Business curriculum and instruction at Association of Collegiate Business Schools and Programs (ACBSP) (Doctoral Thesis) Philadelphia, PA: Saint Joseph’s University.
- Goddard, R. D., Hoy, W. K., & Hoy, A. W. 2000. Collective efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37, 479-507.

- Goldhaber, D. D., & Brewer, D. J. 2000. Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-146.
- Goldhaber, D. D., & Anthony, E. 2007. Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *Review of Economics and Statistics*, 89(1), 134-150.
- Goodwin, L.D., & Leech, N.L. 2003. The meaning of validity in the new Standards for Educational and Psychological Testing: Implications for measurement courses, *Measurement and Evaluation in Counseling and Development*, 36, 181-191.
- Goodwin, L.D. 1997. Changing conceptions of measurement validity. *Journal of Nursing Education*, 36, 102-107.
- Goodwin, L.D. 2002. Changing conceptions of measurement validity: An update on the new Standards, *Journal of Nursing Education*, 41, 100-106.
- Gorin, J.S. 2007. Reconsidering issues in validity. *Educational Researcher*, 36, 456-462.
- Gorin, J. S. 2006. Test design with cognition in mind. *Educational Measurement: Issues and Practice*, 25, 21–35.
- Gorin, J.S. 2005. Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generation. *Journal of Educational Measurement*, 42(4), 351-373.
- Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36, 193-202.
- Graesser, A., Golding, J. M., & Long, D. L. 1991. Narrative representation and comprehension. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *The handbook of reading research* (pp. 171–205). New York: Longman.
- Graves, M.F. 1986. Vocabulary learning and instruction. In E.Z. Rothkopf (Ed.), *Review of Research in Education*, 13, 49-89.

- Greaney, V. 1996. (Ed.). Promoting reading in developing countries: Views on making reading materials accessible to increase literacy levels (pp. 148-163). Newark, DE: International Reading Association.
- Greenwald, R., Hedges, L., & Laine, R. 1996. The effect of school resources on students' achievement. *Review of Educational Research* 66(3), 361-396.
- Guthrie, J.T., & Kirsch, I.S. 1987. Distinctions between reading comprehension and locating information in the text. *Journal of Educational Psychology*, 79, 220-227.
- Guthrie, J.T., Hoa, A.L., Wigfield, A., Tonks, S.M., Humenick, N.M. & Littles, E. 2007. Reading motivation and reading comprehension growth in later elementary years. *Contemporary Educational Psychology*, 32, 282-313.
- Hacker, D.J. 1998. Self-regulated comprehension during normal reading, In Hacker, D.J., Dunlosky, J. and Graesser, A.C. (Eds.) *Metacognition in education theory and practice*. Mahwah, NJ: Erlbaum, 1-24.
- Haladyna, T.M., & Downing, S. M. 1993. How many options is enough for multiple-choice item? *Educational and Psychological Measurement*, 53, 999-1010.
- Haladyna, T. M. 1994. *Developing and validating multiple-choice test items*. Hillsdale, NJ: Erlbaum.
- Haladyna, T. M., & Downing, S. M. 1988. A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 1, 37-50.
- Haladyna, T. M., Downing, S.M., & Rodriguez, M. C. 2002. A Review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Hammadou, J. 1991. Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *Modern Language Journal*, 75, 27-37.
- Han, B., Dai, M., & Yang, L. 2004. Problems with College English Test as emerged from a survey. *Foreign Languages and their teaching*, 179(2), 17-23.

- Harris, D. & Sass, T. 2001. The effects of teacher training on the teacher value added. *Journal of Teacher Education*, 51, 164-169.
- Haynes, M., & Carr, T. 1990. Writing system background and second language reading: A component skills analysis of English reading by native speakers of Chinese. In T. Carr and B. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 375-421), San Diego, CA: Academic Press.
- Heyneman, S. P., Farrel, J. P., & Sepulveda-Stuardo, M. 1981. Textbooks and achievement in developing countries: What we know. *Journal of Curriculum Studies* 13, 227-246.
- Hill, N. E., & Craft, S. A. 2003. Parent-school involvement and school performance: Mediated pathways among socioeconomically comparable African American and Euro-American families. *Journal of Educational Psychology*, 95, 74-83.
- Hock, T.S. 1990. The role of prior knowledge and language proficiency as predictors of reading comprehension among undergraduates. In J.H.A.L. d. Jong and D.K. Stevenson (Eds.). *Individualizing the assessment of language abilities*. Clevedon, PA: Multilingual Matters.
- Hoover, W.A., & Tunmer, W.E. 1993. The components of reading. In Thompson, G.B., Tunmer, W.E. and Nicholson, T., (Eds.) *Reading acquisition processes*. Clevedon, PA: Multilingual Matters.
- Hoy, W. K., Sweetland, S. R., & Smith, P. A. 2002. Toward an organizational model of achievement in high school: The significance of collective efficacy. *Educational Administration Quarterly*, 38, 77-93.
- <http://monusco.unmissions.org>, 2010
- Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. 2009. The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly*, 25, 4-32.
- Hudson, T. 1982. The effects of induced schemata on the “short-circuit” in L2 reading: Non-decoding factors in L2 reading performance. *Language Learning*, 23, 1-31.

- Hudson, T. 1991. Relationships among IRT item discrimination and item fit indices in criterion-referenced language testing. *Language Testing*, 8(2), 160-181.
- Huff, K. 2006. Using item difficulty to inform descriptive score reports. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.
- Hymes, D. 1972. On communicative competence. In J.B. Pride & J. Holmes (Eds.) *Sociolinguistics: Selected readings* (pp. 267-293). Harmondsworth, Middlesex: Penguin.
- Jahangir, S. F., Saheen, N., & Kazmi, S. F. 2012. In-service training: A contributory factor influencing teachers' performance. *International Journal of Academic Research in Progressive Education and Development*, 1(1), 31- 38.
- Jones, P. W. 1992. *World Bank financing of education, lending, learning and development*. London, Routledge.
- Kalibbala, G. 1999. Sustainable textbook provision and utilization in Uganda PETDP for the Government of Uganda.
- Kane, M. T. 2002. Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Kane, M. T. 2004. Certification testing as an illustration of argument-based validation. *Measurement* 2(3), 135-170.
- Kane, M.T. 2006. Validation. In R.L. Linn (Ed.) *Educational measurement* (4<sup>th</sup> ed., pp. 17-64), Westport, CT: Praeger.
- Kane, M.T. 2008. Terminology, emphasis and utility in validation. *Educational Researcher*, 37, 276-282.
- Kane, M. T., 2011. *Validating score interpretations and use: Messick lecture, Language Testing Research Colloquium*, SAGE: Cambridge.
- Kasai, M. 1997. Application of the rule space model to the reading comprehension section of the Test of English as a Foreign Language (TOEFL). Unpublished doctoral dissertation, University of Illinois, Urbana.

- Kasanga, L. A. 2012. English in the Democratic Republic of the Congo. *World Englishes*, 31(1), 48-69.
- Kasanga, L. A. 2010. Streetwise English and French advertising in multilingual DR Congo: symbolism, modernity, and cosmopolitan identity. *Int'l. J. Soc. Lang.* 206, 181–205.
- Katalayi, G. B. & Sivasubramaniam, S. 2013. Careful reading versus expeditious reading: Investigating the construct validity of a multiple-choice reading test. *Theory and Practice in Language Studies*, 3(6), 877-884.
- Katalayi, G. B. 2011. The DR Congo English state examination: Some fundamental validity issues (Context validity evidence), Master Thesis, Department of Language Education, Faculty of Education, University of Western Cape.
- Katz, S., Lautenschlager, G.J., Blackburn, A.B., & Harris, F.H. 1990. Answering reading comprehension items without passages on the SAT. *Psychological Science*, 1, 122-127.
- Kaufman, D. 2004. Constructivist issues in language learning and teaching. *Annual Review of Applied Linguistics*, 24, 303-319.
- Khalifa, H. & Weir, C.J. 2009. *Examining Reading: Research and practice in assessing second language reading*, *Studies in Language Testing* 29, Cambridge: UCLES/Cambridge University Press.
- Kintsch, W.M., & van Dijk, T. A. 1978. Towards a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kintsch, W. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95, 2, 163-182.
- Kintsch, W. 1994. Text comprehension, memory, and learning. *American Psychologist*, 49, 294-303.
- Kintsch, W. 1998. *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

- Kintsch, W. 2004. The Construction-Integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.) *Theoretical Models and Processes of Reading*. 5th Edition, International Reading Association.
- Kintsch, W., Mandel, T.S., & Kozminsky, E. 1977. Summarizing scrambled stories. *Memory and Cognition*, 5, 547-552.
- Klein, M.L., Peterson, S., & Simington, L. 1991. *Teaching reading in elementary grades*. Needham Heights, Mass.: Allyn and Bacon.
- Knoester, M. 2010. Independent reading and the 'Social Turn': How adolescent reading habits and motivation relate to cultivating social relationships. *Networks*, 12(1), 1-13.
- Kunnan, A. J. 1998. An introduction to structural equating modeling for language assessment research. *Language Testing*, 15(3), 295-352.
- Kuusela, H., & Paul, P. 2000. A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology* 113(3), 387-404.
- LaBerge, D., & Samuels, S. J. 1974. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.
- Lau, K. 2009. Reading motivation, perceptions of reading instruction and reading amount: A comparison of junior and senior secondary students in Hong Kong. *Journal of Research in Reading*, 32(4), 366-382.
- Le Potentiel, 17 (July, 2013). <http://www.lepotentiel.cd>
- Le Potentiel, 21 (June 2013). <http://www.lepotentiel.cd>
- Lee, J., & Shute, V. J. 2010. Personal and social-contextual factors in k-12 academic performance: An integrative perspective on student learning. *Educational Psychologist*, 45(3), 185-202.
- Lewin, K. M. 1987. *Education in austerity: options for planners*, Paris, UNESCO.

- Lewin, K., & Stuart, J. 2003. *The Multi-Site Teacher Education Research Project (MUSTER)* DFID, London.
- Lightman, E. J., McCarthy, P. M., Dufty, D. F., & McNamara, D. S. 2007. Cohesion and structured organization in high school texts. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the twentieth International Florida Artificial Intelligence Research Society Conference* (pp. 235-240). Menlo Park, California: The AAAI Press.
- Lin, J., Chu, K., & Meng, Y. 2010. Distractor rationale taxonomy: Diagnostic assessment of reading with ordered multiple-choice items. Paper presented at the annual meeting of American Educational Research Association, Denver, Colorado.
- Lissitz, R. & Samuelsen, K. 2007. A suggested change in terminology and emphasis regarding validity and education, *Educational Researcher*, 36, 437-448.
- Lockheed, M.E., & Verspoor, A.M. 1991. *Improving primary education in developing countries*, ed. associates, Oxford: Oxford University Press for the World Bank.
- Lonsdale, D. 1997. Modeling cognition in SI: Methodological issues. *International journal of research and practice in interpreting*, 2(1/2):91-117.
- Lu, S. 1999. An investigation into EFL reading processes: Reading effectiveness, inference construction, metacognitive strategy. Unpublished M.A. Dissertation, Zhejiang University, China.
- Mandler, J.M., & Johnson, N.S. 1977. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111–151.
- Mandler, J. M. 1978. A code in the mode: The use of a story schema in retrieval. *Discourse Processes*, 1, 14-35.
- Mannesa, S., & Hoyer, S. M. 1991. Reinstating knowledge during reading: A strategic process. *Discourse Processes*, 21(1), 105-130.
- McMilan, J. H., & Schumacher, S. 2006. *Research in Education: Evidence-based inquiry* (6<sup>th</sup> edition), Pearson, Boston.

- McNamara, T.F., & Lumley, T. 1997. The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. *Language Testing*, 14(2), 140-156.
- McNamara, T. F. & Roever, C. 2006. *Language testing: The social dimension*. Oxford, UK: Blackwell Publishing.
- McNamara, D. F., Kintsch, E., Songer, N., & Kintsch, W. 1996. Are good texts always better? Interactions of texts coherence, background knowledge, and level of understanding in learning from text. *Cognition and Instruction*, 14, 1-43.
- McNamara, T.F. 1990. Item Response Theory and the validation of an ESP test for health professionals. *Language Testing*, 7(1), 52-75.
- McNamara, T. F. 1991. Test dimensionality: IRT analysis of an ESP listening test. *Language Testing*, 8(2), 139-159.
- McNamara, D. S. 2001. Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55, 51-62.
- Messick, S. 1989. Validity. In R. L. Linn (Ed.). *Educational measurement* (3rd ed., pp 13-103). New York: American Council on Education? Macmillan.
- Messick, S., 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Messick, S. 1998, Test validity: A matter of consequence. *Social Indicators Research* 45, 35–44.
- Milanović, M. 2011. The construct of reading and its operationalization in the Internet-Based Test of English as a foreign language. *Philologia*, 9, 73-82.
- Mislevy, R.J. 2004. Can there be reliability without “reliability”? *Journal of Educational and Behavioral Statistics*, 29, 241–244.
- Morgan, P.L. & Fuchs, D. 2007. Is there a bidirectional relationship between children’s reading skills and reading motivation? *Exceptional Children*, 73, 165-183.

- Moss, P. A. 1994. Can there be validity without reliability? *Educational Researcher*, 23(2), 5-12.
- Moss, P.A. 2007. Reconstructing validity. *Educational Researcher*, 36, 470-476.
- Mucherah, W., & Yoder, A. 2008. Motivation for reading and middle school students' performance on standardized testing in reading. *Reading Psychology*, 29(1), 214-235.
- Mufwene, S.S. 2010. Globalization and the spread of English: What does it mean to be Anglophone? *English Today* 101, 26(1), 57-59.
- Murray, J. C., Riazi, A. M., & Cross, J. L. 2012. Test candidates' attitudes and their relationship to demographic and experiential variables: The case of oversea trained teachers in NSW, Australia. *Language Testing* 29(4), 577-595.
- Nassaji, H. 2002. Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning*, 52(2) 439-481.
- Nassaji, H. 2003. L2 Vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4), 645-670.
- Nevo, N.1989. Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing* 6(2), 199-215.
- Norris, J. 2009. Purposeful language assessment. *English Teaching Forum*, 38(1), 18-23.
- Nunan, D. 1991. *Language Teaching Methodology*. New York: Prentice-Hall.
- O'Sullivan, B. 2000. Exploring gender and oral proficiency interview performance. *System*, 28(3), 373-386.
- Olson, M. W. 1985. Text type and reader ability: the effects on paraphrase and text-based inferences questions. *Journal of Reading Behavior*, 3, 199- 214.
- Ory, B., & Ryan, K. E. 1993. *Tips for improving: Testing and grading*. Newbury Park, CA: Sage Publications.

- Ostyn, P., & Godin, P. 1985. An alternative approach to language teaching. *The Modern Language Journal*, 69, 346-55.
- Owen, S.V. & Froman, R.D. 1987. What's wrong with three-option multiple choice items? *Educational and Psychological Measurement*, 47(2), 513-523.
- Ozuru, Y., Rowe, M., O'Reilly, & McNamara, D.S. 2008. Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods*, 40(4), 1001-1015.
- Pang, J. 2010. Research on good and poor reader characteristics: Implication for L2 reading research in China. *Reading in a Foreign Language*, 20, 1, 1-17.
- Park, S. 2009. Verbal report in language testing. *The Journal of Kanda University of International Studies*, 2, 287-307.
- Parkes, J. 2007. Reliability as argument. *Educational Measurement: Issues and Practice* 26(4), 2-10.
- Pelayo, I., & Brewer, D. J. 2010. *Teacher quality in education production*. University of Southern California, Elsevier Ltd.
- Peña, E. D., & Quinn, R. 1997. Task familiarity effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools*, 28, 323-332.
- Perfetti, C. A. 1985. *Reading ability*. Oxford University Press, New York.
- Perfetti, C.A. 1997. Sentences, individual differences, and multiple texts: three issues in text comprehension, *Discourse Processes* 23, 337-355.
- Petrina, S. 2004. *Curriculum and instruction for technology teachers*. Available Online: <http://www.cust.educ.ubc.ca/programs/tsed/>
- Phakiti, A. 2003. A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(3), 26-56.

*Programme National d'Anglais*, 1988 DEPS, Kinshasa.

Purpura, J.E., 1998. An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance, *Language Learning*, 47, 289-325.

Purpura, J. E. 1999. *Learner strategy use and performance on language tests: A structural equation modelling approach*. Cambridge, UK: Cambridge University Press.

Radio Okapi, 22 (June 2013). [http://www. Radiiokapi.net](http://www.Radiiokapi.net)

Radio Okapi, 17 (July, 2013). <http://www. Radiiokapi.net>

Rasti, I. 2009. Iranian candidates' attitudes towards IELTS. *Asian EFL Journal*, 11(3), 75-92.

Rayner, K., & Pollatsek, A. 1989. *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Reeve, C. L., Heggstad, E. D., & Lievens, F. 2009. Modeling the impact of test anxiety and test familiarity on the criterion-related validity of cognitive ability tests. *Intelligence* 37, 34–41.

Retelsdorf, J., Koller, O., & Moller, J. 2010. On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, 21(4), 550-559.

Revue de l'Inspecteur de l'Enseignement, 1984, No. 6, Kinshasa.

Revue de l'Inspecteur de l'Eneignement , 2011, No. 1, Kinshasa.

Rivkin, S.G., Hanushek, E. A., & Kain, J. F. 1998. *Teachers, schools, and academic achievements*. National Bureau of Economic Research, Working Papers 6691.

Rodriguez, M. C. 2005. Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice* 24(2), 3–13.

Rogers, T. W., & Harley, D. 1999. An empirical comparison of three- and four – choice items and tests: Susceptibility to test wiseness and internal consistency reliability. *Educational and Psychological Measurement*, 59(2), 234-247.

- Rosenshine, B.V. 1980. Skill hierarchies in reading comprehension, in R.J. Spiro, B.C. Bruce & W.F. Brewer (Eds.) *Theoretical issues in reading comprehension* (p.535-559), Hillsdale, NJ: Lawrence Erlbaum.
- Rouet, J. F., Vidal-Abarca, E., Erbou, A. B., & Miligo, V. 2001. Effect of information search tasks on the comprehension of instructional text. *Discourse Processes*, 3(2), 163-186.
- Royer, J. 1990. The sentence verification technique: A new direction in the assessment of reading comprehension, In S. Legg & J. Algina (Eds.), *Cognitive assessment of language and math outcomes* (pp. 144-191), Norwood, NJ: Ablex.
- Ruddell, R. B., Ruddell, M. R., & Singer, H. 1994. *Theoretical models and processes of reading*. Newark, Delaware: International Reading Association.
- Rumelhart, D.E., & McClelland, J.L. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1), Cambridge: MIT Press.
- Rupp, A., Ferne, T., & Choi, H. 2006. How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language Testing*, 23(4), 441-474.
- Ryan, R. M., & Deci, E. L. 2000. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54-67.
- Salehi, M. 2011. Test taking strategies: Implications for test validation. *Journal of Language Teaching and Research*, 2(4), 850-858.
- Samupwa, M. 2008. Teacher training and work behaviour. *International Journal of Human Resources*, 65, 88-98.
- Sanders, T. J., & Noordman, L. G. 2000. The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37-60.
- Sarnaci, R. F. 1979. An examination of test-wiseness in the cognitive test domain. *Review of Educational Research*, 49(2), 259-279.

- Sarojani, K. & Krishnan, D. 2011. Careful versus expeditious reading: The case of IELTS reading test. *Academic Research International*, 1(3), 25-35.
- Saw, S. M., & Ng, T. P. 2001. The design and assessment of questionnaires in clinical research. *Singapore Med J*, 42, 131–135.
- Schaefer, J. M. 2009. The effects of stem completeness and stem orientation on multiple-choice item difficulty and discrimination. Master Thesis, Department of Psychology, California State University, Sacramento.
- Schedl, M., Gordon, A., Carey, P.A., & Tang, K.L. 1996. *An analysis of the dimensionality of TOEFL reading comprehension items*, (TOEFL Research Reports 53), Princeton, N.J.: Educational Testing Services.
- Schmeiser, C. B., & Whitney, D. R. 1975. Effect of two selected item-writing practices on test difficulty, discrimination and reliability. *The Journal of Experimental Education*, 43, 30-34.
- Scrase, T. J. 1992. Education and cultural reproduction in India: a content analysis of selected textbooks, in R. J. Burns & A. R. Welch (Eds.), *Contemporary perspectives in comparative education*, pp. 295-326 (New York, Garland).
- Sechrest, L., Kihlstrom, J. F., & Bootzin, R. R. 1993. How to develop multiple-choice tests. *APS Observer*, 10-11, 22-24.
- Seidenberg, M. S., 1989. A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Sharon, L. 2004. Coding qualitative data. In M. S. Lewis-Beck, A. Bryman & T. F. Liao (Eds.), *The SAGE Encyclopedia of social science research methods* (138-139), Sage Publications: Thousand Oaks.
- Sheehan, K.M. & Ginther, A. 2001. What do passage-based MC verbal reasoning items really measure? An analysis of the cognitive skills underlying performance on the current TOEFL reading section. Paper presented at the 2000 Annual Meeting of the National Council of Measurement in Education.

- Shepard, L. A. 1993. Evaluating test validity, *Review of Research in Education*, 19, 405-450.
- Shin, S. 2002. Effects of subskills and text types on Korean EFL reading scores. *Second Language Studies*, 20(2), 107-130.
- Shiotsu, T., & Weir, C.J. 2007. The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance, *Language Testing*, 24(1), 99-128.
- Shohamy, E., 1990. Language testing priorities: a different perspective. *Foreign Language Annals*, 23(5), 385-394.
- Sireci, S. G. 2007. On validity theory and test validation, *Educational Researcher*, 36(8), 477-481.
- Sivasubramanian, S. 2004. An investigation of L2 students' reading and writing in a literature-based language programme growing through responding. Unpublished Thesis, University of Nottingham.
- Smith, R. L., & Smith, J. K. (1988). Differential use of item information by judges using Angoff and Nedelsky procedures. *Journal of Educational Measurement*, 25, 259–274.
- Smith, F. 2004. *Understanding reading: A psycholinguistic analysis of reading and learning to read*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Snow, C.E. 2002. *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Snow, C.E. 2003. Assessment of reading comprehension: Researchers and practitioners helping themselves and each other. In Sweet, A. & Snow, C.E. (Eds.). *Rethinking Reading Comprehension*. (pp. 192-206). New York: The Guilford Press.
- Snyder, L. 2010. Reading expository materials: Are we asking the right questions? *Topics in Language Disorders*, 3(1), 39-47.
- Song, M. Y. 2008. Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing*, 25(4), 435-464.

- Spearritt, D. 1972. Identification of subskills of reading comprehension by Maximum Likelihood Factor Analysis. *Reading Research Quarterly*, 8(1), 92-111.
- Stanovich, K. E. 1994. Constructivism in reading education. *The Journal of Special Education* 28, 259-274.
- Stemler, Steve (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). Retrieved October 16, 2013 from <http://PAREonline.net>
- Swaffer, J. 1985. Reading authentic texts in a foreign language: a cognitive model. *The Modern Language Journal* 69(1), 15-34.
- Swales, J.M. 1990. *Genre analysis*, Cambridge: CUP.
- Sylvia, R. D., & T. Hutchinson. 1985. What makes Ms. Johnson teach? A study of teacher motivation. *Human Relations*, 38, 841- 56.
- Taboada, A. Tonks, S.M., Wiogfield, A. & Guthrie, J.T. 2009. Effects of motivational and cognitive variables on reading comprehension. *Reading and Writing*, 22, 185-209.
- Tamir, P. 1993. Positive and negative multiple-choice items: How different are they? *Studies in Educational Evaluation*, 19, 113-325.
- Taylor, A. K. 2005. Violating conventional wisdom in multiple-choice test construction. *Coll Stud J* 39(1), 213-219.
- Taylor, L. C., Clayton, J. D., & Rowley, S. J. 2004. Academic socialization: Understanding parental influences on children's school- related development in the early years. *Review of General Psychology*, 8, 163-178.
- The World Bank. 2005. *Education in the Democratic Republic of Congo: Priorities and options for regeneration*. The International Bank for Reconstruction and Development / The World Bank 1818 H Street, N.W. Washington, D.C. 20433, U.
- The World Bank. 2013. *The World Development Report*. The International Bank for Reconstruction and Development / The World Bank 1818 H Street, N.W. Washington, D.C. 20433, U.

- Thorndyke, P.W. 1977. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77–110.
- Toulmin, S., 1958. *The Uses of Argument*. Cambridge: Cambridge University Press.
- Ulijn, J. M. 1984. Reading for professional purposes: Psycholinguistic evidence in a cross-linguistic perspective. In A. K. Pugh & J. M. Ulijn (Eds.), *Reading for Professional Purposes*, 66-81. London: Heineman.
- Urquhart, A. S. & Weir, C. J. 1998. *Reading in a second language: Process, product and practice*. Essex: Pearson Education Ltd.
- Urquhart, A. S. 1984. The effect of rhetorical ordering on readability in *Reading in a foreign language*, J. Alderson and Urquhart, London, Longman.
- van Dijk, T.A. and Kintsch, W. 1983. *Strategies of discourse comprehension*. New York: Academic Press.
- van Gompel, R.P. 2007. Syntactic parsing, in G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.
- van Steensel, R., Oostdam, R., & van Gelderen, A. 2013. Assessing reading comprehension in adolescent low achievers: Subskills identification and task specificity. *Language Testing*, 30(1), 3-21.
- Violato, C., & Harasym, P. H. 1987. Effects of structural characteristics of stem format of multiple-choice items on item difficulty and discrimination. *Psychological Reports*, 60, 1259-1262.
- Violato, C., & Marini, A. E. 1989. Effects of stem orientation and completeness of multiple-choice items on item difficulty and discrimination. *Educational and Psychological Measurement*, 49, 287-297.
- Vyas, R. & Supe, A. 2008. Multiple-choice questions: A literature review on the optimal number of options. *Natl Med J India*. 21(3), 130-133.

- Vygotsky, L. S. 1978. *Mind in society: The development of higher psychological processes*, (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.), Cambridge, MA: Harvard University Press.
- Wagenmakers, E. J., Steyvers, M., Raaijmakers, J.G., Shiffrin, R.M., van Rijn, H, & Zeelenberg, R. 2000. A Model for evidence accumulation in the lexical decision task. Paper presented at the 33rd Annual Meeting of the Society for Mathematical Psychology.
- Wallace, C. 1992. *Reading*, Oxford: OUP.
- Weir, C.J., & Khalifa, H. 2008a. A cognitive processing approach towards defining reading comprehension, *Cambridge ESOL: Research Notes*, 31, 2-10.
- Weir, C.J., & Khalifa, H. 2008b. Applying a cognitive processing model to Main Suite reading papers, *Cambridge ESOL: Research Notes* 31, 11-16.
- Weir, C. J., Hawkey, R., Green, A., Unaldi, A., & Devi, S. 2008. The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a British university. *IELTS Research Reports*, 9, 97-156.
- Weir, C., Hawkey, R., Green, A., & Devi, S. 2008. The cognitive processes underlying the academic reading construct as measured by IELTS. *IELTS Research Reports*, 9, 157-189
- Weir, C. J. 2005. *Language testing and validation: an evidence-based approach*. Palgrave, Macmillan
- Widdowson, H. G. 1983. *Teaching language as communication*. Oxford: Oxford University Press.
- Wigfield, A. & Guthrie, J.T. 1997. Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology*, 89, 420-432.
- Wilson, M. 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625-636.
- Xue-chun, H. 2006. A new approach to examining validity: Test-taking strategy investigation. *US-China Foreign Language*, 4(11), 88-93.

- Yali, S. & Jiliang, C. 2007. Effects of text type and test type on L2 reading comprehension test performance. *CELEA Journal*, 30(2), 145-153.
- Yalin, S., & Wei, T. 2011. The relative significance of vocabulary breadth and syntactic knowledge in the prediction of reading comprehension test performance. *Chinese Journal of Applied Linguistics*, 34(3), 113-126.
- Yamachita, J. 1999. Reading in a first and a foreign language: a study of reading comprehension in Japanese (the L1) and English (the L2). Unpublished PhD thesis, Lancaster University.
- Yang, M., & Embretson, S.E. 2007. Construct validity and cognitive diagnostic assessment. In J. P. Leighton & M. J. Gier (Eds.), *Cognitive diagnostic assessment for education: Theory and application* (pp. 119-145), New York: Cambridge University Press.
- Young, J.W. 2008. Ensuring valid content tests for English language learners. *R&D Connections*, 8, ETS Research and Development, Princeton, NJ.
- Yousif, A.N., & Shumaimeri, A. 2006. The effects of content familiarity and language ability on reading comprehension performance of low-and high-ability Saudi tertiary students studying English as a foreign language. *Edu. Sci. & Islamic Studies*, 18(2), 1-19.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. 1995. The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 6, 292-297.

# APPENDICES



UNIVERSITY *of the*  
WESTERN CAPE

## Appendix 1

### State examination performance sheet

REPUBLIQUE DEMOCRATIQUE DU CONGO <i>Ministère de l'Enseignement Primaire Secondaire et Professionnel</i>		EXAMEN D'ETAT SESSION 2009	
INSPECTION GENERALE CENTRE NATIONAL DE CORRECTION C.N.C.			
<b>PERFORMANCE PAR EPREUVE ET PAR ETABLISSEMENT - OPTION</b>			
Province :	KASAI OCCIDENTAL 1	Code :	81
Centre :	MASUIKA	Code :	81021
Option :	LATIN-PHILOSOPHIE	Code :	101
ETOP :	TSHONGA INSTITUT	N° Ordre :	02
Gestion :	NON CONVENTIONNEE	Code :	1
Participants 20 dont Garçons : 20 et Fille : 0			
	MAX	MOY	NBR. REUS./JOUR
Jour 1	20	5	1
Jour 2	20	3	0
Jour 3	20	9	12
Jour 4	25	10	5
HS	15	6	5
Moyenne P.S. : 41 %		Moyenne P.Exet : 27 %	
Diplômés : 14, Garçons : 14 et Fille : 0		% de réussite : 70	
DATE :		SIGNATURE DU RESPONSABLE	

## APPENDIX 2

### Weir and Khalifa's (2008) questionnaire

#### **Part 1: Instructions to participants**

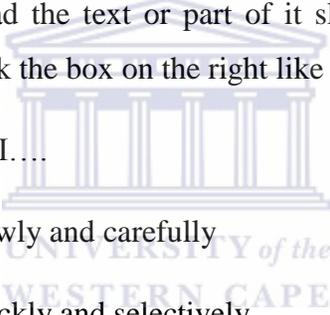
You will have 2 hours to do the test and fill out the questionnaire

Please answer the test questions on the ANSWER sheet provided. After answering each question, please fill out the QUESTIONNAIRE for that question.

#### **Questionnaire Section 1**

In this section of the questionnaire, please describe what you did before you read the test questions. For example, if you read the text or part of it slowly and carefully before reading questions of the test, you should tick the box on the right like this:

Before reading the questions, I....

- 
- a) Read the text or part of it slowly and carefully
  - b) Read the text or part of it quickly and selectively
  - c) Did not read the text

#### **Questionnaire section 2**

After answering each question on the test, please turn immediately to the questionnaire and tick the sentences (1 to 12) that describe what you did when you answered the test question. Then go on to the next test question, and repeat the same procedure until you have answered all the questions

For example, immediately after answering question 1 if you matched words that appeared in the question with exactly the same words in the text, you would tick sentence 1 under Q1. If you also worked out the meaning of a difficult word in the text, you also tick sentence 6.

To find the answer to the question, I tried to...

<b>To find the answer to the question, I tried to...</b>	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1. match words that appeared in the question with exactly the same words in the text								
2. quickly match words that appeared in the question with similar word in the text								
3. look for parts of the text that the writer indicates to be important								
4. read key parts of the text such as the introduction and conclusion								
5. work out the meaning of difficult word in question								
6. work out the meaning of a difficult word in the text								
7. use my knowledge of vocabulary								
8. use my knowledge of grammar								
9. read the text or part of it slowly and carefully								
10. read relevant parts of the text again								
11. use my knowledge of how texts like this are organized								
12. connect information from the text with knowledge I already have								

**I found the answer...**

1 within a single sentence								
2 by putting information together								
3 by putting how information in the whole text fits together								
4 I knew the answer without reading the text								
5 I could not answer the question								

### Appendix 3

#### Strategies codes

<b>Code</b>	<b>Strategy definition</b>
<b>S1</b>	Read the whole text carefully
<b>S2</b>	Read the whole text rapidly
<b>S3</b>	Read a portion of the text carefully
<b>S4</b>	Read a portion of the text rapidly
<b>S5</b>	Use my knowledge of letter writing parts
<b>S6</b>	Use my knowledge of letter writing steps
<b>S7</b>	Use my knowledge of grammar
<b>S8</b>	Use my knowledge of vocabulary
<b>S9</b>	Use context clues to figure out the meaning of difficult words
<b>S10</b>	Draw conclusion based on what the paragraph/text implies
<b>S11</b>	Look for the part(s) that the writer has recommended
<b>S12</b>	Consider the options and postpone consideration of the option
<b>S13</b>	Consider the options and focus on a familiar option
<b>S14</b>	Produce my own answer after reading the question and then look at the six options
<b>S15</b>	Use clues provided in the options
<b>S16</b>	Select the option through paragraph/passage overall comprehension
<b>S17</b>	Select the option through elimination of other options
<b>S18</b>	Select the option through guessing

## Appendix 4

### Strategies questionnaire

#### INSTRUCTIONS /INSTRUCTIONS

- You will have 2 hours to do the test and fill out the questionnaire;

*Vous aurez deux heures pour faire le test et remplir le questionnaire;*

- The questionnaire has three sections; you will have to complete section 1 now.

*Le questionnaire a trois sections; vous devez compléter la section 1 maintenant:*

- After you have completed section 1, then turn to the test, read the text and provide answers to questions based on the text.

*Après avoir compléte la section 1, passez directement au test, lisez les questions et repondez aux questions sur le texte).*

- After answering each test question, please fill out the questionnaire for that question. To fill out the questionnaire, you will have to identify and list in order ALL the strategies you remembered you have used to answer that individual question. For example, if to question 1 you first read the text carefully, and then reread it rapidly and finally look at the question options and produce your own answer after reading the question, you must write in the boxes as follows:

*Après avoir repondu a chaque question, veuillez remplir le questionnaire pour cette question. Pou remplir le questionnaire, vous devez identifier et enumerer en ordre TOUTES les strategies que vous avez pu utiliser afin de repondre cette question. Par exemple, si a la question 1 vous vaez d'abord lu le text attentivement , et ensuite vous l'avez relu rapidement; et enfin vous avez parcouru toutes les assertions et produit votre proper reponse après avoir lu la question, vous allez donc ecrire ans les case suivantes:*

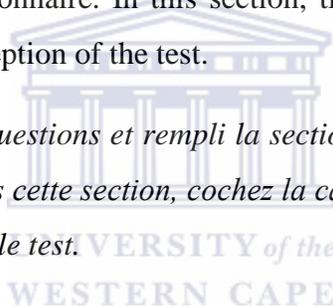
Order of strategies <i>Ordre des strategies</i>	1	2	3	4	5	6	7
Strategies used <i>Strategies utilisees</i>	S1	S2	S14				

- Then go on to the next test question, and repeat the same procedure until you have answered all the questions;

*Ensuite passer a la question suivante, suivez la meme procedure jusqu'a ce que vous repondez a toutes les questions;*

- After you have answered all test questions and filled out section 2 of the questionnaire, then turn to the section 3 of the questionnaire. In this section, tick the box that corresponds to the option that most reflects your perception of the test.

*Apres avoir repondu a toutes les questions et rempli la section 2 du questionnaire, passez onc a la section 3 du questionnaire. Dans cette section, cochez la case qui correspond a l'assertion qui reflète le plus votre perception sur le test.*



## **QUESTIONNAIRE / QUESTIONNAIRE**

### **- Participant demographic details / Informations generales sur le participant:**

1.1. Age / *Age* (Please circle your appropriate age / *Veillez encercler votre age*):

18 – 19 – 20 – 21 – 22 – 23 – 24

1.2. Gender / *Sexe* (Please circle your gender / *Veillez encercler votre sexe*):

Male – Female *Masculin - Feminin*

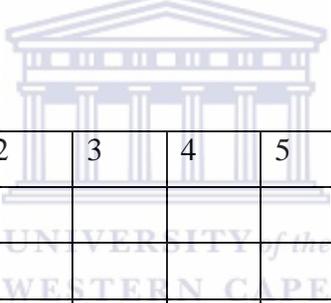
1.3. Subject area (Tick your subject area in the box on the right) :

*Discipline d'etude (Cochez la case correspondant a votre discipline):*

Math-Physics <i>Mathematique- Physique</i>		Bio-Chemistry <i>Biologie-Chimie</i>		Pedagogy <i>Pedagogie</i>		Literary <i>Litteraire</i>	
Com.-Admin. <i>Commerciale- Administrative</i>		Secr.-Computing <i>Secretariat- Informatique</i>		Agriculture <i>Agriculture</i>		Dressmaking <i>Coupe- Couture</i>	

1.4. Languages spoken (Indicate your degree of proficiency on a scale where 1= poor and 5= excellent)

*Langues parlees (Indiquez votre niveau sur une echelle ou 1=niveau bas et 5=excellent niveau)*



Language	1	2	3	4	5
Ciluba					
Swahili					
Lingala					
Kikongo					
French					

### **Section 1: Contextual parameters questionnaire**

#### ***Section 1: Questionnaire se rapportant au contexte***

In this questionnaire, you are requested to tick the box that reflects your views.

*A ce questionnaire, cochez dans la case qui reflète le mieux votre point de vue*

1. How would you rate yourself at English language?

*Comment pourriez-vous vous évaluer votre niveau en anglais?*

1. Very good / *Tres bon*
2. Good / *Bon*
3. Average / *Moyen*
4. Low / *Bas*
5. Very low / *Tres bas*

2. Have you ever written the ESE before?

*Avez-vous déjà présente l'épreuve d'anglais aux exetats avant?*

1. Yes, I have written the ESE before? / *Oui, j'ai déjà présente ca avant*
2. No, I have never written the ESE before / *Non, je n'ai jamais présente ca avant*

3. How often do you read in English at school and at home? Indicate your frequency of reading on a scale where 1=regularly; 2=sometimes; 3= rarely; 4= never

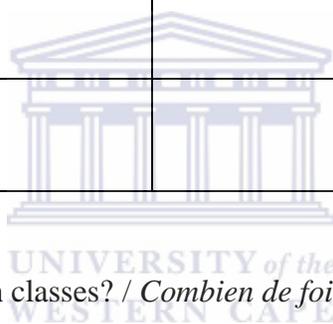
*Combien de fois lisez-vous en anglais a l'école et a la maison? Indiquez votre fréquence de lecture sur une échelle ou 1=regulierement; 2=quelquefois; 3=rarement et 4=jamais*

Frequency of reading		1	2	3	4
1	At school / <i>A l'école</i>				
2	At home / <i>A la maison</i>				

4. Which of the following sources for reading do you use at school and at home?

*Parmi les supports suivants, lesquels utilisez vous pour lire a l'école et a la maison?*

<b>Sources used for reading</b> <i>Supports de lecture</i>		<b>At school /</b> <i>A l'école</i>	<b>At home /</b> <i>A la maison</i>
1	Texts copied from the chalkboard <i>Textes copiés au tableau</i>		
2	Texts photocopied from textbooks <i>Textes photocopiés d'un livre</i>		
3	Textbooks / <i>livre (manuel)</i>		
4	Magazines/newspapers/brochures <i>Magazine/revues/brochures</i>		
5	Books/novels/plays <i>Livre/roman/pièce de théâtre</i>		
6	Internet source <i>Internet</i>		
7	No source <i>Pas de source</i>		



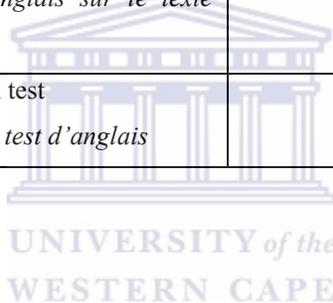
5. How often do you attend English classes? / *Combien de fois assistez-vous au cours d'anglais?*

1. Regularly / *Régulièrement*
2. Sometimes / *Quelquefois*
3. Rarely / *Rarement*
4. Never / *Jamais*

6. For each of the following statements, tick in the box that reflects your attitude to the ESE

*A chacune de ces assertions, cochez la case qui reflète votre attitude vis-à-vis de l'épreuve d'anglais aux examens d'état*

<b>Statements / Assertions</b>		<b>Agree</b> <i>J'accepte</i>	<b>Disagree</b> <i>Je n'accepte pas</i>	<b>Don't know</b> <i>Je ne sais pas</i>
1	All final year students must write an English test <i>Tous les finalists doivent presenter un test d'anglais</i>			
2	The ESE is just a way of failing the candidates to the national test <i>Le test d'anglais est seulement une maniere de faire echouer les candidats au test national</i>			
3	I sometimes feel that I don't need to write the English test <i>Il m'arrive quelquefois de sentir que que je ne dois pas presenter le test d'anglais</i>			
4	I feel that basing the English test on reading comprehension questions only is not fair <i>Je pense que baser le test d'anglais sur le texte seulement n'est pas honnete</i>			
5	I don't mind failing to the English test <i>Je m'en fous meme si j'echoue au test d'anglais</i>			



7. How many hours per week will you like to study English?

*Combien d'heures par semaine preferriez-vous apprendre l'anglais?*

1. one hour per week / *une heure par semaine*
2. two hours per week / *deux heures par semaine*
3. three hours per week / *trois heures par semaine*
4. four hours per week / *quatre heures par semaine*
5. five hours per week / *cinq heures par semaine*
6. more than five hours per week / *plus de cinq heures par semaine*

8. How much time would you like to have to write the ESE?

*Pendant combine d'heures aimeriez-vous presenter le test d'anglais aux exetats?*

1. 1 hour / *1 heure*
2. 1 hour and half / *1 heure et demi*
3. 2 hours / *2 heures*
4. more than two hours / *plus de 2 heures*

9. How many questions would you like the ESE to have?

*Combien de questions voulez-vous que au test d'anglais ?*

1. 3 questions / *3 questions*
2. 5 questions / *5 questions*
3. 9 questions / *9 questions*
4. 12 questions / *12 questions*
5. more than 12 questions / *plus de 12 questions*

10. How many alternatives do you prefer the English state examination to include?

*Combien d'options voudriez-vous que le test d'anglais ait?*

1. 3 alternatives / *3 options*
2. 4 alternatives / *4 options*
3. 5 alternatives / *5 options*
4. 6 alternatives / *6 options*

9. How many parallel forms would you like the ESE to have?

*Combien de series aimeriez-vous que le test d'anglais ait?*

1. Only a single form / *une serie seulement*
2. 2 forms / *2 series*
3. 4 forms / *4 series*

## Section 2: Cognitive-metacognitive strategies questionnaire

In this section, you are requested to read the text and then answer to questions based on it. Immediately after providing answer to each individual item, please write in the boxes the order of strategies you used to respond that individual question. Here are the strategies and their codes (see Appendix 3)

*Dans cette section, lisez le texte et ensuite repondez aux questions basees sur ce texte. Immédiatement après avoir répondu a chaque question, veuillez écrire dans les cases et en ordre les strategies que vous avez utilisees pour reponre a chacune de ces questions. Voici la liste des strategies (voir Annexe 3)*

## Section 3: Your perception of text and test difficulty

### *Section 3: Votre perception de la diffuculte du texte et du test*

In this section, you are instructed to report your perceptions of text and questions difficulty immediately after you have completed the task.

*Dans cette section, vous etes convies a reporter vos perceptions en ce qui concerne la difficulte du texte et du test*

1. Rank your perception of difficulty on a scale from 1 to 5 where 1 expresses the most easy and 5 the most difficult

*Indiquez le degre de difficulte sur une echelle allant de 1 a 5, ou 1 signifie plus facile et 5 plus difficile*

	Statements / <i>Propositions</i>	1	2	3	4	5
1	Text content difficulty / <i>contenu du texte</i>					
2	Test questions difficulty / <i>questions sur le texte</i>					
3	Grammar difficulty / <i>grammaire</i>					

4	Vocabulary difficulty / <i>vocabulaire</i>								
---	--	--	--	--	--	--	--	--	--

2. Tick in each box questions you have found

*Cochez dans chace case les questions que vous aviez trouvee*

Questions I found <i>Les questions que j'ai trouvees</i>	1	2	3	4	5	6	7	8	9
1 Easy questions / <i>Faciles</i>									
2 Difficult questions / <i>Difficiles</i>									
3 Moderate difficult questions <i>Passables</i>									
4 Questions I could not answer <i>Que je ne pouvais pas repondre</i>									

3. How did you perceive your familiarity with the text content?

*Comment aviez-vous percu votre familiarite avec le contenu du texte*

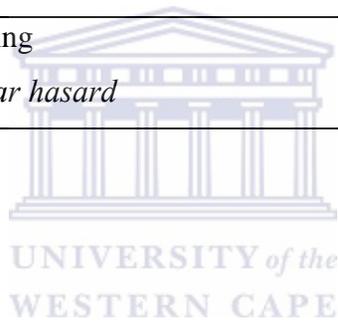
1	Familiar / <i>Familier</i>	
2	Somewhat familiar / <i>Plus au moins familier</i>	
3	Not familiar / <i>Non familier</i>	

4. If you have selected option 6, tick all options that correspond to your choice of this option

*Si vous avez choisi l'option 6, cochez toutes les options correspondant a votre choix de cette option*

1	I selected 6 because there was no best option among the suggested 5 options <i>J'ai choisi 6 parce qu'il n'y avait pas de bonne reponse parmi les 5</i>	
---	--	--

	<i>options suggerees</i>	
2	I selected 6 because I could not find the correct answer <i>J'ai choisi 6 parce que je ne pouvais pas trouver la bonne reponse</i>	
3	I selected 6 because there were more than one correct answer options <i>J'ai choisi 6 parce qu'il y avait plus d'une bonne reponse</i>	
4	I selected 6 because I was confused with the other 5 suggested options <i>J'ai choisi 6 parce que j' avais la confusion entre les autres 5 options suggerees</i>	
5	I selected 6 because I sensed that it could also be part of the correct options <i>J'ai choisi 6 parce que j'ai compris que cette option devais aussi faire partie de bonne reponses</i>	
6	I selected 6 just by guessing <i>J'ai choisi 6 seulement par hasard</i>	



**Appendix 5: Contextual questionnaire administered to final year students**

**Annexe 5: Questionnaire contextual administred aux eleves finalistes**

**1. Participant demographic details / Informations generales sur le participant:**

1.1. Age / Age (Please circle your appropriate age / *Veillez encercler votre age*):

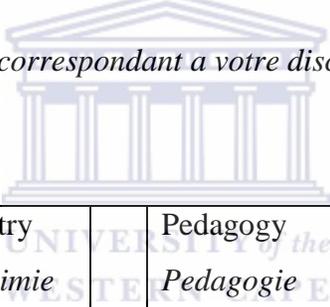
18 – 19 – 20 – 21 – 22 – 23 – 24

1.2. Gender / Sexe (Please circle your gender / *Veillez encercler votre sexe*):

Male – Female *Masculin - Feminin*

1.3. Subject area (Tick your subject area in the box on the right) :

*Discipline d'etude (Cochez la case correspondant a votre discipline):*



Math-Physics <i>Mathematique- Physique</i>		Bio-Chemistry <i>Biologie-Chimie</i>		Pedagogy <i>Pedagogie</i>		Literary <i>Litteraire</i>	
Com.-Admin. <i>Commerciale- Administrative</i>		Secr.-Computing <i>Secretariat- Informatique</i>		Agriculture <i>Agriculture</i>		Dressmaking <i>Coupe- Couture</i>	

1.4. Languages spoken (Indicate your degree of proficiency on a scale where 1= poor and 5= excellent)

*Langues parrees (Indiquez votre niveau sur une echelle ou 1=niveau bas et 5=excellent niveau)*

<b>Language / Langue</b>	1	2	3	4	5
Ciluba					
Swahili					

Lingala					
Kikongo					
French					

## 2. Questionnaire

In this questionnaire, you are requested to tick the box that reflects your views. Then after, you have to explain your choice.

*A ce questionnaire, cochez dans la case qui reflète le mieux votre point de vue*

1. How would you rate yourself at English language?

*Comment pourriez-vous vous évaluer votre niveau en anglais?*

1. Very good / *Tres bon*
3. Good / *Bon*
4. Average / *Moyen*
5. Poor / *Faible*



Comment on/Explain your choices

*Commentez/ Expliquer votre choix*

.....

.....

.....

.....

.....

2. How often do you read in English at school and at home? Indicate your frequency of reading on a scale where 1=regularly, 2=sometimes, 3= rarely and 4= never

*Combien de fois lisez-vous en anglais a l'école et a la maison? Indiquez votre fréquence de lecture sur une échelle ou 1=régulièrement; 2=quelquefois; 3=rarement et 4=jamais*

<b>Frequency of reading</b> <i>Frequence de lecture</i>		1	2	3	4
1	At school / <i>A l'ecole</i>				
2	At home / <i>A la maison</i>				

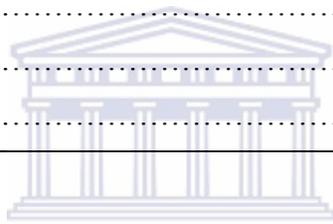
Comment on/Explain your choices  
*Commentez/ Expliquer votre choix*

.....

.....

.....

.....



3. Which of the following sources for reading do you use at school and at home?

*Parmi les supports suivants, lesquels utilisez vous pour lire a l'ecole et a la maison?*

<b>Sources used for reading</b> <i>Supports de lecture</i>		<b>At school</b> <i>A l'ecole</i>	<b>At home</b> <i>A la maison</i>
1	Texts copied from the chalkboard <i>Textes copied au tableau</i>		
2	Texts photocopied from textbooks <i>Textes photocopies d' un livre</i>		
3	Textbooks / <i>livre (manuel)</i>		
4	Magazines/newspapers/brochures <i>Magasine/revues/brochures</i>		
5	Books/novels/plays <i>Livre/roman/piece de theatre</i>		

6	Internet source <i>Internet</i>		
7	No source		

Comment on/Explain your choices  
*Commentez/ Expliquer votre choix*

.....

.....

.....

.....

.....

.....



4. How often do you attend English classes? / *Combien de fois assistez-vous au cours d'anglais?*

- 1. Regularly / *Régulièrement*
- 2. Sometimes / *Quelquefois*
- 3. Rarely / *Rarement*
- 4. Never / *Jamais*

Comment on/Explain your choice  
*Commentez/ Expliquer votre choix*

.....

.....

.....

.....

5. For each of the following statements, tick in the box that reflects your attitude to the ESE

A chacune de ces assertions, cochez la case qui reflète votre attitude vis-à-vis de l'épreuve d'anglais aux examens d'état

Statements / Assertions		Agree <i>J'accepte</i>	Disagree <i>Je n'accepte pas</i>	Don't know <i>Je ne sais pas</i>
1	All final year students must write an English test <i>Tous les finalistes doivent présenter un test d'anglais</i>			
2	The ESE is just a way of failing the candidates to the national test <i>Le test d'anglais est seulement une manière de faire échouer les candidats au test national</i>			
3	I sometimes feel that I don't need to write the English test <i>Il m'arrive quelquefois de sentir que je ne dois pas présenter le test d'anglais</i>			
4	I feel that basing the English test on reading comprehension questions only is not fair <i>Je pense que baser le test d'anglais sur le texte seulement n'est pas honnête</i>			
5	I don't mind failing to the English test <i>Je m'en fous même si j'échoue au test d'anglais</i>			

<p>Comment on/Explain your choices <i>Commentez/ Expliquer votre choix</i></p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p>
---

6. How many hours per week will you like to study English?

*Combien d'heures par semaine préféreriez-vous apprendre l'anglais?*

1. one hour per week / *une heure par semaine*
2. two hours per week / *deux heures par semaine*
3. three hours per week / *trois heures par semaine*
4. four hours per week / *quatre heures par semaine*
5. five hours per week / *cinq heures par semaine*
6. more than five hours per week / *plus de cinq heures par semaine*

Comment on/Explain your choice

*Commentez/ Expliquer votre choix*

.....

.....

.....

.....

.....

7. How much time would you like to have to write the ESE?

*Pendant combien d'heures aimeriez-vous présenter le test d'anglais aux exetats?*

1. 1 hour / *1 heure*
2. 1 hour and half / *1 heure et demi*
3. 2 hours / *2 heures*
4. more than two hours / *plus de 2 heures*

Comment on/explain your choice

*Commentez/ Expliquer votre choix*

.....

.....

.....

.....

.....

8. How many questions would you like the ESE to have?

*Combien de questions voulez-vous que au test d'anglais ?*

1. 3 questions / 3 questions
2. 5 questions / 5 questions
3. 9 questions / 9 questions
4. 12 questions / 12 questions
5. more than 12 questions / plus de 12 questions

Comment on/explain your choice

*Commentez/ Expliquer votre choix*

.....

.....

.....

.....

.....

9. How many alternatives do you prefer the English state examination to include?

*Combien d'options voudriez-vous que le test d'anglais ait?*

1. 3 alternatives / 3 options
2. 4 alternatives / 4 options
3. 5 alternatives / 5 options
4. 6 alternatives / 6 options

Comment on/explain your choice

*Commentez/ Expliquer votre choix*

.....

.....

.....

.....

.....

10. How many parallel forms would you like the ESE to have?

*Combien de series aimeriez-vous que le test d'anglais ait?*

1. Only a single form / *une serie seulement*
2. 2 forms / *2 series*
3. 4 forms / *4 series*

Comment on/explain your choice

.....

.....

.....

.....

.....



**Appendix 6**

**Questionnaire administered to Grade 12 English language teachers**

1. Indicate your highest degree in education:

Degree		
1	Licence (Bed/Honours)	
2	Graduat (BEd)	
3	Diplome d'Etat (National State Certificate)	

2. How long have been teaching English in Grade 12 secondary school? Select a group

Your experience teaching in Grade 6 high school		
1	Less than 3 years	
2	Between 3 years and 5 years	
3	More than 5 years	

3. How often do you use each of these methods to evaluate your students?

Testing methods used		Usually	Sometimes	Rarely	Never
1	Open questions				
2	Multiple-choice questions				
3	Oral questions				
4	Essay questions				
5	Yes-No questions				
6	True-False questions				
7	Debates				

<p>Comment on/select your choices</p> <p>.....</p> <p>.....</p> <p>.....</p>
--

.....

4. How often do you use the following sources of materials for your reading classes?

Materials		Usually	Sometimes	Rarely	Never
1	Books				
2	Magazines				
3	Newspapers				
4	Internet sites				

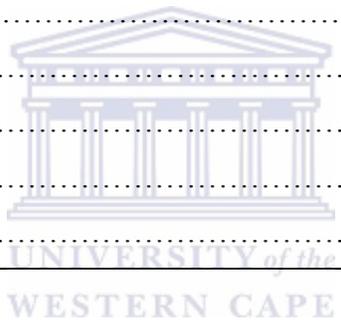
Explain/Comment: .....

.....

.....

.....

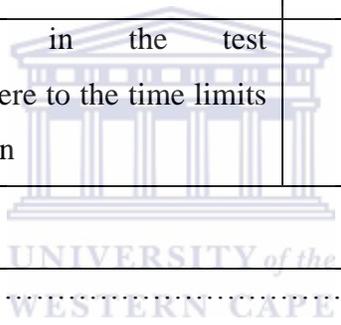
.....



5. What is your view on each of the following statements?

Your view on the following		Agree	Disagree	Can't tell
1	I have all textbooks required for teaching English in Grade 12			
2	My students also have textbooks			
3	I have the English curriculum			
4	I use the English curriculum to design my course outline			
5	The English curriculum needs to be updated			
6	In reading classes, my students have difficulty with vocabulary			
7	In reading classes, my students have			

	difficulty with grammar			
8	In reading classes, my students have difficulty with text content			
9	I am not motivated to teach English			
10	My poor working conditions negatively affect my teaching			
11	The ESE must also test other skills (writing, speaking, and listening)			
12	The actual time allocated for writing the test is too much and therefore must be reduced			
13	Those in charge with test administration do not sometimes respect time limits			
14	Everybody involved in the test administration must adhere to the time limits during test administration			



Explain/Comment ..... ..... ..... ..... .....
---

6. What is the degree of your students' motivation in reading classes and while taking classroom tests?

Student's degree of motivation		Very high	High	Moderate	Low
1	Reading classes				
2	Reading tests				

Explain/Comment ..... ..... ..... ..... .....
---

7. Tick from this list what you think are variables that negatively affect the reading classes in particular and English classes in general

Variables		
1	Large number of students	
2	Lack of students' involvement in English classes	
3	Students' poor level of instruction that does not reflect the grade level	
4	Students' expectation that they will collaborate when they take the ESE	
4	The policy adopted by the school in terms of discipline	
5	Lack of collaboration among teachers	
6	Lack of parents' involvement in their children's studies	
7	Poor quality of English curriculum	
8	Lack of involvement of English teachers in test construction	
9	Lack of teachers' in-service training	
10	Teaching programmes not usually completed	

Explain/Comment ..... ..... .....
---

.....
.....



**Appendix 7**  
**Test materials**

## **Section 1: T1**

### **English text**

At the rural market of Burungangu, Yasekuru sells groundnuts, maize, rice cakes and soya beans. She appears happy and healthy. Her cheerful smile but hides a tragic story.

In 1997, Yasekuru and Kirongozi were overjoyed at the birth of their daughter Asha. At the outset Asha seemed to be in good health. Yet, she stopped gaining weight and contracted one infection to another. At the age of three, Asha died from AIDS (Acquired Immuno-Deficiency Syndrome).

A few years later, Kirongozi who was a taxi driver from Worobe to Elaka also began to get sick. One day he collapsed and was taken to Wango Medical Centre. Doctors could not save him. Yasekuru's husband of eight years died of Aids-related complications.

Yasekuru now lives alone in one room's hut in the suburbs of Burungangu. One might expect that she would be beginning to rebuild her life. She, however, explains: "I have got HIV. I will not get married. I will not have any more children."

Sadly, such experiences are hardly unique in Democratic Republic of Congo (DRC) and Africa. Women can be infected through normal sexual intercourse and are at special risk of contracting HIV. In one hand, they mustn't discuss sexuality. In the other, African men commonly abuse many sexual partners. This is why women risk sex abuse if they refuse. To exclude women from Aids matters does not favor HIV prevention.

### **INSTRUCTIONS**

- Read the above text carefully two or three times and answer the questions based on it.
- Choose the correct answer among those proposed (1, 2, 3, 4, or 5)
- If there is no good answer, choose 6

### **QUESTIONS FROM THE TEXT**

1 The title which best suits the fourth paragraph of the above text is:

1. Yasekuru's sickness
2. Kirongozi's death
3. A tragic experience on aids
4. African women and aids
5. Widow Yasekuru's situation

2 The event that affected deeply Yasekuru's life is found in the :

<ol style="list-style-type: none"> <li>1. fifth paragraph</li> <li>2. second paragraph</li> <li>3. first and fourth paragraphs</li> <li>4. second and third paragraphs</li> <li>5. first paragraph</li> </ol>
<p>3 According to the text, Yasekuru’s last decision for her life was:</p> <ol style="list-style-type: none"> <li>1. HIV prevention</li> <li>2. aids-related complications</li> <li>3. refusal to get married again</li> <li>4. being married by a man with many sexual partners</li> <li>5. Asha’s death of aids</li> </ol>
<p>4 The contrary of “hardly” underlined in the last paragraph of the text can be:</p> <ol style="list-style-type: none"> <li>1. happily</li> <li>2. commonly</li> <li>3. rarely</li> <li>4. joyfully</li> <li>5. unhappily</li> </ol>
<p>5 All the following sentences agree with the text, except:</p> <ol style="list-style-type: none"> <li>1. Yasekuru will not get married because she is aware of her sickness</li> <li>2. Kirongozi suffered from a well-known illness</li> <li>3. Yasekuru’s husband had a lot of sexual partners</li> <li>4. Yasekuru’s husband was a car driver</li> <li>5. Kirongozi died at the eighth year of marriage</li> </ol>
<p>5 One of the following groups of sentences is a paragraph</p> <ol style="list-style-type: none"> <li>1. Kirongozi had a wife. She was called Yasekuru. Men were dying after a long and painful disease</li> <li>2. Kirongozi had a wife who was called Yasekuru. Many people all over the world have died and are still dying of aids. The DRD is in the centre of Africa</li> <li>3. Kirongozi who was Yasekuru’s husband was a taxi bike driver. He died after a long and painful disease. Kabila is the president of the DRC.</li> <li>4. Kirongozi and Yasekuru shared a very nice life. The elections in the DRC will take place at the end of July</li> <li>5. Yasekuru sold a number of items at the market, among them there were maize, rice cakes. She had a daughter called Asha who died of aids.</li> </ol>
<p>6 “She” underlined in the second paragraph is used instead of:</p> <ol style="list-style-type: none"> <li>1. Asha</li> <li>2. Yasekuru</li> <li>3. Kirongozi</li> </ol>

- 4. Woman
- 5. Doctors

**Section 2: T2**

**English text**

That Wednesday morning a large number of drivers drove their car and taxi buses through that irritating black smoke along MABANGA ROAD over the National Commercial School’s fence. Some stopped abruptly. Passengers were puzzled and wondered whether they could continue their trip. Meanwhile, facing the police, flocks of furious young men were feverishly throwing to their opponents any kind of harmful objects: pieces of stones, bricks, iron sticks... Whereas men in dark blue uniform were launching tears’ gaz from their guns in order to disperse those crazy students. Yet, their security equipment could not protect them from the anger of their close neighbours.

Worldwide radio and TV reporters announced that the situation was getting worse at the National Commercial School. A lorry bumped a tree and was badly damaged. Its driver was mortally wounded. APOLOSA, a well-known policeman, was seriously injured. One student died; sexual abuses were committed in ladies’ residential flats. The so-called “strong men” were accused for that offence. Phone cells, golden jewelry and money robbery was also observed. Some demonstrators were caught to be jailed.

The incident broke up because students hardly stand the police presence close to them. Among other reasons, the Congo Technical Institute’s students leaders was deadly beaten in one ladies’ residential flat as he went to pay a visit to his girl-friend two days ago. Therefore, his mates might have promised to invade that school. The Chief District of police then decided the protection of the school from the revenge of the Congo Technical Institute’s students. Unfortunately the belief of the students could not agree with the police.

ANONYMOUS AUTHOR.

Note that:

C.T.I. stands for CONGO TECHNICAL INSTITUTE

N.C.S. stands for NATIONAL COMMERCIAL SCHOOL

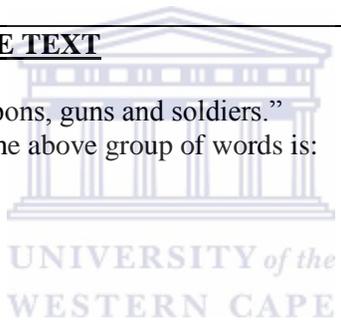
**INSTRUCTIONS**

- Read the above text carefully two or three times and answer the questions based on it.
- Choose the correct answer among those proposed (1, 2, 3, 4, 5).
- If no good answer, choose “6”

**QUESTIONS FROM THE TEXT**

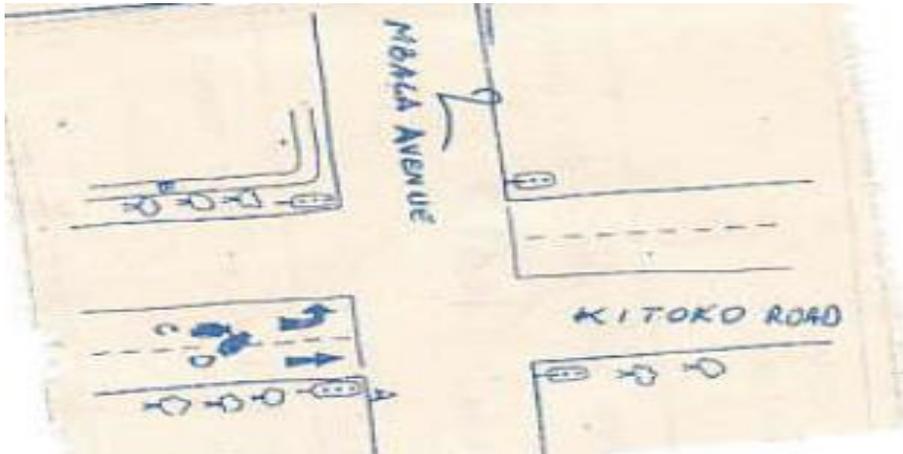
<p>1 The title which best suits the text is:</p> <ol style="list-style-type: none"> <li>1. Trouble in the N.C.S</li> <li>2. Violent disturbance of the peace by the police and the students</li> <li>3. The results of trouble in the N.C.S.</li> <li>4. The accident in the N.C.S.</li> <li>5. The causes of the trouble in the N.C.S.</li> </ol>
<p>2 All the following results of the incident at the N.C.S. are correct, except:</p> <ol style="list-style-type: none"> <li>1. golden jewelry robbery was observed</li> <li>2. sexual abuses were committed</li> <li>3. the lorry driver was mortally wounded</li> <li>4. One student died</li> <li>5. Apolosa was badly beaten</li> </ol>
<p>3 The reason of the incident at the N.C.S. which is a particular concern of the C.T.I.'s students is:</p> <ol style="list-style-type: none"> <li>1. the police anger on Apolosa's injury.</li> <li>2. the probable revenge of the C.T.I.'s students.</li> <li>3. the refusal of the N.C.S.'s students to stand the police presence close to them</li> <li>4. the police intervention to protect the N.C.S.</li> <li>5. The severe harm inflicted to the leader of the C.T.I's students by the N.C.S.'s ones.</li> </ol>
<p>4 The following objects were used by students in the incident at the N.C.S., except:</p> <ol style="list-style-type: none"> <li>1. pieces of stones</li> <li>2. Any harmful objects</li> <li>3. Iron sticks</li> <li>4. tears'gaz</li> <li>5. Pieces of bricks</li> </ol>
<p>5 In the last paragraph of the text, the underlined "his mates" replaces:</p> <ol style="list-style-type: none"> <li>1. The police</li> <li>2. The school teacher</li> <li>3. The N.C.S.'s students</li> <li>4. The C.T.I.'s students</li> <li>5. The lorry driver</li> </ol>
<p>6 All the following sentences agree with the text, except:</p> <ol style="list-style-type: none"> <li>1. iron sticks were launched to policemen</li> <li>2. the driver died as the lorry bumped a tree</li> <li>3. Apolosa was injured by a harmful object</li> <li>4. the Chief District of police ordered to protect the N.C.S.</li> </ol>

<p>5. sexual abuses were committed by policemen</p>
<p>7 Imagine you were present at the moment the incident took place. As a witness, the police might ask you one of the following questions to learn about the killer’s identity.</p> <ol style="list-style-type: none"> <li>1. “Who killed the student?”</li> <li>2. How many people died during the incident?”</li> <li>3. “How old was the dead student?”</li> <li>4. “When did the incident happen?”</li> <li>5. “What are the causes of the incident?”</li> </ol>
<p>8 With reference to the results of the incident, one might suggest the following pieces of advice, except:</p> <ol style="list-style-type: none"> <li>1. avoid violence in every similar case in the future.</li> <li>2. keep away from violence at all costs since it brings more harm than good.</li> <li>3. avoid exchanging views with the other party.</li> <li>4. get in touch each other to look closely at the causes of the incident for a better future.</li> </ol>
<p><b><u>QUESTIONS OUT OF THE TEXT</u></b></p> <p>9 “Tears’ gaz, fire-arms, weapons, guns and soldiers.” The title which best suits the above group of words is:</p> <ol style="list-style-type: none"> <li>1. army</li> <li>2. communication</li> <li>3. transportation</li> <li>4. hospital</li> <li>5. school</li> </ol>



**Section 3: T3**

### English text



On April 18<sup>th</sup> at 9.15 hrs, I was standing on the corner of Mbala Avenue and Kitoko road. I heard a loud crash and turning round I saw two cars stopped in the middle of the road. I did not see the crash itself. When I arrived on the scene the two drivers had got out of their vehicles and were arguing noisily.

Mr. Mbela, 56, an engineer had crashed into the side door of the white Mazda belonging to Ms. Mujinga, 25, a headmistress. There was a witness, Mrs. Mafuta, 72, who was walking along the opposite side of the road with her dog.

Mr. Mbela had rudely accused Ms Mujinga of being careless and not signaling. Ms. Mujinga said she had signaled but she was upset and her account of the accident was not clear. She explained that she had been driving to work and was going to turn left at the traffic lights into Mbala Avenue. I suspected, since it was already 9.15, that Ms Mujinga was late and in a hurry. When I questioned her she admitted this was true.

Mrs. Mafuta also gave her own account of the accident. She seemed to sympathize with Ms Mujinga and supported her. Since she was walking along the pavement her view may have been obscured by the row of trees between the pavement and the road. Mr. Mbela tried to get Mrs. Mafuta to say he was right and it was Ms Mujinga's fault.

After several minutes each one had explained what they had seen and there had been much agreement and contradiction, I finally calmed everyone down and took their names and addresses. Since nobody was injured they continued on their way.

### INSTRUCTIONS

- Read the above text carefully two or three times and answer the questions based on it.
- Choose the correct answer among those proposed (1, 2, 3, 4, 5).

- If no good answer, choose “6”

**QUESTIONS ON THE TEXT**

1 According to the text, the accident happened:

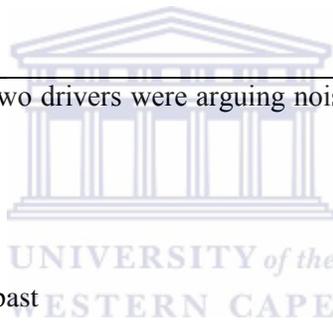
1. in the evening
2. in the afternoon
3. in the morning
4. at 9.15 hrs in the evening
5. at night

2 The word “I” used throughout the text replaces:

1. the author
2. Mr. Mbela
3. Miss Mujinga
4. Mrs. Mafuta
5. the witness

3 The verb “were arguing” in “the two drivers were arguing noisily”, is used to express an action that:

1. happens regularly
2. is going to happen
3. is happening at present
4. is completely finished in the past
5. went on for a moment in the past



4 According to the text:

1. Miss Mujinga is as young as Mr. Mbela
2. Mrs. Mafuta is older than Mr. Mbela
3. Mr. Mbela is younger than Miss Mujinga
4. Mr. Mbela is older than Mrs. Mafuta
5. Mrs. Mafuta is as old as Mr. Mbela

5 According to the above diagram, the accident could have been avoided if:

1. Miss Mujinga had gone straight on
2. M. Mbela had been absent minded
3. Miss Mujinga had turned left
4. Mr. Mbela had gone straight on
5. Mr. Mbela hadn't turned left

6 The following things appear on the diagram, except:

1. traffic lights
2. cars

3. a dog
4. trees
5. arrows

7 Match the letters on the diagram with the characters of the text and choose the correct answer

Letters

- A
- B
- C
- D

Characters

1. Miss Mujinga
2. Author
3. M. Mbela
4. Mrs. Mafuta

1. A2-B3-C4-D1
2. A3-B1-C2-D4
3. A1-B2-C3-D4
4. A2-B4-C3-D1
5. A4-B3-C1-D2

8 Mrs Mafuta gave her own account of the accident, ... she?

1. did
2. didn't
3. doesn't
4. does
5. hadn't



9 "I finally calmed everyone down." The author says he finally.... everyone down

1. calms
2. had calmed
3. calmed
4. would calmed
5. has calmed

#### **SECTION 4: T4**

## English text

Look at large urban centres. They are great cities of the world. We often do not think about how they were originally planned. Or what decisions people took to make them what we call “great”. However, urban planning requires a city to be beautiful, the traffic to move freely, communications and commercial centres to operate effectively and people to live in clean, safe and healthy conditions. So we need organized plans and well-conceived designs at all stages of the city’s birth and development.

The art of designing urban centres is not new. Historians have recorded the plans of many ancient cities. Archaeologists can trace signs of urban planning in the cities such as Mesopotamia and South America, and through Asia. When Alexander the Great walked along the Mediterranean beach, he measured his steps and pointed out where he wanted the commercial centre, the Egyptian city of Alexandria to be built: where the main harbor should be located, where temples, palaces, and homes would rise. This city is a well-known urban centre of commerce, learning, and philosophy founded in 332 B.C. When Alexander planned his city on a north-south, east-west grid, he was practicing urban planning.

In the United States, as in any new country, it took time for the early citizens to develop well-defined designs. The reason is that they wished their future cities look and function well. Not only after gaining independence, the new Americans decided to build their capital on the banks of the Potomac River.

Pierre Charles l’Enfant, a French immigrant, artist, engineer, was asked to design with broad diagonal main avenues to provide light and fresh air for pedestrians. A visitor can find today in Washington, D. C. majestic buildings and monuments, an open-air ceremonial space where we see the domed Capitol building as its centre.

## INSTRUCTIONS

- Read the above text carefully two or three times and answer the questions based on it.

- Choose the correct answer among those proposed (1, 2, 3, 4, 5).
- If no good answer, choose “6”

## QUESTIONS ON THE TEXT

1 The title which best suits the third paragraph is:

1. the practice of urban planning
2. the making of a city
3. the good plans for cities
4. the time required to build a new city
5. the city’s plan modification

2 In which part of the text, can we find the idea “One person can decide on the place where to construct a city?”

1. The first paragraph
2. The second paragraph
3. The fourth paragraph
4. The second and third paragraphs
5. The third paragraph

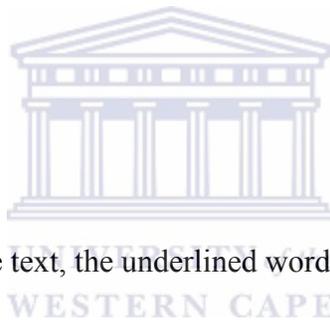
3 According to the plan of the city....



1. Can be cities' designers
2. Must be beautiful
3. Must be designed to look nice
4. Was prosperous
5. Need good designers

4 Few modifications have been made to l'Enfant's design, ...

1. isn't it
2. haven't they
3. does it
4. didn't they
5. doesn't it



5 In the first paragraph of the text, the underlined word "planning" is used as:

1. a noun
2. a past participle
3. a continuous past tense
4. an adjective
5. a verb

6 Mr Bulankete obliges Mr Katakata to come. Mr Katakata has to refuse.

1. Bulankete: "You must come, mustn't you?"

Katakata: "No, I can't".

2. Bulankete: "Do come, can't you?"

Katakata: "No, I won't".

3. Bulankete: "You might come, mightn't you?"

Katakata: "No, I mightn't".

4. Bulankete: "You do come, don't you?"

Katakata: "No, I don't".

5. Bulankete: "Can't you come?"

Katakata: "No, I can't".

7 ...is a thick, stiff kind of paper used for making boxes

1. Wrapping

2. Cartoon

3. Cardboard

4. Tissue paper

5. Envelope



8 Be ..... please. I am busy working

1. annoying

2. sleepy

3. noisy

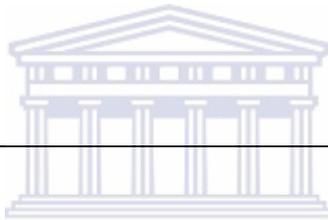
4. quiet

5. boring

9 Choose the correct word for the blank space to give the meaning of the word in brackets. The police look ... the murder

1. up
2. in
3. off
4. away
5. across

### Section 5: T5



#### English text

One day I entered into a bank. I was happy to see nice booking-offices. I got excited. The clerks looked at me. I saw much money which attracted me. Everything enjoyed me in the bank. I attempted to open an account there but I became an irresponsible person. My salary had been raised to fifty dollars a month and I felt that the bank was the only safe place to keep the money. So I went in and looked timidly round at the clerks. I had an idea that a person who wants to open an account must consult the manager. I went up to a window marked "Accountant". The accountant was a tall, cool man. My voice was trembling. I asked him to see the manager alone. The accountant went to call for him. The manager was a grave, calm man. I held my fifty-six dollars in my pocket.

I asked him if he were the manager. He said: "yes". The manager looked at me in some alarm. He felt that I had terrible secret to reveal. He led the way to a private room. I told him that I have come to open an account and I intend to keep all my money in this bank. The manager looked at relieved and serious. I told him that it was a fairly large amount. I added that I propose to deposit fifty-six dollars now and fifty dollars a month regularly.

The manager got up and opened the door. He called on the accountant, Mr. Montgomery, and he said that this gentleman is opening an account; he will deposit fifty-six dollars. I rose and went to the accountant's window and pushed the money at him. He took the money and gave it to another clerk. He made me write the sum on a slip and sign my name in a book. As soon as the money was deposited I wanted to draw a cheque. My idea was to draw out six dollars for present use. Someone gave me a cheque book and someone else began telling me out to write it out. The people in the bank had the impression that I was an invalid millionaire.

They continued to look at me. I wrote something on the cheque and, thrust it at the clerk. He looked at it and asked in surprise: "Are you drawing it all out again?" Then I realized that I had written fifty-six.

**INSTRUCTIONS:**

- Read the above text carefully two or three times and answer the questions based on it.
- Choose the correct answer among those proposed (1, 2, 3, 4, or 5)
- If there is no good answer, choose 6

**QUESTIONS FROM THE TEXT**

1 The title which suits the third paragraph of the text is:

1. Getting contact with people in a bank
2. Introducing the gentleman to the accountant
3. Opening an account in a bank
4. Meeting the manager in a bank
5. Meeting the clerks in a bank

2 According to the text, the clerks were:

1. not sure to keep the money in the bank
2. sure it was a large amount
3. astonished by the act of the client
4. looking at the person who entered in the bank
5. able to put down the money in the bank

3 The underlined word "They" in the third paragraph of the text refers to:

1. Gentleman
2. Accountant
3. Manager
4. People
5. Person

4 I did not push the money to the manager, ... ?

1. did I
2. don't I
3. was I
4. didn't I
5. wasn't I

**THE ENGLISH LETTER**

Ekunde Sales Services (ESS)  
Bunia

DRC

21 June 2006

Dear Mr. Polo Kamba,

In our telephone call of 14 June, we drew your attention to a number of difficulties we have had lately with several large costumers. One of the most serious results of this development has been a cash-flow problem, which means that the payment of a certain number of outstanding accounts may well have to be briefly delayed.

We are terribly sorry that ESS has been unfortunately affected by this development. As you can imagine, we are eager to avoid unnecessary delay. We feel able to settle our debts very soon.

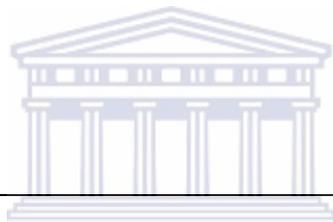
We are sure that it would not be in your interest, if the liquidation of our company becomes necessary.

I promise to keep you informed of further developments in the course of this month.

Yours sincerely

Djugu Djugu

The Manager.



UNIVERSITY of the  
WESTERN CAPE

**QUESTIONS ON THE LETTER**

**INSTRUCTIONS:**

- Read the above letter carefully two or three times and answer the questions based on it.
- Choose the correct answer among those proposed (1, 2, 3, 4, or 5)
- If there is no good answer, choose 6

5 The main idea of the third paragraph of the letter is the:

1. reference to the telephone call
2. content of the telephone call
3. expression of the firm's regrets
4. request for extension of credit
5. reference to dangers of bankruptcy

6 The signature in reply to the above letter may be:

- |                            |                                    |                         |
|----------------------------|------------------------------------|-------------------------|
| 1. Polo Kamba,<br>Customer | 2. Mr. Polo Kamba,<br>Bunia<br>DRC | 3. Dear Mr. Djugu Djugu |
| 4. Mr. Djugu Djugu         | 5. Mr. Polo Kamba                  |                         |

Ekunde Sales Services Bunia DRC	ESS Bunia DRC
7 One of the following parts of the letter does not appear:	
<ol style="list-style-type: none"> <li>1. Salutation</li> <li>2. Heading</li> <li>3. Attention line</li> <li>4. Complimentary close</li> <li>5. Signature</li> </ol>	

### **Section 6: T6**

#### **English text**

Once upon a time, there was a fox. And the fox was very hungry. He hadn't eaten for a long time, and he was just dying for a good meal. So, he went out of his wood and walked towards a farm where he knew there were some geese, chickens and ducks. In a field near the farm the fox came upon a flock of beautiful fat white geese. They didn't hear him coming and when they were trapped, they couldn't get away, the fox was going to eat them. "I am going to eat you!" said the fox. The geese were terrified and they didn't know what to do.

One of the geese who was a little cleverer than the others turned to the fox and said "well, Mr Fox, you've got the better of us this time, you are going to eat us, we have no way of escape. Can we please ask one last favour? Well, when the fox heard this and he was sure the geese couldn't get away he'd give them one last favour, and he said, "all right, one last favour you can have. What do you want?" The goose said "erm ... I'd like to say my prayers". And this seemed a reasonable request to the fox, so he said, you can say your prayers." He sat down and the geese began to say their prayers. Well, when the geese say prayers, they say in very loud voices, and they opened their mouths and cackled and cackled as loudly as they could while the fox sat and listening to their prayers and thought: "what a terrible noise they make saying their prayers."

The noise of their cackling could be heard as far as the farm-house, so what do you think happened? The farmer of course heard the cracklings, knew something was wrong, took his gun, rushed out of the farmhouse, rushed down to the field and there he saw the fox. At the same time the fox saw the farmer and of course he had no time to eat any of the geese, turned round and ran away as fast as he could back to his wood, as hungry as he had left that morning.

#### **INSTRUCTIONS:**

- Read the above text carefully two or three times and answer the questions based on it.

- Choose the correct answer among those proposed (1, 2, 3, 4, or 5)
- If there is no good answer, choose 6

**QUESTIONS ON THE TEXT**

1 One of the following sentences best summarizes the second paragraph

1. Geese cackled prayers loudly
2. A hungry fox found some geese in the field and wanted to eat them
3. The farmer heard the noise, came and chased the fox away
4. A goose begged for a last favour and the fox agreed
5. Geese were eaten by the fox

2 According to the text, geese, chickens and ducks could have been eaten if:

1. the fox had been kind to them
2. they had said the prayers
3. one of the geese had asked for a last favour
4. the fox had been very hungry
5. their request hadn't been reasonable to the fox

3 According to the text

1. The cackles were a usual way of praying for gees
2. The fox was at ease while geese were praying
3. The fox was very wise in allowing the geese to pray
4. The fox was not aware of what was going to happen latter on
5. The crackling was not a message sent to the farmer about the bad situation

4 The following items or things are birds, except:

1. duck
2. goose
3. cow
4. chicken
5. pigeon

5 A goose turned to the fox, ...?

1. didn't it
2. does it
3. was it
4. did it
5. had it

6 One of the following sentences about the above pictures is correct



1. The lady is sitting in front of the television set
2. The radio set is near the flowers bowl
3. The two boys are sitting on a chair
4. The lady is sitting beside the dog
5. The television set is near the door

### **ENGLISH CORRESPONDENCE**

2. The Manager  
 YMCA  
 221, OSWE STREET  
 KINSHASA

1. 262 Zombo  
 Luanda  
 Republic of Angola

3. March 25, 2009

4. Dear Sir,

5. I would be grateful if you would send me details about working as an aupair in September. I would like to take an English course in August.

Would you please send me any details about your school.  
 I am enclosing an international reply coupon.

6. Your faithfully,

7. Helene Seya

7 The part of the letter numbered "2" is called:

1. sender's address
2. salutation
3. inside address
4. complimentary clause
5. body of the letter

8 If you were asked to write a reply to the above letter, here are the steps you should go through. They are given in a jumbled way. Give the correct order

1. write your first draft
2. revise your first draft
3. write an outline
4. assemble all the relevant information and documents
5. write your aim
6. arrange the points in the order of importance
7. write and type your final version

1. 1-3-4-2-5-7-6
2. 1-6-4-2-3-7-5
3. 4-1-2-6-7-3-5
4. 3-7-6-1-2-5-4
5. 5-4-6-3-1-2-7

9 One of the following abbreviations is used before the name of both married and unmarried women

1. Miss.
2. Mr.
3. Ms.
4. Messrs.
5. Mrs.



## **Section 7: T7**

### **English text**

The Democratic Republic of Congo (DRC) has a vast unrealized agricultural capacity. The majority of its land is arable and suitable for farming.

Coffee beans, potatoes can be produced in the high plains of the east and south. The cool temperature and fertile soil of Eastern Highlands favour the cultivation of tomatoes, sweet potatoes, yam and even mediterranean vegetables.

Subsistence farming produces food crops in tropical areas like corn, cassava and rice. Most of commercial crops such as coffee, cocoa, rubber, tea, palm oil and sugar cane are grown in plantations.

With huge potential fish could be a valuable source of food, but it is not exploited. The lakes in eastern and southern regions are a massive reserve of a variety of freshwater species such as tilapia. The river Congo is another important source with major fishing ports in Kisangani and Bandaka.

Great potential for export earnings also exists in the timber industry. More or less two thirds of its territory is covered by an equatorial forest. But most of the forest remains intact because of the lack of viable infrastructure in the country.

### **INSTRUCTIONS**

- Read the text carefully
- Answer each question by choosing the correct answer among those proposed (1 to 5)
- If there isn't any correct answer, choose "6"

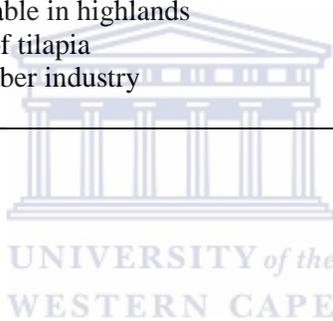
### **QUESTIONS ON THE TEXT**

1 All the following statements about the text are true, except:

1. tomatoes are produced in cool temperature
2. palm oil is grown in plantations
3. mediterranean vegetables are suitable in highlands
4. the port of Kisangani is a source of tilapia
5. equatorial forest is a source of timber industry

2 The title which best suits the text is:

1. forest potential in DRC
2. agricultural exploitation in DRC
3. timber situation in DRC
4. export culture in DRC
5. fish exploitation in DRC



3 Fish is:

1. produced in tropical areas
2. mediterranean vegetable
3. grown in very large plantations
4. commercial vegetable
5. an important source of food

4 The following list contains commercial crops, except:

1. Fish
2. Coffee
3. Cocoa
4. Palm oil
5. Tea

5 The cool temperature and fertile soil favour the cultivation of potatoes, ...?

1. do they
2. are they
3. don't they
4. haven't they
5. have they

## Section 8: T8

### **English text**

In Boyasegeze City, Joyce Gbamogigi's house stands near the lake. It has seven rooms, among them a living room, a dining-room and a kitchen. Here, next to the corner a cupboard is full of cutlery sets, pans and plates of different sizes, hot and cold dishes. Although her room is not very large, it is really comfortable.

At the entrance of her room stands a cutting-table. Close to it she puts a treadle sewing-machine which provides a convenient place to work. At the corner, a large wardrobe is full of up-to-date dresses, skirts, blouses and many clothes' patterns. Through the window the rising sun fills her room every morning until sunset. Near the window the perfumed curtains emit the smell of roses whose scent makes her relax. The soft and warm linens make her sleep peacefully.

The house stands by the railroad. So the noise of trains enters everywhere. In addition music from the stereo usually fills the room and makes her comfortable to stay at home. Consequently, the various articles in her room make her pleased. The room plays then an integral role in her life.

### **INSTRUCTIONS**

- Read carefully the above text two or three times and answer the questions based on it.
- Choose the correct answer among the five proposed (1, 2, 3, 4, 5).
- If no good answer, choose "6"

### **QUESTIONS ON THE TEXT**

- 1 The title which best suits the second paragraph of the text is
1. Gbamogigi's room
  2. Boyasegeze city's buildings
  3. Life in Gbamogigi's room
  4. A peaceful place

5. Life in Boyasegeze city												
<p>2 The idea of “sound of different kinds” is located in the ... of the text</p> <ol style="list-style-type: none"> <li>1. first paragraph</li> <li>2. first and second paragraphs</li> <li>3. first and third paragraphs</li> <li>4. second paragraph</li> <li>5. third paragraph</li> </ol>												
<p>3 Throughout the description of her room, there is no doubt Joyce Gbamogigi is a:</p> <ol style="list-style-type: none"> <li>1. dressmaker</li> <li>2. stewardess</li> <li>3. cook</li> <li>4. student</li> <li>5. music star</li> </ol>												
<p>4 The opposite of “up-to-date” is:</p> <ol style="list-style-type: none"> <li>1. nice fashion</li> <li>2. latest fashion</li> <li>3. out of date</li> <li>4. old fashion</li> <li>5. fair fashion</li> </ol> 												
<p>5 Match the words from box (A) with the actions from box (B) and choose the correct association</p> <table style="width: 100%; border: none;"> <tr> <td style="text-align: center; width: 50%;">(A)</td> <td style="text-align: center; width: 50%;">(B)</td> </tr> <tr> <td>1. kitchen</td> <td>a. studying</td> </tr> <tr> <td>2. lesson</td> <td>b. wearing</td> </tr> <tr> <td>3. music</td> <td>c. cooking</td> </tr> <tr> <td>4. clothes</td> <td>d. sleeping</td> </tr> <tr> <td>5. bedroom</td> <td>e. listening</td> </tr> </table> <ol style="list-style-type: none"> <li>1. 1c - 2b - 3a - 4d - 5e</li> <li>2. 1e - 2a - 3c - 4b - 5d</li> <li>3. 1b - 2c - 3d - 4c - 5e</li> <li>4. 1c - 2d - 3a - 4c - 5e</li> <li>5. 1d - 2b - 3c - 4d - 5e</li> </ol>	(A)	(B)	1. kitchen	a. studying	2. lesson	b. wearing	3. music	c. cooking	4. clothes	d. sleeping	5. bedroom	e. listening
(A)	(B)											
1. kitchen	a. studying											
2. lesson	b. wearing											
3. music	c. cooking											
4. clothes	d. sleeping											
5. bedroom	e. listening											

## **APPENDIX 8: Information sheet for student-participants**

My name is Godefroid B. Katalayi. I am Senior Lecturer at “Institut Supérieur Pédagogique de Kananga” where I have been lecturing at the Department of English for nearly twenty years. At present, I am doing my PhD in Language Education, Faculty of Education, University of the Western Cape, South Africa. I am conducting my research in the assessment of high/secondary school finalist students at the State Examination, with a particular focus on the validity of the English test. The title of my research is “*Fundamental validity issues of an English as a foreign language test: A process-oriented approach to examining the reading construct as measured by the DR Congo English state examination*”. The main objective of this study is to explore different variables that potentially threaten the validity of the English state examination and suggest ways to improve its quality.

The data to be collected must provide information on the different reading strategies you utilize to read the text and answer questions based on it. Therefore, if you agree to take part in this study, I will give you an English state examination test together with a questionnaire in which you are required to select from a strategies list the different strategies you have used to answer individual test items.

Although I really need your participation in this investigation, this participation is based on your free consent and you are guaranteed the right to withdraw from research at any stage and this without any consequence. I assure you that you have right to anonymity in that your names will not appear on the questionnaires or they will not be mentioned in the writing up of the findings. Also, I will make sure that your responses from the test and questionnaire will remain confidential.

If you agree to participate please sign the attached consent form. If you have any query, do not hesitate to contact me at [godefroidkatalayi@gmail.com/](mailto:godefroidkatalayi@gmail.com) +27827583456; or contact Professor Sivakumar Sivasubramaniam, Supervisor and Head of Language Education Department, Faculty of Education, University of the Western Cape, South Africa, Email: [ssivasubramaniam@uwc.ac.za](mailto:ssivasubramaniam@uwc.ac.za)

## Annexe 8

### ***Fiche d'information pour les élèves participants à l'étude***

*Je m'appelle Godefroid Katalayi. Je suis Chef de travaux à l'Institut Supérieur Pédagogique de Kananga où j'enseigne au département d'anglais il y a près de vingt ans. À présent, je suis entraîné de faire mes études doctorales en 'Language education' à la faculté des sciences de l'éducation, University of the Western Cape. Je conduis mon étude sur l'évaluation des élèves finalistes du secondaire aux examens d'état; particulièrement sur la validité du test d'anglais. Mon étude s'intitule: "Fundamental validity issues of an English as a foreign language test: A process-oriented approach to examining the reading construct as measured by the DR Congo English state examination". L'objectif principal de cette est d'explorer les différentes variables qui menacent la validité du test d'anglais et suggérer des stratégies pour améliorer la validité de cet test.*

*Les données qui devront être collectées devront donner des informations sur les stratégies que vous allez utiliser pour lire le texte et répondre aux questions sur le texte. Ainsi donc, si vous consentez de participer à cette étude, vous allez faire un test d'anglais qui sera accompagné d'un questionnaire auquel vous devez indiquer toutes les stratégies que vous aurez utilisées pour répondre à chacune des questions sur le texte.*

*Bien que j'ai besoin de votre participation à cette étude, je dois vous informer que vous êtes libre de participer à cette étude et vous avez le droit de vous retirer à n'importe quel moment de cette investigation. Je vous garantis que votre identité ne sera pas reprise sur le questionnaire ou dans le rapport final. De même, vos réponses au questionnaire et au test seront strictement confidentielles.*

*Si vous acceptez de participer à cette étude, vous devrez donc signer la fiche de consentement en annexe. Si vous avez besoin d'information supplémentaire, vous pouvez me contacter à l'adresse suivante: [godefroidkatalayi@gmail.com](mailto:godefroidkatalayi@gmail.com)/ +27827583456; ou encore contacter mon promoteur, le Professeur Sivakumar Sivasubramaniam à l'adresse suivante: [ssivasubramaniam@uwc.ac.za](mailto:ssivasubramaniam@uwc.ac.za)*

**APPENDIX 9: STATEMENT BY THE RESEARCHER**

I, the undersigned, have accurately read out the information sheet to the participant, and to the best of my ability;

I have made sure that the participant understands what he/she is expected to do;

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability;

I confirm that the participant has not been coerced into giving consent, and the consent has been given freely and voluntarily.

Researcher \_\_\_\_\_

Signature: \_\_\_\_\_

Date \_\_\_\_\_

Day/month/year



**ANNEXE 9: DECLARATION DU CHERCHEUR**

*Moi, Godefroid Katalayi, reconnais avoir lu et explique la fiche d'information pour les eleves participants a cette etude;*

*Je me suis rassure que chaque participant a cette etude a compris ce que j'attend de lui/d'elle;*

*Je confirme que chaque participant a eu l'occasion de poser des questions relatives a cette etude; et que j'ai repondu a toutes ces questions correctement;*

*Je confirme aussi qu'aucun participant n'a ete contraint de participer a cette etude, et que chaque participant a consenti librement et volontairement.*

*Le Chercheur* \_\_\_\_\_

*Signature:* \_\_\_\_\_

*Date* \_\_\_\_\_

*Jour / moi / annee*



**APPENDIX 10: STUDENT-PARTICIPANTS' INFORMED CONSENT**

**Research Title:** Fundamental validity issues of an English as a foreign language test: A process-oriented approach to examining the reading construct as measured by the D.R. Congo English state examination

**Researcher: Godefroid B. KATALAYI**

**Contact details: +27837583456; [godefroidkatalayi@gmail.com](mailto:godefroidkatalayi@gmail.com)**

I, undersigned, as a student of .....High School, I have read the foregoing information.

I have had the opportunity to ask questions about it and any questions I have asked have been answered to my satisfaction.

I consent voluntarily to be a participant in this study.

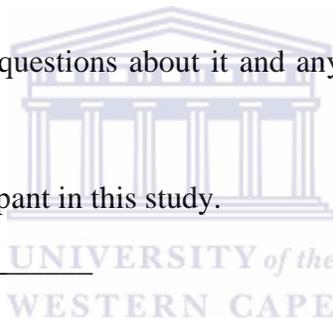
Name of Participant \_\_\_\_\_

Signature of Participant \_\_\_\_\_

Date \_\_\_\_\_

Day/month/year

*Jour/mois/annee*



**APPENDIX 10: FICHE DE CONSENTEMENT DE L'ELEVE-PARTICIPANT**

***Sujet de la recherche:*** *Fundamental validity issues of an English as a foreign language test: A process-oriented approach to examining the reading construct as measured by the D.R. Congo English state examination*

*Je soussigne, eleve a l'ecole secondaire ....., reconnais avoir lu toutes les informations relatives a cette etude;*

*J'ai eu l'occasion de poser des questions relatives a ces informations; et j'ai ete satisfait de toutes les reponses posees a ces questions.*

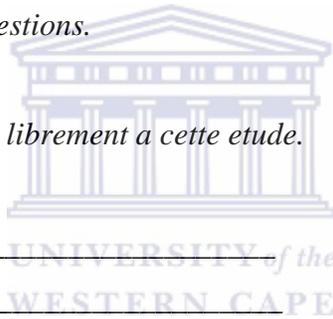
*Ainsi donc je consens de participer librement a cette etude.*

*Nom du Participant:* \_\_\_\_\_

*Signature du Participant:* \_\_\_\_\_

*Date:* \_\_\_\_\_

*Jour / mois / annee*



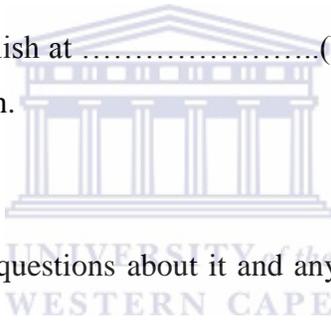
**APPENDIX 11: TEACHER-PARTICIPANTS INFORMED CONSENT**

**Research Title:** Fundamental validity issues of an English as a foreign language test: A process-oriented approach to examining the reading construct as measured by the D.R. Congo English state examination

**Researcher: Godefroid B. KATALAYI**

**Contact details: +27837583456; [godefroidkatalayi@gmail.com](mailto:godefroidkatalayi@gmail.com)**

I, undersigned, as a teacher of English at .....(Name of the school) High School, I have read the foregoing information.



I have had the opportunity to ask questions about it and any questions I have asked have been answered to my satisfaction.

I consent voluntarily to be a participant in this study.

Name of Participant \_\_\_\_\_

Signature of Participant \_\_\_\_\_

## APPENDIX 12

### LETTER TO PROVINCIAL EDUCATION SERVICES

The Head  
Provincial Education Services  
PO Box 742  
Kananga  
Kasai Occidental Province  
DR Congo

Dear Sir/Madam

I request your permission to use the 2011 and 2012 English state examination papers for research purposes. Please find attached the information form for research participants in order to have further information on my study purpose and significance and the reasons why I request your consent.

I will appreciate it if this request deserves your careful attention.

Yours sincerely

Godefroid B. Katalayi  
PhD Candidate in Language Education  
Faculty of Education  
University of the Western Cape  
Tel: +2837583456  
E-mail: [godefroidkatalayi@gmail.com](mailto:godefroidkatalayi@gmail.com)

**ANNEXE 13**

**LETTRE ADRESEE A L'INSPECTION PROVINCIALE DE L'ENSEIGNEMT**

*A l'Inspecteur Provincial de l'EPSP  
Kananga, Kasai Occidental, RD Congo.*

*Kananga, le \_\_\_\_/\_\_\_\_/2012*

*Madame l'Inspectrice,*

*Pour des raisons de recherche, j'ai le plaisir de demander votre permission pour utiliser les carnets des examens d'etat, epreuve d'anglais, pour les editions ci après: 2008, 2009, 2010 et 2011. Vous trouverez en annexe, pour votre attention, ma fiche d'information ou j'ai explique le bien fonde de ma recherche.*

*Tres sinceres salutations*

*Godefroid B. Katalayi*

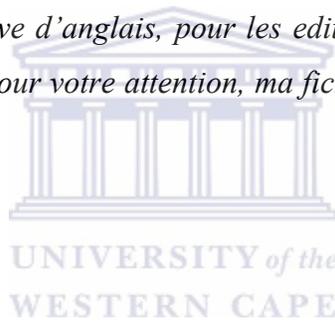
*PhD Candidate in Language Education*

*Faculty of Education*

*University of the Western Cape*

*Tel: +2837583456*

*E-mail: [godefroidkatalayi@gmail.com](mailto:godefroidkatalayi@gmail.com)*



**APPENDIX 13**

**LETTER TO SCHOOL PRINCIPALS**

The Principal,  
(School name)  
Kasai Occidental Province  
Kananga  
DR Congo

Date:

Dear Sir/Madam

For research purpose, I request your permission to test Grade 12 students. The test will be administered during 2 hours and will consist of the 2011 English state examination material. If you consent, I will make sure that during this time, the normal progress of courses in the school will not be disturbed, and also I am ready to adhere to the school code of conduct. Please find attached the information form for research participants in order to have further information on my study purpose and significance and the reasons why I request your consent.

I will appreciate it if this request meets your consent.

Yours sincerely

Godefroid B. Katalayi  
PhD Candidate in Language Education  
Faculty of Education  
University of the Western Cape  
Tel: +2837583456  
E-mail: [godefroidkatalayi@gmail.com](mailto:godefroidkatalayi@gmail.com)

**ANNEXE 13: LETTRE ADRESSEE AUX CHEFS D'ETABLISSEMENTS**

*A Monsieur le Prefet de l'Etablissement \_\_\_\_\_*

*B.P. \_\_\_\_ Kananga, Kasai Occidental,*

*RD Congo*

*Kananga, le \_\_\_\_/\_\_\_\_/2012*

*Monsieur/ Madame le Prefet,*

*Pour raison de recherch , je vous prie de me permettre d'administrer un test d'anglais aux eleves finalistes de votre ecole. La duree de ce test est de deux heures et le materiel utilise est le test d'anglais issu del'epreuve des exetats 2011. Si vous marquez votre consentement, je vais me rassurer que ma presence au sein e votre ecole ne pourra pas perturber le deroulement normal des enseignements. A cet effet, je prendrai aussi l'engagement de respecter scrupuleusement le code de bonne conduite de votre ecole. Vous trouverez en annexe, pour votre attention, ma fiche d'information ainsi que le bien fonde de ma recherche.*

*Tres sinceres salutations.*

*Godefroid B. Katalayi*

*PhD Candidate in Language Education*

*Faculty of Education*

*University of the Western Cape*

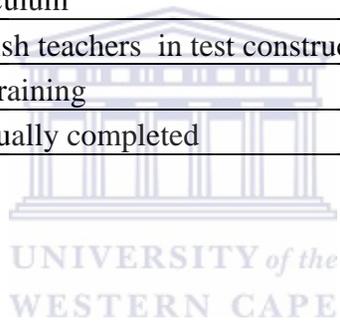
*Tel: +2837583456*

*E-mail: [godefroidkatalayi@gmail.com](mailto:godefroidkatalayi@gmail.com)*

**APPENDIX 14: TEACHER-PARTICIPANTS REPORT OF VARIABLES THAT NEGATIVELY AFFECT READING CLASSES AND PERFORMANCE ON THE ESE**

N=27

<b>Variables that affect reading classes and performance on the ESE</b>		<b>N</b>	<b>%</b>
1	Large number of students	21	77.8
2	Lack of students' involvement in English classes	17	63.0
3	Students' poor level of instruction that does not reflect the grade level	12	44.4
4	Students' expectation that they will collaborate when they take the ESE	15	55.6
5	The policy adopted by the school in terms of discipline	14	51.9
6	Lack of parents' involvement in their children's studies	11	40.7
7	Poor quality of English curriculum	18	66.7
8	Lack of involvement of English teachers in test construction	5	18.5
9	Lack of teachers' in-service training	25	92.6
10	Teaching programmes not usually completed	27	100.0

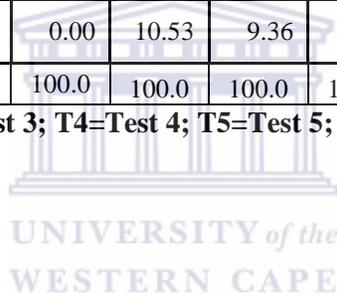


**APPENDIX 15: STUDENT-PARTICIPANTS' PREFERENCES FOR THE NUMBER OF HOURS PER WEEK FOR LEARNING ENGLISH IN BOTH GENERAL AND TECHNICAL SCHOOLS**

**N=496**

<b>Number of hours per week for learning English</b>	<b>T1</b>	<b>T2</b>	<b>T3</b>	<b>T4</b>	<b>T5</b>	<b>T6</b>	<b>T7</b>	<b>T8</b>	<b>Mean</b>	<b>SD</b>
1 hour per week	6.67	9.38	13.41	3.51	6.54	20.00	58.33	53.57	21.43	21.93
2 hours per week	38.67	23.44	19.51	24.56	23.36	25.71	41.67	25	27.74	7.93
3 hours per week	16.00	26.56	20.73	21.05	22.43	17.14	0.00	21.43	18.17	8.02
4 hours per week	12.00	17.18	18.29	12.28	14.95	8.58	0.00	0.00	10.41	7.12
5 hours per week	16.00	21.88	28.06	28.07	23.36	25.71	0.00	0.00	17.89	11.70
more than 5 hours per week	10.66	1.56	0.00	10.53	9.36	2.86	0.00	0.00	4.37	4.93
<b>TOTAL</b>	<b>100.0</b>	<b>100</b>								

**Note: T1=Test 1; T2=Test 2, T3=Test 3; T4=Test 4; T5=Test 5; T6=Test6; T7=Test 7; T8=Test8**



**APPENDIX 16: TEACHER-PARTICIPANTS' VIEWS ON SOME STATEMENTS  
RELATED TO ENGLISH LANGUAGE TEACHING AND ASSESSMENT**

Your view on the following		Agree		Disagree		Can't tell		Total	
		N	%	N	%	N	%	N	%
1	I have all textbooks required for teaching English in Grade 12	2	7.4	25	92.6	0	0.0	27	100.0
2	My students also have textbooks	2	7.4	25	92.6	0	0.0	27	100.0
3	I have the English curriculum	11	40.8	16	59.3	0	0.0	27	100.0
4	I use the English curriculum to design my course outline	15	55.6	12	44.4	0	0.0	27	100.0
5	The English curriculum needs to be updated	18	66.7	3	11.1	6	22.2	27	100.0
6	In reading classes, my students have difficulty with vocabulary	25	92.6	2	7.4	0	0.0	27	100.0
7	In reading classes, my students have difficulty with grammar	27	100.0	0	0.0	0	0.0	27	100.0
8	In reading classes, my students have difficulty with text content	24	88.9	3	11.1	0	0.0	27	100.0
9	I am motivated to teach English	12	44.4	15	55.6	0	0.0	27	100.0
10	My poor working conditions negatively affect my teaching	23	85.2	4	14.8	0	0.0	27	100.0
11	The ESE must also test other skills (writing, speaking, and listening)	14	51.9	10	37.0	3	11.1	27	100.0
12	The actual time allocated for writing the test is too much and therefore must be reduced	11	40.7	8	29.63	8	29.6	27	100.0
13	Those in charge of test administration do not sometimes respect time limits	12	44.4	13	48.15	2	7.4	27	100.0
14	Everybody involved in the test administration must adhere to the time limits during test administration	14	51.9	11	40.74	2	7.4	27	100.0

**APPENDIX 17: STRATEGIES PATTERNS AND ITEM DIFFICULTY FOR INDIVIDUAL ITEMS**

Careful reading at global level			
Skills	Test questions	Strategy patterns	ID
1. Establishing text or paragraph global com-prehension  Items: T1.1; T2.1; T4.1; T5.1; T7.2;T8.1	T1.1	1. S1- S3- S212-S4	.4
		2. S1-S3-S12-S4	
		3. S1-S3-S12-S4- S17	
		4. S1-S3-S13	
		5. S1-S2-S13	
	T2.1	1. S1-S2	.5
		2. S1-S2-S12	
		3. S1-S2-S14	
	T4.1	1. S3-S4-S18	.1
		2. S3-S4-S13	
	T5.1	1. S1-S3	.6
		2. S10-S13	
		3. S1-S2-S3-S4- S12-S15	
		4. S1-S3- S15	
		5. S1-S3- S12-S18	
	T7.2	1. S1-S2-S12-S18	.1
		2. S1-S2- S18	
	T8.1	1. S1-S2-S3-S12-S16	.5
		2. S1-S3-S12-S18	
3. S1-S4-S12			
<b>Mean item difficulty</b>			<b>.4</b>
<b>Standard Deviation</b>			<b>0.20</b>
<b>Strategies patterns across test questions: 20</b> (1) S1=17 [85 %]; (2) S2=8 [40 %]; (3) S3=9 [45 %]; (4) S4=4 [20 %] S12=10 [50 %]; S18=5 [25 %]; S13=5 [25%]			
2. Establishing accurate comprehension of explicitly stated	T2.3	1. S8-S11-S13-S12	.5
		2. S11-S12	
		3. S11-S12	
	T4.3	1. S2-S3- S13	.2

main ideas and supporting details across sentences  Items: T2.3; T4.3; T6.1		2. S2-S3- S12	
		3. S2-S3-S10	
	T5.5	1. S1-S2-S3-S10-S13	.4
		2. S1-S2-S3-S12-S17	
		3. S1-S2-S3-S12	
	T6.1	1. S1-S4-S11-S3- S12-S16	.2
		2. S1-S11-S3-S12-S18	
	3. S1-S11-S3- S12		
<b>Mean item difficulty</b>			<b>.3</b>
<b>Standard Deviation</b>			<b>0.17</b>
<b>Strategies patterns across test questions: 10 patterns</b> S1= 6 [60%]; S2=3 [30 %]; S3=9 [90 %]; S12=5 [50 %]; S13=3 [30 %]; S10=2 [20 %]			
3. Making text/paragraph inference  Items: T1.3; T2.8; T3.1; T5.2; T6.2; T6.3; T8.3	T1.3	1. S2-S3-S13	.6
		2. S2-S3-S12-S17	
		3. S2-S3-S12-S15	
	T2.8	1. S13	.5
		2. S10	
	T3.1	1. S1-S4-S14	.5
		2. S1-S4	
		3. S1-S4-S18	
	T5.2	1. S2-S3-S4- S10-S16	.4
		2. S3-S4-S12-S16	
		3. S3-S4-S12-S17	
	T6.2	1. S2-S7- S12-S16	.3
		2. S4-S12	
		3. S2-S12	
	T6.3	1. S2-S8- S18	.3
	2. S2-S13		
	3. S2-S15		
T8.3	1. S1-S10-S13	.8	
	2. S1-S10-S13		
	3. S1-S10-S17		
<b>Mean item difficulty</b>			<b>.5</b>
<b>Standard deviation</b>			<b>0.18</b>
<b>Strategies patterns across test questions: 20 patterns</b> S1= 6 [30%]; S2=9 [45 %]; S3=6 [30 %]; S4=7 [35%]; S13=4 [20 %]; S12=7 [35 %]; S18=2			

		[10 %]; S15=2 [10%]; S10=5 [25 %]; S16=3 [15 %]; S17=3 [15%]		
Careful reading at local level			1.	
1. Establish accurate comprehension of explicitly stated main idea or supporting details within a sentence	No items found			
2. Identifying or understanding lexis  T1.4; T3.6; T5.7; T6.4; T6.6; T6.7; T6.9; T8.4; T8.5	T1.4	1. S11-S7- S12	.3	
		2. S11-S8 –S13		
	T3.6	1. S8-S13	.7	
		2. S8-S17		
	T5.7	1. S1- S14	.3	
		2. S1- S12-S18		
	T6.4	1. S8-S13-S18-S15	.8	
		2. S8-S12-S15		
	T6.6	1. S8-S7-S13	.5	
		2. S7-S17-S13		
		3. S7-S8-S12		
	T6.7	1. S1-S2- S11-S14- S17	.5	
		2. S2-S1-S11-S12		
	T6.9	1. S8	.5	
	T8.4	1. S1-S2-S3-S12-S16	.6	
		2. S1-S3-S12-S17		
	T8.5	1. S8-S15-S17	.3	
		2. S8-S17		
	<b>Mean item difficulty</b>			<b>.5</b>
	<b>Standard deviation</b>			<b>0.18</b>
<b>Strategies patterns across test questions: 18</b>				
S11=4 [22 %]; S7=4 [22 %]; S8=13 [72 %]; S12=7 [39 %]; S13=5 [28 %]; S17=7 [39%]; S1=5 [28 %]; S3=3 [17 %]; S11=2 [11 %];				
3. Understanding syntax	T1.7	1. S11-S7-S13-S17-S14	.6	
		2. S14-S11-S7-S13		
		3. S11-S14		
		4. S11-S15-S18		

	T2.5	1. S3-S7-S14-S17 2. S11-S14 3. S11-S7-S13-S18	.3
	T2.7	1. S7 2. S7-S8-S14 3. S12	.8
	T3.2	1. S2-S7-S14-S12-S17 2. S3-S7-S14-S18	.5
	T3.3	1. S3-S7-S14-S13-S17 2. S3-S7-S13	.5
	T3.4	1. S3-S7-S13 2. S3-S7-S12-S17	.5
	T3.8	1. S7-S14-S13 2. S4-S7-S13-S14	.8
	T3.9	1. S7-S14-S13-S17	.5
	T4.4	1. S7-S14-S17-S18 2. S4-S7-S13	.8
	T4.5	1. S11-S7-S14- S13-S12-S18	.6
	T5.3	1. S11-S14 2. S11-S7-S13-S17 3. S11-S15-S17 4. S11-S14	.8
	T5.4	1. S7-S14-S17-S18 2. S14-S7-S13	.7
	T6.5	1. S7-S14-S13-S17	.5
	T6.6	2. S8-S7-S13-S12 3. S7-S17-S13	.5
	T7.5	1. S7-S13-14-18	.8
	<b>Mean item difficulty</b>		<b>.6</b>
	<b>Standard deviation</b>		<b>0.16</b>
	<b>Strategies patterns across test questions: 32</b> S11=11[34 %]; S7=26 [81%]; S13=20[63 %]; S17=18[56%]; S12=6[19%]; S14=16[50%]; S15=2[6%]; S3=7[22%]; S8=2[6%]; S2=1[3%]		
<b>Expeditious reading at global level</b>			
1. Skimming to find the title that	T1.1	1. S1- S3- S12-S4 2. S1-S3-S12-S4- S17	.4

best fits the text/paragraph		3. S1-S3-S13	
		4. S1-S2-S13	
	T2.1	1. S1-S2	.5
		2. S1-S2-S12	
		3. S1-S2-S13	
	T4.1	1. S3-S4	.1
		2. S3-S4-S17	
		3. S3-S4-S18	
		4. S3-S4-S13	
	T5.1	1. S1-S3	.6
		2. S10-S13	
		3. S1-S2-S3-S4	
		4. S1-S3- S12-S17	
	T7.2	1. S1-S2-S12-S17	.1
	2. S1-S2-S17		
T8.1	1. S1-S2-S3-S12-S16	.5	
	2. S1-S3-S12-S17		
	3. S1-S3-S12-S18		
<b>Mean item difficulty</b>			<b>.4</b>
<b>Standard deviation</b>			<b>0.20</b>
<b>Strategies patterns across test questions: 20</b>			
(1) S1=15 [75 %];(2) S2=8 [40 %]; (3) S3=13 [65 %]; (4) S4=7 [35 %]; S12=8 [40 %]; S17=6 [30 %]; S13=5 [25%]; S18=2 [10 %]			
2. Skimming the text or paragraph to locate needed information	T1.2	1. S1- S17	.6
		2. S1- S13	
		3. S1-S3-S4-S13-S17	
	T4.2	1. S1-S4-S10-S13	.7
		2. S1-S16-S13	
		3. S4-S16-S13	
	T8.2	1. S1-S13	.7
	2. S1-S3-S4-S13		
<b>Mean item difficulty</b>			<b>.7</b>
<b>Standard deviation</b>			<b>0.06</b>
<b>Strategies patterns across test questions: 8</b>			
S1=7[88%]; S17=2[25%]; S13=7[88%]; S3=2[25%]; S4=4[50%]; S10=1[13%]; S16=2[25%];			

3. Search reading to quickly locate implicit information	T1.3	1. S2-S3-S13	.6
		2. S2-S3-S12-S17-S15-S10	
	T3.1	1. S1-S4-S14	.5
		2. S1-S4- S14-S17	
		3. S1-S4	
	<b>Mean item difficulty</b>		
<b>Standard deviation</b>			<b>0.07</b>
<b>Strategies patterns across test questions:5</b> S2=2[40%];S3=2[40%];S13=1[20%];S12=1[20%];S17=2[40%];S15=1[20%]; S1=3[60%]; S4=3[60%]; S14=2[40%]. [20%];			
4. Search reading to quickly locate explicit pieces of information scattered through the whole text	T1.5	1. S2-S16-S17-S18	.3
		2. S2-S17- S12-S16	
		3. S1-S2-S17	
	T2.2	1. S2- S16-S18-S17	.2
		2. S2-S12	
	T2.4	1. S3-S13-S12	.3
	T2.6	1. S2-S18-S16-S17	.3
		2. S2-S17	
		3. S1-S2-S17	
	T3.4	1. S3-S7-S13-S17	.5
		2. S3-S7-S13	
	T3.7	1. S15-S17-S13-S18	.7
		2. S18	
	T4.3	1. S2-S3-S13-S17-S12-S10	.2
		2. S2-S3-S13	
		3. S2-S3-S10	
	T6.2	1. S2-S7-S12-S16-S17-S18	.3
		2. S2-S7-S13	
		3. S2-S12	
T6.3	1. S2-S8-S17-S16	.3	
	2. S2-S13-S15		
T7.1	1. S1-S2-S17	.4	
<b>Mean item difficulty</b>			<b>.4</b>
<b>Standard deviation</b>			<b>0.15</b>
<b>Strategies patterns across test questions: 22</b> S2=16[73%]; S16=6[27%]; S18=6[27%]; S17=14[64%]; S12=6[27%]; S1=4[18%];			

	S3=6[27%]; S13=8[36%]; S7=4[18%]; S15=2[9%]; S8=1[5%]		
<b>2. Expeditious reading at local level</b>			
1. Scanning	T1.4	1. S11-S7-S8-S12-S13	.3
		2. S11-S7-S8-S12	
		3. S11-S18	
	T1.7	1. S11-S7-S8-S13-S17-S14	.6
		2. S14-S11-S7-S13	
		3. S11-S14	
	T2.5	1. S3-S7-S14-S17	.3
		2. S11-S14	
		3. S11-S7-S13-S17	
	T3.2	1. S2-S7-S14-S17-S12	.5
		2. S3-S7-S14-S17-S18	
	T4.5	1. S11-S7-S14-S13-S12-S17-S18	.6
	T5.3	1. S11-S14	.8
		2. S11-S15	
3. S11-S15-S17			
T7.3	1. S4-S3-S17	.7	
	2. S3-S4-S15		
T7.4	1. S4-S3-S13-S17	.3	
	2. S4-S3-S16/28-S18		
	<b>Mean item difficulty</b>		<b>.5</b>
	<b>Standard deviation</b>		<b>0.20</b>
	<b>Strategies patterns across test questions: 19</b>		
	S11=12[63%]; S7=9[47%]; S8=2[11%]; S12=5[26%]; S13=5[26%]; S18=4[21%]; S17=9[47%]; S14=8[42%]; S3=6[32%]; S4=4[21%]; S15=2[11%]		

**Note:** S1=Read the whole text carefully; S2= Read the whole text rapidly; S3= Read a portion of the text carefully; S4=Read a portion of the text rapidly; S5=Use my knowledge of letter writing parts; S6=Use my knowledge of letter writing steps; S7=Use my knowledge of grammar; S8= Use my knowledge of vocabulary; S9=Use context clues to figure out the meaning of difficult words; S10= Draw conclusions based on what the paragraph/text is about; S11=Look for parts of the text the writer has recommended; S12=Consider the options and postpone consideration of the option; S13=Consider the options and focuses on a familiar option; S14=Produce my own answer after reading the question and then look at the options; S15=Use clues provided in the options; S16=Select the option through passage/paragraph overall comprehension; S17=Select the option through elimination of other options; S18=Select the option through guessing.

T1.1=Test 1, Item number 1; T2.5=Test 2, Item number 5; T3.2=Test 3, Item number 2; T7.8=Test 7, Item number 8; etc.

ID=Itemdifficulty.

**APPENDIX 18: MULTIPLE-CHOICE STEM FOCUS, COMPLETENESS AND ORIENTATION FOR THE 56 TEST ITEMS**

TEST	Test item	Stem focus		Stem completeness		Stem orientation		Item diff.
		focused	unfocused	Complete	Incomplete	Positive	Negative	
T1	T1.1	x		x		x		
	T1.2	x		x		x		
	T1.3	x		x		x		
	T1.4	x		x		x		
	T1.5		x	x			x	
	T1.6		x	x		x		
	T1.7	x		x		x		
	TOTAL	5	2	7	0	6	1	
T2	T2.1	x		x		x		
	T2.2		x	x			x	
	T2.3		x	x		x		
	T2.4		x	x			x	
	T2.5	x		x		x		
	T2.6		x	x			x	
	T2.7		x	x		x		
	T2.8		x	x		x		
	T2.9	x		x		x		
	TOTAL	3	6	9	0	6	3	
T3	T3.1	x		x		x		
	T3.2	x		x		x		
	T3.3	x		x		x		
	T3.4		x	x		x		
	T3.5	x		x		x		
	T3.6	x		x			x	
	T3.7		x	x		x		
	T3.8	x			x	x		
	T3.9	x			x	x		

	TOTAL	7	2	7	2	8	1	
T4	T4.1	x		x		x		
	T4.2	x		x		x		
	T4.3		x		x	x		
	T4.4	x			x	x		
	T4.5	x		x		x		
	TOTAL	4	1	3	2	5	0	
T5	T5.1	x		x		x		
	T5.2		x	x		x		
	T5.3	x		x		x		
	T5.4	x			x	x		
	T5.5	x		x		x		
	T5.6	x		x		x		
	T5.7	x		x		x		
	TOTAL	6	1	6	1	7	0	
T6	T6.1		x	x		x		
	T6.2		x	x		x		
	T6.3		x	x		x		
	T6.4		x	x			x	
	T6.5	x			x	x		
	T6.6		x	x		x		
	T6.7	x		x		x		
	T6.8		x	x		x		
	T6.9	x		x		x		
	TOTAL	3	6	8	1	8	1	
T7	T7.1		x	x			x	
	T7.2	x		x		x		
	T7.3		x	x		x		
	T7.4		x	x			x	
	T7.5	x			x	x		
	TOTAL	2	3	4	1	3	2	
T8	T8.1	x		x		x		
	T8.2	x			x	x		

	T8.3	x		x		x		
	T8.4	x		x		x		
	T8.5		x	x		x		
	TOTAL	4	1	4	1	5	0	

**Notes: T1=Test 1; T2=Test 2, T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8**

**T1.1=Test 1, Item number 1; T2.5=Test 2, Item number 5; T3.2=Test 3, Item number 2; T7.8=Test 7, Item number8;**



**APPENDIX 19: DISTRACTORS' WEIGHTS ACROSS THE 8 TESTS**

Item	Choices	T1		T2		T3		T4		T5		T6		T7		T8	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1	1	8	11	35	55	3	4	5	9	0	0	9	26	10	21	14	50
	2	7	9	6	9	9	11	5	9	13	12	3	9	4	8	0	0
	3	11	15	4	6	38	46	10	18	66	62	5	14	6	13	8	29
	4	28	37	1	2	32	39	27	47	8	7	6	17	20	42	6	21
	5	20	27	18	28	0	0	5	9	13	12	6	17	7	15	0	0
	6	1	1	0	0	0	0	5	9	7	7	6	17	1	2	0	0
	<b>Total</b>	<b>75</b>	<b>100</b>	<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>	<b>57</b>	<b>100</b>	<b>107</b>	<b>100</b>	<b>35</b>	<b>100</b>	<b>48</b>	<b>100</b>	<b>28</b>	<b>100</b>
2	1	2	3	19	30	45	55	4	7	13	12	0	0	7	15	0	0
	2	16	21	7	11	7	9	39	68	40	37	11	31	25	52	4	14
	3	4	5	2	3	5	6	0	0	41	38	6	17	4	8	5	18
	4	48	64	4	6	6	7	7	12	7	7	6	17	4	8	0	0
	5	3	4	14	22	19	23	7	12	6	6	12	34	1	2	19	68
	6	2	3	18	28	0	0	0	0	0	0	0	0	7	15	0	0
	<b>Total</b>	<b>75</b>	<b>100</b>	<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>	<b>57</b>	<b>100</b>	<b>107</b>	<b>100</b>	<b>35</b>		<b>48</b>	<b>100</b>	<b>28</b>	<b>100</b>
3	1	14	19	11	17	5	6	4	7	7	7	6	17	7	15	21	75
	2	0	0	7	11	3	4	19	33	0	0	3	9	0	0	5	18
	3	44	59	29	45	18	22	13	23	13	12	9	26	5	10	0	0
	4	9	12	17	27	14	17	0	0	87	81	12	34	2	4	0	0
	5	4	5	0	0	42	51	11	19	0	0	5	14	34	71	2	7
	6	2	3	0	0	0	0	10	18	0	0	0	0	0	0	0	0
	<b>Total</b>	<b>75</b>	<b>97</b>	<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>	<b>57</b>	<b>100</b>	<b>107</b>	<b>100</b>	<b>35</b>	<b>100</b>	<b>48</b>	<b>100</b>	<b>28</b>	<b>100</b>
4	1	16	21	15	23	3	4	7	12	79	74	2	6	14	29	3	11
	2	25	33	8	13	38	46	47	82	0	0	2	6	3	6	2	7
	3	15	20	3	5	11	13	0	0	15	14	29	83	6	13	6	21
	4	5	7	18	28	21	26	3	5	13	12	2	6	5	10	16	57
	5	4	5	4	6	3	4	0	0	0	0	0	0	2	4	1	4
	6	10	13	16	25	6	7	0	0	0	0	0	0	18	38	0	0
	<b>Total</b>	<b>75</b>	<b>100</b>	<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>	<b>57</b>	<b>100</b>	<b>107</b>	<b>100</b>	<b>35</b>	<b>100</b>	<b>48</b>	<b>100</b>	<b>28</b>	<b>100</b>
5	1	13	17	5	8	27	33	39	68	6	6	18	51	10	21	10	36
	2	10	13	2	3	0	0	3	5	8	7	2	6	0	0	6	21
	3	26	35	20	31	22	27	5	9	27	25	5	14	36	75	2	7
	4	5	7	16	25	11	13	6	11	13	12	6	17	2	4	3	11
	5	6	8	11	17	11	13	0	0	40	37	4	11	0	0	0	0
	6	15	20	10	16	11	13	6	7	13	12	0	0	0	0	7	25

	Total	75	100	64	100	82	100	57	100	107	100	35	100	48	100	28	100
6	1	24	32	5	8	4	5			40	37	3	9				
	2	0	0	22	34	2	2			7	7	18	51				
	3	0	0	7	11	60	73			13	12	6	17				
	4	4	5	11	17	0	0			34	32	0	0				
	5	47	63	0	0	3	4			13	12	3	9				
	6	0	0	19	30	13	16			0	0	5	14				
	<b>Total</b>	<b>75</b>	<b>100</b>	<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>			<b>107</b>	<b>100</b>	<b>35</b>	<b>100</b>				
7	1	45	60	48	75	16	20			0	0	6	17				
	2	9	12	0	0	1	1			46	43	0	0				
	3	0	0	0	0	5	6			27	25	18	51				
	4	19	25	9	14	54	66			13	12	6	17				
	5	2	3	7	11	6	7			0	0	2	6				
	6	0	0	0	0	0	0			21	20	3	9				
	<b>Total</b>	<b>75</b>	<b>100</b>	<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>			<b>107</b>	<b>100</b>	<b>35</b>	<b>100</b>				
8	1			8	13	12	15					0	0				
	2			0	0	64	78					12	34				
	3			33	52	3	4					5	14				
	4			7	11	3	4					3	9				
	5			0	0	0	0					12	34				
	6			16	25	0	0					3	9				
	<b>Total</b>			<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>					<b>35</b>	<b>100</b>				
9	1			63	98	3	4					6	17				
	2			1	2	39	48					4	11				
	3			0	0	29	35					11	31				
	4			0	0	4	5					3	9				
	5			0	0	7	9					11	31				
	6			0	0	0	0					0	0				
<b>Total</b>			<b>64</b>	<b>100</b>	<b>82</b>	<b>100</b>					<b>35</b>	<b>100</b>					

Note: T1=Test 1; T2=Test 2, T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test

**APPENDIX 20: FREQUENCY OF OPTION 6 ACROSS THE EIGHT TESTS**

Test items	T1		T2		T3		T4		T5		T6		T7		T8	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Item 1	1	1	0	0	0	0	5	9	7	7	6	17	1	2	0	0
Item 2	2	3	18	28	0	0	0	0	0	0	12	34	7	15	0	0
Item 3	2	3	0	0	0	0	10	18	0	0	0	0	0	0	0	0
Item 4	10	13	16	25	6	7	0	0	0	0	0	0	18	38	0	0
Item 5	15	20	10	16	11	13	0	0	13	12	0	0	0	0	7	25
Item 6	0	0	19	30	13	16			0	0	5	14				
Item 7	0	0	0	0	0	0			21	20	3	9				
Item 8			16	25	0	0					3	9				
Item 9			0	0	0	0					0	0				

Notes: T1=Test 1; T2=Test 2, T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8



**APPENDIX 21: ITEM DIFFICULTY AND PARTICIPANTS' PERCEPTIONS OF ITEM DIFFICULTY FOR THE 56 ITEMS ACROSS THE EIGHT TESTS**

Test	Test item	Item difficulty	Participants' perceptions of item difficulty						Total	
			Easy		Difficult		Moderate		N	%
			N	%	N	%	N	%		
T1 N=75	T1.1	.4	12	16	26	35	37	49	75	100
	T1.2	.6	53	71	9	12	13	17	75	100
	T1.3	.6	47	63	10	13	18	24	75	100
	T1.4	.3	6	8	52	69	17	23	75	100
	T1.5	.4	14	19	47	63	14	19	75	100
	T1.6	.6	50	67	7	9	18	24	75	100
	T1.7	.6	35	47	17	23	23	31	75	100
T2 N=64	T2.1	.6	4	6	47	73	13	20	64	100
	T2.2	.2	12	19	22	34	30	47	64	100
	T2.3	.5	7	11	18	28	39	61	64	100
	T2.4	.3	15	23	6	9	43	67	64	100
	T2.5	.3	11	17	7	11	46	72	64	100
	T2.6	.3	9	14	17	27	38	59	64	100
	T2.7	.8	13	20	11	17	40	63	64	100
	T2.8	.5	17	27	12	19	35	55	64	100
	T2.9	.9	29	45	9	14	26	41	64	100
T3 N=82	T3.1	.5	9	11	22	27	51	62	82	100
	T3.2	.6	12	15	27	33	43	52	82	100
	T3.3	.5	48	59	14	17	20	24	82	100
	T3.4	.5	52	63	13	16	17	21	82	100
	T3.5	.3	18	22	49	60	15	18	82	100
	T3.6	.7	44	54	17	21	21	26	82	100
	T3.7	.7	8	10	57	70	17	21	82	100
	T3.8	.8	52	63	14	17	16	20	82	100
	T3.9	.5	17	21	20	24	45	55	82	100
T4 N=57	T4.1	.1	4	7	35	61	18	32	57	100
	T4.2	.7	7	12	39	68	11	19	57	100
	T4.3	.2	9	16	18	32	30	53	57	100
	T4.4	.8	37	65	7	12	13	23	57	100

	T4.5	.7	10	18	12	21	35	61	57	100
T5 N=107	T5.1	.6	7	7	71	66	29	27	107	100
	T5.2	.4	9	8	67	63	31	29	107	100
	T5.3	.8	59	55	20	19	28	26	107	100
	T5.4	.7	62	58	19	18	26	24	107	100
	T5.5	.3	26	24	24	22	57	53	107	100
	T5.6	.4	24	22	59	55	24	22	107	100
	T5.7	.3	7	7	73	68	27	25	107	100
T6 N=35	T6.1	.2	6	17	20	57	9	26	35	100
	T6.2	.3	5	14	22	63	8	23	35	100
	T6.3	.3	10	29	6	17	19	54	35	100
	T6.4	.8	23	66	6	17	6	17	35	100
	T6.5	.5	23	66	8	23	4	11	35	100
	T6.6	.5	9	26	5	14	21	60	35	100
	T6.7	.3	7	20	7	20	21	60	35	100
	T6.8	.3	3	9	25	71	7	20	35	100
	T6.9	.3	5	14	7	20	23	66	35	100
T7 N=48	T7.1	.4	7	15	29	60	12	25	48	100
	T7.2	.2	7	15	31	65	10	21	48	100
	T7.3	.7	11	23	7	15	30	63	48	100
	T7.4	.3	12	25	9	19	27	56	48	100
	T7.5	.8	28	58	10	21	10	21	48	100
T8 N=28	T8.1	.5	3	11	19	68	6	21	28	100
	T8.2	.7	5	18	5	18	18	64	28	100
	T8.3	.8	7	25	5	18	16	57	28	100
	T8.4	.6	5	18	6	21	17	61	28	100
	T8.5	.3	2	7	20	71	6	21	28	100

Notes: T1=Test 1; T2=Test 2; T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8

T1.1=Test 1, Item number 1; T2.5=Test 2, Item number 5; T3.2=Test 3, Item number 2; T7.8=Test 7, Item number 8; etc.

**APPENDIX 22: ITEM DIFFICULTY FOR THE 56 TEST ITEMS ACROSS THE EIGHT TESTS**

Test	Test item	Item difficulty
T1 N=75	T1.1	.4
	T1.2	.6
	T1.3	.6
	T1.4	.3
	T1.5	.3
	T1.6	.4
	T1.7	.6
T2 N=64	T2.1	.5
	T2.2	.2
	T2.3	.5
	T2.4	.3
	T2.5	.3
	T2.6	.3
	T2.7	.8
	T2.8	.5
	T2.9	.3
T3 N=82	T3.1	.5
	T3.2	.5
	T3.3	.5
	T3.4	.5
	T3.5	.3
	T3.6	.7
	T3.7	.7
	T3.8	.8
	T3.9	.5
T4 N=57	T4.1	.1
	T4.2	.2
	T4.3	.2
	T4.4	.8
	T4.5	.6



T5 N=107	T5.1	.6
	T5.2	.4
	T5.3	.8
	T5.4	.7
	T5.5	.4
	T5.6	.5
	T5.7	.3
T6 N=35	T6.1	.2
	T6.2	.3
	T6.3	.3
	T6.4	.8
	T6.5	.5
	T6.6	.5
	T6.7	.5
	T6.8	.3
	T6.9	.5
T7 N=48	T7.1	.4
	T7.2	.1
	T7.3	.7
	T7.4	.3
	T7.5	.8
T8 N=28	T8.1	.5
	T8.2	.7
	T8.3	.8
	T8.4	.6
	T8.5	.3



**Notes: T1=Test 1; T2=Test 2, T3=Test 3; T4=Test 4; T5=Test 5; T6=Test 6; T7=Test 7; T8=Test 8**

**T1.1=Test 1, Item number 1; T2.5=Test 2, Item number 5; T3.2=Test 3, Item number 2; T7.8=Test 7, Item number 8; etc.**

## APPENDIX 23:

### STUDENT-PARTICIPANTS SELECTED RESPONSES TO THE CONTEXTUAL QUESTIONNAIRE

#### Question 1: Students' skills in English language

❖ *Participants who reported their proficiency level in English to be very low or low*

**Respondent 1:** I don't understand English, I can't speak it; it is a difficult course.

**Respondent 2:** Really, English is a very difficult language for me. I don't understand what my teacher says. When she asks me the question, I am unable to say something.

**Respondent 3:** This is a very difficult language, they write it in one way and you must pronounce words in another way. This is too confusing.

**Respondent 4:** I don't like English. The pronunciation is different from the French pronunciation. When you are given a text, you find two or three words with the same vowel; but when you read these three words in the same way, the teacher tells you that it is wrong. Sometimes my friends laugh at me after reading a word because I have badly pronounced it.

**Respondent 5:** It is not our language. I speak Ciluba, Lingala, Swahili and French. Do I really need to learn again this language (English)? It is very difficult. I can't be good at it.

**Respondent 6:** English is a very difficult subject for me. I cannot speak it.

**Respondent 7:** This is a French speaking country. I don't care if I cannot speak English

**Respondent 8:** English is a difficult language because it demands much effort to pronounce it. It is not like French.

**Respondent 9:** I am not good at English. It is too late for me to cope with it.

❖ *Participants who reported their proficiency level in English to be either average or good*

**Respondent 1:** I understand almost everything, and I try to speak a bit. I am comfortable with reading and writing. I think I am good at English.

**Respondent 2:** I like English because I can read a text and understand it. I can even write a letter in English. I cannot say that I am bad at English.

**Respondent 3:** I have no problem in English classes. I am a good learner.

**Respondent 4:** I am not too bad at English. I understand a bit; but I have problems to speak it.

**Respondent 5:** I have never failed in English. I can say that I am good.

**Respondent 6:** I like English. I am good at it.

## Question 2: Reading frequency at school and at home

❖ *The participants who reported to rarely read at school*

**Respondent 1:** We have just 2 hours per week; how can we read regularly?

**Respondent 2:** First, I would say that even in class we do not really read. We just have 2 hours per week for English. I think we can read more if we have much time.

**Respondent 3:** It takes 1 hour to copy the text that is written on the board. If it is a long text, when are we then going to read?

**Respondent 4:** Our teacher is too lazy; he takes much time to analyze one text. He takes 2 sessions for writing it on the chalkboard, 2 sessions for teaching vocabulary on the text; 2 or 3 sessions for teaching grammar on the same text; 1 or 2 sessions for text comprehension and discussion. After, he takes 1 session for dictation of a paragraph of the same text, and 1 session for essay writing. So, 1 text may take us 2 months.

**Respondent 5:** My teacher of English rarely comes to teach us. When he comes,

he takes so much time to tell us long stories of his personal life.

**Respondent 6:** He [the teacher of English] has another job in a Pakistani shop where he serves as a translator. He rarely comes to class and whenever he comes to teach; he just sits down and asks us to copy the text.

**Respondent 7:** If we had had textbooks, we could usually read at school. But in our school, there are no textbooks. The texts are written on the chalkboard and this is demotivating.

**Respondent 8:** My English teacher is extremely lazy. We don't read at school.

**Respondent 9:** Last year, we read only three texts. This year, we have only read two texts. Yet, the state examination is approaching. What shall we be capable to do in the English test?

❖ *2. Participants who reported that they never read at home*

**Respondent 1:** I don't read in English at home because I don't have textbooks. I don't have anything to read.

**Respondent 2:** I don't read at home because I have nothing to read.

**Respondent 3:** I cannot spend my time to read a text that I believe I have copied it with so many mistakes. I think it is important that we use printed texts. These texts can motivate us to read at home. When the teacher gives the book [teacher book] to one student and asks him to write the text on the board; the student does not pay attention to write correctly the text.

**Respondent 4:** Those texts like 'the discovery of radium', 'Uncle Bernard in London', I don't understand them. I can't spend my time to read them at home.

**Respondent 5:** I don't have time to read at home. When I come back from school, I quickly rush to the market to sell some plastics in order to get some money to pay my studies.

**Respondent 6:** I don't read at home. I am not motivated to read in English.

**Respondent 7:** I don't have textbook. How can I read at home? The textbooks we use at school is kept by the school principal in his office. He says if we take them with us at home we are going to dirty them.

**Respondent 8:** No need to read at home when you don't understand what you are reading.

**Respondent 9:** I have nothing to read at home.

### Question 3: Reading sources used at school and at home

#### ❖ *Participants who reported not reading either at home or at school*

**Respondent 1:** What am I going to read? I don't have textbooks or books. What can I read? I don't know why the educational authorities do not provide the school with textbooks for English like what happens for French where we use a textbook. It is as if English is not important.

**Respondent 2:** One of our classmates writes the text on the board; and this is with many spelling mistakes. So I cannot spend my time to read a text that I know is written with many spelling mistakes.

**Respondent 3:** We do not read frequently at school; why do I need to read at home? I am not interested in reading at home.

**Respondent 4:** I prefer to use my time to read the accounting course; I don't have time to read English. This is not my language, and I don't need to know it. Where shall I go with English?

**Respondent 5:** I don't read at school. How can I read at home?

**Respondent 6:** I prefer to focus on other subjects that I understand. English is not my favorite.

**Respondent 7:** We do not have any resource for reading. I am not motivated in

English.

**Respondent 8:** We have nothing to read.

**Respondent 9:** I am not motivated for reading. The texts we use are not relevant.

**Respondent 10:** There is nothing interesting to read in English.

#### **Question 4: Attendance to English classes**

##### *❖ Participants who regularly attended reading classes*

**Respondent 1:** English language is very important; the world is in rapid change and I must have at least a basic knowledge of English.

**Respondent 2:** There is nothing I can do today without some knowledge of English.

**Respondent 3:** They [the government] want to make English a second international language beside French; so I must make effort to speak it.

**Respondent 4:** They are telling us that at university, English is a compulsory subject in the first year of all faculties. They say that by 2015, students must be able to study different university subjects in French as well as in English. So, I must be prepared now.

**Respondent 5:** What can you do today if you do not know English? To get a good job at all these United Nations organizations that are working here to bring peace in the country, you need to be skillful in English. So I want to be good at English to get a well-paying job in one of these organizations.

**Respondent 6:** I like English because I know that if I am good at English, I will easily get a good job.

**Respondent 7:** I know that English is an important subject. If you are not good at it, you can be in trouble at the state examination.

**Respondent 8:** My success at the state examination also depends on my

performance on the English test; so I must attend the English class

**Respondent 9:** I don't want to compromise my luck. I must attend the English classes.

❖ *Participants who were attending English classes either sometimes or rarely*

**Respondent 1:** I would like to be good at English; but it is since Grade 9 that I was disappointed by the English teacher. He was not a good teacher, he was not motivated and he was coming to teach rarely; then I lost all my motivation. Now, I feel I cannot catch up whatever I can do.

**Respondent 2:** I don't see why I can regularly attend to English classes when my teacher [of English] rarely comes to teach.

**Respondent 3:** Whenever we have English class, I just go out. This is a very difficult course. We don't have textbooks. The text is written on the chalkboard with so many mistakes. I don't feel any interest to attend the English classes regularly.

**Respondent 4:** I think we need some materials for learning English. The use of these materials can motivate us to participate. I have lost all interest in English because of the way it is taught. I don't understand anything.

**Respondent 5:** I don't know if in English the same word can be pronounced differently. Today our teacher pronounces a word in one way; tomorrow he pronounces the same word in another way; few days later another pronunciation. Then what can you learn from such confusion?

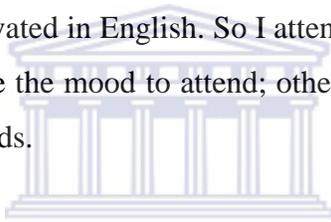
**Respondent 6:** I am not motivated to attend English classes. If I sometimes attend these classes, it is because we are obliged to do so. However, whenever I have an opportunity to skip these classes, I

don't hesitate.

**Respondent 7:** The Principal is very strict. If we don't attend English classes, he sends us for our parents. This is stressful because when your father comes to school, the Principal tells him many stories; and when you go back home, you find yourself in serious trouble with parents. So, I do my best to sometimes attend English classes in order to avoid trouble. Anyway, I am not motivated to attend English classes.

**Respondent 8:** I don't feel I need to attend English classes. Why can I attend English classes when I don't understand this subject? However, because we are forced to attend these classes that is why I do my best to attend some classes.

**Respondent 9:** I am not motivated in English. So I attend English classes when I feel that I have the mood to attend; otherwise, I just stay out and chat with friends.



#### **Question 5: Preferences of the number of hours per week for studying English**

WESTERN CAPE

##### *❖ Participants who preferred to study English in the time planned in the curriculum*

**Respondent 1:** English as an important subject. If we want to be able to speak, read and write it, we need many hours per week to learn it.

**Respondent 2:** I want more time than five hours [per week] because this enables me to become good at English and stand a chance for grabbing some job opportunities.

**Respondent 3:** I am for an increase of the time for learning English. Two hours are not enough to achieve good communication skills. Yet, if one is not able to communicate, he cannot pretend to now English.

**Respondent 4:** English is an important subject. We need to study it five hours a week as instructed.

**Respondent 5:** I don't find any problem to study English five hours a week.

❖ *Participants who had preferences for a reduced time to learn English*

**Respondent 1:** Five hours are too much. It is also too boring. This time must be reduced to two hours as in other technical schools.

**Respondent 2:** This is a difficult course. Five hours are too much; why not one or two hours; this will be enough.

**Respondent 3:** English subject is not as important as Math. I think we need to reduce the time for learning English to two hours so that we can have much time to learn Math subject.

**Respondent 4:** English is not an easy subject. It is good that the time to learn it be simply reduced.

**Respondent 5:** We need much time for French and Mathematics; not English.

**Respondent 6:** English is a headache. There is no need to study it five times a week.

**Respondent 7:** In other streams, there are only two English classes. Why do we have five classes? This is not fair.

**Respondent 8:** It does not make sense to spend so much time to study English. First our teacher of English is very lazy. Most of the time he comes late to class and when he is there, he spends much time for chatting with us. The English classes are just a waste of time. Better reduce the number of hour for English and increase the hours of the main subjects.

**Question 6: Attitudes towards the ESE**

❖ *Participants who were of the view that all Grade 12 students must write an English test*

**Respondent 1:** This is an important subject. We must write it.

**Respondent 2:** It is part of the curriculum; and we must write it. We have no choice.

**Respondent 3:** If you refuse to write this test, you lose 9 per cent; and your chance to get your certificate becomes reduced.

**Respondent 4:** It is good to be evaluated by some people other than your teachers. I feel proud to write the English test.

**Respondent 5:** An external exam is the only way to ensure that what we were taught is what is requested to be taught. I believe the English test is a good measure for that.

**Respondent 6:** As English is a school subject; we must also write it.

**Respondent 7:** I don't see any problem writing an English test

**Respondent 8:** It is part of the curriculum. We have no choice

❖ *Participants who were not of the view that all Grade 12 students must write an English test*

UNIVERSITY of the  
WESTERN CAPE

**Respondent 1:** Why do we need to write an English test when they know we are going to fail? Why don't they make it an optional subject? All of us don't need English. How many of those who teach English have gone to an English country? We can live without this language, and they don't need to make it a must.

**Respondent 2:** It would be good for us not to write the English test. We can just study English as a school subject; but not write it.

❖ *Participants who were of the view that the ESE was a way of failing candidates in the national test*

**Respondent 1:** I am sure I am not going to get any marks in this test [ESE]. I

am working hard in other subjects to catch up. I'm afraid I may not get my certificate with good marks and be denied to start university by next year.

**Respondent 2:** They know that English is difficult; but they have included nine questions [9 percentages] for the English test. We are going to fail.

**Respondent 3:** I am afraid I might get my certificate with a low score because of this [English] test.

**Respondent 4:** They know that most students are not good at English; and that we do not like it. Still, they design exams that are very difficult. This is to make us fail.

❖ *Participants who were of the view that they sometimes felt they did not need to write an English test*

**Respondent 1:** If I were given a choice, I would not write the English test.

**Respondent 2:** I don't know if I can just ignore writing this test; maybe this cannot minimize my chance.

**Respondent 3:** I really don't feel the need to write an English test.

**Respondent 4:** I don't know what I can get if I don't write this test. Maybe I just to ignore it.

❖ *Participants who were of the view that the ESE must test not only reading but also other skills*

**Respondent 1:** Why only reading? It is as if reading is the only thing we must know.

**Respondent 2:** If the ESE includes sections of questions on grammar and vocabulary; we will be able to pass it quite easily. But reading [comprehension] questions are difficult because you cannot understand

the text the way it is written. You can have your understanding and when you look at alternatives you choose one you think relates to your understanding; yet this is a wrong answer. But with a question on grammar, we all understand it in the same way. It is good to include other questions on grammar and vocabulary.

**Respondent 3:** I want the English test to include also questions on grammar and vocabulary; and not only on text comprehension.

**Respondent 4:** I don't understand why the test is only based on the text. Maybe it must include other things.

**Respondent 5:** Reading is not the only thing [skill] we need to know when we learn English. Why not also speaking?

#### **Question 6: Preferences for the number of hours to write the ESE**

##### *❖ Participants who preferred two hours or more*

**Respondent 1:** Two hours are not much. This is not our language. We need much time to write the test, maybe there can be some kind of assistance and if the time is reduced; we will not be able to get this assistance and collaborate.

**Respondent 2:** With 4 different alternate forms, we need much time so that we can check who has got the same form with me. After each of us has answered the test questions, we can therefore try to compare our answers.

**Respondent 3:** As English is a very challenging subject, the only one way to pass it is to organize ourselves in order to get answers from outside. I am sure that if we are given much time; we can do so.

**Respondent 4:** We need more time for the English test. It is a difficult subject.

**Respondent 5:** A test like English requires that we have much time.

**Respondent 6:** For the English test, there is some benefit in more time than in less time.

**Respondent 7:** Two hours for writing the English test are at our advantage.

**Question 7: Preferences for the number of questions to be included in the ESE**

❖ *Participants who preferred the actual number of questions to be reduced*

**Respondent 1:** When the ESE has 9 questions, this number of questions accounts for 9 % of the total marks. Since we are not sure to get many questions right, a reduced number of questions is advantageous to us; I can fail in English but still get my certificate with good marks.

**Respondent 2:** Imagine the ESE included only two or three questions. If I don't get these questions right, I cannot worry because my chances to get my certificate are still intact.

**Respondent 3:** Few questions are good because my final score to the state examination will not be affected.

**Respondent 4:** Why so many questions for a subject I don't understand. I think few questions are good.

**Respondent 5:** I prefer the ESE to include less than the actual nine questions because I view nine questions as too much. When you fail the test, you risk to fail the state examination

❖ *Participants who preferred the ESE to include more than the actual number of 9 or 5 questions*

**Respondent 1:** 9 [nine] questions are not enough. How can we learn English for 4 years and at the end they ask 9 questions to check our ability? This cannot provide a clear indication of our abilities. We need many questions.

**Respondent 2:** Those people cannot expect to include everything we have learned in 5 questions. This is not good. More questions are

needed.

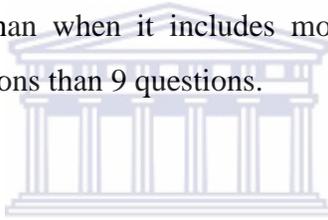
**Respondent 3:** I believe that I stand a chance to pass the ESE if it includes more questions than it does now. With few questions, you take all risks to get most of them either right or wrong. I am not for taking such a risk.

**Respondent 4:** I like the ESE to include more than 20 questions. This gives more chance to pass this test.

**Respondent 5:** More questions are good. I want the ESE to include more than the actual number of five questions.

**Respondent 6:** English is a main subject. The ESE must include many questions.

**Respondent 7:** I think it is too risky to fail the ESE when it includes few questions than when it includes more questions. So, I prefer more questions than 9 questions.



#### **Question 8: Preferences for the number of alternatives to be included in the ESE**

WESTERN CAPE

##### **❖ *Participants who preferred a three alternative test***

**Respondent 1:** 3 [three] alternatives are good because they do not confuse us. So many alternatives create much confusion.

**Respondent 2:** 3 [three] alternatives are good. We can easily guess the correct answer. With 6 alternatives, you cannot easily guess.

**Respondent 3:** With many options, it becomes difficult for us get the correct option. But with only three options, it is easy.

**Respondent 4:** Many alternatives are confusing.

**Respondent 5:** Three alternatives are good. There is no confusion.

**Respondent 6:** It is easy for me to guess the correct answer with three alternatives than with six alternatives

**Respondent 7:** With six alternatives, some of them look similar and this is

sometimes confusing. I prefer three alternatives.

**Question 9: Preferences of the number of parallel forms to be included in the ESE**

❖ Participants' preferences for the single form test

**Respondent 1:** I want a single form. This enables us to easily compare answers. With 4 alternate forms, you must be careful to look at your neighbor's answer sheet.

**Respondent 2:** I like a single form test because it is easy for us to collaborate.

**Respondent 3:** These people want only to make our lives difficult. With four alternate forms, it becomes difficult to know who has the same exam as you. Therefore, you cannot have any help.

**Respondent 4:** When we study English, we are in the same class. The teacher asks the same questions to everybody. Those who succeed are those who deserve to succeed. But when students who have studied under the same conditions are given different questions, this is no longer fair because some [parallel] forms might be easier than other [parallel] forms. So if all candidates are given the same form; the assessment of these students is fair.

**Respondent 5:** Why so many complications? When we write school examinations, we are all given the same questions. Why do we need to have different questions in the state examination? Do they want to say that school examinations are not good?

**Respondent 6:** A single form test is good. We can compare our answers.

## APPENDIX 24:

### TEACHER-PARTICIPANTS SELECTED RESPONSES TO THE CONTEXTUAL QUESTIONNAIRE

#### Question 1: Methods used for classroom assessment

##### ❖ *Participants who regularly used open questions*

**Respondent 1:** It is easy for me to construct open questions. I do not have time to construct multiple-choice questions.

**Respondent 2:** It just takes me few minutes to design a test when I use open questions.

**Respondent 3:** I use open questions because they give me the opportunity to know which students have understood the course materials and which ones have not understood. With the MCQ, you cannot tell that. Many students easily pass answers to their colleagues.”

**Respondent 4:** Open questions are good because they enable to check which students have understood the course materials and which ones have problems.

**Respondent 5:** With open questions I am able to identify areas where my students have problems so that I can plan corrective tasks.

**Respondent 6:** Open ended questions are the best way to assess students’ text comprehension.

**Respondent 7:** These types of questions are easy to construct. I like them.

**Respondent 8:** They are not only easy to design; but they provide a good indication of learners’ achievement,

**Respondent 9:** Although it takes time to mark these questions, they are challenging to the students. Thus, they provide an good picture of

their proficiencies.

❖ *Participants who used the multiple-choice questions*

**Respondent 1:** I know that in classroom assessment, we must initiate our students to the MCQ because the ESE is in an MCQ format.

**Respondent 2:** I must familiarize my students with multiple-choice questions. Otherwise, they will look confused when they write the ESE.

**Respondent 3:** My students should be able to identify all tricks used in the multiple-choice questions. This is the only way they can build confidence and develop positive attitude towards the ESE.

**Respondent 4:** I use the old ESE edition tests to familiarize my students with the MCQ technique. This builds their confidence when they write the test.

**Respondent 5:** I use the MCQ to familiarize my students with the ESE.

❖ *Participants who never used the MCQ*

**Respondent 1:** I never use the MCQ. They are very difficult to construct; and they need so much time. I think I am not prepared to do that. Also, they give an unfair advantage to weak students.

**Respondent 2:** I never use the MCQ because I don't have samples of ESE test papers. The time I get some of these ESE papers, I will use them.

**Respondent 3:** I don't like them [MCQ]. They don't provide a good indication of students' abilities.

**Respondent 4:** I never use them. They are difficult to construct. Besides, they make students lazy.

**Respondent 5:** They are easy to answer. I don't use them.

## Question 2: Sources of materials used for reading classes

**Respondent 1:** Since I have been teaching English, the school does not have textbooks. I have managed to photocopy some texts from 2 [two] textbooks and I have to write them on the chalkboard for the students to copy in their notebooks.

**Respondent 2:** We do not have a school library; we do not have internet connection at school; where can I get the text? I have my textbook but my students do not have any copy; so I have photocopied some texts I have found good for students. Photocopied texts are the only one reading material I use in reading classes.

**Respondent 3:** There is a huge problem of textbooks. What I do is to search for some relevant materials from technical brochures and magazines and I make some copies of these materials. This is not easy because the majority of students cannot afford these photocopied texts. But at least, these texts have the advantage to provide the learners with visual support for reading.

**Respondent 4:** My school has only five (5) copies of the textbook. Yet, I have 53 students in my class. It becomes difficult to use these five copies in reading classes. The only solution I have is to ask one student to write the text on the board and the other students to copy the text from the board. This is not without consequence. It takes so much time to complete writing the text and also it is too demotivating as some students just refuse to copy the text from the board.

**Respondent 5:** I don't have a large class; therefore, four students share one textbook.

## Question 3: Views on some statements

❖ *I have all textbooks required for teaching English in Grade 12*

**Respondent 1:** Normally, there are four prescribed textbooks that can be used in Grade 12. These textbooks are: Cartledge Book 4, English for Africa 6, Today's English, and Go for English. Of these textbooks, I only have one: English for Africa. The school does not have the other textbooks.

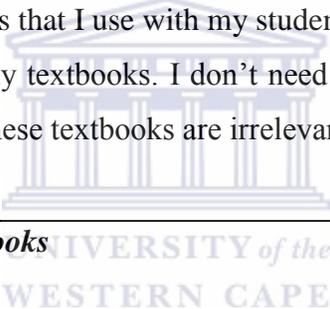
**Respondent 2:** I have not any of the required textbooks. I have photocopied some texts selected from the textbooks from a colleague.

**Respondent 3:** I have only one textbook; Cartledge Book 4. And this is the only textbook I use to teach reading.

**Respondent 4:** I have two textbooks: English for Africa 6 and Cartledge Book 4. However, I don't use these textbooks because they are old-fashioned. Furthermore, the texts included are more general than specific to the subject area of my students. So I have made my own collection of some technical texts that I use with my students.

**Respondent 5:** I don't have any textbooks. I don't need them because the text passages included in these textbooks are irrelevant to my students' needs.

❖ *My students also have textbooks*



UNIVERSITY of the  
WESTERN CAPE

**Respondent 1:** Since I have been teaching English, I don't remember a single time when my students have used textbooks. There are no textbooks in my schools and nobody cares about that.

**Respondent 2:** Three to five students share one textbook. After the class, I get these textbooks back and have them kept in the Principal's office. So, my students don't have textbooks to read at home.

**Respondent 3:** My students don't have textbooks. This is really a reading problem. How can I expect them to be good readers when they don't have reading materials?

**Respondent 4:** My students don't have textbooks. Nevertheless, I have photocopied some texts selected from the different textbooks and my students are using these texts as reading materials.

❖ *I have the English curriculum*

**Respondent 1:** It is really a pity that, as an English teacher, I do not have the curriculum. This document is very important because it contains all the topics that teachers must use to teach their students. Yet, when I asked for it to the school principal, he told me that this document was not part of documents delivered by the provincial government and that I had to check with English teachers from other schools to get it and print a copy for myself.

**Respondent 2:** I have an English curriculum; but there is little relationship between the curriculum content and the textbook content. So it becomes difficult to teach all the required topics suggested in the curriculum because there is not any support for some topics.

**Respondent 3:** I don't have an original copy of the English curriculum; but I have copied relevant topics planned in the curriculum. I use these notes to plan my English course.

UNIVERSITY of the

❖ *I use the English curriculum to design my course outline*

**Respondent 1:** When I was having my English methodology module, the lecturer repeatedly insisted that before we graduate, we should make sure that we have all necessary documents required for use by the English teachers. Unfortunately, the Department of English was not able to provide us with these documents at that time. Upon graduation, I realized that I had none of them and it is not now that I can expect to get them in my school since what counts more for the school principal is to find teaching materials for the main subjects; not for English that they view as minor subject.

**Respondent 2:** Since I don't have a copy of the English curriculum, I only rely on my old notes to design the course outline.

**Respondent 3:** At the beginning at each school year, I borrow a curriculum from a

friend who is teaching English in a neighboring school; and I therefore design my English course.

❖ *The English curriculum needs to be updated*

**Respondent 1:** Our curriculum suffers from old age. Nowadays, things are advancing, the world is progressing and changes are necessary. The curriculum needs to be adapted to reality and this cannot be expected with the curriculum that was elaborated in the eighties when computers and internet were not known by most Congolese scholars.

**Respondent 2:** In teaching general English, most authors suggested are situated in a certain time and space. Take for example Chinua Achebe, Cyprian Ekwensi or Amos Tutuola. Their writings belong to a certain epoch of independent Africa. If it is a good idea that our students must be initiated to reading African literary writers in English, we must look at the content of the themes developed by these writers. I think that the list of the writers must be updated along with the themes developed in order to be closer to learners' real life. I don't know, maybe there are writers who have addressed the problems of corruption, war, insecurity or human rights that plague the African states nowadays. Why not use their writings if there are some.

**Respondent 3:** The English curriculum is too old. Just when you look at it; you find many inconsistencies. Its content hardly relates to the current textbooks. Also, the suggested methods for teaching are old-fashioned. This document is very mute when it comes to the assessment of learning.

**Respondent 4:** It is a pity that the government can't look into this problem. They want English to be taught as a subject and that it must be

spoken by our students; but at the same time, they ignore to design an appropriate curriculum that can translate their wishes. This document is not to me a curriculum. It is a pity that we use it; and we mustn't expect better results.

❖ *I am not motivated to teach English*

**Respondent 1:** I wish I were not a teacher. I don't feel I can do what I am supposed to do.

With no resources, no empowerment, teaching activity becomes simply a waste of time.

**Respondent 2:** I am really disappointed. When I enrolled at university to become a teacher in the early eighties, I had many expectations. I was dreaming to be a model for my family as well as my country. But what I am today is nothing just a good-for-nothing man. I have no motivation for teaching.

**Respondent 3:** With populated classrooms, no teaching supports, no motivation from the learners, no discipline at school, I feel that my place should not be here. If I am an opportunity to do other things, I won't hesitate to grab it.

❖ *My poor working conditions negatively affect my teaching*

**Respondent 1:** My salary does not enable me to make ends meet. I have to do other jobs to make more money. As a consequence, I don't have time to do properly classwork.

**Respondent 2:** With the current poor working conditions of Congolese teachers, there is nothing the government can expect from us. In my case, I rarely give tasks to my students because I won't be able to mark them. Yet, since English is a foreign language, learning can be enhanced through the design and use of many tasks that require the learners to use the language. Surely, what I am doing today is not teaching.

**Respondent 3:** Teaching conditions in the Congo are unacceptable. I confess that I do not prepare my lessons because I have to do other jobs to get extra money. Otherwise, my wife and children are going to starve.

❖ *The ESE must also test other skills (writing, speaking, and listening)*

**Respondent 1:** Why do we need to reduce English proficiency to reading comprehension? It is as if other skills are of no use. I think other skills must be tested to ensure that students are capable to be competent in English.

**Respondent 2:** You see what happens now. Other skills are neglected and even in course planning, these other skills deserve no attention.

**Respondent 3:** Language is speech; not writing. So, reading does not reflect proficiency in English language.

**Respondent 4:** Since comprehension is basic to language, it does not matter to base the ESE on reading comprehension.

**Respondent 5:** If other skills like speaking and writing are tested, most students will fail the test.

**Respondent 6:** It will be too costly to attempt to test oral skills like listening and speaking. This cannot be feasible.

❖ *The actual time allocated for writing the test is too much and therefore must be reduced*

**Respondent 1:** 2 [two] hours are too much for a 9 question test; likely, 1 hour and half is too much for only 5 questions. The reality is that many students do not use all this time to write the test; instead, they use much of this time for struggling to communicate among each other.

**Respondent 2:** The majority of Grade 12 students have proficiency deficit in English. I believe students need much time to read the text passage as many times as they can so as to comprehend it. So, the actual time is proportional to the test demands and students' skills.

**Respondent 3:** Our students are really struggling with reading. They need much time.

**Respondent 4:** Since English is a challenging subject, I think students need much time. This will help them to read the text and answer all the questions. Otherwise, they will leave some questions unanswered.

**Question 4: Students' motivation in reading classes**

**Respondent 1:** The majority of my students are not motivated in my course. They say that English is a difficult subject. Some of these students don't regularly attend English classes.

**Respondent 2:** Many students do not attend my classes. When I come in, I find that some students have left the class and they are in the school yard chatting. These students really don't have interest in English course.

**Respondent 3:** I am always concerned about my students' poor participation in my classes.

**Respondent 4:** My students are lazy. They don't push. They expect things to be easy; but at the same time they don't work hard.

**Question 5: Some variables that negatively affect the reading classes in particular and English classes in general**

❖ *Large number of students*

**Respondent 1:** How can a small venue not equipped with enough desks accommodate 70 learners? It is very difficult to attend to each student and provide assistance. When you try to assist some, the others feel abandoned and therefore shift their focus to other things. As a result, you find out that most students have not been able to read the text and complete the reading task as requested.

**Respondent 2:** My class is very populated. I can't check how most of my learners read the text.

**Respondent 3:** I have 51 students; but I have only 10 textbooks. It becomes very difficult to teach reading in good conditions. More than five students agglutinate around one textbook and this usually creates chaos.

**Respondent 4:** I have many students; and I can't assist each and every student. I think that few students could learn better than the actual number of students in my class.

❖ *Lack of students' involvement in English classes*

**Respondent 1:** Some of my students do not regularly attend English classes. They repeatedly say that English is difficult; and this has as consequence poor proficiency level in English. Those who attend regularly are generally good at English. So I think the only way to be good at English is to attend English classes regularly.

**Respondent 2:** Only about half of my students regularly attend my classes. What can we then expect from students who do not regularly attend the English classes? These are the students that are among those who repeat the year two, three or four times.

**Respondent 3:** Many students attend my classes, but only few actually participate in the lesson. The majority of students remain very quiet and almost ignore me during my classes.

❖ *Students' poor level of instruction that does not reflect the grade level*

**Respondent 1:** I think that many of my students who are in Grade 12 have a very low level in English as well as in other subjects. I can state that they do not deserve to be in Grade 12.

**Respondent 2:** The level of my student is very low. It does not reflect the level of Grade 12 students.

**Respondent 3:** My students don't know anything in English. They do not deserve to be in the final year.

❖ *Students' expectation that they will collaborate when they take the ESE*

**Respondent 1:** Many students are lazy. They do not attend my classes. They maintain that they will collaborate in the testing centre.

**Respondent 2:** It is a pity to see that the majority of final year students do not regularly attend the English classes. They repeatedly say that English is a difficult subject; and they will find a way to cope with the English test.

❖ *The policy adopted by the school in terms of discipline*

**Respondent 1:** Discipline is partly responsible of final year students' poor results. They lack respect for the Principal and the teachers. They think they have achieved.

**Respondent 2:** There is no discipline at school. Final year students attend classes when they want to. This seriously affects their performance on the national test.

❖ *Lack of collaboration among teachers*

**Respondent 1:** Each teacher stays in his corner. There is no collaboration between teachers. As a consequence, when you take a decision related to

students' discipline behaviours, there is no support from other teachers. Therefore, you find isolated and you therefore drop the decision.

**Respondent 2:** It is a pity that English teachers do not meet. We don't have an association that can help address issues related to the teaching of English in our town. Yet, if we were united and we were collaborating; we could easily overcome many issues such as those relating to the teaching materials and the use of multiple methods in English teaching.

❖ *Lack of parents' involvement in their children's studies*

**Respondent 1:** Many parents do not get involved in their children's education?

**Respondent 2:** It is a pity to see that the majority of parents are not interested in their children's education. They think that their responsibility is limited at home and when the child comes to school it is only the school's responsibility.

**Respondent 3:** Frankly, there is no collaboration between the school and the parents.

❖ *Lack of involvement of English teachers in test construction*

**Respondent 1:** Why is the ESE constructed by people who do not actually teach? We need to be involved in ESE construction because we know the reality better than do the inspectors who construct the ESE.

**Respondent 2:** It is unfair to keep teachers away from the construction of the ESE. They are the people who know better the students and the actual learning problems. They must be involved in one way or another.

❖ *Lack of teachers' in-service training*

**Participant 1:** To the best of my knowledge, since 2001 when I graduated and started teaching English; there has never been any training organized for English teachers who are currently practicing. How can teachers keep on with new developments in language teaching? Everything I know of teaching English is what I had learned at university.

**Participant 2:** English is not the language used in this country. It is important that those who teach it be regularly trained. This training programme may focus on sending teachers for some time to an English speaking country even those which are neighbor to us like Zambia, Tanzania or Uganda. This may give them the opportunity to learn how language is actually used. Sometimes we just teach bookish English and not actual English; and this does not benefit to learners.

**Participant 3:** I have been teaching English since 1980. Since then, I remember having participated in a training session only two times. The first time was in 1984 when the British Council in Kinshasa organized a training session on the communicative approach to language teaching. The second time was in 1986 when the Peace Corps came to organize a seminar on the teaching of literature. That is all. The past 20 years no seminar has been organized. This has a big consequence on the use of teaching methods. There is a big confusion in our country depending on where you completed your studies. Some teachers pretend they have been better trained because they use the communicative approach. Others repeatedly support that the audio-lingual approach is best especially in situations like our country where the communicative approach cannot fit because of lack of appropriate resources. Still others support the direct method and they say it produces good results. Some even still use the Grammar Translation method, especially in technical schools. Myself I usually support that I don't teach the method, I teach the language. If there were regular seminar organized for teachers, we could not have this confusion. We could know what we are all expected to do.

❖ *Teaching programmes not usually completed*

**Participant 1:** I have been teaching English for the past 13 years and I do not

remember a single time when I completed all the topics scheduled in the curriculum. There are always so many interruptions. Sometimes it is teachers who interrupt working because they are not paid; or other times they interrupt teaching because they ask an increase of salary which the government refuses to make. At other times students regularly interrupt courses for weeks or months because they ask for a reduction of school fees but the educational authorities refuse to attend to their demands. Negotiations take too long before school starts again.

**Participant 2:** Although we are at the centre of the country and we have never been directly affected by war, insecurity and violence that characterize the country; we still face the consequences of this instability. Many children who were displaced by war in the Kivu come here to start school sometimes only two months before the end of the school year. Since there is no special programme designed to respond to such situations, you put these children in the same class with those who started at the beginning of the school year. You realize that you have almost two or three different groups in one class and you find out that you are actually teaching three groups at different speed. As a consequence, you end up by not completing the entirety of topics scheduled in the curriculum. When next year these students pass to the next grade, the English teacher does not recapitulate the previous year's materials; instead, he just starts the programme for that grade. With this gap, students hardly catch up.

**Participant 3:** With the past general elections and the chaos that has resulted from these elections, schools were closed for more than two

months. When we finally resumed school, there was nothing arranged by educational authorities to catch up with the time that was lost. As a result, we just try to do what we could do, and many lessons foreseen in the curriculum could not be taught.



## Appendix 25

### SELECTED STUDENT-PARTICIPANTS PROFILE

Participant Code number	Age [years]	Gender	Subject area	Languages spoken and proficiency level*				
				Ciluba	Swahili	Lingala	Kikongo	French
1	19	M	Math-physics	3	2	5	3	4
2	18	M	"	5	1	4	1	4
3	19	M	"	5	1	2	1	3
4	22	F	"	4	4	4	1	4
5	18	M	"	5	1	3	1	4
6	18	F	"	5	1	3	1	4
7	22	M	"	5	4	4	2	5
8	19	M	"	3	5	3	1	5
9	22	M	"	5	5	4	1	5
10	19	F	"	5	1	2	1	4
11	20	M	"	4	5	4	2	4
12	18	M	Bio-Chemistry	5	5	4	1	4
13	23	M	"	3	2	5	5	4
14	19	M	"	3	1	5	3	4
15	18	F	"	5	2	4	1	4
16	19	M	"	5	1	2	1	3
17	22	F	"	5	3	4	1	4
18	18	F	"	5	1	3	3	4
19	18	M	"	5	1	3	1	4
20	22	M	"	5	4	4	1	5
21	19	M	"	3	3	3	1	5
22	22	F	"	5	5	4	2	5
23	19	M	Literary	5	1	2	1	4
24	20	M	"	4	5	4	1	4
25	18	M	"	4	3	4	1	4
26	23	M	"	3	2	5	4	4
27	19	F	"	3	2	5	3	4
28	18	M	"	5	3	4	2	4
29	19	M	"	5	3	2	1	3
30	22	F	"	5	4	4	1	4
31	18	M	"	5	2	3	1	4

32	18	M	"	5	1	4	2	4
33	22	F	"	5	5	4	1	5
34	19	M	Comm.-Adm.	4	5	3	2	5
35	22	M	"	5	3	4	1	5
36	19	M	"	5	1	2	1	4
37	20	M	"	4	5	4	2	4
38	18	M	"	5	4	2	1	4
39	23	M	"	3	2	5	5	4
40	19	M	"	3	2	3	3	4
41	18	M	"	4	3	4	1	4
42	19	M	"	5	1	2	4	3
43	22	F	"	5	4	4	1	4
44	18	M	"	5	2	5	1	4
45	18	M	"	5	1	3	5	4
46	22	M	"	3	3	4	1	5
47	19	M	"	3	5	3	4	5
48	22	F	Pedagogy	5	3	4	1	5
49	19	F	"	5	1	2	2	4
50	20	M	"	4	4	4	1	4
51	18	M	"	5	5	4	1	4
52	23	M	"	5	2	3	5	4
53	19	M	"	4	2	5	3	4
54	18	M	"	5	1	4	1	4
55	19	M	"	5	5	5	1	3
56	22	M	"	5	4	4	2	4
57	18	M	"	5	1	3	1	4
58	18	M	"	5	2	2	1	4
59	22	M	Agriculture	5	4	4	1	5
60	19	M	"	2	5	3	1	5
61	22	F	"	5	5	4	1	5
62	19	F	"	5	1	2	2	4
63	20	M	"	4	5	2	1	4
64	18	M	"	5	5	4	1	4
65	23	M	"	3	2	5	5	4
66	19	F	Dressmaking	5	2	5	3	4
67	18	F	"	5	1	4	1	4
68	19	F	"	5	3	2	2	3
69	22	F	"	5	4	4	1	4
70	18	F	"	5	1	5	1	4
71	18	F	"	3	3	3	1	4
72	22	F	"	5	4	2	1	5

73	19	M	Sec.-Comp.	3	5	3	1	5
74	22	F	"	5	5	4	1	5
75	19	F	"	5	1	2	3	4
76	20	M	"	4	2	5	1	4
77	18	M	"	4	5	4	1	4
78	23	M	"	3	2	5	5	4
79	18	M	"	5	3	4	1	4
80	23	M	"	3	2	2	5	4

**\*1= poor; 2= adequate ; 3= good; 4= very good ; 5= excellent**



**Appendix 26**

**TEACHER-PARTICIPANTS PROFILE**

<b>Participant code</b>	<b>Age [years]</b>	<b>Qualification</b>	<b>Teaching experience [years]</b>
1	32	B.Ed	5
2	28	B.Ed	2
3	46	B.Ed/Hon.	4
4	57	B.Ed/Hon.	15
5	61	No degree	37
6	27	B.Ed	3
7	34	B.Ed/Hon.	7
8	42	B.Ed	17
9	52	B.Ed	23
10	27	B.Ed/Hon.	2
11	29	No degree	3
12	35	B.Ed/Hon.	5
13	32	B.Ed	7
14	27	B.Ed	3
15	63	No degree	32
16	28	B.Ed/Hon.	2
17	30	B.Ed	3
18	51	No degree	25
19	45	B.Ed	5

20	37	B.Ed/Hon.	5
21	31	B.Ed	3
22	33	B.Ed/Hon.	2
23	25	B.Ed	2
24	53	B.Ed/Hon.	17
25	48	No degree	20
26	51	B.Ed	20
27	44	B.Ed	21

