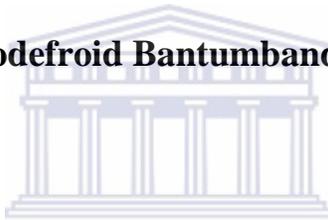


**THE DR CONGO ENGLISH STATE EXAMINATION:  
SOME FUNDAMENTAL VALIDITY ISSUES**

**Evaluating Context Validity Evidence**

**By**

**Godefroid Bantumbandi Katalayi**



**Thesis presented in fulfillment of the requirements for the degree  
of Masters in Language Education in the Faculty of Education at  
the University of the Western Cape, South Africa**

**Supervisor: Prof. Sivakumar Sivasubramaniam**

**November 2011**

## **KEY WORDS**

State examination

English state examination

Validity

Validation

Context validity

Context validity evidence

Text-based test

Multiple-choice text

Test takers

Test constructors



## DECLARATION

I, the undersigned, declare that “The DR Congo English State Examination: Some Fundamental Validity Issues (Evaluating Context Validity Evidence)” 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: -----November, 2011

-----

## ABSTRACT

The test context is of paramount importance in language testing as it provides an understanding of the kind of tasks to be included in the test, how these tasks are executed by the test takers and how they can be efficiently administered. The objective of this study was to investigate the extent to which the context of the DR Congo English state examination (ESE) is valid and to come out with some useful suggestions that are likely to improve its validity. Two basic theories, the modern validity theory and the schema theory, informed this study. Weir's (2005) socio-cognitive framework was used to build the validity argument for the evaluation of the English state examination.

A mixed method was used where the research design consisted of the combination of both qualitative and quantitative data during the collection and analysis stages. The content document analysis method was used to examine the content of the different state examination papers so as to identify the main features of the test, and the statistic (descriptive) method was used to quantify observations identified in the state examination papers and to evaluate the context validity of the ESE. Three techniques were used to collect the research data: the questionnaire, the test, and the interview.

Three main findings of this study were reported: (1) the conditions under which the ESE tasks are performed and the relevance of these tasks to the test domain and characteristics are still far to contribute to the quality of evaluation of high school finalist students; (2) the extent to which the ESE includes tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task is globally good; (3) the conditions under which the test takes place are poor and these conditions affect the validity of test scores.

The study recommends the test developers to approximate test tasks to those students have been exposed to in classroom situations and those they are likely to encounter in real life. It also recommends all the people involved in the administration of the test to adhere to high ethical standards.

## ACKNOWLEDGEMENTS

*Some trust in their war chariots and others in their horses, but we trust in the power of the Lord our God (Psalms 20: 7).*

I wish to express my gratitude to Professor Sivakumar Sivasubramaniam for the supervision of this thesis. His advice and guidance have been valuable to the completion of this work.

I am very grateful to Dr. Caroline Kerfoot who started the supervision of this thesis before the coming of Prof. Sivasubramaniam to the University of the Western Cape. Her contribution to the completion of this thesis has been highly valuable. Without her advice, support and encouragement, I would have not been able to achieve this study.

Special acknowledgements are addressed to Professor Biselela Tshimankinda from “Ecole Doctorale, Universite Pedagogique Nationale, DR Congo” who assisted me in the development of the proposal of this study before I came to the University of the Western Cape. His advice laid the foundation to this thesis.

This study would have not been possible without the sacrifice and support of my family. Special acknowledgments are addressed to my wife Angelique Kasokome for her love, advice, encouragement and support; and to my children for their love and confidence in me.

Finally, I am sincerely grateful to all those who, in one way or another, contributed to the completion of this work. I would like to mention particularly Clement Bula Basuayi for his true friendship and his assistance to find me a way to this university; Dr. Justin Munyakazi for his assistance and encouragement; and Professor Zubeida Desai, Education Faculty Dean, for her financial support.

## ACRONYMS

<b>AERA</b>	American Educational Research Association
<b>ANOVA</b>	Analysis of variance
<b>APA</b>	American Psychological Association
<b>ASL</b>	Average sentence length
<b>ASW</b>	Average word length
<b>DR Congo</b>	Democratic Republic of the Congo
<b>ESE</b>	English state examination
<b>ESL</b>	English as a second language
<b>GEPT</b>	General English proficiency test
<b>GS</b>	General Subjects
<b>IELTS</b>	International English language testing system
<b>L2</b>	Second language
<b>MC</b>	Multiple-choice
<b>NCME</b>	National Council on Measurement in Education
<b>SAT</b>	Scholastic Aptitude Test
<b>SE</b>	State examination
<b>TOEFL</b>	Test of English as a foreign language
<b>TS</b>	Technical subjects
<b>TSAR</b>	Technical subjects: Agriculture and related subjects
<b>TSCR</b>	Technical subject: Commercial and related subjects

**TSD**

Technical subject: dressmaking

**TSMR**

Technical subjects: Mechanic and related subjects



## **LIST OF TABLES**

Table 401: Sample size

Table 501: Respondents' views on the rubric clarity

Table 502: Task requirements and strategies use

Table 503: Scoring criterion: 1 question = 1 point and 1 %

Table 504: Combination of test scores and other components

Table 505: Moderation of test scores with classroom assessment scores

Table 506: Ordering of items in ESE

Table 507: Respondents' views on the time allocated to the ESE

Table 508: Percentage of text length in three textbooks in use

Table 509: Patterns of text length at ESE for both general and technical subjects

Table 510: Respondents' views on text length

Table 511: Difficulty estimate for the 7 ESE texts

Table 512: Respondents' views on the uniformity of administration conditions

Table 513: Respondents' views on the issue of possessing ancient ESE papers

Table 514: Respondents' views on students' access to test content before test administration

Table 515: Quantitative and qualitative evaluation of ESE context validity

## **LIST OF FIGURES**

Figure 501: Stakeholders' views (agree/strongly agree) on task purpose

Figure 502: Grouping of ESE items in testlets: views of stakeholders

Figure 503: Respondents' views on the assumption that the time allocated to take the ESE is too much

Figure 504: Text types as used in both general and technical subjects

Figure 505: Concentration of texts regards their length

Figure 506: Quantitative evaluation of the ESE task setting

Figure 507: Quantitative evaluation of the ESE task demands

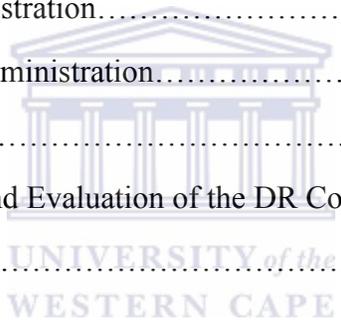


# CONTENTS

Key words.....	i
Declaration.....	ii
Abstract.....	iii
Acknowledgements.....	iv
Acronyms.....	v
List of tables.....	vi
List of figures.....	vii
Contents.....	viii
Chapter 1 Introduction.....	1
1.1. Background.....	1
1.2. Motivation.....	2
1.3. Problem statement.....	4
1.4. Research question.....	6
1.5. Research objective.....	7
1.6. Significance of the study.....	7
1.7. Outline of the study.....	8
1.8. Basic concepts.....	9
Chapter 2 Theoretical Framework.....	12
2.1. Modern validity theory.....	12
2.1.1. Background: Traditional view of validity and validity evidence.....	12
2.1.2. Modern view of validity and validity evidence.....	14
2.1.3. Validity in language testing.....	16
2.1.3.1. Unified validity theory and theories of language use.....	16
2.1.3.2. Context validity evidence of a text-based test: Weir's framework for test validation.....	18
2.1.3.2.1. Task setting.....	20
2.1.3.2.2. Task demands.....	23
2.1.3.2.3. Test setting and administration.....	26
2.2. The Schema theory.....	26

Chapter 3 Literature Review.....	32
3.1. Reader variables.....	32
3.1.1. The reader schemata and background knowledge.....	32
3.1.2. Reader skill.....	40
3.2. Text variables.....	45
3.2.1. Text topic and content.....	45
3.2.2. Text type or genre.....	47
3.2.3. Text organization.....	48
3.2.4. Text syntax.....	50
3.2.5. Text vocabulary load and difficulty.....	51
3.2.6. Text length.....	53
3.2.7. Text readability.....	55
Chapter 4 Methodology.....	58
4.1. Research approach.....	58
4.2. Research methods.....	59
4.3. Data Collection.....	61
4.3. 1. Population.....	61
4.3.2. Sampling.....	63
4.3. 3. Data collection instruments.....	64
4.3.3.1. Questionnaire.....	64
4.3.3.2. Interview.....	68
4.3.3.3. Tests.....	70
4.4. Data Analysis.....	71
4.4.1. Data coding.....	71
4.4.2. Frequency analysis and cross-tabulation.....	72
Chapter 5 Data Presentation and Analysis.....	74
5.1. Task setting.....	74

5.1.1. ESE rubrics.....	75
5.1.2. ESE task purpose.....	77
5.1.3. Known criteria.....	80
5.1.4. Order of items.....	84
5.1.5. Time allocation.....	88
5.2. Task demands.....	90
5.2.2. Text types.....	91
5.2.3. Text length.....	100
5.2.4. Text content.....	104
5.2.5. Text difficulty and linguistic structure.....	116
5.3. Setting and administration.....	127
5.3.1. Uniformity of administration.....	127
5.3.2. Security.....	129
Chapter 6 Findings and Evaluation of the DR Congo Context Validity.....	135
6.1. Findings.....	135
6.1.1. Task setting.....	135
6.1.1.1. ESE rubrics.....	135
6.1.1.2. Task purpose.....	136
6.1.1.3. Known criteria.....	137
6.1.1.4. Order of items.....	137
6.1.1.5. Time allocation.....	139
6.1.2. Task demands.....	140
6.1.2.1. Text types.....	140
6.1.2.2. Text length.....	141
6.1.2.3. Text content.....	141
6.1.2.4. Text difficulty and linguistic structure.....	142
6.1.3. Task setting and administration.....	143



6.2. Evaluating the context validity of the ESE.....	144
Chapter 7 Conclusion, Suggestions for Test Development, Limitations, and Further Research.....	151
7.1. Conclusion.....	151
7.1.1. Task setting.....	152
7.1.2. Task demands.....	153
7.1.3. Setting and administration.....	154
7.2. Suggestions for test development.....	155
7.3. Limitations.....	158
7.4. Further research.....	159
Reference List.....	160
Appendix 1 Questionnaire submitted to high school finalists before they present the test.....	176
Appendix 2 Questionnaire submitted to high school finalists after they have presented the test.....	179
Appendix 3 Questionnaire submitted to language teachers.....	183
Appendix 4 Questionnaire submitted to English language inspectors.....	186
Appendix 5 Questionnaire submitted to other stakeholders.....	188
Appendix 6 Interview with A., a high school finalist.....	189
Appendix 7 Interview with W., a high school finalist.....	195
Appendix 8 Interview with M., a head inspector.....	196
Appendix 9 Interview with S., a language teacher.....	199
Appendix 10 Example of an expository test.....	200
Appendix 11 Example of argumentative test.....	201
Appendix 12 Interview with Prosper Lituli, General Inspector of Education..	

## CHAPTER 1 INTRODUCTION

In this chapter, I propose to outline the background to the study before I provide the motivation for the study and state the research problem. I also discuss the research aim in order to address the research question. This will help me explain the significance of this study, clarify some key concepts and outline the content of different chapters that make up this study.

### 1.1. Background

In the Democratic Republic of Congo (DR Congo), high school finalist students (6<sup>th</sup> grade) are required to take a national exit test in order to get a state “diploma” (certificate). This certificate enables them either to go to university or to start a professional life. This exit test is called the “examen d’etat” (state examination). Although the content of this national test depends on the students’ main subject, the test consists of four subtests: the science subtest (including mathematics, physics and chemistry components), the general cultural knowledge subtest (including geography, history and philosophy components), the main subject subtest (depending on the student’s main subject) and the language subtest (including French and English). This research investigates some validity issues that pertain to the English test.

The DR Congo English state examination (ESE), as well as the other subtests that make up the national test, is externally set, marked and moderated. Students in the whole country take the test the same day, under the same controlled conditions and within a specified time limit. The ESE is a text-based test of the multiple-choice type, with 3, 5, or 9 item questions depending on the student’s subject.

As regards ESE domain, the ESE is an achievement test. Its purpose is to measure test takers’ English language proficiency in situations and tasks that reflect the kind of tasks performed by the same students in the classroom as well as in real life situations. Therefore, the ESE is supposed to include tasks that are similar to those that are described in the English language national curriculum as well as

authentic tasks; that is, those that reflect the kind of task test takers are likely to perform in real life situations.

The participants to the ESE are Grade 6 high school students who are males and females. The number of candidates who have participated in the 2011 state exam session in June was 485,741 according to the education service sources as broadcasted on media.

The normal age of a 6<sup>th</sup> grade high school student in Congo is 18 years. However, since there are many factors that account for students' progress in both primary as well as high school, there are always some candidates who are more than 18 years of age.

As for the exposure to the English language, I hasten to mention that in DR Congo, English is taught as a foreign language beside French which is the official language. Therefore, the setting of English is *almost* only the instructional milieu; that is, the classroom where formal instruction takes place. In this context, students' exposure to English language is limited and such limited exposure is reflected through the design of test tasks. Participants in the ESE are students who have been learning English as a foreign language since grade 3 high school. The amount of time devoted to the learning of English as determined by the national curriculum is five (5) hours a week in general subjects and 2 hours a week in technical subjects. Note here that the general subjects refer to schools whose main subjects are arts, mathematics, physics, biology, and pedagogy while technical subjects are schools that organize jobs related subjects such as mechanics, electricity, electronic, dressmaking, etc.

## **1.2. Motivation**

As a lecturer at a “Teachers’ Training College” in DR Congo for seventeen years, I had continuously, together with most of my colleagues, wondering about the poor performance of high school finalist students in the national test as well as about poor performance of new registered students who have matriculated from

high schools. Poor performance on the national test was reflected in many reports that indicated that between 1995 and 2010, only 55 percent to 65 percent of candidates could get their national certificate (Revue de l'Inspecteur, 2010). Poor performance of new registered students at university was reflected in many dropouts of students especially in their first year. For instance, at Kananga Teachers Training College, only about 60 percent of first year students could pass in the second year (Rapport Academique de l'ISP/Kananga, 2010). Although no serious and comprehensive empirical studies were conducted to account for the causes of poor performance on the national test, disparate reports pointed to various factors that appeared to account for the low performance (for instance unqualified teachers, poor socio-economic conditions, ineffective school leadership) and to the evaluation policy adopted by the Congolese government (testing methods, evaluation standards; etc.). As for high dropout rate of new registered students at university, some reports pointed, among other causes, to the poor quality of high school national evaluation by suspecting the test to produce scores that could not reflect test takers' performance.

In 2005, I participated in the national education commission whose task was to identify major problems that undermine the evaluation of high school students to the national test and make some recommendations. Among the problems, the commission identified that since the national test was instituted in 1967, no validation studies were conducted to assess the effectiveness of the tests. Therefore, the commission highly recommended the need to conduct empirical studies in all curriculum disciplines in order to validate the existing tests. I therefore decided to research on the English test so as to investigate into those issues that pertain to its validity, that is, the extent to which the scores generated by this test reflect the test takers' performance. But since validation involves an extended research program that can hardly be conducted by an individual and during an academic time span, my concern in this study is limited to a single validity aspect of the English test, its context validity. I agree with McNamara (2000) when he argues that there is no perfect test that is valid for all purposes

and uses. However, test users have the responsibility to make sure that the test they use is appropriate for their own institutional or contextual needs.

### **1.3. Problem statement**

Since the DR Congo ESE is a multiple choice text-based test, there are potential issues that are due to the specifics as well as essence of such tests and those issues are likely to threaten its validity. First, there are issues that pertain to the testing of language skills selectively. As Power (2010) notes, there are serious consequences of choosing to test some aspects of language proficiency and not others. Second, there are issues that pertain to reading as a skill and a construct. Since reading is a cognitive process that involves an interaction between a reader and a text, all that the reader brings with him/her and the text have an impact on test performance (Alderson & Banerjee, 2000; Bachman, 1990; Bachman & Palmer, 1996; Weir, 2005). Therefore, since all these variables (text variables and reader variables) interact, measurement issues become complicated as they tend to reflect, as put by Fletcher (2006), the complex and multidimensional nature of reading comprehension. Third, it is important to consider the format by which reading comprehension is tested. The use of multiple choice (MC) format to test text comprehension addresses two basic issues. First, the issue that pertains to the reading construct that is assessed by the test. Rupp *et al.* (2006) sustain that there is no evidence that asking test takers to respond to text passages with MC questions induces response processes that are similar to those that test takers would draw on reading in non-testing situations. Second, there is the issue that pertains to different strategies that are used by test takers to answer test questions. Some of these strategies include the logical elimination of distractors, the informed guessing, and the role of test taker's prior knowledge (Sheeham & Ginter, 2001). These strategies, if not controlled by an appropriately designed test, are likely to affect test scores, thus test validity.

The above issues that pertain to a text-based test like the ESE can be overcome by a careful evaluation of the test. Such evaluation is made through validation studies

where an argument is developed to support the adequacy and usefulness of test scores. Yet, if such validation studies are not conducted, we cannot be sure that the test used is valid, that is, it is likely to produce scores that reflect candidates' performance.

As observed by Weir (2005), validity is multifaceted and different aspects of evidence are needed to support any claims for the validity of test scores. Weir (2005) advises that the more comprehensive the approach to validation, the more evidence collected on each of these aspects of validity, the more secure we can be in our claims for the validity of a given test. These aspects are: (a) Context validity (the extent to which the contextual characteristics of the test task and its administration are situationally fair and appropriate to the test takers), (b) Theory-based validity (the extent to which the test includes tasks that reflect the cognitive skills and strategies test takers are likely to deploy in reading the text and in providing answers to questions based on the text), (c) Scoring validity (the extent to which the test produces scores that reflect the test takers' level of language knowledge and skills), (d) Criterion-referenced validity (the extent to which the test scores correlate to other measures of the same construct it measures) and (e) Consequential validity (the extent to which the test influences positively or negatively curriculum developers, textbooks designers, language teachers and test takers).

In light of the above mentioned points, my study questions the first aspect of validity, that is, the context validity of the ESE. The next two validity aspects (theory-based validity and scoring validity) will be dealt with in my further research.

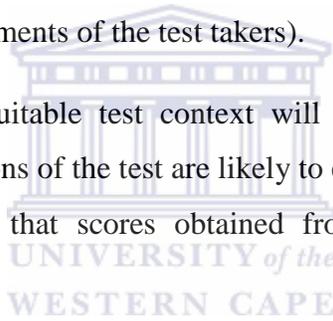
The test context is of paramount importance in language testing since it provides an understanding of the kind of tasks to be included in the test, how these tasks are executed by the test takers and how these tasks can be efficiently administered. Weir (2005), Bachman and Palmer (1996) argue that the test context should provide a suitable space for assessing test takers language abilities. The research problem of this study focuses on the capacity of ESE test developers

and administrators to provide a suitable context for test takers to perform the test in such a way that scores generated reflect their abilities.

Suitable test context will be provided if test tasks are performed under conditions that are relevant to test domain and test characteristics (test takers' age, their linguistic background, the experience they have with the test, clarity of test rubrics, purpose of the test task, criteria to be used in the test marking, order of items and tasks, time allocated for completion the task, etc.).

Suitable test context will also be provided if the test is designed in such a way that it includes tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task (text length, information in the text, lexical, structural and functional items appropriate for the target situation requirements of the test takers).

Last, but not least, suitable test context will be provided if the setting and administration conditions of the test are likely to enable the test takers perform the tasks in such a way that scores obtained from the test reflect their actual performance.



#### **1.4. Research question**

The issues discussed so far lead me to believe that validity is a matter of degree, that is, not all, not any. Therefore, this study addresses the following main research question: “To what extent is the context of the DR Congo ESE valid?”

On the basis of this main research question, I wish to address the following sub-questions:

- (1) To what extent are the conditions under which the test tasks are performed relevant to test domain and test characteristics?
- (2) To what extent the ESE includes tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task?

- (3) To what extent are the ESE setting and administration conditions likely to enable the test takers perform the tasks in such a way that scores obtained from the test reflect their actual performance?

### **1.5. Research objective**

The objective of the study is to investigate the extent to which the context of the DR Congo English examination is valid so as to offer some useful suggestions that are likely to improve its validity.

### **1.6. Significance of the study**

- This study explores an important concern in contemporary education debates; that of designing appropriate measurement tools for assessing students' abilities;
- The topic of this research is believed to be of important social value at this time in DR Congo, where the test used by the government has never been validated, and where the need for validation studies is more than crucial;
- This study is of practical importance to DR Congo education officials and testing body as it will serve as a starting point to developing a rationale assessment policy for high school finalists, not only for English, but also for other subjects;
- This study gives the opportunity to other researchers from 'developing' countries to look at the assessment policies of their countries with critical lens and to conduct similar research for developing rationale assessment policy proposals;
- This study opens up the possibility for further debate between academics, testing theorists, test constructors and the general public.

## **1.7. Outline of the study**

### **Chapter one**

This chapter provides the research background, the motivation for the study, the research problem, the research questions, the research objective, the significance of the study, and the key concepts used in this research.

### **Chapter two**

This chapter describes the two theories that inform this study: the modern validity theory and the schema theory. The modern validity theory is examined through a background survey of the traditional validity theory; then the application of this theory to language testing and how it can be used for context validation of a text-based test. The schema theory is described in relation to its impact on test processing and comprehension.

### **Chapter three**

This chapter explores the relevant literature that pertains to this study. This literature relates to two main trends: those that tap to the context validity of text-based tests and those that relate to the impact of background knowledge of both text processing and comprehension and their impact on test performance.

### **Chapter four**

This chapter explains the mixed method that is used in this study as research design. The two basic methods (content document analysis and descriptive statistic methods) are explained. The different instruments used for collecting data and how they were actually used are explained. The study population is defined, the sampling techniques clarified and the methods of treatment of data is explained.

### **Chapter five**

This chapter analyzes the research data. The data are analyzed with regards to the three research sub-questions.

### **Chapter six**

This chapter presents research findings and addresses the vital issues that underpin the context validity of the English state examination.

### **Chapter seven**

This chapter concludes this study by summarizing the research findings, giving suggestions for the improvement of the quality of the English state examination, highlighting the limitations of the study and suggesting further research.

## **1.8. Basic concepts**



The following section clarifies key concepts that make up the framework of this study. I believe that the understanding of these concepts aims to provide the readers with the conceptual framework of this study. These concepts include: test, testing, assessment, DR Congo state examination, DR English state examination, test validity, test validation, and context validity.

1. *Test/ testing*: In EFL context, a language test is a device that tries to assess how much has been learned in a foreign language course, or some parts of the course (Oller, 1979). Language testing is, as put by Inbar-Lourie (2008), a science that deals with issues pertaining to the design and use of tests as a mean for determining language proficiency. Yoshida (2006) gives a more general definition of language testing. For him, testing is a way to systematically measure a person's ability or knowledge, and it is formalized as a set of techniques or procedures.

2. *Assessment* is perceived to be an overarching term used to refer to all methods and approaches to testing and evaluation whether in research studies or in educational contexts. Norris (2000) defines language assessment as the process of using language tests to accomplish particular jobs in language classrooms and

programs. In language assessment, we first gather information in a systematic way with the help of language testing tools; then we may make interpretations of test results. Thus, on the basis of these interpretations, we make a decision or take action. In this perspective, language assessment appears as the entire process of test use; therefore, it helps us to be informed on the decisions we make and the actions we take in language education.

3. *DR Congo SE*; is a national exit test that is designed to evaluate the 6<sup>th</sup> grade high school students' achievement. Successful students get a "certificate" that enables them either to go to university or to start a professional life.

5. *DR Congo ESE*: is a subtest of the SE. This test is designed to evaluate the 6<sup>th</sup> grade high school students' achievement in English. The test is a text-based test of multiple choice type.

6. *Test validity*: I am of the view and belief that Mesick's position on test (1989, p. 41) validity can be helpful in understanding what this could mean in the larger domain of my study. As observed by him, 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

This definition entails some of the following considerations, as stated by Sireci (2007): (a) validity is not a property of a test; rather, it refers to the use of a test for particular purpose. This means that tests themselves are not valid or invalid, but we validate the use of a test score; (b) we need multiple sources of evidence in order to evaluate the utility and appropriateness of a particular test; (c) if the use of a test is to be defensible for a particular purpose, sufficient evidence must be given to defend the use of the test for that purpose.

7. *Test validation*: As pointed by Lane (1999) and in the newest edition of the American Educational Research Association (AERA), American Psychological Association (APA), and National Council of Measurement in Education (NCME)

*Standards for Educational and Psychological Testing* (2011), test validation is a process that involves the accumulation of evidence to support the proposed test score interpretations and uses. The process of accumulating evidence to support the validity of test score interpretations starts prior to the development of an assessment. As pointed out by Messick (1989), validation is a continuous process and it begins with a construct in search of appropriate assessment instruments and procedures.

One important issue related to the validation process is that the process of accumulating evidence is not to be understood as a checklist procedure where the test developer looks at test development standards and checks which standards the test conforms to. Rather, it is a process that involves the development and evaluation of a coherent validity argument for or against proposed test score interpretations and uses (Haertel, 1999; Messick, 1989; Moss, 2007, Kane, 2008).

8. *Context validity*: This could be understood as the extent to which the test uses tasks that are representative of the larger universe of tasks of which it is assumed to be a sample. Context validity relates to both the content of the materials taught in classroom situations and the kind of tasks the candidate is likely to perform outside the classroom and his future life.

## CHAPTER 2: THEORETICAL FRAMEWORK

I propose to discuss in this chapter the two basic theories that inform this study. These two theories are the modern validity theory and the schema theory. The first theory pertains to test validity and the second pertains to reading skill.

### 2.1. Modern validity theory

In this section, I offer a discussion on the modern validity theory. This discussion is built around two sections: first, a background of validity theory in relation to the traditional view of validity and validity evidence, and then a discussion on the modern view of validity and validity evidence.

#### 2.1.1. Background: Traditional view of validity and validity evidence

Validity is a theoretical concept that has undergone different meanings over time. Traditionally, validity was defined as “the degree to which a test measures what it claims, or purports, to be measuring” (Brown, 2000; p. 231). It was subdivided into three types: content validity, criterion-referenced validity and construct validity (Brown, 2000; Brualdi, 1999).

**Criterion referenced validity** usually included any validity strategies that focused on the correlation of the test being validated with some outside measures of the same objectives or specifications. In other words, this validity type aimed to demonstrate that the test scores are systematically related to one or more outside criteria. In terms of achievement test, like an exit test for example, criterion-referenced validity could refer to the extent to which a test can be used to draw inferences regarding students’ achievement (Brualdi, 1999).

**Content validity** was understood as the degree to which a test accurately samples from the course of study or domain. This has been termed the ‘real life’ approach to validity, in which different test tasks should be “a representative sample of tasks from a well-defined target domain” (Bachman, 1990, p. 310). Underhill, 1987, cited by Fulcher (1999) argues that content validity is the same as asking whether the test content reflects the syllabus of a course, the aims of a program of study, or the needs as set out in a needs analysis. Therefore, the validation of a test must rely to a greater extent on the test designer’s intuitive knowledge of the implicit objectives of the program. Content validity is often evaluated by examining the plan and procedures used in test construction. As suggested by Brualdi (1999), relevant questions to content validity include the following: Did the test development procedure follow a rational approach to ensure that the test includes appropriate items? Did the process ensure that the collection of items would represent appropriate skills?

**Construct validity** was traditionally defined as the experimental demonstration that a test is measuring the construct it claims to measure (Brown, 2000). As observed by Moss (2007), construct validity served the aims of inferring

the degree to which the individual possesses some hypothetical trait 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 through studies that check on the theory underlying the test (p. 471).

Studies of construct validity of a test aim to seek the theory underlying the test. This process involves two steps: first, the researcher deduces the relevance for the theory in order to propose hypotheses that he/she can make regarding the behavior of persons with high or low scores. Second, the researcher gathers data to test these hypotheses; and third, the researcher makes inferences as to whether the theory is adequate enough to explain the data collected. Such an experiment likely takes the form of a differential-group study where test scores are compared for two groups: one that has the construct and another that does not have the construct. If the group with the construct performs better than the group without the construct, the result is then said to provide evidence for the construct validity

of the test. Brown (2000) gives an alternative strategy that is called the ‘intervention study’. This strategy consists of testing a group of students who are weak in a given construct; then teach this group that construct and then test the group again. If non- significant difference is found between the results of the two test administrations, the test will be said to have construct validity.

### **2.1.2. Modern view of validity and validity evidence**

By the end of the 1970s, the view initially articulated by Loevinger (1957, p. 636) cited by Kane (2001) that “since predictive, concurrent and content validities are all essentially ad hoc, construct validity is the whole of validity from a scientific point of view” became widely accepted. The construct validity model came to be seen, not as one of the three kinds of validity evidence, but as a general approach to validity, including content and criterion-based evidence, reliability as well as the wide range of methods associated with theory testing (Messick, 1989). This is the birth of the modern view of validity theory at the beginning of 1980s 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*.

Messick (1989, 1996) argues that the traditional view of validity theory that considered validity as three-faced (criterion-related, content, and construct) is fragmented and incomplete especially because this view fails to take into consideration the evidence of value implication of score meaning as a basis for action and the social consequences of score use. For him (1989, p. 41), validity refers not just 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)

This definition entails some considerations of which one pertains to this study: we need multiple sources of evidence in order to evaluate the utility and appropriateness of a particular test.

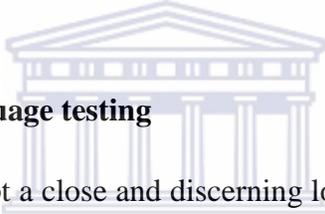
Since we agree that tests are neither totally valid nor invalid, we need to establish validity, as suggested by Shepard (1993), through a procedure that includes the conditions of the test taker preparation, test administration, and so forth. However, since every test use involves inferences and interpretations, any validation study requires the combination of logical argument and empirical evidence needed to support those inferences. Landy (1986) cited by Shepard (1993) compares this process of test validation to traditional hypothesis testing. Anastasi (1986) cited by Shepard (1993) concurs that construct validity subsumes both content validity and criterion-related validity requirements. In this perspective, the logic of test development should be checked by different analyses using real-life criteria, and the validity of criterion measures must also be evaluated.

Another characteristic of modern conception of validity is that it 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. And this view of construct validity as superordinate concept to all other validity types is considered as a major development in testing circles around the world. Since these three types of validity were traditionally referred to as the holy trinity (see for example Guion, 1980) as researchers had to combine the three aspects in their studies and the validity took on the character of religious orthodoxy, there are still some aspects of all these validity types in the new “construct validity”.

Messick (1998, p. 37) explains that the unified theory implies a single validity for test interpretation and use. What is singular in the unified theory of validity is that “all validity is one kind, namely construct validity”, and that the other traditional types of validity are “complementary forms of evidence to be integrated into an overall judgment of construct validity”; and he 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”.

The unified concept of validity is a complex concept. Messick (1989, 1996, 1998) lists six distinguishable aspects of validity that he highlights as features with complementary evidence. As put by Brualdi (1999) these six aspects conjointly function as general validity criteria or standards for all educational and psychological measurements. These six aspects should be viewed as interdependent and complementary forms of validity evidence and not viewed as separate and substitutable validity types. These are content, substantive, structural, generalizability, external, and consequential aspects of construct validity.



### **2.1.3. Validity in language testing**

In this section, I attempt a close and discerning look at validity theory through the linguistic lens. The section concentrates on the use of the unified validity theory in interaction with the different theories of language use. I will then describe the framework I have used in this study to validate the ESE.

#### **2.1.3.1. Unified validity theory and theories of language use.**

The first researchers in language testing 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. They successfully incorporated the modern view of construct validity, and they could also add some dimensions that affect test development in the real world. For them, test usefulness is to be understood as a macro concept that is made up of six major concepts that are: construct validity, reliability, consequences, interactiveness, authenticity and practicality. Therefore, for them, a good test has these qualities: 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 are measured and weighed in relation to each other.

Shohamy (1990) argued that a test validation agenda should be defined in terms of utility, feasibility, and fairness. Test utility relates to what extent a test serves the practical information needed of a given audience. Test feasibility relates to the ease of administration in different contexts; and finally test fairness relates to whether tests are based on material which test takers are expected to know. Shohamy's argument of test validity provides to this study insights in the rationale of test context as an aspect of validity evidence.

Indeed, it has been acknowledged that language testing should not be confined to the knowledge of how to write test items that will discriminate between the 'strong' candidates and the 'weak' candidates. Rather, it requires an understanding of what language is, and what it takes to learn and use language, and this is the focus for establishing ways of assessing people's abilities. Such relationship between testing and language theories has been highlighted by Bachman (1991) when he put forward the view that a significant advance in language testing is the development of a theory that considers language ability to be multi-componential, and that acknowledges the influence of the test method and the test taker characteristics on test performance. He describes what he calls an interactional model of language test performance that includes two major components: language ability and test method. In this model, language ability consists of language knowledge and metacognitive strategies and test method includes characteristics of the environment, rubric, input, and expected response. This model has become to be known as the Bachman model (Bachman, 1990; and Bachman & Palmer, 1996). This model has been extensively used for developing and validating language tests. Many researchers still use it today. Although I do not use Bachman's model in this study, this model still provides insights to this study on the ground that test context focuses on the consideration of the testing environment, the language input of the test, test method facets, task characteristics as well as the test administration conditions. As put by Alderson and Banerjee

(2002), Bachman's model remains very useful as the basis for test construction and evaluation. This model is a development of testing theories in conjunction with applied linguistic thinking by Hymes (1972) and Canale and Swain (1980). Other subsequent models (Weir's model [2005] for example) have built on Bachman's model.

### **2.1.3.2. Context validity evidence of a text-based test: Weir's framework for test validation**

In this section, I propose to describe the framework I have used in this study to evaluate the context validity of the DR Congo ESE. This framework is called Weir's (2005) evidence-based validation framework.

The premise behind Weir's framework is that test developers should work to generate evidence of the validity of the tests they have been using, and such evidence should be provided from a variety of perspectives.

Weir's model is built on five components that make up the different aspects of validity evidence. These components are the context validity, the theory-based validity, the scoring validity, the consequential validity and the criterion-related validity (Weir, 2005). As indicated by O'Sullivan (2005), there is a symbiotic relationship between context validity, theory-based validity and scoring validity; and these three validity aspects constitute the "construct validity".

The most noticeable innovation in Weir's framework is the use of the concept "context validity" in place of the traditional concept "content validity". Weir (2005, p. 56) starts his argument by pointing:

The last decade of the twentieth century saw a general decline in the prestige of psychometric, statistically-driven approaches to testing. In its place there has been a growing interest in the importance of *context*, in defining domain of use *performance* conditions and *operations*. (p. 56)

Weir (2005) believes that the important role of context as a determinant of communicative language ability is paramount. The context must be acceptable to

both the test takers and the test developers as a suitable milieu for assessing students' language abilities. Also, a conscious effort should be made to build into tests as many real-life conditions as possible.

The concept of context validity calls for an effort that test developers and administrators should make to approximate the test situation to authenticity, although, as Douglas (2000), O'Sullivan (2004) and Weir (2005) acknowledge, full authenticity of setting is not attainable neither in the classroom nor in the test situations. In this perspective, context validity could pertain to the extent to which the test reflects both the content of the materials taught in classroom situations and the kind of tasks the candidate is likely to perform outside the classroom and his/her future life.

Weir (2005) calls his framework a socio-cognitive framework since he believes that the best way to design good language tests is by taking into consideration the test takers cognitive factors and integrating them in the social context of language education. In this respect, since each language skill (listening, speaking, reading, and writing) calls for a specific framework as cognitive factors differ in the processing of each skill, Weir has developed his framework into four component frameworks. However, since the DR Congo ESE is a text-based test, only one aspect of Weir's framework is relevant to this study: "the socio-cognitive framework for validating reading tests".

In this framework, the first step in investigating the context validity of a text-based test is the investigation of the test takers characteristics. Weir (2005) argues that test takers' characteristics are important moderator variables in so far as they affect test performance. For him, the construction of the different test tasks should take into consideration the test population and the target language use situation.

O' Sullivan (2000) provides a useful literature on test taker characteristics that are likely to affect test performance. These are: the physical and physiological characteristics such as short-term ailments, age, sex, vision, hearing, and speaking; the psychological characteristics such as personality, memory, cognitive

styles, affective schemata, concentration; and the experiential characteristics of the test taker such as educational background and test taker experience with the test. All these characteristics are included in what is called “test domain”.

These test takers’ characteristics enable us to address two key questions while validating a language test: (1) in what ways does the test put candidates at their ease? and (2) are the candidates sufficiently familiar with what they have to do in the test?

Context validation focuses on the following aspects: task setting, task demands, and the test setting and administration.

#### **2.1.3.2.1. Task setting**

The first step in the evaluation of the context validity of a language test is the description of the test tasks on the basis of the test domain and the test takers’ characteristics. Task setting includes the analysis of test rubrics, task purpose, test criteria, the order of items and the test timing.

1. Test rubrics: test rubrics consist, according to Bachman and Palmer (1996) and Weir (2005), of facets that specify how test takers are expected to proceed in taking the test. They are guidelines that enable the assessment process of communicating expectations; providing ongoing feedback and grading (Reddy, 2007; Andrade & Du, 2005; Holmes & Smith 2003; and Moskal, 2000).

Weir (2005, p. 57) advises that clear instructions should be provided to the test taker so that he/she cannot misinterpret the task. “Rubrics should be comprehensive, explicit, brief, simple and accessible”. In brief, the rubrics should be accurate and accessible.

2. The task purpose: Jamieson *et al.* (1999, p. 15) define the task purpose as “the reason why we engage in tasks”. This definition stresses the importance of the purpose for which students have been using English in classroom situations. In the scope of this study, since the ESE is an achievement exit test, and assuming

that most of the reading that high school students do is heuristic, we expect the ESE to include tasks that reflect the following three components of reading, as suggested by Enright *et al.* (2000): (a) reading for global comprehension (b) reading for basic comprehension, and (c) reading to learn. Weir (2005) advises that text passages and activities should reflect as closely as possible those that the candidates have been exposed to, or are likely to meet in their future target language. Van Dijk (1985), Goldman (1997) and Enright *et al.* (2000) argue for the selection of test tasks that profile test takers' expectations so that they can determine the knowledge and strategies to deploy during task processing.

Weir (2005) advises that test takers should be given a clear unequivocal idea in the rubrics of what the requirements of the task are so that they can choose the most appropriate strategies and determine what information they are to target in the test text in comprehension activities. Having a clear purpose of the task will facilitate goal setting and monitoring, two key meta-cognitive strategies in language processing.

3. **Known criteria:** the knowledge of what candidates are expected to do in the test task. Weir (2005) advises test developers to clearly give, prior to the examination, assessment criteria to the candidates. Published information about how the tasks are scored, including the different steps for scoring and how the item scores are combined into test score should be readily available. Weir believes that if certain criteria are not to be used in the marking, this will have an effect on both planning and execution mechanisms in the cognitive processing involved in task completion. In brief, I believe that marking criteria should be made explicit for the candidate.

4. **Weighting:** the assignment of maximum points to each test item or subtest. Weir (2005) stresses the importance of advising candidates on the weighting of test items, tasks or subtests so as to enable them allocate their time and attention accordingly. Similarly, if any of the marking criteria are to receive different weighing, then test takers need to know this and allocate time and attention for

monitoring their output accordingly. In brief, there should be a justification for any weighting of different test components.

5. Order of items: the ordering of questions and the rationale in adopting a particular ordering is of crucial importance. If candidates are given a text passage and are asked to read it carefully and provide answers to questions based on it, then it is advised that the questions based on the text follow a serial order as theory suggests that this is the way we construct meaning (Urquhart & Weir, 1998). Further to this, Weir (2005) advises test developers to separate into distinct 'testlets' those items that focus on expeditious, quick and efficient reading strategies from those catering for more intensive careful reading. Within the expeditious reading section, we may also separate surveying for gist from scanning for specific information items. The order of items in each section must reflect the way such skills and strategies are developed in normal processing for the particular reading purpose. In brief, test developers should make sure that the items and the tasks in the test are in justifiable order.

6. Time constraints: the amount of time needed to process the text and provide answers to questions based on it. The test developer should sequence the text and tasks in such a way that candidates are capable to complete satisfactorily the tasks within the time allocated to them. Too little time may result in creating stress in the candidates and too much time may result in creating conditions of collaboration and cheating. Also, the time to be spent on each task should be clearly indicated on the test paper and the proctors should make sure that test takers comply with this instruction. In brief, the test developer should make sure that for each part of the test, there is enough time for preparation and completion of the task.

#### **2.1.3.2.2. Task demands**

Task demands refer to the extent to which the test includes tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task.

Weir (2005) advises test constructors to take into consideration the discourse mode that is used for answering the test (writing, reading, or speaking), the text length (brevity), the nature of information in the text (its relevance to the appropriateness of the test), and the content knowledge required for completing task (if the text content relates to the candidates' general knowledge and subject matter).

Urquhart and Weir (1998) argue that test developers have to decide what types of texts are appropriate for a particular test population through needs analysis of the students' target situations, and careful examination of the text and the tasks. As said previously, they argue that texts should be authentic, that is, they should either be taken from the target language use situation or possess characteristics of target language use texts.

Alderson (2000) argues that the knowledge of how texts are organized, the kind of information to be expected and the knowledge of how the information is signaled as well as how the change of content might be marked are factors that impact on reading performance.

Weir (2005) believes that the criteria for the selection of the text will depend on the purpose of the test and the intended target audience and that for any text type one will have to select, these organizational features should be taken into consideration:

- (a) The way a text is structured would seem to affect the ease with which it is processed. Coherent texts, for example, are easier to understand at the main ideas level than incoherent texts (Kintsch & Yarborough, 1982; cited by Weir, 2005);
- (b) The way texts are organized into sections is likely to enhance the writing of test items on surveying for gist and understanding the main ideas;
- (c) Grammar, cohesion and rhetorical structure need to be suitable as they impact on reading process.

In brief, the discourse mode should be appropriate to the skills or strategies being tested.

Four variables are to be considered in task demands: the text length, the nature of information in the text, the content knowledge, and the input-output relationship.

### 1. Text length

The length of a text on the basis of which a test is based has an impact on test performance. Engineer (1977), cited by Weir (2005) argues that by using texts greater than 1,000 words in length, one can measure different abilities, such as discourse rather than syntactic or lexical abilities only. In the same vein, Weir (2005) suggests the use of either single longer texts or multiple short texts.

The question of text length is crucial since we have to know if the test constructor can write enough comprehension items on the type of questions he/she wants from a given text. Single short texts raise some of the following considerations: skills like skimming and search reading might not be likely to be tested. Rather, only intensive careful reading skills will tend to be tested, with particular focus on scanning. Also, if texts are short, we are likely to question if the few items included in the test and the scores obtained are likely to serve for generalizing the students' reading ability.

We align with Weir (2005) when he believes that text length potentially has an additional effect in terms of the executive resources that will be called into play in cognitive processing. In general, the longer the text test takers are presented with, the greater the language knowledge that might be required to process it.

### 2. Nature of information in the text

Weir (2005) stresses the importance of the nature of information contained in the text. He distinguishes abstract texts (ethics, love, etc.) from concrete texts (description of a house, a tree, etc.). He advises that the nature of information in

the text should be relevant to the target situation requirements of the student being tested.

### 3. Content knowledge required

The content knowledge required to complete test task is likely to affect the way the task is executed. The notion of the content knowledge guides us to select texts that are suitable in terms of genre, rhetorical patterns as well as appropriate level of specificity. In this perspective, a test should not be biased or favor one group of the test takers on behalf of others. Urquhart and Weir (1998) advise that the content of the text should be sufficiently familiar to candidates so that candidates of a requisite level of ability have sufficient existing schemata to enable them to deploy appropriate skills and strategies to understand it.

### 4. Input/output

There are sets of linguistic variables that are applicable to both task input and task output. The first variable is the lexical variable. It has been demonstrated that texts with more high frequency vocabulary tend to be easier than texts with low-frequency vocabulary. Because of this, Weir (2005) advises that the lexical items in both the test and the text should be appropriate to the level of the test takers.

The second variable is the structural variable. As Alderson (2000) argues, texts with less complex grammar tend to be easier than texts with complex grammar; at least for foreign language learners. That is why Weir (2005) advises test developers to ensure that texts selected contain grammatical structures that are at the appropriate level of the test takers.

The third variable is the functional variable. The test developer needs to take into account the functional purpose of the text as well as the effect intended by the text writer (convincing, advising, describing...). Such functions in the test should be appropriate to the level of the test takers.

### ***2.1.3.2.3. Test setting and administration***

As put by Weir (2005) primary considerations affecting test validity are the circumstances under which the test takes place. For high stake tests, like the DR Congo English state examination, conditions of test administration should be similar across different sites; otherwise the processing will differ, affecting the reliability of scores generated by the test. The following variables affect test administration:

1. Physical conditions: the actual place as well as the appropriate materials needed for administering the test. Non-distressing or adverse physical conditions are likely to bias the test.
2. Uniformity of administration: Weir (2005) believes that a constant testing environment where the test is conducted according to detailed rules and specifications is necessary so that testing conditions are the same for all test takers. If all candidates cannot conduct the task in the same conditions then the test context validity is affected.
3. Security: this involves limiting access to the specific content of a test to those who need to know it for test development, test scoring, and test evaluation. In this respect, test items are not to be published; unauthorized copying is forbidden by any test taker or anyone otherwise associated with the test (Weir, 2005).

## **2.2. The Schema theory**

Having addressed the first part of my theoretical orientation in the previous section, I will now move on to focus on the schema theory. I believe that this theory has many important implications for this study.

Also called the theory of background knowledge of the reader, the schema theory is associated with cognitive psychology. Cook (1997, p.86) defines the concept “schema” (schemata in plural) 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”. According to schema theory, 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. In other words, the reader brings some knowledge to the text and this knowledge affects text comprehension. In this perspective, the schema theory is a theory about knowledge and about how knowledge is represented and how that representation facilitates text comprehension.

Of all the theories that establish the impact of reader’s schema on text comprehension, the most influential one is the theory of Kintsch and Van Dijk as developed by Kintsch (1988, 1998, and 2004), and Van Dijk & Kintsch (1983). This theory has been used as a starting point in developing other theories and models. Kintsch and Van Dijk (1983) developed a construction-integration model to explain the role of knowledge in discourse comprehension. For them, general knowledge about words, syntax, the world, spatial relations, in short, general knowledge about anything influences text processing and comprehension. They argued that knowledge provides part of the context within which a text is processed for understanding. The context is thought of as a kind of filter through which people perceive the world. Brief, knowledge makes understanding process smart: it keeps it on the right track and avoids exploring blind alleys (Kintsch, 1988, 1998; Van Dijk & Kintsch, 1983). They termed their model “construction-integration” because it combines a construction process in which a text base is constructed from the linguistic input as well as from the reader’s knowledge base, with an integration phase in which the text base is integrated into a coherent whole (Kintsch, 1988, 1998, and 2004, Van Dijk & Kintsch, 1983 ).

Indeed, it is worth noting that the schemata theory is one of the main theories of reading comprehension that focus on the cognitive process of reading and that stress the interaction between the reader’s prior knowledge and the text. In this perspective, as put by Kitao (2003) and Gunning (2006), such interaction results in comprehension since each schema is filed in an individual compartment and stored there. Thus, in attempting to comprehend a text, the reader relates the new

information contained in the text to the existing information he/she has stored the mind. This is the reason why the schema theory is considered as the theory of background knowledge or prior knowledge. Prior knowledge refers to all knowledge a reader has accumulated through his/her life. This knowledge influences reading process, product and recall.

Indeed, many reading models have been developed on the basis of the schema theory. I can mention at this stage, for instance, Laberge and Samuels's (1974) "Model of Automatic Information and Processing" that emphasizes internal aspects of attention as crucial to text comprehension, or Rumelhart's (1985) "Interactive Model" which contends that information from several knowledge sources are considered simultaneously, and that when information from one source is deficient, the reader will rely on information from another source. This correlates to what Stanovich (1980) had previously termed "Interactive Compensatory Model". Other models include the "Sociocognitive Processing Model" (Ruddell, Ruddell & Singer, 1994) that states that the role of the classroom's social context and the influence of the teacher on the reader's meaning negotiation and construction are central; and Rosenblatt (1994)'s "Transactional Model" that considers both the reader and the text as two aspects of "a total dynamic situation", and therefore posits that meaning does not reside ready-made in the text or in the reader, but it rather comes into being during the transaction between the reader and the text. All these models provide insights to this study as they help understand the extent to which the ESE includes texts that ask test takers to activate their schemata to process and understand them, as well as to answer questions based on them.

As put by Jannuzi (1997), the application of the schema theory to foreign language learning goes back to works by Widdowson (1983), and Carrell and Eisterhold (1983) as their works have been part of English language teaching, especially in approaches to reading. Oller (1995) also demonstrated that the schema theory could facilitate discourse comprehension; hence communication and language acquisition.

Two notions are widely used in the schema theory: formal schemata and content schemata. Formal schemata, also called knowledge of the language (For example, rhetorical patterns, story grammar, narrative script that guide expectations in our attempt to understand a meaningful piece of writing), relates to the impact of the reader's knowledge of the language on text processing (Carrell, 1983). Carrell argues that poor knowledge of the language negatively impacts on the reader's comprehension of the text. The same view is shared by Perfetti (1989) and Rayner (1990) when they argue that structural knowledge has a facilitative effect on reading.

Apart from structural knowledge, vocabulary knowledge has long been recognized to impact on reading process. Read (2000) assumes that the reader's vocabulary knowledge impacts on text comprehension. The more a reader has good vocabulary knowledge, the better he/she will comprehend the text. This means that poor readers struggle with reading partly because of poor vocabulary knowledge. In the same line of thought, Chou (2011), Qian (2002), Joshi & Aaron (2000) and Martin-Chang & Gould (2008) have demonstrated in different studies that vocabulary is significant in helping students to read and understand reading passages.

As for content schemata, Carrell (1983) relates them to the knowledge of the world (or cultural knowledge), including the subject matter of the text. Jannuzi (1997) argues that since some schemata are culturally specific, culture helps to determine our life experiences and how we make meaning of them. In this perspective, content schemata are subdivided into background knowledge and subject matter knowledge. Rumelhart (1980, 1985), who conducted studies on content schemata, argued that the reader needs some knowledge about the content of the passage in order to comprehend it efficiently. The background knowledge helps the reader to process a familiar text. It seems obvious that a reader who knows only little on the topic of a text will be likely to process it with some difficulties than a reader who is familiar with the topic. This argument is sustained by Liu *et al.* (2009) who examined the impact of content knowledge on Test of

English as a Foreign Language (TOEFL) reading performance. Their objective was to determine whether prior knowledge gained either through studies in an examinee's academic subject or from his/her immersion in particular culture has a facilitating effect on performance to the TOEFL reading performance. Among the various results reported, they could find that content schemata suggested a possible interaction between prior knowledge and the type of item although this impact could vary across passages.

However, the schema theory is not without critics. It has been criticized on the ground that it attempts to explain how new information is integrated with previous information, but fails to explain how the new information is handled. Also, as put by Alderson (2000), the schema theory does not lead to explicit definitions or predictions of comprehension processes. Carver (1992) is even more critical when he argues that schema theory applies not to normal reading, but to study reading and memorizing. He claims that this theory applies only when materials to be read are relatively difficult, such as the material used with college-level students. He goes on claiming that there is no effect for prediction activities on test performance, and that the effect of prior knowledge on text comprehension is only the effect of general reading ability. Other researchers (Brantmeir, 2005; Hammadou, 2000; Pulido, 2004) reported limitations to the schema theory and its impact on background knowledge. They could find that background knowledge may not help improve reading comprehension, especially in longer passages.

However, in the scope of this study, I believe that since the schema theory helps to understand some of the variables that impact on reading process, it should be understood in order to design and validate the DR Congo ESE. In the context of my study, it is known that almost all students entering a high school, prior to learning English, know at least two languages: a local language (mother tongue) and a lingua franca, which is, a language of the community in which the student evolves. On top of these languages, they are required to learn French as the medium of instruction. In the last four grades of high school, these students have to learn English, a foreign language whose use is strictly limited to classroom

contexts. Therefore, since both teachers and students are not much exposed to English language, and since the ESE is text-based, the understanding of the schema theory is likely to help the test constructors select appropriate texts that can be efficiently processed by the test takers. In this perspective, it stands to reason that students' background knowledge as well as knowledge of the language should be recognized as impacting factors on text comprehension, hence, on test performance. Therefore, test constructors need to use this knowledge to facilitate students' performance to the test rather than its absence to inhibit performance.



## CHAPTER 3 LITERATURE REVIEW

After having described the two theories that inform this study in the previous chapter, I am going to explore, in this third chapter, the relevant literature that pertains to this study. I have organized this literature to cover two main trends: literature that relates to the impact of background knowledge of the reader on both text processing and comprehension and their impact on test performance and literature that taps to the characteristics of texts used in reading assessments.

My argument for such organization is motivated by the need to demonstrate that reading is an interaction between a reader and a text and that performance to a text-based test like the DR Congo ESE results from such interaction; that is the capacity of the reader to process the text and construct meaning.

### 3.1. Reader variables

Reader as an individual affects reading process and product in a number of ways. The knowledge he/she possesses, the purpose he/she has for reading, the different strategies he/she employs to decode the text, his/her individual skills affect the reading process and product.

#### 3.1.1. The reader schemata and background knowledge

Research studies on the effect of background knowledge on text comprehension and test performance are abundant and varied. In the scope of this study, three aspects of these research assume particular relevance : (a) those that highlight the danger of selecting a text that can generate different interpretations by test takers depending on their respective backgrounds; (b) those that focus on the necessity to select texts that require test takers to recourse to their background knowledge to better perform in the test, and finally (c) those that highlight the danger of test takers' too much familiarity with the text that can result in responding to test questions without necessarily reading the text.

### 3.1.1.1. *Research that highlight the danger of selecting a text that can generate different interpretations by test takers depending on their respective backgrounds*

These studies are among the earliest studies documenting the effects of background knowledge on text comprehension. Anderson *et al.* (1977) investigated with two groups of college sophomores with significantly different sets of background knowledge: 50 % female music majors and 50 % male physical education majors. These two groups of participants were given a text thought to be understood as each participant belongs to one group or another. The results of the study indicated that 72 % of the music majors' responses and 64 % of the physical education majors' responses were consistent with understanding the passage as relevant with their respective academic knowledge.

In the same lines of thought, Reynolds *et al.* (1982) investigated with two groups of students (8<sup>th</sup> grade) with different cultural backgrounds (urban and rural) to whom they administered the same passage and the respondents were tested on their comprehension. Again, results indicated that the "urban" group interpreted the story in a way that was consistent with their cultural knowledge while the "rural" group came to an almost another interpretation that was consistent with their cultural knowledge.

The conclusions of these two studies demonstrate that passage meaning is not totally text-based and that individual background knowledge in interaction with words and sentence processing can lead to significantly different interpretations of the passage. These findings support the idea that reading comprehension is as much about integrating new information into what readers already know (their background knowledge) as it is about properly identifying words (Reynolds, 2002). These two studies, which assume prominence and substance in a study of this nature, inform us that texts that are capable of generating different interpretations depending on student's background knowledge might turn to be problematic in that they might fail to provide unique information to all the test takers. Therefore, it is generally recommended to create contexts that are clear

enough to avoid such different interpretations of the same text as such texts are likely to yield results that are not an indication of students' text comprehension.

*3.1.1.2. Research that focus on the necessity to select texts that require test takers to recourse to their background knowledge to better perform in the test*

There are numerous studies on the facilitating effect of background knowledge on performance to reading tests. Prior knowledge that is likely to advantage test takers on reading tests comes from two sources, as suggested by Liu *et al.* (2009): knowledge gained from training in the major subject of study, and knowledge gained from being immersed in a given culture.

Levine and Haus (1985) investigated the effect of background knowledge on reading comprehension test performance with students of Spanish as a foreign language. Their subjects were asked to read an authentic report of a baseball game from a Spanish language newspaper. A 12-item multiple choice test of reading comprehension consisted of two types of questions: explicit and implicit. Background knowledge of baseball was assessed by means of a 9-item multiple choice questionnaire. Subsequently, 52 of the original subjects were identified as having "limited knowledge" of baseball and 38 as having "high knowledge" for a total of 90 subjects. The analysis of the data and results obtained indicated that (a) background knowledge was a significant factor that affected reading comprehension across the two types of questions when the textual material was at the "instructional" reading level; and (b) background knowledge could be more important than language level in comprehending such material.

Droop and Verhoeven (1999) investigated the role of cultural background knowledge on the reading comprehension of third graders acquiring literacy in Dutch as a first and second language. Their subjects were given three types of texts: texts referring to Dutch culture, texts referring to the cultures of immigrants from Near Eastern countries (i.e. Turkey and Morocco), and neutral texts. Within

each type of text, a distinction was made between two levels of linguistic complexity. They used verbal protocols and metacognitive questionnaires, and after the analysis of the data, the results indicated a facilitating effect of cultural familiarity for reading comprehension. As for the linguistic complexity variable, they could find that the effect of background knowledge for students from the minority group was restricted to linguistically simple texts, and this was due to their limited knowledge of the target language, Dutch.

Burgoyne *et al.* (2011) investigated the role of background knowledge with second language learners of English by relating to the difficulties of their monolingual pairs. Sixteen students learning English as an additional language and sixteen monolingual students were matched for word reading accuracy. Their subjects were assessed using a standard measure of reading comprehension and an experimental measure of reading comprehension for which relevant background knowledge was taught before assessing understanding. Results indicated lower levels of reading comprehension for second language students, compared to their monolingual pairs and second language students showed specific difficulties with both literal questions and questions requiring the interpretation of a simile. These results indicate that lack of sufficient background knowledge is a factor that contributes to inadequate text understanding and poor performance.

In the same vein, Hammadou (1991) investigated the inference strategies that are used by students while reading and he found that background knowledge was a factor that impacted students' comprehension process. He concluded that beginning readers use more inference strategies than advanced readers, and this is an indication that reader's background knowledge affects text comprehension.

Kendeou and Broek (2007) investigated with university undergraduate students enrolled in scientific courses the effect of prior knowledge and text structure on reading performance on scientific texts. Using a think aloud method in experiment 1 and a reading time method in experiment 2, they found that their subjects could adjust text processing as a function of interaction between prior knowledge and

text structure, and that the readers' memory for the text was affected by differences in their prior knowledge regardless of the structure of the text. These results replicate previous research which found a facilitating impact of background knowledge on processing scientific texts (Kendeou, Rapp, & van den Broek, 2004; Kendeou & van den Broek, 2005; van den Broek, Rapp, & Kendeou, 2005; Qian 1999).

I believe that these conclusions are relevant to this study as they provide useful insights on the criteria of evaluation of the different texts used in the ESE as regards their appropriateness to the test takers' background knowledge. That is the reason why I agree with Alderson (2000) when he argues that readers will find it easy to read texts in areas they are familiar with, for example, those they have studied, than those which they are not familiar with. For him, topic familiarity is expected to have facilitating effects on text comprehension and test performance.

However, other studies that aimed to establish the facilitating relationship between background knowledge and reading comprehension came out with inconclusive results. These studies could find out that the good performance in the reading test could not be always achieved by the expected group of test takers. Krekeler (2006) gives some reasons that accounted for these inconclusive results: (a) the difficulty to measure the extent to which topic familiarity impacts on test scores since there are too many other textual variables (such as readability, genre, and topic) that interact with background knowledge, as reported in findings by Alderson and Urquhart (1985) and Birjandi *et al.* (2002), Erickson & Molloy (1983) and Clapham (1985); (b) the test method used contributed to variations within the background effect, as reported by Koh (1985), Peretz & Shoham (1990); (c) the effect of background knowledge on test performance varied according to the test taker's level of proficiency, as reported by Alvermann & Hynd (1989), Ridgway (1997) and Chen & Graves (1995).

These results appear to confirm the argument that many variables (reader variables, text variables and test variables) interact in the text processing and task

completion; and that it is this interaction between these variables that results in reading test performance. Therefore, any attempt to single out a given factor for the purpose of accounting for examinee's performance is neither necessary nor desirable.

*3.1.1.3. Research that highlight the danger of test takers' too much familiarity with the text that can result in responding to test questions without necessarily reading the text.*

Although reader's background knowledge is crucial in text processing and comprehension, it should be used in conjunction with the reader's ability to read new information. As put by Tian (2006), the results of reading comprehension tests should reflect the combination of reader's background knowledge and information gained from the texts. In this perspective, attention has been called on the need for test takers to read the text passage in order to successfully answer questions based on it (Tuinman, 1974, Pyrczak, 1975, Pyrczak & Axelrod, 1976). This is what we call passage dependency of reading comprehension items. This issue is relevant as we have to examine the extent to which some test takers are able to correctly answer all or some reading comprehension test items without the passage on which the questions are based. Such questions that can be answered correctly without reading the text are likely to affect the validity of test scores, as claimed by Bernhardt (1983).

Studies that investigate the issue of passage dependency of reading comprehension test items are numerous. In an earlier study, Pyrczak (1974) demonstrated that college students with a great deal of experience in test taking strategies could identify appropriate answers in an MC reading test without reading the information contained in the text and this with greater chance of success. He viewed such finding as a serious threat to the validity of reading comprehension tests. The underlying assumption that could be made was that

what is measured with the passage removed is something other than text comprehension.

Tuinman (1974) proposed the passage dependency index to determine the degree to which reading test items could be answered without the text. He investigated with groups of students to whom he administered five major standardized reading tests. The results showed that, although there were varying levels of passage dependency among tests, none contained an optimal number of passage dependent items. He demonstrated that test takers were able to answer from 32 to 50 % of the multiple choice items without reading the text.

Perkins and Jones (1985) used two commercial reading tests that they administered to two groups of 44 international students in two different conditions: with and without the stimulus passage. The results indicated that the passage dependency index was negative for the two tests, showing low passage dependency, and they concluded that the majority of test items were assessing background knowledge and not information contained in the text.

Katz *et al.* (1990) analyzed students' scores on the SAT and they could find that the results correlated highly with performance in both groups: those who were given the passage and those who were not given the passage. They concluded that performance on the reading section of the SAT seemed to depend on factors that did not relate to the comprehension of the accompanying passages.

Tian (2006) investigated with Taiwanese students the passage dependency of the reading section of the General English Proficiency Test (GEPT) and the Test of English as a Foreign Language (TOEFL). Results indicated no significant difference in passage dependency between items from the two tests, although items from the GEPT had higher passage dependency index than those from the TOEFL.

All these studies provide evidence that many evaluation procedures do not take into consideration the number of test items that can be answered commensurately without reading the text passage. In this perspective, I agree with Bernhardt

(1983) that if reading tests are passage independent and that students do not have to read carefully to perform at criterion level, the testing system defeats the purpose of reading instruction. Valette (1977) alerts us to the problem of background knowledge and passage independence. She argues that reading tests should not turn into logical thinking and problem-solving simply because the candidate relies on his/her background knowledge, not on the text passage to successfully provide answers to the test. This is a very serious challenge that many test developers face, especially for designing reading test for students of specialized subjects. Hence, one should be careful with reading tests based on texts that are too specialized since these might test subject matter knowledge rather than reading ability; therefore, failing to discriminate good students from poor students. This is a danger for such tests because results reported on test scores might be biased given that they appear to reflect more subject matter knowledge than reading comprehension knowledge.

That is why I tend to hold with Alderson (2000) that the development of tests of reading for specific purposes, usually subject-related, is an area where text and background knowledge effects are crucial, and with Hale (1988) that a student's academic discipline (a crude measure of background knowledge) sometimes interacts with test content in determining performance.

To conclude this point on schema theory and its impact on reading and test performance, I believe, at this stage, that the test developer needs to bear in mind that the text on which the test is based should be chosen with careful care. It should not be too unfamiliar so as not to enable test takers deploy their background knowledge and successfully process the text and answer questions based on it, nor too familiar so as to result in correctly responding to all or some test items without reading the text.

However, reader's background knowledge is not enough to help in the processing of the text and the building of its meaning. One other variable is the reader skill to read and comprehend the text.

### **3.1.2. Reader skill**

There is no doubt that that students have different reading abilities as it has been observed that some students have the skills to process a text and understand it with more ease than others.

In reading literature, the understanding of the reader skill to process the text and understand it has generally been understood in the distinction between two dichotomous terms: good versus poor readers. Although I will be using these two concepts in this study, other synonymous dichotomous concepts are widely reported in the literature. These include skilled/unskilled readers, successful/unsuccessful readers, proficient/less proficient readers, or fluent/non-fluent readers.

There is substantial literature on good and poor readers. In the scope of this study I need to examine the distinction between good and poor readers as regards to three levels: linguistic, cognitive and metacognitive levels. The linguistic level relates to the formal linguistic knowledge (vocabulary, syntax and discourse) of the reader. The cognitive level relates to the reader's capacity to recourse to his background knowledge and various strategies to process and understand the text. Finally, the metacognitive level relates to the reader's ability to monitor and control the different reading strategies deployed.

Research studies that relate to reader's language knowledge and its effect on L2 text processing and comprehension have generally pointed to the positive association between these two variables. Nassaji (2003), in a study of higher – level text processing skills in reading comprehension in English as a second language, reported that lower-level processes (word recognition) and higher-level processes (syntax and semantic) could contribute significantly to the distinction between good reader and poor reader. He concluded that poor readers are slower in word recognition and weak at rapid and automatic syntactic processing. Word recognition was found to account for the difference between good readers and poor readers in a similar study by Parry (1991) who revealed that in respect to

vocabulary growth, good readers guess but simply read much more, thus exposing themselves to many more words in meaningful contexts.

At lexical and syntactic levels, Barnett (1986) examined English-speaking readers' abilities to process and understand a French text and she reported that readers' recall significantly increased in accordance with their vocabulary and syntactic proficiency levels. She concluded that both readers' vocabulary and syntactic levels affect text processing and comprehension.

Chen (1998) also reported that poor L2 readers are weak in processing more complex ambiguous sentences, and this weakness results from their lack of syntactic knowledge in L2 which constrains their reading comprehension. Similar findings were reported in a study by Liu and Bever (2002) who reported that good readers do not exhibit apparent effort to use syntactic analysis in their comprehension processes as they are capable to process sentences in a quick and subconscious manner.

Other studies that relate to the impact of the reader language knowledge on reading performance have looked at inference generation during text processing and comprehension. For example, Lu (1999), Barry and Lazarte (1998) reported that L2 readers' language proficiencies have a substantial impact on inference generation in reading. This conclusion is replicated by Hammadou (1991) when he found 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 proficiencies.

Still other studies that look at the interaction between reader language proficiency and reading performance have focused on discourse organization and its contribution to text processing and comprehension. For example, Carrell (1985; 1987; 1992) reported that good readers have the ability to recognize the discourse structure of a text much better than poor readers; and that such recognition helps them to process and understand the text.

Research studies that relate to the reader cognitive ability and its effect on L2 text processing and comprehension recognize that good readers are those who are strategic readers, that is, those who use reading strategies to process and 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 aspects are integration, recognition of aspects of text structure, use of general knowledge, personal experiences as well as associations, and finally response to extensive versus reflexive modes. She explained 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, however, readers focus on the author's ideas expressed in the text instead of relating the text to themselves personally and affectively. Therefore, 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 for the impact of reader metacognitive ability on text processing and comprehension, we agree with Pang (2010) when he argues that comprehension monitoring competence is particularly crucial in L2 context. Metacognition is the control readers execute on their ability to understand a text (Block, 1992). Metacognition involves thinking about what one is doing while reading. Klein *et al.* (1991) argue that good readers attempt the following strategies while reading: identifying the purpose of reading before reading, identifying the form or type of the text before reading, thinking about the general character and features of the form or type of the text (for instance, they try to locate a topic sentence and follow supporting details towards a conclusion); 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. Good readers attempt to form summary of what they have read. Metacognitive knowledge enables 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) in describing 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 this perspective, when someone reads a text, an understanding of the text is created in the reader's mind. Kintsch and Van Dijk (1983) sustain 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 ([http://www. Utwente. NI/cw/theorieenoverzicht/Theory % 20 clusters...](http://www.Utwente.Nl/cw/theorieenoverzicht/Theory%20clusters...)). 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.

Block (1992) provides a useful review of metacognition and its relation to reading. For her, metalinguistic awareness plays a part in learning to read. Good readers are aware of how they control their reading and they can verbalize this

awareness. They can be sensitive to inconsistencies in text and usually tend to use meaning-based cues to evaluate whether they have understood what they have read or not. She examined the monitoring process involved in reading with ESL readers in United States colleges and concluded that control of the various stages of the monitoring process depends on reading abilities. Prior to her research, Duff *et al.* (1987) demonstrated in a study with adult ESL readers how some good readers are aware of the mental processing involved in reading. Subsequent to Block research, Alderson *et al.* (1997) found that metalinguistic knowledge is distinct from linguistic knowledge. This evidence culminates in the view that the ability to answer questions about the language of a text is unlikely to relate to the ability to understand the text.

In the scope of this research, the metalinguistic as well as metacognition knowledge are factors that are likely to inform the test developer in designing and evaluating tests, especially for students within a very complex multilinguistic and multicultural context. A text-based test like the DR Congo ESE needs not only to consider test takers' linguistic knowledge, but it also needs to integrate their capacity to go beyond their linguistic knowledge and deploy their metalinguistic and metacognitive skills in order to process the text and comprehend it. Such integration is likely to generate assessments that are likely to reflect the students' performance. This is obviously a challenge that necessitates skill, knowledge and theory on the part of test constructors.

As an implication for testing, it stands to reason that since it is admitted that test takers have different abilities in understanding texts, the test constructors should be careful while deciding on the text on the basis of which the test is based. They should ensure that the majority of the test takers should be able to read the text and understand it, regardless their differences in reading skills. They should avoid one important bias: that of test takers failing to efficiently perform to the test not because of test questions' difficulty, but because of the difficulty for them to read and comprehend the text.

### **3.2. Text variables**

I have previously said that text comprehension is an interaction between the reader and the text. This is the main reason why I have stressed, in the preceding section, the relevance of the literature that point to the impact of reader's background knowledge on text comprehension and test performance. This section explores the literature that stresses the impact of the characteristics of the text on performance. I believe that some aspects of text impact on the reading process by making it either easy or difficult to understand. These textual aspects are text topic and content, text type (or genre), text organization, and text syntactic complexity.

#### **3.2.1. Text topic and content**

Text topic and content affect how readers process and comprehend texts. Liu (2011) and Alderson (2000) assume that abstract texts are harder to understand than texts that describe real objects, events or situations. The more concrete, imaginable and interesting, the more readable is the text. Also, texts that relate to the daily activities and familiar topics are likely to be easily processed by readers than those that relate to unfamiliar events. Besides, the quantity of information in the text affects understanding and recall, so does the density of propositions, and the extent to which information is stated explicitly in the text has an impact on recall.

Claphan (1996) elaborated this issue when she investigated the relationship between language ability of students taking the IELTS test of reading for academic purposes and their ability to understand texts in and out of their own subject disciplines. She discovered two linguistic thresholds. The first one: at a score of roughly 60 per cent on her grammar test represented a level of linguistic knowledge below which students were unable to understand even in their own subject discipline. The second: at a score of roughly 80 per cent on the same test, represented a level of linguistic knowledge below which students had little

difficulty reading texts outside their own discipline. The crucial area in which subject knowledge could facilitate understanding of texts within one's own subject area was 60-80 per cent on the test. From these findings, she concluded that it is only with more specific texts that background knowledge has a significant effect on text comprehension. This means that the more specific a text, the more important is the contribution of background knowledge; and the less specific a text, the more important the contribution of language proficiency.

Hock (1990) examined whether familiarity with text content or level of language proficiency was a predictor of ability in reading comprehension. In all subjects under study (medicine, law and economy) she found that comprehension of a discipline-related text could be predicted by both knowledge of the subject area and by language level. Similar results were reported in a study by Yousif & Shumaimeri (2006) who investigated the effects of content familiarity and language ability on reading performance with EFL Saudi students. Their subjects had to perform two different types of text qualified as "familiar" and "unfamiliar". The results indicated that familiarity with text content in conjunction with students' language ability had significant effects on reading performance. They concluded that content familiarity had a facilitative effect on students' reading comprehension. Chang (2006)'s study came out with similar results.

Other studies that reported the impact of test topic and content on text comprehension and test performance include Alderson and Urqhart (1985) who reported that on relatively easy texts, linguistic proficiency might be sufficient enough to answer test questions adequately whereas difficult text might require more subject matter knowledge, or higher linguistic proficiency or both. And Hale (1988) who examined performance on TOEFL reading texts and established that students in the humanities/social sciences and biological/physiological sciences performed better on passages related to their groups than on other passages. Hale concluded that TOEFL test developers were justified in seeking to maintain a

balance of reading passages across the humanities/social sciences and the biological/physical sciences, in order to counter any possible bias.

As implication for test design, it is advised to base the test on texts whose content and topic relate to the students' experience and environment. A text-based test on "computing and cybernetic" will be harder to process for the majority of Congolese students since the computer is not available in most Congolese schools and that the majority of high school students have no access to the internet. A test constructed with such a bias will not likely be a good measure of students' performance.

These studies are relevant to my research as they provide insights on the kind of texts that should be included in a testing session and the relevance of their content to test takers. I want to remind what I said previously: that good context can be built in the test when test developers make an effort to build into tests as many real-life conditions as possible.

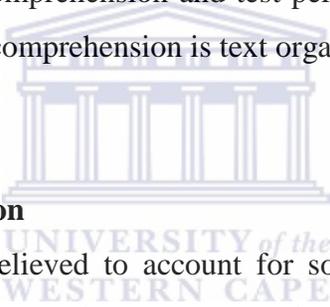
Nevertheless, text topic and content is not a sufficient variable to account for test takers' capacity to process the text, understand it, as well as better perform to the test. Another textual variable is text type (or genre).

### **3.2.2. Text type or genre**

Text type or genre has also long been considered a variable that affects reading test performance. Texts are generally classified into narrative, descriptive, argumentative or expository types; and each of these types has at least one or more major text structures, such as classification, comparison/contrast, cause/effect, and problem/solution, with the information presented from more than one perspective or point of view (Cohen & Upton, 2007). As put by Olson *et al.* (1980, 1981, 1982) cited in Liu (2011), there is a difference in the reading process according to the type of text read. It is evident that certain topics are associated with certain types of texts.

Alderson (2000) argues that what causes difficulty in text to process is less the actual content than the way it is written: its style or the features that make one text different from another. Both Liu (2011) and Alderson (2000) contend that expository texts are harder to process than narrative texts perhaps because of the greater variety of relationships among text units, and possibly due to greater variety of content. This is because narrative texts appear to induce visualization in the reader as part of the reading process, as readers report “seeing” scenes in their heads when they read a history text. But here again, it is worth noting that different readers are likely to visualize different scenes depending on their prior experience and different expectations.

However, neither the text content nor the text types are the only textual variables that account for text comprehension and test performance. One relevant variable of text processing and comprehension is text organization.



### **3.2.3. Text organization**

Text organization is believed to account for some of the problems the reader encounters to process texts. The way paragraphs relate to each other, the way the relationship between ideas are signaled or not signaled, and the way different organizations within one text type may lead to different comprehension outcomes in readers have been reported to affect reading test performance.

There are some forms of text organization that make the text readable. For instance, Urqhart (1984), in a study on the effect of chronological and spatial ordering in text, showed that texts organized according to the sequence of events could be read faster and these texts were easier to understand than those whose temporal sequencing was distributed. Also, it was generally agreed that texts with a consistent spatial organization were easier to understand and recall than texts without specific organization.

In the same perspective, Beck *et al.* (1982) concluded that coherent texts are easier to comprehend than less coherent texts. This is the case of texts that present

facts with little explanation of relationships between them as they force readers to make inferences. They argued that texts written in such a way as to expose the reasoning that connects a cause to an event, and an event to a consequence are easier to understand than those that fail to make the causal sequence more clear. The same conclusion is reached by Alderson (2000) when he sustains that when a text is manipulated into good or bad organization, comprehension is affected by poor organization, but enhanced by good organization.

Kintsch and Yarbrough, cited in Alderson (2000), tested the rhetorical structures such as classification, illustration, comparison and contrast as well as procedural description and definitions. Their results were corroborated by McKeown *et al.* (1992) when they assumed that text coherence best facilitates comprehension when the content is moderately unfamiliar and that coherent texts enable readers with relevant background knowledge to understand the text better.

Kobayashi (2004) investigated the impact of text organization and response format on L2 learners' performance in reading comprehension tests. Using Bachman's (1990) model of test method facets and Bachman and Palmer's (1996) framework, she concluded that text organization and test format had significant impact on the students' performance on reading comprehension tests, and that the two variables interacted significantly.

However, a study by Freebody and Anderson (1983) came out with mixed results. They investigated the effects of cohesion on understanding and recall reported that the absence of connectives does not seriously damage text comprehension. For them, the effects of cohesion are weak because readers can make bridging inferences. They concluded that cohesion is not a key variable to readability. Nevertheless, they could admit that cohesion, together with connectives, interacts with text topic to impact on text processing.

The above discussion of the literature that relates to text organization is relevant to this study as this literature provides insights to understand how ESE texts are organized and how such organization is likely to enable the test takers process the

texts and answer questions based on them in such a manner inferences on scores generated by such tests may be considered appropriate to the DR Congo testing context.

#### **3.2.4. Text syntax**

The syntax of the text is also a variable that has been investigated as factor that affects text comprehension and processing. Bentin and Linberman (1990), and Scarborough (1991) claimed that an absence of grammatical knowledge or lack of processing ability interferes with higher level text comprehension. Under the Structural Deficit Hypothesis, the acquisition of syntactic structures is staged and gradual with inherently simpler structures preceding more complex ones in language development. As put by Martohardjono *et al* (2005), it is the more complex structures that poor readers have more difficulty with.

Nuttal (1996) reported that syntax (long sentences and difficult grammar), could block comprehension even when vocabulary is familiar. She sustains that insufficient knowledge of cohesive devices such as reference words, conjunctions, substitutions, and ellipsis can significantly decrease comprehension levels.

Berman (1984) investigated the opacity and heaviness of the constituent structure of sentences which make it difficult for readers to comprehend texts. He focused on the reader's capacity to recognize the basic constituents of subject-verb-object (S-V-O), and noun-verb-noun (N-V-N) relations. He concluded that it is difficult to say with precision that particular syntactic structures always cause reading problems.

McNamara (2007) argued that readers with deficits in sentence-level processing may have access to a rich network of a schema-based knowledge structures or metacognitive comprehension monitoring skills that enable them to compensate for weak skills in lower linguistic processes. She assumed that syntactic processing contributes to differential reading ability through its application of proposition assembly and integration process.

The above brief discussion of the literature that relates to text syntax provides insights to understand the structure ESE texts and how such structure contributes to test takers' performance.

Another textual variable closely related to text syntax is text vocabulary load and difficulty, and this variable is discussed below.

### **3.2.5. Text vocabulary load and difficulty**

The vocabulary load and difficulty in a text is another linguistic variable to comprehension. Analyses of research findings reveal that if a reader is to become very good at comprehending what he/she reads, he/she must meet two main learning requirements. First, he/she must know words and second, he/she must be able to reason with physical text (<http://www.designedinstruction.com/learningleads/reading-vocabul...>). It then clearly appears that the first recommendation referred to relates to vocabulary whereas the second recommendation relates to text comprehension. As an implication for learning, Brett *et al.* (1996), Medo and Ryder (1993) suggest that vocabulary instruction should be tied to the specific materials being read, particularly when employed prior to reading.

Indeed, research literature on how vocabulary knowledge can be enhanced in readers is abundant and varied. Daniels (1994), and Dole *et al.* (1995) argued that repetition and restructured exposure of readers to key word meaning provide the reader with opportunities to repeatedly explore situations, texts, and contexts where key vocabulary words have similar or dissimilar meanings. As we might assume, repetition is especially effective when working with vocabulary items that are likely to appear in many contexts; and also through multiple exposures in authentic contexts that extend beyond class sessions. This relevance of context has also been stressed by Dole *et al.* (1995) when they argue that high-interest and relevant contexts contain special meaning to the key words that a reader needs to learn, and this yields positive results. In this perspective, we may think that using

rich context may promote vocabulary learning since interesting contexts highly motivate students and assist the teacher to connect vocabulary tasks to active involvement in context learning. Another strategy is suggested by Gordon *et al.* (1992). For them, effective restructuring, especially with low-achieving students, should involve the direct modification of the text materials being used; for example, replace hard with easy words or phrases.

Indeed, research has consistently shown the effect of vocabulary load and difficulty on text comprehension for L2 readers. Williams and Dallas (1984) contended that topic unfamiliarity could not be compensated for by easy vocabulary. Therefore, both difficult vocabulary and low familiarity reduce comprehension. They also claimed that texts with difficult vocabulary did not become easier if more familiar topics were used, and vice-versa. They also focused on the use of idiomatic expressions in texts and concluded that the meanings of idiomatic expressions could make text comprehension difficult especially for L2 readers.

Belinda (2008) investigated the impact of idioms on text comprehension with Latino and non-Latino students. She concluded that there is a positive relationship between idiom comprehension and reading comprehension. She assumed that since reading comprehension and idiom comprehension share many of the same linguistic processes, idioms comprehension may provide a sound perspective for investigating reading comprehension.

Levorato *et al.* (2004; 2007) investigated idioms comprehension in students with different reading comprehension skills. According to their hypothesis, the reader's text comprehension skills should predict his/her ability to understand idiomatic meanings. Yet, it is generally admitted that idiom comprehension requires readers to go beyond a simple word-by-word comprehension strategy and to integrate figurative meaning into contextual information. The result of their study showed that the ability to understand a text could predict reader understanding of idioms only in appropriate contexts. They also tested whether possible improvements in readers' comprehension skills might produce an increase in figurative language

understanding. The results showed that readers whose general comprehension skills were good improved their performance on an idiom comprehension test.

The above discussion of the literature that relates to text vocabulary load and difficulty provides insights that will enable me to evaluate the ESE texts with regards to their vocabulary difficulty and the extent to which such difficulty is likely to impact on test takers' performance.

### **3.2.6. Text length**

With respect to the relationship between text length and performance to the reading test, Alderson (2000, p. 108) maintains that “a problem all reading test developers face is how long the texts should be on which they base their tests. Text length is surprisingly an under researched area”.

Engineer (1977) cited in Alderson (2000) and Cohen & Upton (2007) found that when texts longer than 1,000 words were used, the ability that could be measured changed. They all concluded that longer texts could allow testers to assess more study-related abilities and to reduce reliance on sentential processing abilities that might tap syntactic and lexical knowledge more than discourse-processing abilities. Engineer also believed that the ability to identify the main idea of a long text might be qualitatively different from the ability to identify the main idea in a shorter text.

Mehrpour and Riazi (2006) investigated the impact of text length on reading comprehension. Three reading comprehension passages accompanied by 30 items were reduced to two-thirds of their original lengths. The truncated version, along with the original version, was randomly administered to 100 male and female college students of English (50) and non-English majors (50). Although participants performed better on the shortened version, the ANOVA test revealed that the length of the text did not have any statistically significant effect on test performance.

To conclude this point, a question is worth addressing: which text do we need to use, long texts or short texts? Indeed, the use of longer texts in testing situations reflects the situation where students have to read and study long texts. Short texts, on the other hand, are likely to introduce bias from its restricted range of information contained. This argument elicits an interesting question on the definition of a long/ short text. In teaching situations, what might be a long text for one setting might be a short text for another setting. The criteria of definition of text length are unclear and they depend on many variables that are likely to interact. For example, a one page English text for a Congolese high school finalist might appear to be a long text than it might appear for a South African high school finalist. Also, what might be termed a long text for a Grade 10 student might be considered a short text for a Grade 12 student. Therefore, the matter of text length on which a reading text should be based should be taken with much caution. The test constructor should only bear in his mind that the length of the text should be of crucial importance while designing a reading test.

The above discussion of literature related to text length is relevant to this study as it provides insights that will enable me to evaluate to which extent the length of ESE texts is appropriate for the target situation requirements of the Congolese high school students.

### **3.2.7. Text readability**

Text readability is also a variable that affects reading process and performance. Readability, sometimes called “text difficulty”, is a concern for those who advocate the selection of appropriate texts for both teaching and testing purposes. We agree with Fulcher (1997, p. 498) when he argues that “giving students reading material that is too difficult is damaging to the learning process, and demotivating to the student”. Readability has been defined by Du Bay (2006) as the ease with which a text can be read and understood.

Indeed, research has long been concerned with the identification of relevant features that make a text readable. A number of methods to measure text readability have been developed. Most of these methods are based on multiple correlation analysis where researchers have selected a number of text properties (such as average number of syllables per word and average number of words per sentence), then test subjects are asked to grade readability of various texts on a scale and results are correlated to make decisions on whether the text is readable or not. Among these methods, mention can be made of the Flesch formula, the Dale-Chall formula, the Gunning Fog formula, and the Fry Readability Grap, etc. Let us say something on one of these methods, the flesch formula, a method that has been used in this study.

The Flesch formula is a readability method that aimed at predicting the difficulty of adult reading material (Flesch, 1943, 1945; cited by Fulcher, 1997). In this method, the notion of prediction is central to the measurement of text difficulty. Flesch uses an abstract scale from 0 to 100, with more difficult texts having lower scores, that is, scores less than 50 (the average score) and easier texts having higher scores than 50. Flesch formula was initially linked to United States grade levels with 0 equivalent to the 12<sup>th</sup> grade and 100 equivalent to the 14<sup>th</sup> grade. The formula reads:

$$\text{Reading ease score} = 206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$$

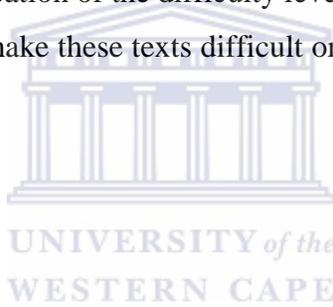
Where: ASL = average sentence length (number of words divided by number of sentences);

ASW = average word length in syllables (number of syllables divided by number of words).

As put by Klare (1963) and Chall (1958), cited by Fulcher, 1997, the Flesch Reading Ease Formula is the one most widely used and also the one most tested and reliable.

However, although the Flesch Reading Ease Formula formula has been qualified to be highly accurate and reliable for grading the readability of existing texts, it is limited in its use since it cannot create a text or modify an existing one. Yet, this latter aspect is of very crucial importance since the outcome of a text that is reported to be difficult for readers is either to modify it or to create a new one. Another limitation of this formula is that the two variables used, that is, word (vocabulary) and sentence (grammar), are not the only variables that account for reading difficulty. Last, but not least, this formula appeals to the statistical sophistications to correlate different variables, yet it is not evident that these statistical procedures can be easily applied by many language teachers.

The above discussion of text readability is relevant to this study as it provides insights on the identification of the difficulty levels of a range of ESE texts and on the main aspects that make these texts difficult or less accessible to the Congolese high school finalists.



## Conclusion

This chapter has focused on the literature that pertains to reading in general and the testing of reading in particular. I have said that reading as a cognitive process is affected by many factors that relate both to the reader and to the text. These factors are complex and interact in such a way that the reading process and product have to be understood as the result of all these complex interactions. The factors or variables that affect reading and that pertain to the reader include the reader's schemata, that is, his/her background knowledge (linguistic knowledge as well as knowledge of the world), his/her skills, metalinguistic knowledge and purpose for reading.

As for reading variables that pertain to the text, I have stressed the impact of the text topic and content on the reading performance, the text type and genre, the text organization and the text readability. What is important to say here is that all these

text variables interact not only among themselves but also with the reader variables to impact on text comprehension.



## CHAPTER 4. METHODOLOGY

Research is the process of collecting, analyzing, and interpreting data in order to understand a phenomenon (Leedy and Ormrod, 2001; Williams, 2007). The research process is systematic in that defining the objective, managing the research data, and communicating various findings occur within a framework and in accordance with existing guidelines. To this end, I propose to use this chapter to discuss or describe the procedures that I have used in carrying out this research; that is from collecting data, analyzing them and reporting findings.

The methodological procedures developed in this chapter will focus on the selection of my research approach, the choice of the appropriate research methods, the different procedures and instruments (techniques) of collecting data and the methods of analyzing data that I decided to employ in my study.

### 4.1. Research approach

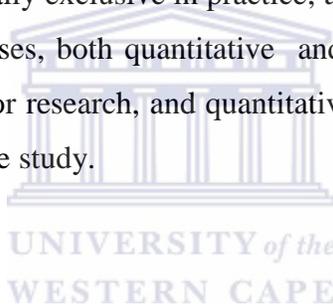
This research uses a mixed method approach in that it uses both qualitative and quantitative approaches. The term “mixed method” is used in this study to refer to all procedures of collecting and analyzing both qualitative and quantitative data in the context of a single study (Driscoll *et al.* 2007).

As clarified by Creswell (2003), a true mixed method research involves a concurrent and sequential use of not only research methods and data, but also the whole design process of analyzing data as well as reporting results. This is not the model used in this study. This study uses a partial mixed method by using the concurrent model in the data collection stage and the merging model in the data analysis stage.

The research design consists of a combination of data forms during the collection stage. The main data (the state examination papers) were collected concurrently

with other data (questionnaire, interviews). This is the concurrent mixed method design and this design is used in this study to validate one form of data with the other forms, as suggested by Creswell and Plano Clark (2007). The mixed method model that is used during the data analysis stage is merging. The content of the state examination papers are analyzed both quantitatively and qualitatively; the questionnaire are analyzed quantitatively whereas the interviews are analyzed qualitatively. The merging model is also used at the reporting of findings and evaluation stage.

I hold with Patton (1990), Banfield and Cayago-Gicain (2006), and Marsland *et al.* (2001) that although quantitative and qualitative research paradigms do not share the same ontological and epistemological assumptions, they do not need to be considered as mutually exclusive in practice; as each method involves different strengths and weaknesses, both quantitative and qualitative methods constitute alternative strategies for research, and quantitative as well as qualitative data can be collected in the same study.



#### **4.2. Research methods**

This study uses mainly two research methods: the content document analysis method and the statistic (descriptive) method.

Content document analysis is a research tool used to determine the presence of certain words or concepts within texts or sets of texts. It has been defined by Stemler (2001) as a technique for making inferences by objectively identifying specified characteristics of texts.

This method is useful to this study as it enables us to examine the content of the different state examination papers so as to identify the main features of the ESE: order of test items (whether they are sequenced in a justifiable order or not), the text length (long, short or moderate text), the type of texts used (description, comparison, contrast, narration, argumentation, etc.), the content of the different

texts used (whether text content is familiar to the test taker or not), as well as the text linguistic structure.

One relevant issue in the use of content document analysis methodology in this study was, as put by Altheide (1996), to identify the specific research problem under study and the contribution of the content analysis of test papers to provide answers to the research problem. I agree with Altheide (1996) when he argues that qualitative document analysis relies on the researcher's interaction and involvement with documents selected for their relevance to a research topic. It is at this level that I clearly postulated that a thorough examination of ESE test papers could enable me to determine to what extent the conditions under which the ESE test tasks are performed relevant to test domain and test characteristics, and also to what extent the ESE includes tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task.

Another issue in the use of content document analysis methodology was the selection of the unit of analysis. At this stage, I had first to familiarize myself with the document for analysis (ESE) so as to be able to decide on what could be the unit of analysis. Such familiarity was achieved through extensive reading of collected test papers. As defined by Trochim (2008, p.6), the unit of analysis is "the major entity that you are analyzing in your study". Two units of analysis were therefore selected: the first unit of analysis was "each ESE test paper from the corpus" and the second unit of analysis was "some selected specific ESE test papers from the corpus". The first unit of analysis was selected in order to examine the general characteristics and patterns of ESE test content (test rubrics, task purpose, order of test items, text length, text content, text difficulty and linguistic structure). As for the second unit of analysis, we selected some specific ESE test papers from the corpus to illustrate some types of test tasks. More specifically, two narrative texts were selected from ESE papers to investigate whether most items follow the story line of the text or not; and also if test questions call for operations test takers are likely to use to construct text meaning.

This is also the case of six texts that were selected from the corpus to investigate the three basis patterns that are (a) texts that include general topics and are likely to enable the test takers to deploy their background knowledge to read the text and understand it, (b) texts that include general topics, but that are likely to advantage one group of candidates on behalf of another group of candidates; and (c) texts that do not relate to the candidates' background knowledge. This second unit of analysis was also selected when I had to determine the difficulty levels of a range of ESE texts as I selected from the corpus 7 ESE sample texts that were examined and analyzed and the Flesch reading index calculated for each of the texts.

The descriptive method is a basic research method that examines the situation as it exists in its current state (Williams, 2007). This study uses simple descriptive statistics to analyze respondents' questionnaires, to quantify observations identified in the state examination papers and to evaluate the context validity of the ESE. Such simple descriptive statistics mainly consists of the use of frequency tables as well as cross-tabulation. Although I sometimes allude to the effect of certain variables on task performance, I have not used statistical analyses such as *t*-tests, correlation and regression, as well as multi-faceted Rash analysis to examine the effect of different independent variables on test performance.

### **4.3. Data Collection**

I consider this step as crucial because collecting credible data is one of the main concerns of this research in that analyses and findings to be reported will largely depend on the credibility of the collected data.

#### **4.3. 1. Population**

The first stage in collecting my research data was the identification of research population and the provision of clear definition of this population. This is what is

called target population. In the scope of this research, the target population includes:

- English state examination papers from 1990 to 2008. These are the English national test papers that contain the text passage and the test questions;
- High school finalists of 2008, 2009 and 2010. This group includes students who were in 6<sup>th</sup> grade (the final grade) and who presented the national test in 2008, 2009 and 2010. This population group is subdivided into two subgroups: high school finalists before the test, and high school finalists after the test.
- English language teachers: These are people who actually teach English in the final grade (6<sup>th</sup> grade). This population group includes teachers with at least three years of experience in teaching in this grade.
- English language inspectors: refers to people who have the mission to supervise school activities, to train the teaching personnel and to evaluate high school finalists.
- Other stakeholders: any person who has direct or indirect implication in the education matters, and generally who has or has had a child (or children) in 6<sup>th</sup> form high school. The criterion of selection in this group was specified to people with a certain education level and directly or indirectly involved in education matters. Most participants targeted belonged to one of these occupational categories: school principals, teachers, lawyers physicians, and pastors.

I collected my research data in Kananga, a capital city of the Kasai Occidental (Western Kasai) province, DR Congo. I collected them between April 2007 and December 2007. These data were supplemented the first time between June and November 2009, and the second time between June and July 2011.

As for data types, this research uses data that fall into two categories: first, the existing data that were collected from official sources. These data include the English state examination papers. These data were provided to me by the provincial inspection of education services in Kananga and the general inspection of education services in Kinshasa. These are the most important data type of this research since they provide comprehensive details on the content of the English state examination. The advantage of having these data types as the primary data for this research is that these “archival data”, to use Hyndman’s term (2008, p.27), provide valid and reliable information upon which the analyses will be based; therefore, they do not need to go through prior testing for validation.

Second, the new data, and these were collected through different collection methods. These are all data collected from high school finalist students, English language teachers, English language inspectors, and other stakeholders.

#### **4.3.2. Sampling**

I have used two types of sampling in this research: the probability sampling and the non-probability sampling. The probability sampling was one in which each element of the population had a non-zero probability of selection in the sample. I used this sampling type with most of the collected data (the English state examination papers, high school finalists, English language teachers, and other stakeholders). With the probability sampling, I used two techniques for inclusion in the sample: the simple random sampling and the stratified random sampling. In the simple random sampling technique, each element of the population had an equal probability to be selected, and in the stratified random sample, I identified for some populations relevant population characteristics and then I grouped members into homogeneous non-overlapping strata based on the identified characteristics before I could draw then random samples from each stratum. I used stratified random sampling with English language teachers considering the variable ‘qualification’, and with high school finalists whether they were studying

in towns or rural areas. As for the non-probability sample, it should be noted that some population categories could not be sampled because of their small number and the effectiveness in their providing data. This is the case of English language inspectors that I intentionally selected because of their experience with the issues explored in this research. Therefore, data provided by all the respondents were included in the sample and thoroughly analyzed. The following table presents the samples drawn from the population.

No.	Population group	Population N	Sample size
1.	English state exams	43	30
3.	High school finalists before the test	1,927	1,200
4.	High school finalists after the test	1,115	500
5.	English language teachers	95	60
6.	English language inspectors	20	20
7.	Other stakeholders	385	250

(Table 401: Sample size)

### 4.3. 3. Data collection instruments

I used three techniques to collect the necessary research data: the questionnaire, the test, and the interview.

#### 4.3.3.1. Questionnaire

I developed five questionnaires to elicit data from high school finalists (before and after the test), English language teachers, English language inspectors, and other stakeholders.

As a research instrument for gathering data, I constructed the five questionnaires in such a way that they had to reflect the relationships and constructs developed in

this study. That is the reason why before developing these research questionnaires, I addressed the following questions: what are the objectives of the questionnaires? Who is to be surveyed? What information do I need to get from him/her? Is the information to be provided sufficient? These questions served as a blueprint to designing the content of the questionnaire. It should be noted here that only relevant and valid questions were included. By relevant questions I understand those questions that could elicit information which was reliable; and by valid questions I allude to the kind of questions that could be easily answered by respondents.

As for questionnaire development, I considered the following features: the determination of the aims of each questionnaire, the selection of an appropriate question style, the design, the piloting and the revision of questionnaires, and finally the administration of questionnaires. Actually, I proceeded from information gathering where a kind of probing in which I asked open-ended questions to some target population groups; and then I could use the responses in the development of the questionnaire. Then I tested each questionnaire on subjects who could meet the criteria for the study sample, but who actually did not participate in the major study. The aim of this pretesting stage was to reveal problems related to answering, completing and returning questionnaires. This ultimately minimized low response rates which undermine many research questionnaires. In this process of pretesting, I referred to Faux (2010) by considering four main issues of design: the questionnaire content, the questionnaire form, the instrument and procedures and the process. I provided space to respondents to indicate which questions presented content problems with regard to clarity, specificity, appropriate language, simplicity and relevance. I conducted this second step during piloting. This was the time for me to detect errors and flows made in the construction of the five questionnaires and make revision to construct the final version that I submitted to respondents. In light of this, I wish to direct the attention of my readership to the following description of the five research questionnaires.

#### **4.3.3.1.1. Research Questionnaire 1 submitted to high school finalists before the test**

Aim: The aim of this questionnaire was to elicit information related to high school finalist students' views on the English state examination regard their knowledge on test rubric, their familiarity with the scoring and moderation criteria, their view on the time allocated to complete the test, and the issues related to test administration and security. This questionnaire was submitted to respondents between May and June 2010 (shortly before the state examination). I hasten to point out that that I constructed this questionnaire in French to facilitate respondents to answer it accurately.

Respondents: This questionnaire targeted high school students who were in the 6<sup>th</sup> grade (terminal grade) and who were actively engaged in state examination preparation.

Questionnaire content: (See Appendix 1)

#### **4.3.3.1.2. Research Questionnaire 2 submitted to high school finalists after the test**

Aim : The aim of this questionnaire was to elicit information from high school finalists after they have taken different tests administered to them. This questionnaire related to respondents' views on the English state examination with regard to their knowledge on test rubric, their familiarity with the scoring and moderation criteria, their view on the time allocated to complete the test, the text difficulty level and the issues related to test administration and security. I constructed this questionnaire in French to facilitate respondents answer it accurately.

Respondents: This questionnaire targeted high school finalists immediately after they have taken the English national test.

Questionnaire content: (See Appendix 2)

#### **4.3.3.1.3. Research questionnaire submitted to English language teachers**

Aim: The aim of this questionnaire was to elicit information about English language teachers' views on the students' familiarity with the test rubrics, task purpose and evaluation criteria, their views on the time allocated to write the test, on the text length and on the text difficulty level. This questionnaire also aimed to elicit information on respondents' views on test administration conditions and security.

Respondents: This questionnaire targeted only 6<sup>th</sup> grade English language teachers with at least three years of experience and with or without any qualifying degree in English language teaching. Besides, the questionnaire targeted both English language teachers of urban as well as rural areas.

Questionnaire content: (See Appendix 3)

#### **4.3.3.1.4. Research questionnaire submitted to English language inspectors**

Aim: Like the questionnaire submitted to language teachers, the aim of this questionnaire was to elicit information about English language inspectors' views on the students' familiarity with the test rubrics, task purpose and evaluation criteria, their views on the time allocated to write the test, on the text length and on the text difficulty level. This questionnaire also aimed to elicit information on respondents' views on test administration conditions and security.

Respondents: This questionnaire targeted English language inspectors.

Questionnaire content: (See Appendix 4)

#### **4.3.3.1.5. Research questionnaire submitted to other stakeholders**

Aim: The aim of this research questionnaire was to elicit other stakeholders' views on the students' familiarity with the test rubrics and evaluation criteria, and their views on test administration conditions and security.

Respondents:

This questionnaire targeted other stakeholders. This questionnaire was constructed in French, the official language.

Questionnaire content: (See Appendix 5)

#### **4.3.3.2. Interview**

I used some interviews in this study as a secondary instrument to collect data. Kvale (1996, p.14) defines interview as “an interchange of views between two or more people on a topic of mutual interest”. For Keats (1997), it is a controlled conversation in which the interviewer tries to obtain information from his/her interlocutor, the interviewed. From this definition, an interview is a human interaction for knowledge production. In research, an interview as a data collection technique is a systematic way of talking and listening to people in order to collect data. It is together with observation one of the most used techniques in qualitative research. This conversation usually takes the form of series of questions. Although the mode of interaction is verbal, non-verbal messages are also interpreted. The interview is an interesting technique because it involves cognitive, affective as well as social processes.

In this research study, I conducted some interviews with some respondents either to triangulate data generated by other primary sources (ESE papers, tests, questionnaires) or to generate some information that the primary sources could not provide me with.

Indeed, there are three main types of interviews that are extensively developed in the literature (Corbetta, 2003 and 2004; Gray, 2004; Bryman, 2001 and Kajornboon, 2005; to cite only few). These types are the structured, the semi-

structured and the unstructured interviews. The key feature with all these three types is that they differ on the degree of structuring. Thus, in a structured interview, there is a set of predefined questions and these questions are asked in the same order to all respondents. A semi-structured interview, on its part, uses predefined questions, but in the course of interview, the interviewer has certain room to adjust the order of questions planned and he/she has the opportunity to probe other questions. Finally, in the unstructured interview, neither the questions nor the answers are determined in advance.

In the scope of this research, I adopted the unstructured interview as the instrument to generate data from respondents. With Punch (1998), I believe that the unstructured interview is the appropriate way to understand complex behavior of people without imposing any a priori categorization which might limit my field of inquiry. My decision to adopt this type is governed by my epistemology and my research objectives. I do believe that to better understand the basic issues related to the complex problems of DR Congo high school finalists' test context, I need to approach these issues through both my perspective as well as stakeholders' perspectives. And stakeholders include here these people who contribute to decision-taking in the evaluation field (mainly inspectors) and those who are directly affected by the decisions (students). In this perspective and siding with Denzin (1989), I could not make any beforehand. That is why I could follow my respondents' narration and could, on the basis of that, generate questions spontaneously based on the reflection of their narration.

Nevertheless, it is worth noting that stating that in unstructured interview there are no predefined questions does not mean that I conducted the interviews in a random and non-directive way; on the contrary, the format of the interviews was loosely guided by a list of questions called "aide-memoire" or agenda, as suggested by Briggs (2000) and McCann and Clark (2005). This was a broad guide to topics that might be covered in the interview rather than the questions I actually asked. I used aide-memoire in order to have a certain degree of consistency across different interview sessions.

#### **4.3.3.2.1. Interview with the high school finalists**

Aim: The aim of this interview was to elicit information about how high school finalists manage to answer some specific test items that are based on a text unfamiliar or too familiar to them, whether or not they possess archives of ESE papers, and if yes, why and how they get these papers. Also, the issue of cheating, collaboration and test disclosure was part of the interview, for candidates who have presented the test.

Respondents: This interview targeted three high school finalists before the test and three high school finalists after the test.

Interview transcription: (See Appendices 6 and 7)

#### **4.3.3.2.2. Interviews with language inspectors**

Aim: The aim of this interview was to elicit information from inspectors on how the state examination is administered and the challenges that pertain to its administration.

Respondents: These interviews targeted two head inspectors

Interview transcription: (See Appendix 8)

#### **4.3.3.2.3. Interview with language teachers**

Aim: The aim of this interview was to elicit information about whether or not language teacher possess the old test papers; if yes why and how they get these papers.

Respondents: This interview targeted two high school finalist English teachers selected on the basis of their experience in teaching English in the last grade and their familiarity with the national test.

#### **4.3.3.3. Tests**

Research studies in education have the specificity to use tests as research instrument (Ngongo, 2007). In this research study, I use tests as primary data. I selected different tests from the ESE papers and I administered these tests to groups of respondents (as a classroom final test) in order to gather information on the text content, text genre, text length, text linguistic structure, text difficulty, and text organization. Since each of these tests contained specific information it intended to target, I will briefly present each of these tests in the data analysis whenever each of them is used.

#### **4.4. Data Analysis**

In this section, I briefly present the procedure I used to analyze the collected data. This procedure starts with how data coding process was conducted and then how the data were analysed using frequency tables and crosstables.

##### **4.4.1. Data coding**

Data coding was the first stage in my 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). As put by Sharon (2004) coding facilitates the organization, retrieval, and interpretation of data and leads to conclusions on the basis of that interpretation.

I developed different codes for my instruments. For questionnaires and tests, I finalized the codes as these instruments were completed. I developed a code book whose importance, as stated by Shenton (2004), could allow this study to be repeated and validated, and to make methods transparent by recording analytical thinking used to devise codes and to allow comparison with other studies.

For missing data, I developed categories such as “refused,” “not applicable,” “missing,” “don’t know”. Code values ranged from 1 to 3. Code **1** corresponded to data from ESE papers, code **2** corresponded to data from Questionnaires, and

code **3** corresponded to data from tests. Within each code, I used some values, and these values also ranged from 1 to 5. For example, as I said previously, I had developed 5 questionnaires; thus, the questionnaire submitted to high school finalist students was coded **21**, to English language teachers **22**, to language inspectors **23** and to other stakeholders **24**. The questionnaire I submitted to high school finalists was of two types: I submitted one to the respondents before they took the test, and the other was to be completed immediately after they had taken the test. So, I created two sub-codes: **21a** and **21b** respectively. For students who participated in the investigation, I used identifiers that anonymized them but still could reveal information to me.

#### **4.4.2. Frequency analysis and cross-tabulation**

I did the analysis of the research data by making use of frequency analysis and cross-tabulation. Frequency analysis involves constructing a frequency distribution. The frequency distribution is a record of the number of scores that fall within each response category. The frequency distribution, then, has two elements: (1) the categories of response, and (2) the frequency with which respondents are identified with each category.

Crosstabulation should be understood as a combination of two (or more) frequency tables arranged such that each cell in the resulting table represents a unique combination of specific values of crosstabulated variables. I used crosstabulation to examine different frequencies of observations that belong to specific categories on more than one variable. By examining these frequencies, I could be able to identify different relations between crosstabulated variables. It should be mentioned here that I could crosstabulate only categorical (*nominal*) variables (variables with a relatively small number of different meaningful values).

**Conclusion:**

In this chapter, I have described the methodology I used to collect the data and analyze them. The approach design I used in this study is the mixed method and the two research methods used are the content document analysis and the descriptive statistics. The main research data are the ESE papers and they are supplemented by questionnaires, interviews and tests. The study population includes the test papers, the high school finalists, language teachers, language inspectors and other stakeholders.



## CHAPTER 5 DATA PRESENTATION AND ANALYSIS

This chapter investigates the context validity evidence of the ESE. It will answer the following research question: “To what extent is the context of the DR Congo ESE valid?”

To answer this main research question, I need to address the following sub-questions:

- (1) To what extent are the conditions under which the test tasks performed relevant to test domain and test characteristics?
- (2) To what extent does the ESE include tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task?
- (3) To what extent are the ESE setting and administration conditions likely to enable the test takers perform the tasks in such a way that scores obtained from the test reflect their actual performance?

This chapter evaluates the context validity of ESE with regard to task setting, task demands and task administration. Task setting will involve the evaluation of the validity aspects of the following variables: ESE rubrics, task purpose and evaluation criteria; the ordering of ESE items and time allocation. Task demands will involve the evaluation of text types or genres, text length, text content and the text difficulty level. Finally, task administration will involve the evaluation of the validity of aspects related to uniformity of administration and the ESE security.

### 5.1. Task setting

In this first section, I focus on the first research sub-question: “To what extent are the conditions under which the test tasks are performed relevant to test domain and test characteristics?”

It should be said, from the outset, that my first step in the evaluation of the ESE context validity was the description of the test task on the basis of the test domain and the test takers' characteristics. Task setting includes the analysis of test rubric, purpose, criteria, the order of items and the test timing.

### **5.1.1. ESE rubrics**

These include the test organization, the time allocation and the instructions. As for test organization, the content document analysis of ESE papers reveals that the English test is printed on the same test sheet with the French test for candidates in general subjects, and on the same test sheet with the French test and the cultural knowledge test for candidates in specific subjects. The ESE has two main parts: the first is a text passage that candidates are asked to read carefully and understand before answering the questions and the second part relates to the questions themselves. Here, these questions in most cases are based on the text.

With regard to time allocation, the content document analysis of ESE papers reveals that the time allocated to the "language test" depends on the subject group. This time is four hours, for candidates in general education who have to answer 20 questions for the two subtests where English has 9 questions and French 11 questions; and four hours for fifteen questions in technical schools group 1 where English has 5 items, French 10 items and the cultural knowledge 7 items; and finally 3 hours for fifteen questions in technical schools group 2. It should be known that there is no specific time allocated to each subtest, but we may infer, on the basis of number of questions for each subtest, that the English test is to be completed in two hours in general options, one hour in technical options, group 1 and forty-five minutes in technical options, group 2. It is worth noting that this is only an estimate since the real amount of time for the completion of each subtest demands careful investigation as there are many factors that account for such partition.

With regards to test instructions as the third type of test rubric, the analysis of test instructions as indicated on the ESE papers reveals that there are two types of instruction in the ESE: the first is the general set of information that concerns all the subtests, and this rubric is on the first page of the SE papers. This rubric is the same across all test sessions. The rubric content lists ten instructions that relate to the place where candidates are required to write and the kind of information they should write (candidates' personal identity and details concerning the school code, testing center code as well as the subject). There are also instructions that relate to where and how candidates' answers to test questions should be written, and a set of warnings against any candidate attempting to collaborate or cheat and further consequences he/she is likely to face. These instructions are clear and comprehensive, but they are written with a very small size Time Roman 6 and candidates have to struggle to read them.

The second type is the specific instructions that relate to the ESE. It is important to mention that the text starts with no instruction; some words are underlined but nothing is said to provide candidates with the reason why these words are underlined. It is only when the candidate is answering the test questions that he/she may realize that the underlined words or phrases are simply some vocabulary words or structures on which questions are set. To introduce questions, there are three instructions: (a) an instruction related to the task that the candidate has to perform before answering the questions: "*Read the text carefully and answer the questions*", (b) an instruction related to how candidates have to select the appropriate correct option from a list of five options numbered 1, 2, 3, 4, 5; and (c) an instruction related to what the candidate has to do in case there is no correct option from the five suggested, that he/she has to write 6.

We have also considered what the different respondents to the research questionnaires have reported about their views on whether or not the test provides instructions that enable the test taker to understand what he/she is expected to do. Table 501 presents (in percentage %) the views of finalist 1 (before taking the

test), finalists 2 (after taking the test), English language teachers, English language inspectors and other stakeholders on ESE rubrics.

Test provides clear instructions	Finalists 1	Finalists 2	Language teachers	Language inspectors	Others stakeholders
Strongly agree	35.58	44.20	71.67	75.00	31.60
Agree	33.58	35.60	26.67	25.00	37.20
Disagree	13.25	13.40	0.00	0.00	11.60
Strongly disagree	1.08	2.00	0.00	0.00	6.80
Can't tell	10.17	3.40	1.67	0.00	10.00
Not answered	6.33	1.40	0.00	0.00	2.80
Total	100.00	100.00	100.00	100.00	100.00

Table 501: Respondents' views on ESE rubrics

From this table, it can be noticed that most respondents appear to assess the ESE rubric positively. 69 % of finalists 1 either strongly agree (35.58 %) or agree (33.58 %) that the test has good rubrics. Finalists' views after taking the test are even more positive as 79.80 % either strongly agree (44.20 %) or agree (35.60 %). English language teachers, since they are very familiar with the test as they use it as part of their teaching materials, are very positive as 71.67 % strongly agree and 26.67% agree that the ESE provides instructions that enable the test taker to understand what he/she is expected to do. This is also the case with English language inspectors as they constitute an essential body in the planning, designing, administration and evaluation of the state examinations in general, and the ESE in particular. Other stakeholders' views on the ESE rubrics are also positive, with 31.60 % who strongly agree and 37.20 % who agree with the proposition.

### 5.1.2. ESE task purpose

The research questions I addressed in the evaluation of ESE task purpose are the following: *Is the purpose of the ESE made unequivocally clear for the test takers? Is it an appropriate purpose?*

The first step to answer the above question is the examination of the ESE rubrics. A content analysis of the ESE rubrics indicates that there is only one instruction

that points to the task purpose. This instruction is: “*Read the above text carefully two or three times and then answer the questions based on it*”. On the basis of this instruction, two observations are worth making: (1) the task purpose pertains to reading for comprehension (‘...carefully/two or three times...’) in that test takers should comprehend the text to be able to answer questions based on it; (2) there is no precise (specific) instruction that points to specific cognitive processes test takers are likely to deploy to find discrete information in the text. Yet, different reading skills are actually deployed by test takers. The content analysis of test papers indicates that there are test questions that ask for comprehending explicitly stated information, those that ask for comprehending implicitly stated information, as well as test questions that ask for identifying words and phrases appropriate to the context, inferring meaning of unfamiliar words or identifying the topic of the text (or paragraph).

It will be helpful for us to take a look at the other data provided by respondents through a questionnaire. The question 2 asked these respondents to circle the number that reflects their viewpoint on a five –point scale. The question reads: “From the requirements of the tasks, I could (test takers can) use appropriate strategies to read the text and answer the test questions”. Table 502 presents the proportions (in %) of respondent for each value.

Task requirement and strategies used	Finalists 2	Language Teachers	Language Inspectors
Strongly agree	7.00	20.00	30.00
Agree	55.60	36.67	40.00
Disagree	24.20	21.67	15.00
Strongly disagree	1.80	10.00	0.00
Can't tell	6.20	8.33	10.00
Not answered	5.20	3.33	5.00
Total	100.00	100.00	100.00

Table 502: Task requirements and strategies use

From this table, it appears that many respondents (55.60 %, 36.67 %, 40.00 % for high school finalists after the test, language teachers and language inspectors,

respectively) agree that the task requirements enable the test takers to deploy appropriate strategies to read the text and process the task. Language inspectors, whose role in test preparation, construction, administration and report is very crucial, are even very convinced when 30.00% of them strongly agreed and 40.00 % agreed that instructions are appropriate to help the test taker process efficiently the task. The following figure presents these proportions:

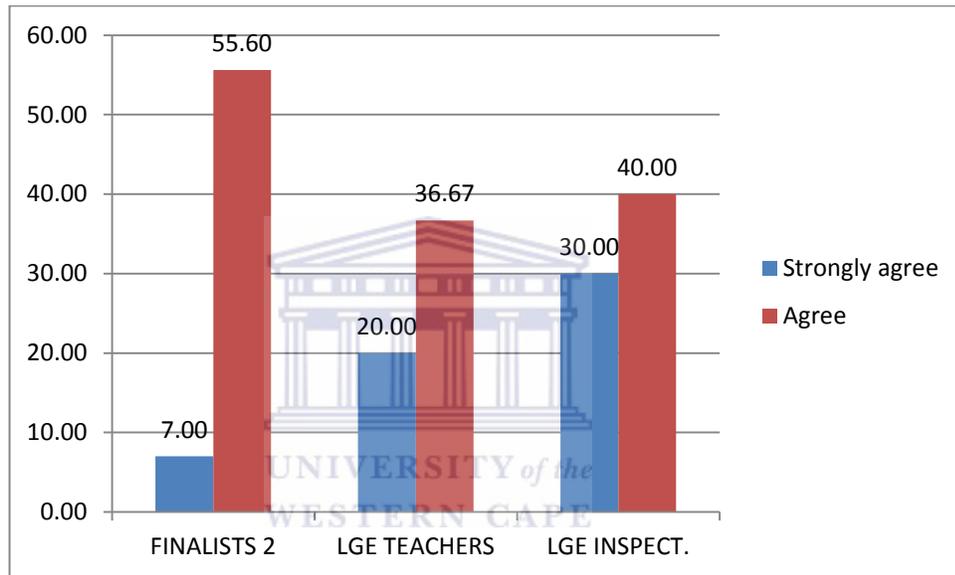


Figure 501: Respondents' views (agree/strongly agree) on task purpose

If we consider the respondents who either agree or strongly agree with the proposition, we will find that most of the language inspectors (70 %), many teachers (56.67%) and test takers (62.60 %) reported that the test rubrics provide instructions that enable to read the text and process the task.

### **Known criteria**

The question addressed for this variable is: *Are the criteria to be used in the marking of the ESE made explicit for the test takers?*

To answer this question related to scoring and assessment criteria variable, the first step is to examine the ESE rubrics. There is a rubric that indicates the scoring criteria, and this is indicated on the first page of the exam paper. This rubric reads: *“Each question is marked out of 1 point; and 1 point equals 1 %”*. Therefore, for the ESE, test takers work out of 9 points (also 9 %) for candidates from general subjects whose test includes 9 questions; and 5 points (also 5 %) for candidates from technical subjects group 1 whose test includes 5 questions; and 3 points (also 3 %) for candidates from technical school group 2 whose test includes 3 questions.

The second step is the examination of moderation procedures as indicated in “Revue de l’Inspecteur de l’Enseignement” (1982, No 6). Inspectors, school principals and teachers are required to vulgarize these procedures so that students know how they will be assessed to the national test.

According to these procedures, all subtest marks for each candidate are summed up with the practical test marks and the French “dissertation” marks. For instance, suppose that candidate X’s total mark in the subtests gives 41/85, and his/her score to the French “dissertation” is 3/5 while his score to the practical test is 7/10, his general mark will be  $41 + 3 + 7 = 51/100$ . The moderation process has three steps: (1) for each group (considered here as all candidates in the same class), all scores of individual candidates are summed up in order to get a global score. Suppose that a group A is made of 5 candidates who have taken the test, and each candidate has respectively got the following score: 51, 53, 53, 56, 61. This makes  $51 + 53 + 53 + 56 + 61 = 274$ . Then this global group score is compared with the global score of the same group in classroom assessment. Suppose that each candidate score in classroom assessment is respectively the following: 56, 54, 60, 60, 51. This also makes a global score of  $56 + 54 + 60 + 60 + 51 = 281$ . The next step (2) is: the test score for the group is subtracted from the

class assessment score for the same group first and for each candidate next. This will give, on the basis of the above example,  $274 - 281 = -7$ ; and for candidate X:  $51 - 56 = -5$ ; Y:  $53 - 54 = -1$ ; Z:  $53 - 60 = -7$ ; V:  $56 - 60 = -4$ ; and candidate W:  $61 - 51 = 10$ . As it can be seen, differences can be positive or negative. Positive difference means that the group or the individual candidate has performed better to the national test than to the classroom assessment; while a negative difference means that the individual candidate or the group has performed worse to the national test than to the classroom assessment. It should be noted here that positive differences are more preferable than negative differences. And step (3), a range for decision making is set. For instance, it can be agreed, prior to any moderation, that for any positive difference, the candidate will be attributed the higher score from his/her two scores. Looking at the above example, candidate W will get his/her diploma with 61 % since his/her difference is positive. However, for negative difference, a comparison will be made between the individual candidate difference score and the group difference score. Looking at the example, candidate Y whose difference is -1 is likely to be better scored than student Z whose difference is -7 simply because he/she is closer to the positive values than Z. In all cases, the final results to be reported depend on many factors, that is, these conventional moderation criteria, together with the political decision since the global results should not be below a certain range (for example, national results below 55 % of success are sometimes not politically accepted).

To evaluate the extent to which these scoring and assessment criteria are known and available to high school finalists prior to examination, we analyzed data from three questions (question 3, question 4, and question 5) of the research questionnaires. Respondents were asked to circle the number that reflects their viewpoint on a five –point scale. The question 3 reads: “For each correct answer I provide (the candidate provides) to question, I know that I (the candidate) get(s) 1 point; thus 1 %”. The following table gives the proportion (in %) of each category of respondents for each value.

Scoring criteria: 1 quest, 1 point, 1%	Finalists 1	Finalists 2	Language Teachers	Language Inspectors	Others
Strongly agree	54.50	60.40	68.33	75.00	39.20
Agree	41.75	37.80	31.67	25.00	40.80
Disagree	0.58	0.40	0.00	0.00	0.80
Strongly disagree	0.83	0.80	0.00	0.00	0.80
Can't tell	2.17	0.60	0.00	0.00	11.60
Not answered	0.17	0.00	0.00	0.00	6.80
Total	100.00	100.00	100.00	100.00	100.00

Table 503: Scoring criterion: 1 question = 1 point and 1 %.

The question 4 is phrased: “Marks I (candidates) get to the state examination are combined with marks I (candidates) get from the practical test and French “dissertation”. The following table gives the proportion (in %) of each category of respondents for each value.

Combination of SE marks & class marks	Finalists 1	Finalists 2	Language Teachers	Language Inspectors	Others
Strongly agree	6.08	4.20	21.67	35.00	15.20
Agree	22.75	25.40	40.00	45.00	27.60
Disagree	20.50	24.20	15.00	20.00	28.40
Strongly disagree	23.42	23.20	18.33	0.00	13.20
Can't tell	24.50	21.60	5.00	0.00	12.80
Not answered	2.75	1.40	0.00	0.00	2.80
Total	100.00	100.00	100.00	100.00	100.00

Table 504: Combination of test scores and other components

The question 5 is phrased: “Scores I (candidates) get in the SE will be moderated with my (candidates) classroom scores in order to get the final score”. The following table gives the proportion (in %) of each category of respondents for each value.\

Moderation of scores	Finalists1	Finalist 2	Language Teachers	Language Inspectors.	Others
Strongly agree	17.75	18.80	18.33	60.00	10.80
Agree	31.33	20.20	65.00	40.00	31.60
Disagree	10.42	15.80	11.67	0.00	26.40
Strongly disagree	6.58	2.60	1.67	0.00	17.20
Can't tell	30.75	40.80	3.33	0.00	8.00
Not answered	3.17	1.80	0.00	0.00	6.00
Total	100.00	100.00	100.00	100.00	100.00

Table 505: Moderation of test scores with classroom assessment scores

The above three tables give a clear indication that the respondents generally have the knowledge of the criteria on the basis of which students are evaluated at the national test. The scoring criterion (1 question = 1 point and this = 1 %) that is written on the first page of the test paper rubric is known by almost all respondents: (a) 54.50 % of finalists before the test strongly agree and 41.75 % agree (making a total of 96.25 %) with the proposition; (b) 60.40 % of finalists after the test strongly agree and 37.80 % agree (making a total of 98.20 %) with the proposition; (c) 68.33 % of English language teachers strongly agree and 31.67 % agree (making a total of 100.00 %) with the proposition; and (d) 75.00 % of English language inspectors strongly agree and 25.00 % agree (making a total of 100.00 %) with the proposition. Nevertheless, relatively lower scores are reported by “other stakeholders” category (with 39.20 % who strongly agree and 40.80 % who agree, making a total of 80.00 %). This relatively lower score compared with the “other stakeholders” category scores may be due to the heterogeneity of this group as different professional categories are included in.

The second criterion that relates to the combination of test scores and classroom assessment scores points to mixed results. As shown in table 504, English language inspectors in their majority (35.00 % strongly agree and 45 % agree, thus a total of 80.00 %) reported to know that test scores are combined with the practical test scores and the French dissertation scores; and many teachers (21.00 % strongly agree and 40.00 % agree, thus a total of 61.00 %) reported to know the

combination of scores procedures. However, only a minority of test takers (6.08 % of finalists before the test strongly agrees and 22.75 % agree, and 4.20 % of finalists after the test strongly agree and 25.40 % agree) reported to know that the scores they will get will be the combination of scores from three components: the test proper, the practical test and the French dissertation. Finally, the heterogeneous group of other stakeholders presents the result that is halfway between those who know the criterion (15.20 % strongly agree and 27.60 % agree, making a total of 42.80 %) and those who do not know it (28.40 % disagree and 13.20 % strongly disagree, making a total of 41.60 %). There is also one striking issue to note here: the “can’t tell” percentage in the finalist category (both before and after the test) which amounts 30.75 % for respondents investigated before taking the test and 40.80 % for those investigated after taking the test.

The moderation of test scores with classroom assessment scores criterion presents results that are also mixed. While language teachers in their majority (18.33 % strongly agree and 65.00 % agree, thus a total of 83.33 %) and all the English language inspectors (60.00 % strongly agree and 40.00 % agree, thus a total of 100.00 %) reported to know the moderation process, only almost a half (17.75 % strongly agree and 31.33 % agree, thus a total of 49.00 %) of the candidates before the test reported to know how their scores were moderated to get the final grade; and few candidates after the test (18.80 % strongly agree and 20.20 % agree, thus 39.00 %) had the knowledge of this criterion. The heterogeneous group “Other stakeholders” reported results that are mixed between those who have knowledge of the criterion (10.80 % strongly agree and 31.60 % agree, making a total of 42.40 %) and those who do not know this moderation criterion (26.40 % disagree and 17.20 % strongly disagree, making a total of 43.20 %).

### **5.1.3. Order of items**

The questions addressed for this variable are: *“Are items in testing test comprehension in a justifiable order? Are they grouped in a subtest in such a way*

*that reading is normally carried out in a linear, incremental fashion as a picture of the whole text?”*

The first step to answering this question is to examine the ESE papers. Indeed, a careful content analysis of the ESE papers has revealed that there is no logical ordering of questions in relation to the information provided in the text. Except for the first question which generally asks the test takers to provide a suitable title to the text or paragraph, other questions are ordered in such a way that the kind of information needed to answer them is placed at any place of the text. To take just an example of the ESE test in general education, question 1 relates to general comprehension of the text and candidates are asked to select the title that best suits the text, but question 2 asks for the test takers to scan a specific information that is located in the last paragraph of the text while question 3 asks students to use some clues in order to find out the meaning of a word located in the first paragraph. There is no linear ordering of questions in such a way that they relate to the way information is provided in the text.

Another aspect of ordering relates to the grouping of questions in subtests. The ESE presents all the nine questions, one after another without specifying which questions belong to which subtests. I hasten to suggest that the test is not divided in subtests although it is evident that it tests different reading constructs: there are questions on text comprehension, questions on grammar knowledge which are based on the text or sometimes not; and there are some questions that relate to vocabulary use which are based on the text or are out of the text. Sometimes there are questions on writing and on communication which may relate to the text or not. All these language skills and aspects are not labeled in the test and it is left to the student's capacity to infer whether such an item question tests a certain skill. More importantly, there is no rational use of the skills. Each test session has particular types of skills it covers, and this also varies according to different options. The following table illustrates the ordering of items in the 2007, 2008 and 2009 edition tests. At this juncture it will be helpful to assign a set of short explanations/ indicators as a way of qualifying the ordering represented by

numbers from 1-6. 1 deals with identifying the topic; 2 deals with comprehending explicit information; 3 deals with comprehending implicit information; 4 deals with inferring meaning of unfamiliar words; 5 deals with identifying relationships between sentences; 6 deals with understanding academic vocabulary.

Items	2007	2008	2009
1	1	1	1
2	2	3	3
3.	2	4	2
4.	6	2	2
5.	5	3	2
6.	3	6	3
7.	5	2	5
8.	3	4	2
9.	6	2	4

Table 506: Ordering of items in ESE

This table shows how there is an ordering of item for each particular test; and this ordering is neither justified nor systematic.

The second step has been to analyze data provided by stakeholders. Asked their view on the grouping, and/or ordering of test items, both language teachers and inspectors in their majority disagree with the proposition “Test items are grouped according to different items they test”; and this can be observed in the following figure:

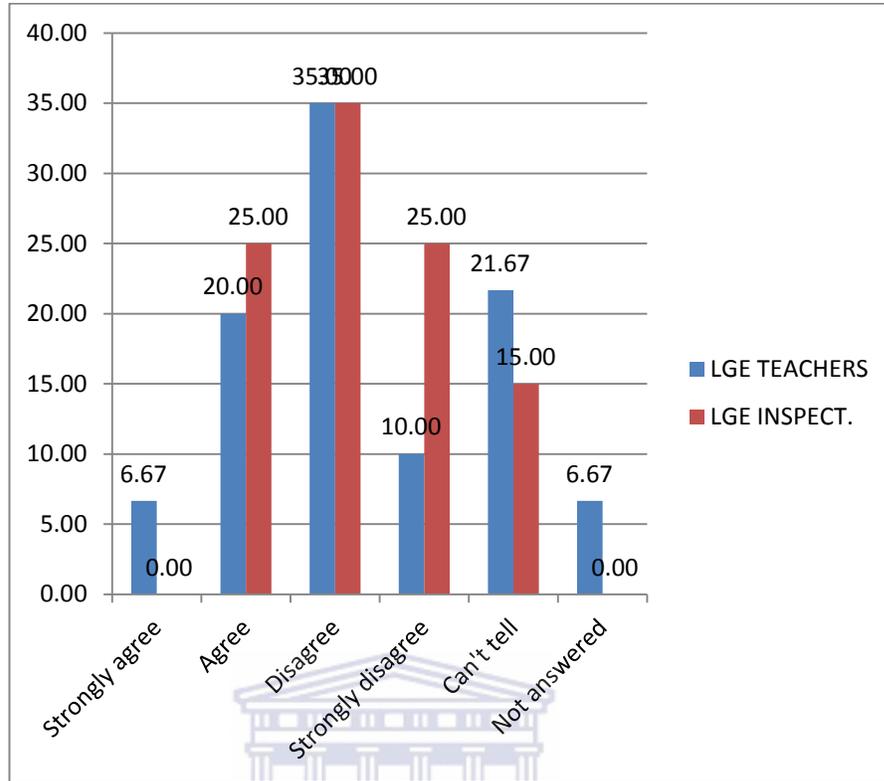


Figure 502: Grouping of ESE items in testlets: views of stakeholders.

In this chart, the majority of language inspectors (35.00 % disagree and 25.00 % strongly disagree; thus a total of 60.00 %) appear to believe that the ESE items are not grouped in testlets or ordered in a logical order. On the contrary, less than half of language teachers (45.00 % where 35.00 % and 10.00 %) could agree with the proposition. The explanation that I might offer here is that a good portion of teachers (21.67 %) “can’t tell” whether the test items flow in a certain logical order or not. I believe that this proportion view can be justified by the fact that to be able to answer this question requires certain knowledge of distinguishing between different skills that make up reading as well as to be capable to tell which item tests which skill; and this might be difficult to some respondents to tell. Even a relatively good portion of language inspectors (15.00 %) fall in that “can’t tell” category.

#### 5.1.4. Time allocation

The question addressed for this variable is “*Is the time allocated to reading the text passage and answering questions based on it appropriate?*” In other words, “*Can the test tasks be performed satisfactorily in the time allocated?*”

To investigate the extent to which time allocated to high school finalist is efficient in the completion of test tasks, I will first direct focus at the time provided for the TOEFL, a well-documented test, in relation to the tasks to be performed by candidates, and then compare this time to ESE allocated time. My choice of TOEFL test as a reference is motivated by the reason that some of the DR Congo students also take the TOEFL test immediately after having succeeded in the national test; and that I can use this comparison as part of my argument in the evaluation of the ESE.

The reading section of TOEFL consists of 3-5 passages, each approximately 700 words in length, with 50 questions for an approximate time of 60-100 minutes (Enright *et al.* 2000). The ESE, on the contrary, consists of 1 passage of approximately 150-250 words in length, with 9 questions for an approximate time of 90-120 minutes. This clearly shows that there is more than necessary time provided for the test takers to take the ESE.

Evidence I have used to investigate the extent to which the provision of much time in the ESE may or may not impact on performance was an experiment I made with three groups of respondents to which I administered the 2008 session test as part of classroom assessment. The first group was allocated 120 minutes, the second group 90 minutes and the third group 60 minutes. The test consisted of a reading passage of 167 words in length, followed by 9 questions based on it. I corrected the test for the three groups and correlated their scores. I found that: (1) the students in the three groups could provide answers to all the 9 test questions; (2) the score mean, 4.41, 4.39 and 4.38 for the first group, second group and third group respectively, were not significantly different.

To triangulate these findings, a question was asked to high school finalists after the test to report about their view whether time allocated to the ESE was enough to complete the task or not. The same question was asked to language teachers and inspectors and it reads: “The time allocated to take the language test (4 hours in general education and 3 hours in technical education) is.... (too much, enough, or not enough). The following table indicates the respondents’ views on this question:

ESE Timing: 4h/3h	Finalists	Language Teachers	Language Inspectors
Too much	2.20	20.00	35.00
Enough	68.80	53.33	65.00
Not enough	27.20	26.67	0.00
Can't tell	1.20	0.00	0.00
Not answered	0.60	0.00	0.00
Total	100.00	100.00	100.00

Table 507: Respondents’ views on the time allocated to the ESE

This table indicates different views on the time allocated to the ESE. While almost all student respondents (96.00 %) reported that the time is either enough (68.80 %) or not enough (27.20 %) for the completion of the test, only few teachers (20.00 %) and inspectors (35.00 %) admitted that time allocated to test completion is too much as this figure better illustrates this:

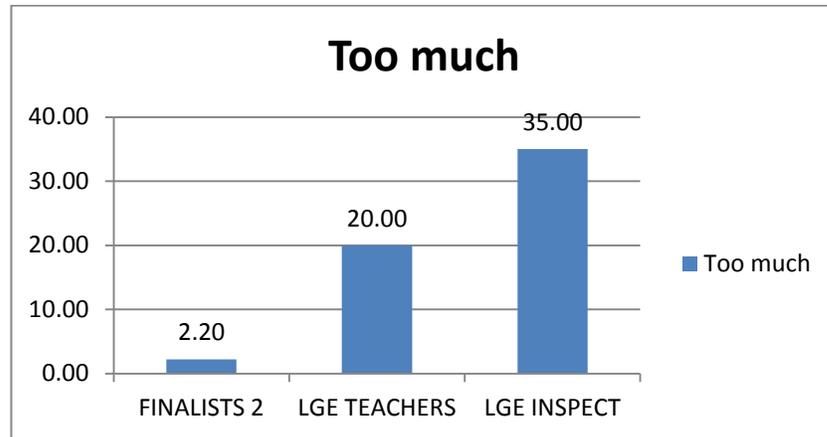
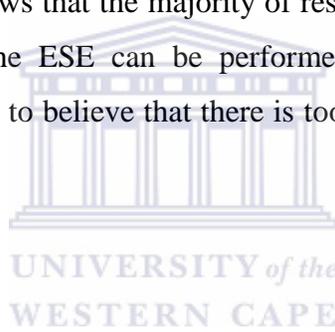


Figure 503: Respondents’ views on the assumption that the time allocated to take the ESE is too much

This figure clearly shows that the majority of respondents do not seem to support my conclusion that the ESE can be performed in lesser time than provided although I am justified to believe that there is too much time provided to take the ESE.



## 5.2. Task demands

This second section investigates the second research sub-question: “*To what extent does the ESE include tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task?*”

To answer this question, the discussion will focus on the analysis of ESE text types, text length, text content, and text difficulty and linguistic structure.

### 5.2.2. Text types

The questions addressed are: “*Are the text types appropriate for high school finalists? Is there a relationship between the text type selected and operations being assessed?*”

In order to answer this question, I will focus my discussion on the analysis of the content of the ESE papers regards the different text types that are used. Bachman (1990), Young and Weir (1998) and Manxia (2008) distinguish between narration, argumentation and exposition types (also called genres). But the DR Congo national curriculum (Programme National d’Anglais, 1988) insists on the need to include another type that relates to the students’ capacity to read and understand some practical simple materials such as letters, memos, notices, invitations, especially in commercial and related subjects. Therefore, I will analyze the text types from this perspective: narration, argumentation; exposition and practical simple materials. The following figure presents the % of text types as used in ESE paper samples for both general subjects and technical subjects.

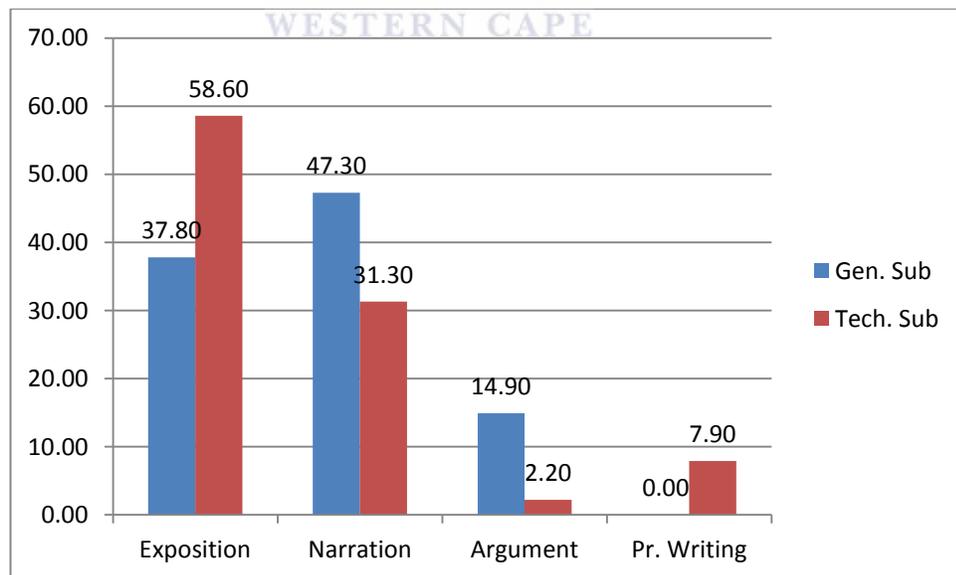
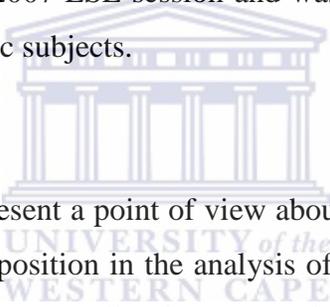


Figure 504: Text types as used in both general and technical subjects

From this figure, it appears that the ESE uses all the three text types (exposition, narration and argumentation) along with practical writing in technical subjects. In general education (subjects), almost half (47.30 %) of texts that are used are narratives. Then come expository texts that account for 37.80 % and argumentative text come last with only 14.90 %. In technical subjects, on the contrary, expository texts are the most used (with 58.60 %), followed by narrative texts with 31.30 %; and practical writing (especially in commercial and related subjects 7.90 % and argumentative texts are the least used with only 2.20 %).

Expository texts are those that primarily serve to inform the readers. This is the most often used text type in the teaching textbooks in DR Congo. Appendix 10 is an example of an expository text that uses description as an expository mode. This text is taken from the 2007 ESE session and was used to assess candidates from electricity and electronic subjects.



Argumentative texts present a point of view about a topic and provide supporting evidence in favor of a position in the analysis of the topic (Enright *et al.*; 2000). Argumentative texts are generally characterized by diction that may be personal in tone, or by the use of some vocabulary that points to an attitude or perception about the topic. The text in appendix 11 is an example of an argumentative text taken from 2008 ESE test session to assess candidates from agriculture and related subjects.

Finally, narrative texts tell a story. An example of narrative text will be provided in the following point related to the extent to which the text types relate to different operations being assessed. The question is: “*Do text types used in the ESE relate to different operations being assessed?*” To answer this question, we have randomly selected the 2007 edition test for general education. The test is based on the following 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".

This text type is narration. Thus the elements of narrative genre, as put by Mandler & Johnson (1977), Stein & Glenn (1979) and Stein & Policastro (1984) are the setting (which includes the characters and the story context) and episodes to reach a goal or solve a problem (which include the initiating event, internal response, attempt, the consequence, and the reaction). As put by Enright *et al.* (2000), these descriptors of story elements constitute a story grammar and are considered critical components for comprehension of narrative texts.

In this text, (a) the setting (which includes the story context and characters) is: (1) the context: the story starts with setting the time of the event in the opening sentence (**That Wednesday morning...**) and the place where the event took place (...**along MABANGA ROAD over the National Commercial School's fence**) and (2) the characters who are the **taxi drivers**, the **passengers**, the **police**, the **students** and **Apolosa**. Since it is a narration with a reporting feature, the writer, who is anonymous, speaks at the place of the characters.

Episode 2 is the response to the problem. 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 reaction to an abnormal situation/conflict the drivers face.

In episode 3, the writer describes the event/incident proper: “...**flocks of furious young men were feverishly throwing to their opponents any kind of harmful objects: pieces of stones, bricks, iron sticks...**”, “...**men in dark blue uniform were launching tears' gaz from their guns...**”, and “...**to disperse their close neighbours.**” In this episode, the writer uses different rhetorical devices to make the description vivid and maintain suspense. Adjectives such as “**furious**”, “**young men**”, “**harmful objects**”, and “**crazy students**” help the reader to visualize the event and to have a correct picture of its intensity. The adjective “**close neighbours**” visualizes the location of the two groups (the school is located next/opposite to the police station).

In episode 4, the author describes the consequences of the incident in the second paragraph. Various rhetorical devices are used to amplify the incident (“**Worldwide radio and TV reporters announced...**”; “the so-called **strong men...**” or to describe (“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** was also observed. Some demonstrators were caught to be **jailed.**”

(c) Flashback: the author uses this narrative technique to help readers discover the cause of the incident, and he/she does this in the concluding paragraph: “The incident broke up because ...”

As an evaluation, this text follows the story line of most narratives as described by Mandler & Johnson (1977), Stein & Glenn (1979) and Stein & Policastro (1984). It can be seen as a highly helpful/effective example of a reading passage that can better evaluate test takers’ achievement since the same kind of narrative texts are found in the three teaching manuals that are in use in the teaching of English in 5<sup>th</sup> and 6<sup>th</sup> grades. Nevertheless, an issue still needs to be investigated: to evaluate the extent to which the kind of questions based on this text reflect this story line used by the writer to construct the story and the readers to process and comprehend the narrative.

To answer this question, I will need to look at the kind of operations that are being performed by test takers. Since this text is a narrative using a description mode; hence, two important things are the story line and the picture of the story. Readers will achieve good understanding of such a text genre when they read for details (through scanning and understanding of explicitly stated information). Therefore, we expect comprehension questions on this text to be based on these two operations. Concretely, scanning will call for items focusing on: the narrative setting (where the story takes place, when it takes place, and who are involved). These might logically be operationalized in **questions 1, 2, and 3**. Likewise, scanning will call for items that relate to each story episode.

**Question 4** will logically focus on the initiating event and the test writer can either focus on adjectives that are used to amplify the incident (**large number, irritating, black smoke**) or on the identification of the problem.

**Question 5** will relate to the second episode of the story, which is the response to the problem. The test writer may ask the test takers to identify, through reading for explicitly stated information, the emotional reaction of “**drivers**” and

“passengers”. Hints to this question will be these two phrases: “**Some [drivers] stopped abruptly**” and “**Passengers were puzzled and wondered ...**”

**Question 6** will relate to the incident itself as described in the third episode. The test writer has the choice between focusing on identifying the two opponent groups (“**flocks of furious young men**” and “**men in dark blue uniform**” by guessing the appropriate words to which these two descriptions refer (“**students**” and “**police**”); or again (**question 7**), by identifying different instruments used by each group to challenge each other (“**any kind of harmful objects**”, “**tear gaz**” ...) or again (**question 8**), there is an opportunity to call for test takers’ capacity to identify names of these “harmful objects” in a vocabulary word list question.

**Question 9** will relate to episode 4 and the test writer has the choice between asking test takers to identify (through scanning) the rhetorical devices that are used to amplify the incident (“**Worldwide** radio and TV reporters announced/ a lorry ... was **badly** damaged/...the driver **mortally** wounded/ **sexual abuses**...”)) or (**question 10**) simply asking test takers to identify, through scanning, the various consequences of the incidents.

**Question 11** will relate to the last paragraph whether the author uses the flashback technique to describe the cause of the incident. The test taker can ask examinees to identify, through scanning for details, the cause of the conflict between the students and the police; or again (**question 12**), the reason why the police was sent to the school (“The Chief District of police then decided the protection of the school from the revenge...”)).

However, the main question at this stage is to know whether or not the 9 ESE test items that are based on this text relate to the operations test takers perform to construct the story line of the passage. In other words, do the items based on this text relate to those I have suggested? And do they follow the linear and incremental order that I have established from test question 1 to test question 12? To answer this question, let us look at the ordering of the 9 items as presented in the ESE paper.

1. The title which best suits the second paragraph of the text is:
  1. Trouble in the N.C.S
  2. Violent disturbance of the peace by the police and the students
  3. The results of trouble in the N.C.S.
  4. The accident in the N.C.S.
  5. The causes of the trouble in the N.C.S.
2. All the following results of the incident at the N.C.S. are correct, except:
  1. golden jewelry robbery was observed
  2. sexual abuses were committed
  3. the lorry driver was mortally wounded
  4. One student died
  5. Apolosa was badly beaten
3. The reason of the incident at the N.C.S. which is a particular concern of the C.T.I.'s students is:
  1. the police anger on Apolosa's injury.
  2. the probable revenge of the C.T.I.'s students.
  3. the refusal of the N.C.S.'s students to stand the police presence close to them
  4. the police intervention to protect the N.C.S.
  5. The severe harm inflicted to the leader of the C.T.I's students by the N.C.S.'s ones.
4. The following objects were used by students in the incident at the N.C.S., except:
  1. pieces of stones
  2. Any harmful objects
  3. Iron sticks
  4. tears'gaz
  5. Pieces of bricks
5. In the last paragraph of the text, the underlined "his mates" replaces:
  1. The police
  2. The school teacher
  3. The N.C.S.'s students
  4. The C.T.I.'s students
  5. The lorry driver
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 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 policemen
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.
  1. "Who killed the student?"
  2. How many people died during the incident?"
  3. "How old was the dead student?"
  4. "When did the incident happen?"
  5. "What are the causes of the incident?"

8. With reference to the results of the incident, one might suggest the following pieces of advice, except:

1. avoid violence in every similar case in the future.
2. keep away from violence at all costs since it brings more harm than good.
3. avoid exchanging views with the other party.
4. get in touch with each other to look closely at the causes of the incident for a better future.

QUESTION OUT OF THE TEXT

9. “car, lorry, drivers, passengers and car-park.”

The title which best suits the above group of words is:

1. Army
2. Communication
3. Transportation
4. Hospital
5. School

### Comments on the 9 questions

A content analysis of the above nine questions appears to suggest the following:

**Question 1** calls for global comprehension in order to identify the topic of the text. Since the topic is not stated, test takers will struggle to read for global comprehension by skimming the whole text as well as the different text parts (paragraphs). This operation will be performed in conjunction with reading for specific details (scanning) to construct the story line. I believe that not stating the text title but asking examinees to skim read the text and identify the title is an operation which is likely to create problems for many test takers.

Also, a careful look at the 5 options reveals that there is more than one option that is correct. Since the text is a narrative organized in a cause /effect mode, either the cause (option 5) or the effect (option 3) or both (option 1) may stand as a title to this text.

As an evaluation of this item, it is evident that the operations to be processed by test takers to answer do not follow the story line that help to construct the text; besides, it cannot be included within the framework of the kind of items that I

have suggested previously. Another reason is that it is ambiguous since it opens the possibility of more than one correct answer.

**Question 2** relates to the consequences of the incident as described by the author in the second paragraph. Here, the test writer fails to follow the story line since the information the question addresses relates to the fourth episode, which, according to our previous argument, should be addressed in question 9. Nevertheless, since this item calls for scanning, an operation that relates to this kind of text that is built around details, it is a good item and needs only to be shifted toward the end position.

**Question 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 the last test question. Nevertheless, this item calls for scanning for specific details; therefore, it is a good item; and it only needs to be put at the correct order.

**Question 4** relates to the vocabulary words (objects) used by the characters in the incidents; as indicated in episode 3 of the story line (see paragraph 2). This question should have been the 6<sup>th</sup> question. As for the operation it calls for answering it, it is a good question since it requires test takers scan the text for specific information.

**Question 5** relates to the last paragraph as the test writer asks test takers 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 test takers' linguistic knowledge. Only, the order of this item on the question list is problematic.

**Question 6** calls for both skimming and scanning skills to get information from the text as the question points to the whole text; not to a specific part of the text. This is a weak question since test takers are likely to use many operations to get the answer. Its scope goes beyond the scope of narrative texts that are essentially

based on reconstructing the story line of the text through scanning for specific details.

**Question 7** calls for the test takers' metacognitive skill of prediction since they have to make inference on the basis of their global comprehension of the text. Also, this question has a communicative function that relates to the kind of materials used in teaching and learning situations. It is a good item.

**Question 8** points to the text purpose: to narrate with the objective of giving advice. This is a sound question.

**Question 9** is out of the text. I am not against the idea of setting some questions out of the text, but we believe that a single question is not appropriate. A section of 4-6 questions would be useful, especially if these questions focus on another skill such as writing skill.

In order to sum up the discussion so far, we could say that as the answer to the first question, the ESE uses different types of texts that are appropriate for high school finalists. As the answer to the second question, it appears that most items in this ESE sample test do not follow the story line of the text; besides, some questions do not call for operations test takers are likely to use to construct text meaning. Yet, there is evidence that narrative texts typically have a hierarchical structure, and that readers are sensitive to such structure so that when this structure is used to guide comprehension and recall, both are facilitated and test performance is enhanced (Shin, 2002; Glenn, 1978; Carrell, 1985).

### **5.2.3. Text length**

The question that I need to address for this variable is: *“Is the text length appropriate for the target situation requirements of the students being tested?”*

The national curriculum is mute as regards the length of texts on which the ESE is based. However, on the basis of the three textbooks widely in use in 6<sup>th</sup> grade, three patterns of text length are distinguished as indicated in this table:

Textbook/Length	English for Africa	Today's English	Go for English
100-150 words	23.70	13.50	19.00
150-200 words	31.75	30.00	35.00
200-300 words	22.50	34.00	25.00
300-500 words	20.30	12.50	17.00
Over 500 words	1.75	10.00	4.00
TOTAL	100.00	100.00	100.00

Table 508: Percentage of text length in three textbooks in use

As this table indicates, the length of the text in the three textbooks in use for teaching English in 6<sup>th</sup> grade varies considerably within each textbook as well as across textbooks. If we consider what can be termed longer texts, there is only one textbook “Today’s English” that has only 10.00 % of texts beyond 500 words; while the other two textbooks (English for Africa, Go for English) almost have no longer texts. These three textbooks use almost half of their texts that are between 100 and 200 words. For instance, “English for Africa” uses 55.45 % (23.70 % between 100-150 words and 31.75 % between 150-200 words), “Today’s English” uses 43.50 % (13.50 % between 100-150 words and 30.00 % between 150-200 words) and “Go for English” uses 52.00 % (19.00 % between 100-150 words and 35.00 % between 150-200 words). The following figure visualizes the most important concentration of texts as regards their length. Note the EFA stands for “English for Africa”; TI stands for “Today’s English” and GFI stands for “Go for English”.

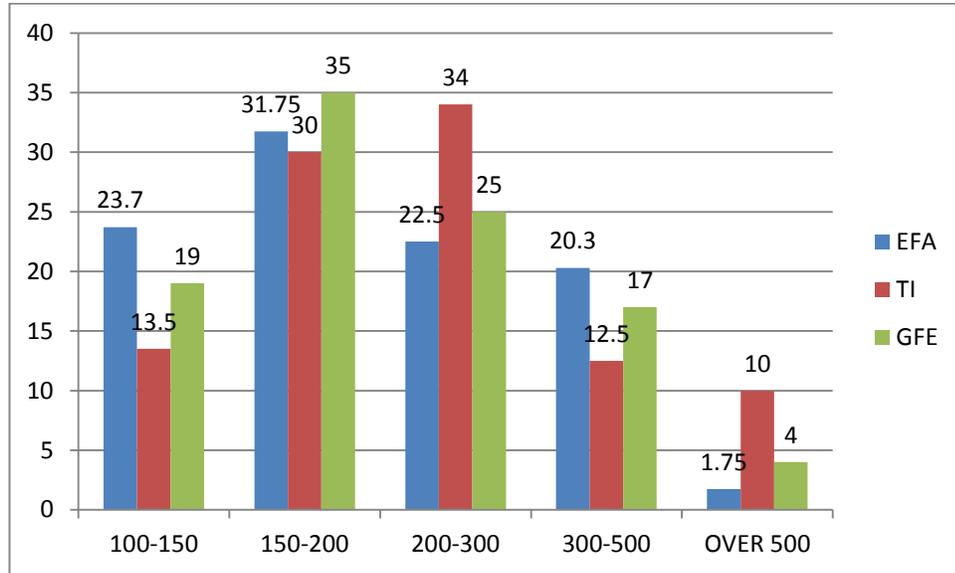


Figure 505: Concentration of texts regards their length

From this figure, it appears that most texts in use for teaching are between 150 and 300 words. “English for Africa” is the textbook that uses the most important percentage of shortest texts while “Today’s English” uses the most important percentage of longest texts.

An examination of text length in the ESE papers reveals that there are only single texts that are used and that there is no specific length for these single texts. The following table indicates text length patterns (in n- words) as used at the ESE.

Length	100-150	150-200	200-300	300-500	Over500	Total
Gen.Educ	27.00	43.00	19.00	9.00	2.00	100.00
Tec.Educ	61.00	39.00	0.00	0.00	0.00	100.00

Table 509: Patterns of text length at ESE for both general and technical subjects

As this table indicates, the length of the ESE texts varies across subjects (general or technical subjects). Texts are longer in general subjects (28.00 % between 200 and 500 words, and 2.00 % over 500 words) than they are in technical subjects (no text over 200 words). This indicates that all texts in use in technical subjects are short, whereas in general subjects, texts are fairly long. However, compared to

the text length used for teaching, we find out that the ESE use texts that are far shorter than those students use in classroom instruction. For example, in ESE all texts used in technical subjects are between 100 and 200 words while for teaching, students are exposed to texts that sometimes exceed 500 words.

However, stakeholders do not agree with me that the ESE uses short texts. The following table presents the views of high school finalists after the test, language teachers, and language inspectors regards the test length.

Text length	Finalists 2	Language Teachers	Language Inspectors.
Very long	3.80	3.33	0.00
Long	42.60	15.00	15.00
Of moderate length	47.20	60.00	60.00
Short	2.40	21.67	25.00
Can't tell	3.20	0.00	0.00
Not answered	0.80	0.00	0.00
Total	100.00	100.00	100.00

Table 510: Respondents' views on text length

As this table indicates, only 25.00 % of inspectors, 21.67 % of language teachers and 2.40 % of high school finalists reported that the ESE uses short texts. The majority of respondents reported that the ESE texts are either of moderate length (60.00 % for both teachers and inspectors, and 47.20 % for high school finalists) or they are long texts (15.00 % for both language teachers and inspectors; and 42.60 % for high school finalists). It might be possible to explain these different results by pointing out that the respondents who reported that the test is of moderate length may have compared the length of many tests used for teaching (that ranges between 150 and 300 words) to the length of those used for testing (that ranges between 100 and 200 words) that fairly correlates to some extent. If this argument seems plausible, I could argue that the length of texts used for teaching in DR Congo is not an appropriate standard; therefore, one cannot use these texts as a point of reference. At this juncture it will be helpful to consider

the following text that has been used in the 2005 session test for dressmaking subject candidates.

The headman allowed his wife to behave in such a way that she did not work. He engaged other wives to do all the work for her so that she could not hoc, pound, cook and even draw water. She sat on her seat of honor and wove baskets. After her husband's death, as is the custom, she became the headman's brother's wife. One day, the two other wives asked her to crush some millet because they were tired. She got angry, but when she was crushing it on the stone, she sang:  
"When I was married, I sat on a seat and wove baskets". As she sang, water began to flow from her, the other wives laughed at her singing, but soon, the water flooded all the village and the people sank and died. **(145 words)**

This 145 words passage is in the interval of text between 100-150 words which includes 61.00 % of texts used in technical subjects (see table 509 previously). Although this text is a narrative, it fails to follow the story line that characterizes most narratives. Besides, it does not include details on the basis of which the test taker can process different operations to understand the text and answer questions based on it. Also, I believe that any test writer will hardly write more than five sound comprehension questions based on this text.

#### **5.2.4. Text content**

The question I will address here is: *"To which extent is the text content appropriate for the target situation requirements of the students being tested? In other words, is the test taker familiar with the text content?"*

This question is very relevant to the Congolese context where candidates of related subjects are grouped and given the same test. For instance, candidates from agriculture, horticulture, veterinary, nutrition, forestry, and fishing subjects are grouped and assigned the same test; so are those from commercial, computing and secretary subjects. Since a single test is submitted to a fairly heterogeneous group, the test writer should ensure that the selection of text content is not likely to advantage one group on behalf of another; thus biasing the test results. Also, another serious issue to consider is the selection of content that is not too familiar

so as to allow the candidate to answer some or all the questions without necessarily reading the text (Buck, 2001).

To investigate this issue, I have used a procedure comprising two steps: first, the identification of the groupings of different subjects since different groups are assigned different tests; and then the content analysis of each test regarding the text content.

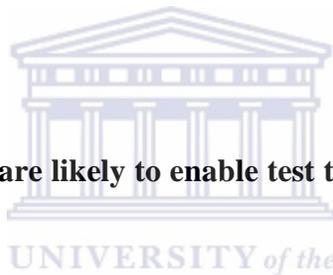
As for groups' identification, the content analysis of the ESE papers reveals that two main groups are distinguished: the general subjects group and the technical subjects group. The general subject group consists of candidates whose main subjects include: (1) Latin-Philosophy, (2) Mathematics-Physics, (3) Biology-Chemistry, (4) Greek-Latin, (5) Latin-Mathematics, (6) General Pedagogy and (7) Physical Education. Candidates belonging to these 7 major subjects take the same English test.

The technical subjects group consists of 6 subgroups: **Subgroup 1** includes candidates whose main subjects are (1) Commercial –Administration, (2) Secretariat, (3) Commercial-Computing, (4) Secretariat-Computing. **Subgroup 2** includes candidates whose main subjects are (1) Plastic arts, (2) Dramatic arts, and (3) Music. **Subgroup 3** includes candidates whose main subjects are (1) Hostess, (2) Hostelery and (3) Housing. **Subgroup 4** includes candidates whose main subjects are (1) Agriculture, (2) Horticulture, (3) Veterinary, (4) Nutrition, (5) Forestry, (6) Fishing, and (7) Agro-Forestry. **Subgroup 5** includes candidates whose main subjects are (1) General mechanic, (2) Machine-Tools mechanic, (3) Electricity, (4) Electronic, and (5) Building. Finally, **subgroup 6** includes candidates whose main subject is dressmaking.

The second step consists of the content analysis of different texts used in different groups and subgroups. In general subjects, although the authors of different texts do not mention the text title, reading passages are built around the following themes: science, western civilization, tourism in some African countries, animal behavior, history and origin of money, detective, law in traditional society,

western culture, sport, the use of underground, food, population growth in developing countries, road accident, AIDS, social life and problems, and problems of building towns.

The analysis of these text contents reveals three basis patterns: (1) texts that include general topics and are likely to enable the test takers to deploy their background knowledge to read the text and understand it. This group includes the majority of texts (65 %); (2) texts that include general topics, but that are likely to advantage one group of candidates on behalf of another group of candidates. This group includes texts that account for 25 %. The last group of texts (3) includes those texts that do not relate to the candidates' background knowledge. This group accounts for 10 % of all texts. The following ESE texts illustrate these three patterns:



**Pattern 1: Texts that are likely to enable test takers deploy their background knowledge.**

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.

(Taken from the 2006 ESE paper)

To investigate the relationship between text content familiarity (background knowledge) and test performance, I assigned this text to two groups of 2010 high school finalists as a classroom final examination. The two groups were selected according to urban-rural variable. And within each group I considered the sex variable. My objective was to investigate whether students from one group might be more advantaged by this text content than another group. Concretely, I hypothesized that the students from the urban could perform better than those from rural areas since the text contains information urban students are likely to be familiar with, and that they could deploy their schemata to read the text and process the task with ease. As for the sex variable, I hypothesized that male students could perform better than female students since the former are more informed than the latter on HIV. Using the correlation coefficient, and the ANOVA test on group scores, the results indicated that there was not a significant difference as regards urban-rural variable as well as the male-female variable. These results indicate that this text content is authentic and familiar to all the candidates in that they recourse to their schemata to read the text and understand it.

**Pattern 2. Texts that are likely to advantage one group of candidates on behalf of another group of candidates.**

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.

(Taken from the 2001 ESE paper)

To investigate the extent to what this text is likely to advantage one group of test takers on behalf of another, I administered, under controlled conditions, this text to two groups of high school finalist students as part of their final classroom examination. The variable manipulated was "rural-urban" variable. My objective was to investigate whether students from one group might be more advantaged by this text content than another group. Concretely I hypothesized that students from the urban area could be more advantaged by the content of this text than those from the rural area. After the correlation of group scores and the use of ANOVA test, I could draw the following conclusions: (1) the mean scores of the two groups was significantly different: the group from the urban area outperformed the group from the rural area; (2) students belonging to the group from urban area could have deployed their schemata to process the text and understand it since the text relates to the kind of information that is familiar to them. They should have activated their background knowledge to comprehend this text; (3) students from the rural area could have found the text content very "strange" and thus, have failed to visualize the storyline that is important to understand the story details and successfully answer to questions based on it.

**Pattern 3: Texts that do not relate to test takers' background knowledge.**

Most large cities now have an underground railway, like the French "Metro". When you need to change trains, you usually have to walk along bare, uninteresting corridors. Sometimes, if you are lucky, there will be buskers; people playing music in the hope of earning few money. They will be very pleased if you throw a few coins into the hat which they had placed on the ground in front of them, but they will go on playing happily, even if you don't. These buskers give us a moment of pleasure as we rush for trains. It is a pity, then, that the authorities are against busking, and do not like to enjoy it. Busking may be breaking the law, but I think that most people are glad to have a little music to brighten their busy lives.

(Taken from the 2004 ESE paper)

To investigate the extent to which this text does not relate to the test takers' background knowledge, I administered it to two groups of students under the same conditions. Two variables were manipulated here: the urban-rural variable and the gender variable. After the analysis of group scores and different correlations; I could conclude the following: (1) All the two groups poorly performed in the test and that neither the urban-rural variable nor the gender variable could differentiate group performance; (2) the text could have been unfamiliar to students so that they could not activate their schemata to read and comprehend it.

In technical subjects there are different themes depending on the subject group. In Commercial and related subjects, for example, text contents relate to these themes: domestic economy, parts of a letter, banking, insurance, and business documents. In Arts related subjects, themes relate to painting and the design of houses and other valuable objects, and to ceramic. In Agriculture and related subjects, text themes pertain to the description of animals, nutrition and related problems, dinosaurs, importance of domestic animals, importance of planting trees, production of groundnuts, collecting and sharing water, and to the importance of forest preservation. In Electricity and related subjects, texts focus on themes that relate to source of electricity, machine tools, milling machine, extra high tension. Finally, in dressmaking subject, themes relate to sewing shop

description and equipment, parts of a sewing machine, clothes, dresses, and weaving.

The content analysis of ESE texts for technical subjects reveals three basic patterns: (1) texts that relate to test takers' subject knowledge, and the test takers use both their linguistic knowledge and subject knowledge to process the task and answer the test questions. This pattern accounts for 40 % of the texts; (2) texts that are too familiar to test takers; therefore, some test items can be answered without reading the text. This pattern accounts for 25 % of the texts; and (3) texts that relate to one group of test takers, and this group is likely to be advantaged on behalf of the other group. This pattern accounts for 35 % of the texts.

**Pattern 1: Texts that relate to test takers' subject knowledge and the test takers use both their linguistic knowledge and subject knowledge to process the task and answer the test questions.**

The following text is taken from commercial and related subjects 2007 ESE test.

Usually the lack of money is a universal problem for travelers. There is a variety of ways to carry it. Each way has both advantages and disadvantages. So a combination of two or three ways is advisable to respond to financial circumstances and to destination.

Traveler's cheques will be replaced if lost or stolen, theoretically within 24 hours. In this case you pay 1 to 1.5 per cent of the value of the cheques. But usually, you get a better rate when cashing them. Travelers must be sure to carry dollars cheques in the American countries.

Carry a small amount of foreign currency for taxi, porters, telephone calls, snacks until you can get to a bank. Most United Kingdom banks need advance notice of your requirements, otherwise, change sterling at the airport or at the port. Some countries, in particular Greece, restrict the amount of their currency that you can import. You should also carry some sterling for necessary expenses when you return.

Post cheques are also useful for travelers. Each cheque, when accompanied by a Postcheque card can now be used to draw up to one hundred sterling in local currency from 90,000 post offices in most Europe and around the Mediterranean as well as Hong Kong, the Bahamas and Japan.

Credit cards such as Access (linked to Mastercard in the United States and Eurocard in Europe) and Barclaycard (linked to Visa) are accepted in nearly 5 million bank operations. Note that Barclaycard is stronger in France, Spain and Italy, whereas Access is most useful in Germany and the United States. The travelers need also Charge cards. Among them are American Express and Diners Club which are less widely accepted than credit cards.

Eurocheques can be used to withdraw local currency as well as pay for hotels, restaurants, garages, and other European services. The cheques, made out to the exact amount you require, are then debited to your account in the same way as a domestic cheque.

Individual cheques can be cashed for up to one hundred sterling or the equivalent in local currency. There is no limit to the number of cheques you can use to make a purchase. You pay around three point fifty sterlings for the card and there is also a commission of one point twenty five per cent on the value of the transactions, plus a 30 pence handling fee per cheque.

(Taken from the 2007 ESE paper; Commercial and related subjects)

To investigate the extent to which this test is likely to require both linguistic and subject knowledge in order to read and understand it, I administered it as a classroom final test to high school finalists in three groups. The first group was made of students from Commercial-Administration subject; the second was made of students from Commercial- Computing subject, and the third group made of students from Secretariat-Computing subject. It should be noted here that all these three groups of students have a course on Banking and Insurance as part of their curriculum. After the analysis of results and the correlations of group mean scores, I could find the following:

(1) Most students reported that they had recourse to their subject matter knowledge to read the text and provide answers to questions based on it. This can be explained by the fact that all these three groups have courses in which “banking and insurance” are taught as part of the curriculum; (2) There was no significant difference that could be attributed to the different groups. This could be seen through the group mean scores (5.75, 5.86 and 5.81 for group 1, 2 and three respectively); (3) Students with good linguistic skills were those who performed better regardless the group to which they belonged. This establishes evidence that both background knowledge and linguistic knowledge are two factors that interact in text comprehension.

**Pattern 2: Texts that are too familiar to test takers; therefore, some test items can be answered without reading the text**

The following text illustrates this pattern. It is taken from the 1997 ESE of commercial and related subjects.

The unit of currency in the first Democratic Republic of Congo was 'Franc'. It was issued from the 1961 monetary reform. Many other reforms have taken place during the last decades.

Since 1967, the monetary council of the Democratic Republic of Congo decided to introduce 'Zaire' as the new currency. This was a powerful currency; 1 zaire equaled 2 US dollars. But the continual printing of bank notes of the new currency led to political conflicts, embezzlement and starvation: this gave some food for thought to a group of people who decided to fight for the good of everybody.

Since the liberation of the country on 17 May 1997, Congolese people hope things are changing in the near future, especially about matters. In his speech to the nation on June 1997, His Excellency President Laurent D Kabila announced that another reform is taking place very soon.

(Taken from the 1997 ESE test paper)

To investigate the extent to which this text is too familiar to candidates and that test takers could answer some questions without reading the test, I administered this test to two groups of high school finalists as a final classroom test. The first group was made of students from Commercial-Administration subject, those for whom the test was intended in the 1997 ESE session. Since I had suspected this text to pertain to the general history of DR Congo, I administered it to another group that was made of students from Mathematics-Physics students. The objective was to investigate if familiarity with the DR Congo history by the two groups could lead to too much familiarity with the information in this text, therefore, to the ability to answer all (or some of) the items without necessarily relying on the text. The following are questions based on this text:

1. The Congolese Franc was adopted by the :
  1. president of the Democratic Republic of Congo
  2. general commissioners
  3. Prime Minister
  4. monetary council
  5. government of the Democratic Republic of Congo
2. The second monetary reform occurred:
  1. in May 1980
  2. during the national conference
  3. in June 1967

4. during the Second World War
5. during the colonial epoch
3. The continual printing of bank notes in the Democratic Republic of Congo showed that the:
  1. government controlled the economy
  2. country was in economic crisis
  3. politicians loved their country
  4. bank governor worked hard
  5. country was well rich
4. Zaire currency was one of the most powerful currencies in the world because:
  1. it had many coins
  2. one zaire equaled 2 dollars
  3. only bank notes were used
  4. the bank had many notes
  5. the president of the country was very powerful
5. The liberation of the country led people to hope that:
  1. there will be more printed notes
  2. their living conditions will not be improved
  3. money matters will change
  4. the government will lose control of the economy
  5. there is no need for a new currency

After the analysis of results and the correlations of group mean scores, we could find the following: (1) All the two groups well performed in the test (8.11 and 7.98 for Commercial-Administration and Mathematics-Physics, respectively); (2) The group mean score difference 0.13 was not significant. The higher mean score observed with students from Commercial-Administration could be explained by their subject knowledge since monetary concerns are part of their course curriculum; (3) All the two groups found the text to be too familiar to them for two reasons: first, the text pertains to the history of DR Congo which is part of the high school curriculum; and second, the text content relates to daily concerns of all Congolese citizens; (4) three students interviewed after the tests reported that they could answer some questions without necessarily reading the text. Except question 1, most respondents pointed to the other questions. Here are two comments from two respondents:

**Respondent 1:**

I could choose 2 [option] to question 4. You know, we all know. Our parents usually tell us... 1 zaire was 2 dollars and they could buy a lot of food with only

5 zaires. In history class we also studied. I could not spend time to read the text; the answer was there.

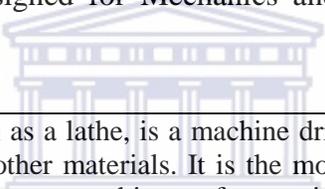
Asked how many times he had read the text before he could read the questions and answer them, another respondent told us:

**Respondent 2:**

You know, we all wanted the new government to change things. The Franc [Congolese Francs] has no value; everyday it loses its value. We are all tired. They [the government] do not do anything... I could read the text 1 time and choose [option] 3 to question 5. It is easy!

**Pattern 3: Texts that relate to one group of test takers and this group is likely to be advantaged on behalf of the other group**

The following text designed for Mechanics and related options illustrates this pattern:



A machine-tool, such as a lathe, is a machine driven by power that cuts, shapes or finishes metal or other materials. It is the mother of all machines. From the machine-tool flows every object of our industrialized world, such as automobiles, air planes, diesel locomotives, washing machines...  
Apart from machine-tools, the engineer who is the guiding force behind the machine-tool uses other tools to do his job. Some of them are spanners, calipers, a hammer, a screw driver, pliers and micrometer. Each tool is used for a special purpose. A spanner for instance is used for loosening a nut.  
(Taken from 2005 ESE test paper)

To investigate the extent to which this text is likely to advantage one group of students on behalf of other groups, I administered it as a final classroom test to 5 groups of students that belong to the General mechanic and related subjects: General mechanic, Machine-Tools mechanic, Electricity, Electronic, and Building. My objective was to investigate if the text content could be familiar to all these 5 groups of students, and if the students from the all groups could perform equally.

After the analysis of results and the correlations of group mean scores, I could notice the following: (1) the test contains the kind of information that advantages

two groups of students (those from General mechanic and Machine-Tool mechanic subjects) on behalf of other 3 groups of students (those from Electricity, Electronic and Building subjects); (2) The group mean scores for the first two groups ( 7.61 and 7.63 for General mechanic and Machine-Tool mechanic respectively) is higher than the group mean scores for the last three groups (6.28, 6.35 and 6.03 for Electricity, Electronic and Building respectively) and this can be partly due to more familiarity with text content for the first two groups and less familiarity with test content for the last three groups. These results appear to provide the evidence of bias since such test based on a text that advantages some groups on behalf of other groups is not likely to be a fair measure of students' performance. Interviewed on how they found the text content, one student who did not find the text to be familiar to him said: *"You see, it is not easy to understand these things: calipers, pliers, eh... loosening. These things are difficult"*. Another respondent was even more explicit when he could said: *"I'm going to be [a]building engineer, not a mechanic"*. This reaction can be understood when one looks at the following item that was one of the test question:

One of the following tools is used for holding or pulling out small things such as nails or for bending or cutting wire.

1. Spanner
2. Screwdriver
3. Calipers
4. Pliers
5. Hammer

This question provides a noticeable evidence of bias. Since the text does not mention the instrument that is normally used for holding or pulling out small things such as nails or for bending or cutting wire, students from Building, Electricity and Electronic cannot know the appropriate instrument for such job since it is not part of the kind of tools they use in their workshops.

### 5.2.5. Text difficulty and linguistic structure

To investigate the extent to what the ESE use texts of appropriate difficulty level, I wish to address the following questions: *“What are the difficulty levels of a range of ESE texts? What makes these texts difficult or less accessible?”*

Some steps are considered to answer the above questions. The first step is to determine the difficulty levels of a range of ESE texts. I selected a corpus of 7 ESE sample texts from both general subjects (3) and technical subjects (4). Then I calculated the Flesch reading index for each of the texts before we could report the predicted range of ESE text difficulty.

As said in the third chapter, the Flesch formula is an abstract scale which runs from 0 to 100, with more difficult texts having lower scores (<50) and the easiest texts having higher scores (>50). The formula reads:

$$\text{Reading ease score} = 206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$$

Where: ASL = average sentence length (number of words divided by number of sentences);

ASW = average word length in syllables (number of syllables divided by number of words).

By applying this formula to the 7 ESE texts from the corpus, I could get the following reading ease score (index): Note that GS= General subjects, TSCR= Technical subject: Commercial and related subjects; TSD= Technical subject: dressmaking; TSAR=Technical subjects: Agriculture and related subjects; and TSMR= Technical subjects: Mechanic and related subjects.

Text code	Flesch reading ease score
1.GS 2004	48.5
2.GS 2007	51.2
3.GS 2010	53.2
4.TSCR 2006	46.8
5.TSD 2007	59.4
6.TSAR 2005	53.8
7.TSMR 2008	58.3

Table 511: Difficulty estimate for the 7 ESE texts

Table 511 demonstrates that all the texts in the corpus are either easy (GS 2010, TSD 2007, TSMR 2008) or of moderate difficulty level (GS 2004, GS 2007, TSCR 2006, and TSAR 2005). Note that texts with reading ease scores around 50 are classified as texts of moderate difficulty; that is, not difficult, not easy.

However, since the Flesch reading ease formula is concerned largely with the function of the length of sentences and the size of words in a text, it does not give a clear indication of the actual text difficulty. Rather, it could only be a predictor of text difficulty. Therefore, to investigate the actual difficulty of the selected 7 texts, I had to use another method. This method required some input from “experts”, who I believed would be able to analyze the texts and report their perceived difficulty on the basis of some criteria. The experts consisted of 3 English language teachers and 2 language inspectors. The 3 language teachers were selected because of their long experience of teaching English in grade 6 and their great familiarity with the ESE test as they were using it as part of their teaching materials. As for the English language inspectors, they were selected on the basis of their involvement in all the evaluation process of the state examination: from preparation to administration and evaluation.

The experts were asked to: (a) sequence the texts in order of perceived difficulty, and (b) indicate the reasons for such sequencing. Rubrics included how clear the text purpose was made for the test takers, sentences length, and clarity of topic

development, high/low information load, lexical density, syntactic complexity, logical connection and signaling devices. For statistical analyses, I used the Spearman rank order correlation between individual experts and I could interpret this coefficient as a measure of inter-expert agreement.

However, since the experts were working individually, the average inter-judge agreement was only 0.27, indicating that they did not rank in order these texts in the same way. Nevertheless, I could find a positive correlation (although low) of 0.33 between the judges' correlation and the Flesch reading ease correlation. This indicated that the experts could find that the 7 reading texts were between easy and moderately difficult texts, as had indicated the Flesch reading ease score.

Although the 5 experts could not rank the texts in the same way, they could agree on the different linguistic aspects that made the texts either relatively difficult or easy. The following text is an example of a text that was reported to be fairly difficult by both the Flesch reading ease formula and the experts. It is coded GS 2004, that is, the 2004 ESE general subject test.

Ought women to have the same rights as men? A hundred years ago, the answer in every country in the world would have been "NO". If you have asked "why not?" you would have been told scornfully and pityingly that women were weaker and less clever than men, and that they had worse characters. Even now, in the twenty first century, there are many countries where women are still treated almost like servants, or even slaves.

It is certainly true that the average woman has weaker muscles than the average man. Thousands of years ago, when men lived in caves and hunted animals for food, strength of body was the most important thing; but now in the twenty first century; brains are more important. Strength of body is still needed for all kinds of works, but the fact that most kinds of work are not well paid shows that the twenty first century doesn't think that muscles are of very great importance.

What about women's brains? Of course, in countries where girls are not given good education as boys they are knowless. But in countries where there is the same education for both, it has been clearly shown there is no difference at all between the brain of the average woman and that of the average man. There have been women judges in Turkey, women ambassadors in America, women ministers in the British government and women university professors in many countries. And among the greatest and strongest rulers of England were Queen Elisabeth and Queen Victoria.

But women can do one thing that men cannot, they can produce children. Because they, and not men do this. They usually love their children more, and better able to look after them since they are more patient and comprehensive towards small children. For this reason many women are happier if they can

stay at home and look after their house and family than if they go out and do the same works as men do. It is their own choice, and not the result of being less clever than men.

All the five experts of the view and belief that the first problem this text (and all other texts from the corpus) has is that it lacks a title. Not providing a text with a title cannot help the readers to predict what they are going to read, and this is likely to render the text comprehension difficult. All of them could indicate that providing a text with a title and subtitles facilitates comprehension. Nevertheless, two experts indicated that it was not a serious problem that the subtitles were not provided because the text is built around a single argument, although it is organized in four distinct paragraphs.

As for the relation between the writer and the audience, most experts appeared to suggest that the writer has a clear idea of the audience in his mind. He/she knows that he/she is writing for high school finalists who have been studying English as a subject only for four years, and whose exposure to the English language is limited to the classroom environment. This can be perceived through the writer's use of clear and explicit style as he/she wants his message to be understood by the test takers.

Nevertheless, I can point to the careless use of the opening sentence "**Ought women to have the same rights as men?**" Although the use of the modal "ought to" is appropriate in this context, it does not reflect the kind of materials students are exposed to in learning situations. One expert admitted that modals such as "ought to" and "needn't to" are part of the curriculum, but most texts in the three textbooks in use do not use them. Therefore, this opening sentence was perceived to be difficult to test takers.

All experts were of the view that the message the writer wants to communicate is clear at the start, but it becomes a bit confusing at the end of the text. Concretely, the text aims to demonstrate that men and women are equal, and this is clearly

shown in the argument contained in the first three paragraphs; but the last paragraph introduces another topic:

But women can do one thing that men cannot, they can produce children. Because they, and not men do this. They usually love their children more, and better able to look after them since they are more patient and comprehensive towards small children. For this reason many women are happier if they can stay at home and look after their house and family than if they go out and do the same works as men do. It is their own choice, and not the result of being less clever than men

All experts wondered whether the text is about the equality of men and women or the superiority of women to men. Therefore, the text does not have one focus, and this is likely to make comprehension difficult, especially to poor and average readers.

As for information load, most experts concurred with my belief that the text requires the readers to retain only short chunks of information in short term memory, and this is good for the test takers. Each paragraph requires the readers to retain one piece of information. Paragraph 1 requires the readers to retain one piece of information: men were believed to be superior to women. Paragraph 2 requires the readers to retain two pieces of information: first, the reason why men were considered superior to women, and then the reason why this view needs to be changed. Paragraph 3 requires the readers to retain one piece of information: examples of what women have achieved. The last paragraph is the most challenging in terms of information load. Readers are required to retain these pieces of information: first, the kind of task women can do, but men cannot; then, the attitude of women (compared to that of men) towards their children; and last, what we have to understand when a woman stays at home to look after her children. Since the last paragraph requires the readers to retain large chunks of information in short term memory, and since the focus of this paragraph overlaps the text purpose, all the experts could agree that this part of the texts was likely to create comprehension problems to the readers.

As for sentence length, the experts were unanimous in their view that the text uses relatively short and simple sentences, with the exception of two sentences that are long and complex. This is the case of this sentence from second paragraph with 33 words:

Thousands of years ago, when men lived in caves and hunted animals for food, strength of body was the most important thing; but now in the twenty first century; brains are more important.

Or again this sentence from the last paragraph with also 33 words:

For this reason many women are happier if they can stay at home and look after their house and family than if they go out and do the same works as men do.

However, although these sentences could be considered as relatively longer than the average sentences, they were perceived as clear and not likely to create problems to an average reader. Therefore, in terms of sentence length, all experts endorsed that the text was well written.

As for lexical density, most experts felt that the text uses not too many words. There is a moderate use of lexis, and there is also a combination of words that readers are likely to infer meaning through context. Nevertheless, experts could not agree whether readers could infer through context the meaning of words such as “ought to”, “scornfully”, “knowless” and “more comprehensive”.

The use of pronouns and their antecedent was also commented by the experts. Although the text uses a limited range of pronouns in an appropriate way, it has been pointed out one example of ambiguous use of pronouns. Compare these two sentences:

(1)...it has been clearly shown there is no difference at all between **the brain** of the average woman and **that** of the average man.

(2) What about women's **brains**? Of course, in countries where **girls** are not given good education as boys **they** are knowless.

If in sentence (1), there is no doubt that the pronoun “**that**” refers to the brain; sentence (2) is ambiguous as to what noun the pronoun “**they**” refers. Some experts reported that the pronoun “**they**” referred to “**brains**” while others reported that it referred to “**girls**”. What is clear here is that the pronoun “they” is likely to mislead readers as it may refer to each of these two antecedents. A close look at this sentence reveals that the reason why some experts could find the pronoun to be ambiguous may be the distance between the pronoun and the word (brain) it relates to.

One point of interest was the use of connectives and transitions to signal different interclausal relationships as well as to create text coherence. The type of semantic relationship between the conditional clause and the main clause as signaled by the conjunction “**if**” is not clearly presented. All experts could agree that conditional clauses create problems of understanding for average and poor readers. This is the case of this sentence from the first paragraph:

If you have asked “why not?” you would have been told scornfully and pityingly that women were weaker and less clever than men, and that they had worse characters.

The use of ‘improbable’ condition in this sentence, coupled with the use of the question “**why not**” in the if-clause, and the unfamiliar lexes “**scornfully/pityingly**” together with the comparative “**weaker and less clever**” makes this paragraph difficult for average readers.

Some experts drew our attention to the use of the conjunction “**but**” in the following sentence:

Thousands of years ago, when men lived in caves and hunted animals for food, strength of body was the most important thing; **but** now in the twenty first century; brains are more important.

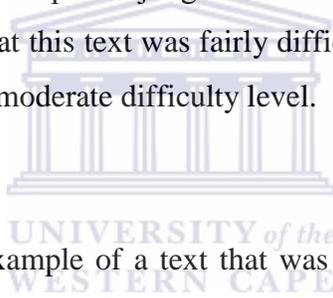
This sentence where the conjunction “**but**” indicates a negative relationship between the two parts of the sentences were reported to be likely to create problems to average readers because of the use of another clause “when men lived in caves” within the two clauses. Most experts agreed that complex sentences like

this one could be a source of many comprehension difficulties in readers. Yet, the following sentence using the same conjunction “but” was reported to be clear for readers:

Strength of body is still needed for all kinds of works, **but** the fact that most kinds of work are not well paid shows that the twenty first century doesn’t think that muscles are of very great importance.

As for the use of transitions, most experts were of the view that the majority of texts in the corpus did not use transitions to create coherence between paragraphs. The use of the conjunction “**But**” to introduce the last paragraph instead of “**However**” to signal a contrasting relationship has been perceived by most experts as misleading to readers.

Nevertheless, both the experts’ judgments and the Flesch reading ease score appeared to indicate that this text was fairly difficult, which can result in the text being seen as a text of moderate difficulty level.



The following is an example of a text that was reported to be easy by both the Flesch reading ease formula and the experts is the following text coded TSD 2007.

When you go to a sewing workshop, mainly a center where young people are trained to become fashion designers, seamstresses and dressmakers, you are likely to be fascinated to see the appropriate equipment. In Mr. Vangu’s workshop there are seven sewing machines. All of them are electric. There are also a dummy, a fashion catalogue, a cutting table, a mirror, an ironing pad, a clothes brush, an electric iron, two tape measures and sewing chalk. Mr. Vangu has five workers: three dressmakers and two tailors. The dressmakers are making skirts, blouse, suits, shorts and hats. The tailors make trousers, shirts, suits for men and shorts for boys. He likes them very much and pays them at the end of the month. The workshop of Mr. Vangu is opened at half past eight and closed at seventeen o’clock. Many people prefer to have their clothes sewed at Mr. Vangu’s workshop.

Like in the preceding text, the title of this text is not provided by the writer. Nevertheless, all the experts could agree that the students can make predictions about what the text is going to be on the basis of the opening sentence: **“When you go to a sewing workshop ...”** As for the audience, it is clear that the writer knows his/her audience which consists of students from dressmaking who are going to become dressmakers or to practice in sewing workshops.

The message the writer wants to convey is clear: how a sewing workshop is organized, and how it functions, who are there, what the different equipments are, and what is made there.

All experts agreed that the information load in the text is moderate. There are not many pieces of information the reader is required to retain in his/her short term memory. In paragraph 1, the reader is required to retain that there are a number of equipments in Mr. Vangu’s workshop. In the second paragraph, the reader is required to retain three pieces of information: the number of workers, what they make, and how Mr. Vangu feels about them. Although there are three pieces of information, all the experts were of the view that the understanding is made easy as these pieces of information are conveyed in short sentences. The last paragraph requires the readers to retain the working hours of the workshop.

All experts could also agree that the text uses sentences which are not long, and this enables readers to comprehend the text. The longest sentence:

When you go to a sewing workshop, mainly a center where young people are trained to become fashion designers, seamstresses and dressmakers, you are likely to be fascinated to see the appropriate equipment

is not difficult to be comprehended by readers since it uses more familiar words and simple structures.

As for lexical density, experts reported that the text uses too limited words and these words are familiar to the readers since they belong to their technical vocabulary.

From the reports made by the experts on the text difficulty of the 7 texts in the corpus, I wish to focus on the following considerations:

(1) The notion of audience is one of the most important factors that accounts for text difficulty. As put by Fulcher (1997), the extent to which the writer has a clear concept of who the text is meant for has an impact on text comprehension. For him, the clearer the notion of audience, the more coherent the text is likely to be, and hence easier to read. Some of the texts in the corpus analyzed did not show a clear sense of audience.

(2) The notion of purpose is also of crucial importance. Some of the texts in the corpus did not have a clear purpose, therefore, the message the writer wished to communicate and the topical focus could become unclear and confusing. Conversely, other texts could have more than one purpose which contradicts each other.

(3) The lack of purpose and topical focus, coupled by the absence of the title and subtitles, was reported to be a serious problem that could not enable average readers to make predictions. Prediction is important in reading as it helps create motivation and arise interest for reading.

(4) In all texts in the corpus, interclausal relationships are expressed simply through juxtaposition of clauses, and no explicit marking was used. Linguistic devices such as ‘therefore’, ‘however’, ‘nonetheless’, as well as conjunctions such as ‘although’, ‘since’ etc... were not used. Yet, as put by Crosson (2008), these devices act as guiding cues that can assist readers’ understanding of how ideas in one clause relate to those in adjacent clauses. An understanding of interclausal relationships is central to reading comprehension (Degand & Sanders, 2002; Graesser et al. 2004; Van Dijk & Kintsch, 1983).

(5) Text writers are often likely to confound linguistic complexity and text readability. It appears that the use of simple and/or simplified linguistic structures aims to make the text readable. Yet, as put by Bernhardt (1991), linguistic

complexity may not be detrimental to comprehension. Rather, syntactic simplicity may decrease text cohesion and thereby hinder comprehension.

To evaluate as to what the extent these conclusions are likely to reflect the test takers' view, I administered these two texts to two groups of high school finalists as part of their final classroom examination. Since I did not want the text questions to influence students' reports on the text difficulty, I asked them to read the text and: (a) write a 100 words paragraph in French explaining what the text is about; and (b), on a four points scale (**1. difficult, 2. Moderately difficult; 3. Easy 4. Can't tell**) select an option that reflects their view on the text difficulty. The following figure presents the text difficulty as perceived by the 2 groups of respondents:

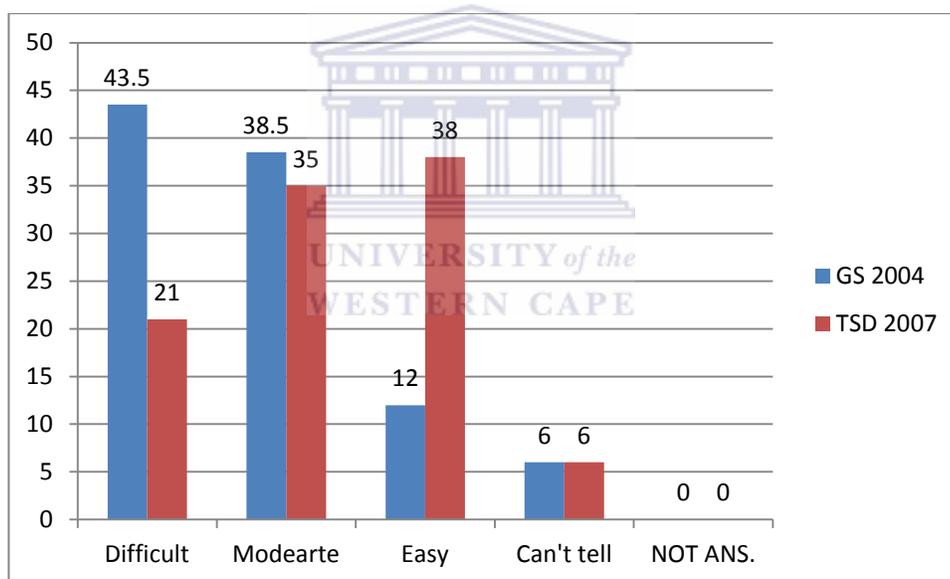


Figure 506: Text difficulty as perceived by respondents

As this chart indicates, views are different within each group and across the two groups. Almost half (43.5 %) of the respondents from general subjects (GS 2004) reported that the text was difficult to read and understand, while only 38.5 % reported that the text was of moderate difficulty level; the conclusion that was reported by the Flesch reading ease score and the 5 experts. Nevertheless, 12 % of respondents in GS 2004 group reported that the text was easy. If these two groups

are considered together (those respondents who have reported that the text is of moderate difficulty [38.5 %] and those who have reported that the text is easy [12.0]), we get a total of 50.5 % (half) of students who could find themselves comfortable with the text.

In technical subject (TSD 2007), only 38 % of respondents found the text to be easy while 35 % found the text to be of relative ease. If these two scores are summed up, we find that 73 % of respondents could find themselves comfortable with the text.

### **5.3. Setting and administration**

This section attempts to answer the third sub-question of our research: *“To what extent are the ESE setting and administration conditions likely to enable the test takers perform the tasks in such a way that scores obtained from the test reflect their actual performance?”*

To answer this question, two variables will be analyzed: the uniformity of administration and the security of the test.

#### **5.3.1. Uniformity of administration**

The question I wish to address for this variable is: *“Is the test administered in the same manner across sites?”*

To answer this question, I conducted an interview with two head inspectors who have been in charge of state examination organization for more than five years. The objective of the interview was twofold: to know if the test is administered

under the same conditions across sites, and to know the challenges that make the uniform test administration difficult.

As for the first issue, all the two inspectors reported that the test was not administrated under the same conditions across sites. One head inspector said:

It is a necessity, and it is desirable to organize the test under the same conditions. This is part of fairness. We have instructions for that; but the reality is quite different. For example, if in towns, like Kananga, we have venues that can accommodate two to three testing centers; this is not the case in rural areas where we are forced to use sometimes churches.

This report illustrates the difficulty to administer the test under the same conditions. This difficulty is due to accommodations in terms of testing venues. But the point related to testing venue was better clarified by another head inspector when she told me:

The big issue is that most testing venues are not in secure conditions; that is, they are not isolated from the public. How can you imagine that a group of people can easily have access to the testing room and be in contact with the test takers? Some testing centers have a security fence and no person from outside can go beyond the fence. Therefore, if in one testing center test takers can work under strictly planned conditions, in other testing centers, test takers arrange to work under conditions that they would like to set up.

A questionnaire administered to stakeholders on this issue appears to corroborate this view. Language teachers, inspectors and other stakeholders admitted that the administration conditions are different across sites. They were asked to select, on a five points scale, the answer that reflects their view on this statement: “Candidates take the state examination in the same conditions”. The following table presents the views of the respondents:

Administrative conditions	Language teachers	Language inspectors	Others
Strongly agree	0.00	0.00	0.00
Agree	0.00	0.00	4.00
Disagree	28.33	60.00	39.20
Strongly disagree	63.33	35.00	51.60
Can't tell	8.33	5.00	4.80
Not answered	0.00	0.00	0.40
Total	100.00	100.00	100.00

Table 512: Respondents' views on the uniformity of administration conditions

As this table indicates, most of our respondents either disagreed or strongly disagreed with the proposition according to which the state examination in general and the ESE in particular is administered in the same conditions. Almost all teachers (91.66 %) either disagree (28.33 %) or strongly disagree (63.3. %) with the proposition. This is the same with language inspectors who, almost in their totality (95.00 %) either disagree (60.00 %) or strongly disagree (35.00 %) with the proposition. Although the group of “others” stakeholders is heterogeneous, almost all of them (90.80 %) either disagree (39.20 %) or strongly disagree (51.60 %) with the proposition.

### 5.3.2. Security

The question I propose for this variable is: “*Is the ESE secure?*”

It is probably not unreasonable to state that the topic of test security can be associated with cheating during test administrations and with disclosure of secure test materials before and after test administrations (Ferrara, 1997).

To investigate the extent to which the ESE is a secure test, two types of data are used: interviews with some stakeholders and a questionnaire submitted to different stakeholders.

(a) Interview

I conducted some interviews with high school finalists before they take the test. These interviews were supposed to help me know whether they had access or possessed some ancient ESE papers or not. Most respondents (8 out of 10) interviewed could admit that they had had access to previous ESE papers and some (4 out of 10) admitted that they did have them in their possession. Asked to give the reason why they had these papers, one of them told me:

You know, we have to look for these papers to be able to know what they ask; and we see how to prepare the test. Sometimes they ask the same questions; and you have such chance to get correct answers.

To the question as to how they got hold of these papers, one of them said:

We usually get them [ESE] from our elders. You know, after I take the test, I pass this to my friend who is in grade 5; and next year after he takes the test he passes it to another one... I will do my best to get out with my test questions; hide it under my trousers.

The result of interviews with language teachers yielded similar results: most language teachers had the ESE papers as part of their teaching materials. One teacher told me:

This is part of my resource library. How can I train students to be successful to the ESE if I can't use these papers as a teaching document? I base my preparation on the papers and I use some test items for classroom assessment.

Interviewed to know if students and teachers are allowed to possess the old ESE papers for classroom teaching and assessment, one of the head inspectors responded in a disapproving manner:

No. No. No one should have access to the test papers. We have a regulation which is clear. When a candidate finishes writing his/her exam, he/she hands in both the answer sheet and the questionnaire papers. But in some testing centers, some candidates make arrangements with some proctors who let them go out with these archives.

These reports from stakeholders indicate that the ESE test is not a secure test since people who are not part of test planning, organization or correction have access to the test archives.

(b) Questionnaire

I administered a questionnaire to stakeholders on the issue that pertains to access to test papers. The following table indicates the respondents' views on whether they could possess some ESE papers or not.

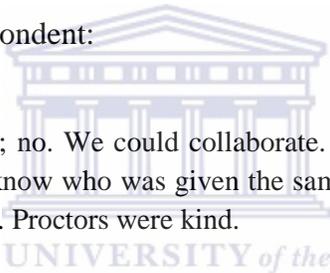
Possession of exam papers	Finalists 1	Finalists 2	Language teachers
Strongly agree	27.25	20.60	20.00
Agree	41.25	33.80	45.00
Disagree	19.75	31.40	20.00
Strongly disagree	9.00	11.80	3.33
Can't tell	2.25	2.40	11.67
Not answered	0.50	0.00	0.00
Total	100.00	100.00	100.00

Table 513: Respondents' views on the issue of possessing previous ESE papers

This table indicates that high school finalists (68.50 % of finalists before the test and 54.4. % of finalists after the test) as well as English language teachers (65.00 %) possess previous ESE papers.

Another issue of great importance that relates to test security pertains to access to test content before the test. We refer here to Ferrara (1997) who uses the concept “cheating” to refer to inappropriately assisting examinees in generating responses to test items during a test administration, and the concept “disclosure” to refer to divulging test topics, questions before test administration.

To investigate the issue of cheating, I interviewed some test takers after they had taken the ESE. The question was to know if they had the opportunity to “cheat” when they were taking the test. To this question, views were shared: some respondents could admit that they could collaborate (passing themselves some answers) while others said they could not. The following is a sequence of interview with one respondent:


 Cheating as such; no. We could collaborate. As soon as we were given the test paper, I tried to know who was given the same test with me. And later we could compare answers. Proctors were kind.

As for the issue of disclosure, a questionnaire administered to stakeholders included a question that asked their views on whether some test takers could have access to test content before they take the test. The following table indicates the views of different respondents on this issue.

Access to test content	Finalists 2	Language teachers	Language inspect.	Others
Strongly agree	2.00	1.67	0.00	6.80
Agree	5.20	5.00	5.00	10.80
Disagree	20.20	38.33	50.00	40.40
Strongly disagree	47.20	45.00	45.00	18.00
Can't tell	8.20	10.00	0.00	17.20
Not answered	17.20	0	0.00	6.80
Total	100.00	100.00	100.00	100.00

Table 514: Respondents’ views on students’ access to test content before test administration

As can be seen from this table, there is some indication that some students have access to the test content prior to the administration of the test. A minority of high school finalists (7.20 %) either agree (5.20 %) or strongly agree (2.00 %) that some students have access to the test content. Likely, 6.67 % of teachers either agree (5.00 %) or strongly agreed (1.67%) with the proposition. Almost the same percentage (5.00) of language inspectors agreed with the proposition. The relatively high percentage (17.60 %) is reported by other stakeholders who either agree (10.80 %) or strongly agree (6.80 %) with the proposition.

In an interview to “AFRIQU’ECHO MAGAZINE” journal of 4 October 2004 (find the interview in appendix 12), the General Inspector of Education, the highest government official whose principal responsibility is to organize the state examination, admitted that cheating and disclosure at the state examination were two phenomena that undermines the credibility of the test. Furnished here is the question he was asked and the answer that he responded with:

Afrique Hebdo Magazine: Quelles sont des mesures prises pour éviter les fuites, la tricherie et la fraude ?

**PLG.** Vous faites mention des fuites, tricherie et fraude lors des épreuves de l’Examen d’Etat ? Les mécanismes que nous avons mis en place sont consubstantiels à l’Examen d’Etat et l’IGE les affinent au fil des sessions. La sériation des items, le mixage inter optionnel dans les salles ainsi, que d’autres procédés en usage procèdent de la logique volontariste d’organiser une évaluation normative, objective garantissant les mêmes chances de réussite à tous les candidats. Comme il est naturel aux élèves les plus faibles à recourir à ces genres de pratiques anti-pédagogiques dues à une maîtrise insuffisante et un déficit des matières enseignées, nous tenons compte de tous ces enjeux car ces épreuves sanctionnent la fin des études secondaires du cycle long. Certains tricheurs organisés ont même développé la forme la plus astucieuse et la plus pernicieuse que représente la substitution des candidats et L’IGE s’y investit en sévissant par l’usage et le contrôle méticuleux des macarons.

Translation:

A.E.M.: What precautions have you taken to avoid disclosure, cheating and fraud at the state examinations?

PLG: You speak of disclosure, cheating and fraud during state examination sessions? We have set up substantial mechanisms at the state examinations and the General Inspection of Education tries to improve them in different sessions. The use of test series (parallel forms), the mixing of candidates from different subjects in the testing rooms, and other procedures used aim to come out with an objective norm-referenced evaluation where all candidates have equal chance of success. As it is common to weak students to recourse to these unethical behavior; and this is due to the insufficient mastery of materials taught. We usually take into account all these factors as these are high school exit tests. Some cheaters have even organized one form of cheating which is very harmful: the substitution of candidates and the General Inspection of Education is working on this to severely punish this by a strict control of candidates' access cards.

This excerpt from the interview with the highest educational official provides evidence for cheating, disclosure and other unethical practices that undermine the evaluation system of the DR Congo state examination in general and ESE in particular.

### Conclusion

On the basis of the content analysis and descriptive statistics, this chapter has analyzed the research data with a focus on the three research sub-questions. Although the main data were made of different ESE papers, other data were used to triangulate the results; these are the questionnaires, the tests and the interviews. The next chapter reports the findings and evaluates the ESE.

## **CHAPTER 6 FINDINGS AND EVALUATION OF THE DR CONGO CONTEXT VALIDITY**

This chapter answers the main research question as well as the research sub-questions. It also presents, on the basis of study findings, the final evaluation of the context validity of the DR Congo English state examination on the basis of study findings.

### **6.1. Findings**

This section presents the study findings on the basis the analyzed data.

#### **6.1.1. Task setting**

Let us address the research sub-question that pertains to this variable: “To what extent are the conditions under which the test tasks are performed relevant to test domain and test characteristics?” To answer to this question, I need to investigate the findings that pertain to the following variables:

##### **6.1.1.1 ESE rubrics**

On the basis of the content document analysis of the ESE rubric, the following two findings assume particular relevance:

(a) Both the general rubrics (those that relate to all subtests of the national test: test organization, time allocation, and test instructions) and specific rubrics (those that relate to the English state examination) are clear, intelligible, comprehensive and explicit so as to enable the test takers perform the test task with much confidence;

(b) While the specific rubrics are brief (“Read the text carefully and answer the questions based on it”, “select the appropriate correct option from a list of five options numbered 1, 2, 3, 4, 5” and “If there is no correct option from the five suggested, write 6”), the general rubrics on the front page of the test paper are

very long; and this may result in some candidates not reading all the instructions completely.

The conclusion to draw from these findings is that the ESE generally provides the test takers with a good rubric that is likely to enable them have a clear indication of what the examiner is asking. Empirical research on test rubrics has indicated the contribution of good rubrics towards improvement in test performance (Reddy, 2007).

#### **6.1.1.2. Task purpose**

From the analysis of the questionnaires administered to respondents, the main findings worth reporting for this variable are:

(a) Except language inspectors, most stakeholders appear to believe that the task purpose as stated by the instruction “Read the text carefully and answer questions based on it” is not clear; it is equivocal since it does not point to a precise (or specific) cognitive process test takers have to deploy while completing the task. Besides, this instruction is stated after the test takers have deployed different cognitive strategies to read the text passage. I think that some test takers might ignore it and straightly proceed to read the questions as a result of test pressure. Therefore, I believe that test takers are not likely to deploy appropriate cognitive strategies to read the text and answer questions based on it; and this may negatively impact on their performance to the test. This conclusion replicates the findings of Van Yperen (2003) who could find a positive relationship between test task purpose and actual performance; and the findings of Pena & Quinn (1997) who reported that students perform significantly better on the familiar test task and that familiar task is sensitive to differentiating good and low language ability students.

(b) Language inspectors generally believe that test purpose is clearly stated in the test rubrics. This finding is in contradiction with what I have just reported here. I believe that language inspectors, since they are actively involved in the different

steps of the national assessment, may have simply thought that the test constructor has done what he/she is supposed to do, that is, to provide clear instructions on task processing.

#### 6.1.1.3. **Known criteria**

This variable presents these mixed findings:

(a) Regarding the scoring criterion, most respondents agree to know that each question is marked out of 1 point; and 1 point equals 1 %”;

(b) As for combination of test scores with other assessment components, while most language inspectors and the majority of teachers either strongly agreed or agreed that all subtest marks for each candidate are summed up with the practical test marks and the French “dissertation” marks to make up the total mark for each candidate, only a minority of test takers reported to know that the scores they will get will be the combination of scores from three components: the test proper, the practical test and the French dissertation;

(c) As regards the moderation of test scores with classroom assessment scores, this criterion presents results that are also mixed. Language teachers in their majority and all the English language inspectors reported to know the moderation process. However, only almost a half of the candidates before the test reported to know how their scores are moderated to get the final grade; and few candidates after the test had the knowledge of this criterion. The heterogeneous group “Other stakeholders” reported results that are mixed between those who had knowledge of the criterion and those who did not know this moderation criterion.

These appear to confirm that test takers do not fully have a clear and unequivocal idea of the criteria on the basis of which they are assessed; therefore, they are likely not to perform better to the test. This assumption aligns with Lenney *et al.* (1983)’s findings as they could find that students who were not given clear and

unambiguous criteria on how they would be assessed could perform badly compared to those who were given clear and unambiguous criteria.

#### 6.1.1.4. Order of items

The major findings that appear to relate to this variable are:

(a) The majority of language inspectors and half of language teachers believe that the ESE items are not grouped in testlets or ordered in a logical order;

(b) A good portion of teachers (21.67 %) and language inspectors “can’t tell” whether the test items flow in a certain logical order or not. We believe that this can be justified by the fact that to be able to answer this question requires certain knowledge of distinguishing between different skills that make up reading as well as to be capable to tell which item tests which skill; and this might be difficult to some respondents to tell.

The conclusion to draw from these findings is that test takers are unlikely to carry out reading in a linear and incremental fashion as a picture of the whole text and this is likely to impact on their performance on test. This conclusion appears to tally with the research findings reported in Plake *et al.* (1983), which points to a positive relationship between items arrangement (that they defined as the physical location of test questions) and impact on test performance; and in the research findings of Gohmann & Spector (1989) who reported that students perform better on a content-ordered exam as they have the possibility to glean information about one question from previous questions and might concentrate better if they do not have to jump from one topic to another.

#### 6.1.1.5. Time allocation

On the basis of the literature, and from results obtained to the test and taking into account different views from stakeholders, the following findings may be reported:

(a) The time allocated to the completion of the ESE is more than necessary. The results from the tests appear to confirm that the ESE can be performed in less time than provided now.

(b) Almost all student respondents (96.00 %) think that the time allocated to write the English test is either enough or not enough.

(c) Only few teachers (20.00 %) and inspectors (35.00 %) admitted that time allocated to test completion is too much.

The conclusion to draw from these findings is that there is evidence that allocating more time does not mean that examinees use all allocated time. This conclusion replicates study findings by Mandinach *et al.* (2002, 2005) who state that many test takers who had been granted extended time could leave at the same time as the standard test takers, and by Bridgeman *et al.* (2003, 2004) who concluded that the provision of extra time to students who do not have effective solution strategies will not prove beneficial; and other studies which reported the same conclusions (Huesman & Frisbie, 2000; Runyan, 1991).

I understand the different point of view of some stakeholders when I consider Bridgeman *et al.* (2004, p. 1)'s view who reported that there is "a common belief among test takers and their families (and even among some school counselors) that giving examinees more time to complete a test could substantially improve their scores".

## **6.1.2. Task demands**

Let us recall the research sub-question: *“To what extent does the ESE include tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task?”*

### **6.1.2.1. Text types**

From the content analysis of the corpus of test papers, the following findings are reported:

(a) The ESE uses all the three text types (exposition, narration and argumentation) along with practical writing in technical subjects;

(b) In general education (subjects), almost half (47.30 %) of texts that are used are narratives, then come expository texts that account for 37.80 % and argumentative text come last with only 14.90 %;

(c) In technical subjects, on the contrary, expository texts are the most used (with 58.60 %), then come narrative texts with 31.30 %; and practical writing (especially in commercial and related subjects 7.90 %) and argumentative texts are the least used with only 2.20 %;

(d) For narrative texts for example, the content analysis of the test papers has demonstrated that most items in this ESE sample test do not follow the story line of the text; besides, some questions do not call for operations test takers are likely to use to construct text meaning. Yet, there is evidence that narrative texts typically have a hierarchical structure, and that readers are sensitive to such structure so that when this structure is used to guide comprehension and recall, both are facilitated and test performance is enhanced (Shin, 2002; Glenn, 1978; Carrell, 1985).

### **6.1.2.2. Text length**

The content analysis of the length of the texts used in ESE papers has revealed the following:

(a) There are only single texts that are used and the length of the ESE texts varies across subjects (general or technical subjects). Texts are longer in general subjects (between 200 and 500 words) than they are in technical subjects (no text over 200 words);

(b) Compared to texts used for teaching, the ESE uses texts that are far shorter than those students use in classroom instruction;

A conclusion I might draw from these findings is that since the majority of ESE texts are between 100 and 300 words, the ESE uses texts that are not appropriate for the target situation requirements of the students being tested. Not only short texts cannot measure test taker's different language abilities, but also they hardly allow the test writer to write the number and type of questions he/she needs. This conclusion replicates Buck (2001)'s findings when she reported that shorter texts tend to focus more on localized grammatical characteristics than discourse skills.

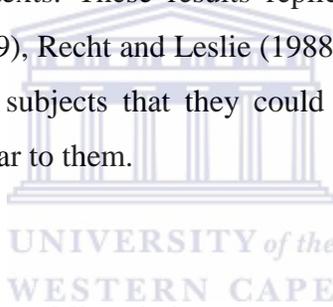
### **6.1.2.3. Text content**

The content analysis of the ESE papers reveals the following findings:

(a) The majority of texts (65. %) are those texts that include general topics and these texts are likely to enable the test takers to deploy their background knowledge to read and understand them. The content of these texts is authentic and familiar to all the candidates in that they recourse to their schemata to process and comprehend them. These results replicate studies by Bernhardt (1991), Urquhart and Weir (1998) and Alderson and Urquhart (1985) who reported that familiar texts enable test takers to deploy their schemata and perform well in the test.

(b) Those texts that include general topics, but that are likely to advantage one group of candidates on behalf of another group of candidates. In this group of texts, students belonging to the advantaged group could perform well to the test as they could deploy their schemata as the text contained information that was familiar to them. These results replicate earlier studies by Alvermann and Hynd (1989), Chen and Graves (1995), Ridgway (1997), Birjandi *et al.* (2000) and Krekeler (2006) who reported the bias of some tests as their content could advantage one group of test takers on behalf of other test takers.

(3) Those texts that do not relate to the candidates' background knowledge. This group included texts that were unfamiliar to respondents. Their performance to the test was poor and we think that they could not activate their schemata to read and comprehend the texts. These results replicate similar findings by Dochy, Segers and Buehl (1999), Recht and Leslie (1988) and Kendeou and Broek (2007) who found with their subjects that they could poorly perform on texts whose contents were unfamiliar to them.



#### **6.1.2.4. Text difficulty and linguistic structure**

Regarding the text difficulty and linguistic structure, the following findings are reported:

(a) The Flesch Reading Ease Formula as well as the experts could indicate that all the texts in the corpus were either easy or of moderate difficulty level.

(b) The majority of respondents from general subjects reported that the text was difficult to read and understand, while only few respondents reported that the text was of moderate difficulty level, the conclusion that was reported by the Flesch reading ease score and the 5 experts. On the contrary, in technical subjects, the majority of respondents could feel themselves comfortable with the text.

(c) All the five experts agreed that the first problem with the texts from the corpus is that they lack a title. They agreed that not providing a text with a title does not

help the readers to predict what they are going to read, and this is likely to render text comprehension difficult. All of them could indicate that providing a text with a title and subtitles facilitates comprehension.

(d) As for sentence length, all the experts unanimously agreed that the texts from the corpus use relatively short and simple sentences, although some cases of complex sentences could be noted. For complex sentences, most experts agreed that these sentences might be a source of many comprehension difficulties in readers.

(e) As for lexical density, most experts could agree that most texts from the corpus used not too many words. There is a moderate use of lexis, and that readers are likely to infer meaning through context.

(f) As for the use of transitions, most experts could agree that the majority of texts in the corpus did not use transitions to create coherence between paragraphs. Also, the use of some conjunctions (like but, however, although) to signal a contrasting relationship has been perceived by most experts as misleading to readers.

(g) Most texts from the corpus were truncated, that is, they were simplified with the objective to facilitate comprehension. This is the reason why in all texts in the corpus, interclausal relationships were expressed simply through juxtaposition of clauses, and no explicit marking was used.

### **6.1.3. Task setting and administration**

The research sub-question for this variable is: *“To what extent are the setting and administration conditions of the ESE likely to enable the test takers perform the tasks in such a way that scores obtained from the test can reflect their actual performance?”*

The analysis of the different data reveals the following findings:

(a) Language teachers, inspectors and other stakeholders admitted that the ESE is not administered under the same conditions across different sites as well as across different testing rooms. This difficulty is due to accommodations in terms of testing venues.

(b) Most respondents indicated that the ESE test is not a secure test since people who are not part of test planning, organization or correction have access to the test archives.

(c) There is some indication that some students have access to the test content before the test administration.

(d) Most respondents acknowledged that there are cases of cheating and collaboration.

These findings appear to confirm that the conditions under which the ESE is administered can pose a serious threat to the validity of this test. I agree with Ferrara (1997) when he argues that cheating and disclosure of secure test materials poses a real threat to the viability of high stake tests and to the integrity of scores produced by these tests; and Cizek (2001) who believes that cheating violates the rules of test administration as it gives an examinee an unfair advantage over other examinees, and this decreases the accuracy of the intended inferences arising from the examinee's test scores or performance.

## **6.2. Evaluating the context validity of the ESE**

This section summarizes the evaluation of the ESE as analyzed in this study. Its aim is to decide, on the basis of analyzed data and reported findings, to what extent the context of the ESE is valid.

To make such decision, I need to revisit the evaluation of discrete variables that have been discussed in the previous section. These variables have been grouped in

three broad categories: task setting, task demands and task administration. Task setting has consisted of the evaluation of the ESE rubric, purpose, criteria, order of items and time allocation. Task demands consisted of the evaluation of ESE text types, text length, text content and text difficulty and task administration has focused on two variables: uniformity of administration and test security. Table 515 summarizes the qualitative and quantitative evaluation of ESE with regards to specific categories and variables.

In this table, the context validity of the ESE is evaluated in relation to three basic components: task setting, task demands and task administration. These three components are included in the first row of the table. Each component is evaluated on the basis of different discrete variables that operationalize it; and these variables are listed under the component heading (in the first row). For instance, the component task setting is operationalized by five variables: test rubric, purpose and criteria, the order of items, and the time allocation. For each discrete variable, quality criteria that operationalize the variable are listed in the second row. For instance, the variable ESE rubric is operationalized by how well the test rubrics are clear, intelligible, explicit, comprehensive, and brief. The third column makes the qualitative evaluation of each discrete variable. This evaluation is made on the basis of results from analyzed data. This qualitative evaluation is done on a five-scale likert where these values are used: very good, good, acceptable, poor and very poor. The quantitative evaluation is used (in the fourth row) to help give the concrete picture of the quality assessment. Therefore, “very good” quality is assigned 5 points and “poor” quality is assigned 1 point. Column five gives the averaged score (over 10) for each component as well as each variable. Finally column six reports the general (global) score (over 100) for each component. Therefore, a global score above 80 indicates very good (very strong) context validity, between 70 and 80 indicates good (strong) validity, between 50 and 70 indicates acceptable validity, between 40 and 50 indicate poor (weak) validity and below 40 indicates very poor validity.

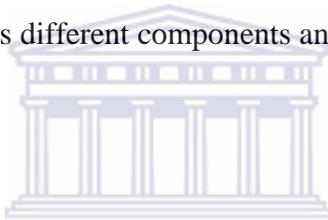
Variables	Characteristics	Evaluation- Ql.	Evaluation-Qt.	AVERAGE/10	TOTAL/100
<b>TASK SETTING</b>					
1. ESE rubric	1. Clarity	very good	5		
	2. Intelligibility	acceptable	3		
	3. Comprehensiveness	acceptable	3		
	4. Explicitness	acceptable	3		
	5. Brevity	good	4		
	<b>AVERAGE</b>	<b>GOOD</b>	18/25= 7.2/10	7.2	
2. ESE purpose	1. Clarity	poor	2	4	
3. ESE criteria	1. Scoring criterion	very good	5		
	2. Combination of scores	acceptable	3		
	3. Moderation of scores	acceptable	3		
	<b>AVERAGE</b>	<b>GOOD</b>	11/15=7.3/10	7.3	
4. Order of items	1. Item ordering	very poor	1		
	2. testlets	very poor	1		
	<b>AVERAGE</b>	<b>VERY POOR</b>	2/10	2	
5. Time allocation	1. Time appropriateness	poor	2	4	
<b>TOTAL Task setting</b>		<b>ACCEPTABLE</b>		<b>24.5</b>	<b>49</b>
<b>TASK DEMANDS</b>					
6. Text types	1. Appropriateness/variety	very good	5		
	2. Relation to operations	poor	2		
	<b>AVERAGE</b>	<b>good</b>	7	7	
7. Text length	1. appropriateness	poor	2	4	
8. Text content	1. Appropriateness to TL	good	4		
	2. Familiarity	acceptable	3		
	<b>AVERAGE</b>	<b>good</b>	7	7	
9. Text difficulty	1. Reading ease	good	4		
	2. Linguistic structure	acceptable	3		
	<b>AVERAGE</b>		7	7	
<b>TOTAL Task demands</b>		<b>GOOD</b>		<b>25/40</b>	<b>62.5</b>
<b>SETTING &amp; ADMINISTRATION</b>					
10. Uniformity	1. Uniformity across sites	poor	2	4	
11. Security	1. ESE papers security	poor	2		
	2. Cheating	poor	2		
	3. Disclosure	poor	2		
	<b>AVERAGE/SECURITY</b>		6/15	4	
<b>TOTAL setting &amp; Adm.</b>		<b>POOR</b>		<b>4</b>	<b>40</b>
<b>TOTAL CONTEXT VALIDITY</b>		<b>ACCEPTABLE</b>			<b>50.3</b>

Table 515: Quantitative and qualitative evaluation of ESE context validity

On the basis of results reported on this table, I wish to make the following conclusion, which can answer the research question: Generally, the ESE has **ACCEPTABLE** context validity. This means that the ESE includes tasks that **fairly** profile the kind of tasks the test takers have been performing in the classroom and/or are likely to perform in their normal life. However, some components have high/better/strong context validity than others. Task demands component presents “**good**” validity (62.5/100), task setting presents “**acceptable**” (49/100) validity, whereas task setting and administration presents “**poor/weak**” validity (40).

However, this conclusion appears to colour the specificity of findings from each component as well as those from each variable. The following lines summarize the findings on the basis different components and variables investigated.

(1) Task setting



This component is generally judged “**ACCEPTABLE**”. Therefore, as the answer to the first research sub-question, we can say that the conditions under which the test tasks are performed are **FAIRLY** relevant to test domain and test characteristics.

However, the evaluation of the different variables that make up this component points to mixed results. Of the five variables that make up this component, two (ESE rubrics and criteria) have been qualified as having “good” validity (7.2 and 7.3), two variables (ESE purpose and time allocation) have been qualified as having “poor” validity (4, and 4) and one variable (order of items) has been qualified as having “very poor” validity (2). The following figure can help illustrate this classification:

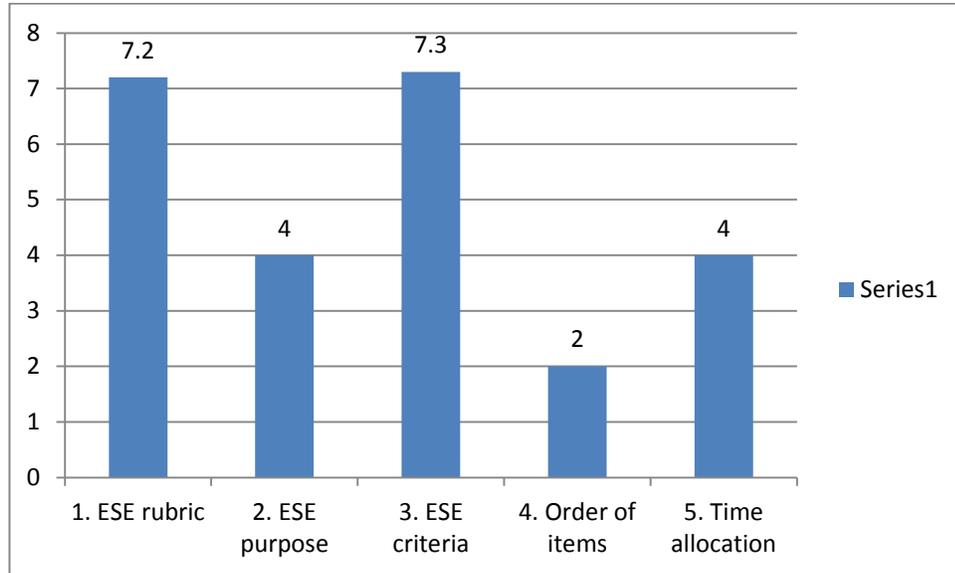


Figure 507: Quantitative evaluation of the ESE task setting

Indeed, within each variable, the evaluation of different characteristics presents some variations. For instance, the ESE rubrics are reported to be of “very good” clarity; but their intelligibility, comprehensiveness and explicitness is reported to be of “acceptable” quality. Likely, the degree to which high school finalists know the evaluation criteria prior to examination varies: data indicate the test takers’ “very good” knowledge of the scoring criterion, while the criteria related to combination and moderation of scores are reported to have an “acceptable” level of knowledge by the test takers.

## (2) Task demands

This component is the best evaluated component (compared to the two others) in regards to the context validity. Therefore, as the answer to the second research sub-question, I can say that the ESE globally includes **GOOD** tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task.

This component is operationalized by four micro-variables that present results characterized by two trends: the first trend includes three variables (text types, text content and text difficulty) that are reported to be of “good” quality (7) whereas the second trend includes one variable (text length) that is reported to be of “poor” quality (4). The following figure can help illustrate the degree of validity of the four variables

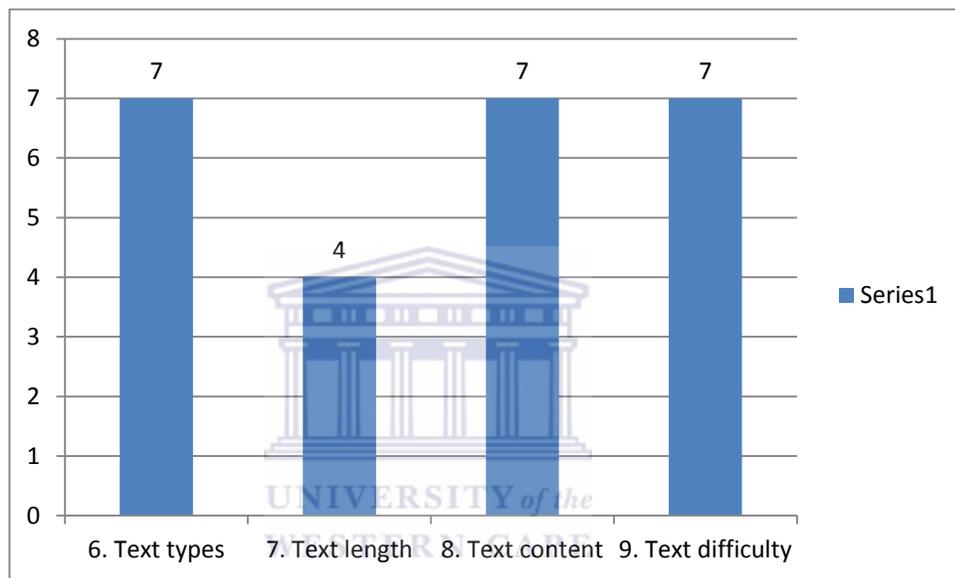


Figure 508: Quantitative evaluation of the ESE task demands.

### (3) Task setting and administration

This component is the worst evaluated of the three components. Therefore, as the answer to the third research sub-question, I can say that the setting and administration conditions of the ESE **ARE NOT** likely to enable the test takers perform the tasks in such a way that scores obtained from the test reflect their actual performance.

This component is operationalized by two variables that are all poorly evaluated: the uniformity of administration and test security.

As a conclusion, the ESE context validity is of an “**acceptable**” degree. It is undermined by (a) the insufficient and inadequate conditions under which the test tasks are performed with regard to their relevance to test domain and test characteristics; and the administration conditions that are not likely to enable the test takers perform the tasks in such a way that scores obtained from the test reflect their actual performance. Nevertheless, there is an effort on the part of the test developer with the inclusion of tasks that profile the linguistic and background knowledge of the test takers as well as those that include varieties of text types and genres that reflect the classroom tasks.



## **CHAPTER 7 CONCLUSION, SUGGESTIONS FOR TEST DEVELOPMENT, LIMITATIONS AND FURTHER RESEARCH**

This chapter concludes this study by summarizing the research findings, offering suggestions for the improvement of the quality of the English state examination, highlighting the limitations of the study and suggesting agendas further research.

### **7.1. Conclusion**

In this concluding section, I propose to present the conclusions that can be drawn from the findings of this study. The objective of this study was to investigate the extent to which the context of the DR Congo English examination is valid so as to spell out some useful suggestions that are likely to improve its validity. The test context is very important in language testing since it provides an understanding of the kind of tasks to be included in the test, how these tasks should be executed by the test takers and how these tasks should be efficiently administered.

The main research question in his study is: “To what extent is the DR Congo ESE valid?” In light of this, I posed three research sub-questions with a view to attempting an informed discussion that could help answer the main research question. They are: (a) To what extent are the conditions under which the test tasks are performed relevant to test domain and test characteristics? (b) To what extent does the ESE include tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task? (c) To what extent are the setting and administration conditions of the ESE likely to enable the test takers perform the tasks in such a way that scores obtained from the test reflect their actual performance?

### 7.1.1. Task setting

With regards to the first sub-question, results have indicated that the extent to which the condition under which the test tasks are performed and their relevance to the test domain and characteristics is judged acceptable. This is a general judgement that eludes particularities of different variables investigated within this macro-variable. More concretely, the instructions given to the test takers in the rubric and the criteria of evaluation variables were evaluated as having “good validity”. To be explicit, the ESE has good rubrics that are clear, intelligible, comprehensive, and explicit for the candidates. However, findings have indicated that if the specific rubric is brief, the general rubric on the front page of the ESE paper is very long; and this may result in some candidates not reading it completely. Also, both the finalist students and other stakeholders have generally the knowledge of the criteria on the basis of which test takers are evaluated at the national test such as the combination of test scores and classroom assessment scores as well as the moderation of test scores with classroom assessment scores.

Two variables (ESE purpose and time allocation) have been qualified as having “poor” validity. Results have indicated that test takers are not given a clear and unequivocal idea in the rubric of what the requirements of the task are so that they can choose the most appropriate strategies and determine what information they are to target in text comprehension and task processing. I am encouraged to hold with Weir (2005) when he argues that the purpose of a reading activity will determine the operations to be performed on it. The more the test taker is aware of it, the likely he/she will select appropriate operations to read the text and process tasks; hence, better perform to the test. The findings have also indicated that the time allocated to the completion of the ESE test is more than needed. Yet, as Bridgeman *et al.* (2004) and Mandinach *et al.* (2005) argue, if too much time is given, test takers will not likely use the extra time to perform the task, and this will impact on their performance.

One variable (order of items) has been qualified as having “very poor” validity. The findings have indicated that ESE test questions do not follow a serial order as evidence suggests that this is the way readers construct meaning. I believe with Weir (2005) when he argues that careful reading is normally carried out in a linear, incremental fashion as a picture of the whole text, and its macro-propositions are built up serially through an understanding of the micro-propositions in the first sentence, then the next, and so on.

As a conclusion to this section, there is evidence to claim that the conditions under which the ESE tasks are performed and the relevance of these tasks to the test domain and characteristics are still far to contribute to the quality of evaluation of high school finalist students.

#### 7.1.2. Task demands

With regards to the second research sub-question, results have appeared to indicate that the extent to which the ESE includes tasks that take into consideration the nature of information in the text as well as the knowledge required for completing the task is judged “good”. Nevertheless, the four variables that operationalize this macro-variable present results that are characterized by two trends: the first trend includes three variables (text types, text content and text difficulty) that are reported to be of “good” quality, whereas the second trend includes one variable (text length) that is reported to be of “poor” quality.

The data appeared to indicate that the ESE uses reading texts across a range of topics and fields, and different types of text materials such as description, comparison and contrast, explanations, argumentation, narration, etc. as used as these are the kind of materials that are either used for classroom teaching or that learners are expected to meet in their future (real) life. However, as for narrative text, the findings have demonstrated that most items in the ESE do not follow the story line of the text; besides, some questions do not call for operations test takers are likely to use to construct text meaning. Yet, there is evidence that narrative

texts typically have a hierarchical structure, and that readers are sensitive to such structure so that when this structure is used to guide comprehension and recall, both are facilitated and test performance is enhanced (Shin, 2002; Glenn, 1978; Carrell, 1985).

With regards to test content, the findings appeared to indicate that the content of ESE reveals three basis patterns: (1) texts that include general topics and these texts are likely to enable the test takers to deploy their background knowledge to read the text and understand it. Such texts with such content have been qualified authentic and familiar to all the candidates in that they could recourse to their schemata to read the text and understand it. (2) Texts that include general topics, but that are likely to advantage one group of candidates on behalf of another group of candidates. Such texts have been judged inappropriate since they are biased. Finally, (3) texts that do not relate to the candidates' background knowledge. These texts have been qualified to be bad for being included in a test.

As for text difficulty, the study findings have indicated that the difficulty level of texts used at the ESE is generally at the level of an “average” reader. However, it has been demonstrated that texts in general subjects are generally perceived to be more difficult than texts in technical subjects.

The study findings have revealed that test length variable had poor validity. The ESE use single texts that are shorter than those used in classroom instruction and it has been concluded that these texts are not appropriate for the target situation requirements of the students being tested. Not only short texts cannot measure test taker's different language abilities, but also they hardly allow the test writer to write the number and type of questions he/she needs.

### 7.1.3. Setting and administration

As for the third and last research sub-question, study findings have shown that this variable has “poor validity” as the circumstances under which the test takes place affect the validity of scores obtained by test takers. Findings have indicated that the ESE is not administered under the same conditions for all test takers

across different sites. Also, the test is not secured in that the findings have indicated cases of cheating and collaboration during the test session. Besides, the test papers archives are not secure as they are available to people who are not part of the evaluation system.

## **7.2. Suggestions for test development**

This section integrates the literature on testing and the study findings into the DR Congo testing context and provides useful suggestions that are likely to improve the context validity of the ESE.

The ESE constructors should ensure that test takers are given a clear and unequivocal idea on the rubrics of what the requirements of the task are so that they can choose the most appropriate strategies and determine what information they are to target in text comprehension and task processing. The instruction that is used (“*Read the above text carefully two or three times and then answer the questions based on it*”) is not sufficient to provide a clear idea of the task purpose. Test constructors should add other precise and specific instructions that point to specific cognitive processes test takers actually deploy to find discrete information in the text. For instance, instructions like “(1) read the first sentence/paragraph and make a mental note of what the text is about, (2) consider your prior knowledge while reading, (3) highlight and/or annotate important words, phrases and sentences, (4) skip unknown words while reading, (5) translate some important words, phrases or sentences in French/your local language, (6) start to answer questions that are easy; (7) Do not circle two options in one item etc.” may be useful instructions that may provide clarity to text comprehension and test performance.

The educational services should ensure that they provide the students with a clear idea of how they will be evaluated and the criteria by which their final marks will be obtained. This information should be available to them and their teachers prior to examination. By doing this, candidates will be fully aware of the criteria by

which they will be assessed, this will likely have a positive effect on both task planning and execution, and on their performance in the test.

Test developer should ensure that test items are ordered so as to reflect the way test takers' skills and strategies are deployed in normal processing when they read the texts in classroom setting or for enjoyment. In this perspective, the test items must be ordered in a linear sequencing and items requesting comprehending explicit information, those requiring comprehending implicit information, those requiring inference etc. are grouped together respectively. This reflects the argument developed by Kintsch (1998) and Urquhart & Weir (1998) who argue that test questions must follow a serial order as evidence suggests that this is the way readers construct meaning, that is, incrementally.

Test developers should ensure that there is **enough** time to complete the task, as put by Weir (2005), Bridgeman *et al.* (2004) and Mandinach *et al.* (2005). I believe that too much time is given to test takers to perform the test task and the test takers do not use the extra time to perform the task. I suggest that either the number of test tasks is increased (number of test questions, for instance) or the time allocated to the test is reduced.

Since written discourse serves many functions such as to inform, entertain, or persuade, I suggest that the ESE focuses on reading texts across a range of topics and fields, and different types of text materials such as description, comparison and contrast, explanations, argumentation, narration, etc. must be used as these are the kind of materials that are either used for classroom teaching or in normal life situations. As observed by Weir (2005), the text should be taken from the target language use situation or possess salient characteristics of target language use texts. In this perspective, as articulated by Urquhart and Weir (1998) and Alderson (2000), there should be a relationship between the text type and the operations used by test takers to process tasks.

The length of the ESE texts should be of a considerable length so as to reflect the length of classroom texts used in the four textbooks in current use. This will

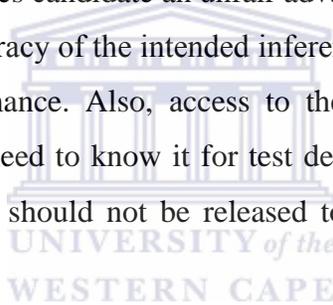
enable to construct as many items as needed and measure different reading abilities, including discourse abilities. Another alternative is that multiple short texts may also be used as they provide the opportunity to avoid bias.

There should be a relationship between text content and the test taker's background knowledge and subject matter knowledge (Douglas, 2000). This interaction helps to make inferences about a test taker's capacity to use language appropriately. Good tests of reading will ensure that candidates have been assessed for their ability to understand the text. Thus, test developers must take into account background knowledge of the test takers. However, since candidates of different related subjects are grouped and submitted the same test, test writers should ensure that the selection of text content is not likely to advantage one group on behalf of another; thus biasing the test results. Also, the test developer should ensure to select a content that is not too familiar so as to allow the candidate to answer some or all the questions without necessarily reading the text. Unfamiliar texts should be avoided as they do not permit test takers to activate their schemata to read and comprehend them.

The texts on the basis of which the test is based should be accessible to the test takers. That is to say, the language of the texts must be highly relevant. The lexical items and grammatical structure must be appropriate for the level of the candidates. Also, the test takers must be able to understand the interclausal relationships in order to create meaningful and coherent relations across the text. The test constructors should provide a title to the text. Not providing a text with a title is likely to render comprehension difficult as the test takers will fail to predict what they are going to read, and this does not reflect the way students read texts in classroom situations or in daily life. Test developers should use texts that are natural in their organization; that is, they should avoid truncated texts that do not use transitions and connectives to create coherence between paragraphs. Whenever transitions and connectives are used, test developers should ensure that these connectives and transitions properly signal different interclausal relationships and create text coherence in a way not to create ambiguity in test

takers. Text writers should avoid confounding linguistic complexity and text readability. Although the use of simple and/or simplified linguistic structures aims to make the text readable, linguistic complexity may not be detrimental to comprehension. Rather, syntactic simplicity may decrease text cohesion and thereby hinder comprehension. As stated by Mehrpour & Riazi (2006), Oh (2000) and Yano *et al.* (1994), input should be modified in the direction of elaboration rather than by artificial simplification.

As regards the issue related to test administration, all the people involved in the administration of the test should adhere to high ethical standards. Cheating and disclosure of secure test materials poses a real threat to the validity of the test and to the integrity of scores produced by the test. Cheating violates the rules of test administration as it gives candidate an unfair advantage over other candidates, and this decreases the accuracy of the intended inferences arising from the examinee's test scores or performance. Also, access to the content of the test should be limited to those who need to know it for test development, test scoring, and test evaluation. Test items should not be released to be accessible to test takers or teachers.



### **7.3. Limitations**

This study is far from perfect given its context and setting. Among its many limitations, I wish to mention the following:

- The number of test papers actually analyzed was reduced, although the corpus selected was fairly large. This limits the validity of my study findings and calls for the need to validate these findings in a major study;
- Although the ESE is a national test, I selected the respondents from one province; yet the DR Congo has eleven provinces and I think that the responses from different questionnaires might have reflected the specific local realities of the respondents;
- Since I could not interview test constructors, it has been difficult to understand the reason why some of the problems reported in this study could not be

overcome. The national context of the test might have provided me with useful insights to my argument.

#### **7.4. Further research**

The validation of a language test is a hard process that one cannot pretend to complete in a single study and within an academic study time limit. It is in fact a program that should engage all stakeholders as well as testing experts and researchers. Evaluating one aspect of test validity evidence might prove ineffective if the other aspects of test validity are not evaluated. As suggested by Weir (2005), to validate a language test, one needs to build an argument on all the five aspects of validity evidence that are: (1) context validity, (2) theory-based validity, (3) scoring validity, (4) consequential validity and (5) criterion-referenced validity. This study has established the first aspect of validity evidence; and I am going to establish the next two types of validity evidence in my further research. My objective will be to evaluate the theory-based validity evidence and the scoring validity evidence of the ESE by demonstrating (a) to what extent the ESE uses tasks that reflect the cognitive skills and strategies high school finalists are likely to deploy in reading the text and in providing answers to questions based on the text; and (b) to what extent the ESE produces scores that reflect high school finalists' level of knowledge and skills in English. Therefore, I expect other researchers to explore the remaining two aspects of validity: (a) consequential validity (the extent to which the ESE influences positively or negatively the language teachers in their use of teaching and testing methods and materials, the 6<sup>th</sup> grade students learning strategies, the course content, etc.), and (b) criterion-referenced validity (the extent to which the ESE scores correlate to other measures of the same construct it measures). It is only by investigating all these other aspects that we can expect to come out with an entire validity evidence of the DR Congo English state examination.

## REFERENCE LIST

- AERA, APA, & NCME (2011). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Alderson, C. J., & Banerjee, J. 2002. Language testing and assessment (Part 2). *Language Teaching*, 35, 79-113.
- Alderson, C.J. 2000. *Assessing reading*. Cambridge: Cambridge University Press.
- Altheide, D.L. 1996. Process of qualitative document analysis, *Sage Research Methods Online*, <http://srmo.sagepub.com/> Accessed 17 June 2011.
- Alvermann, D.E., & Hynd, C.R. 1989. Effect of prior knowledge activation modes and test structure on non-science majors' comprehension of physics. *Journal of Educational Research*. 35 (10): 1091-110, Rowley, MA: Newbury House.
- Anderson, R.C., Reynolds, R.E., Schallert, D.L., & Goetz, E.T. 1977. Frameworks for comprehending discourse. *American Educational Research Journal*, 14, 367-382.
- Andrade, H. & Du, Y. 2005. Student perspectives on rubric referenced assessment. *Practical Assessment, Research and Evaluation*, 10 (3): 256-263.
- Bachman, L. F. & Palmer, A. 1999. *Language testing in practice*. Shanghai: Shanghai Foreign Language Education Press.
- Bachman, L. F. 1990. *Fundamental considerations in language testing*. Oxford, Oxford University Press.
- Bachman, L. F. 1991. What does language testing have to offer? *TESOL Quarterly*, 25 (4): 671-704.
- Banfield, G. & Cayago-Gicain, M.S. 2006. Qualitative approaches to educational evaluation: a regional conference-workshop. *International Education Journal*, 7 (4): 510-513.

- Barnett, M.A 1986. Syntactic and lexical/semantic skill in foreign language reading: Importance and interaction. *The Modern Language Journal*, 70: 343-349.
- 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.
- Belinda, F. 2008. Idiom comprehension in bilingual and monolingual adolescents. Ph D. dissertation, University of South Florida.
- Bentin, S., Deutsch, A. & Linberman, I. 1990. Syntactic competence and reading ability in children. *Journal of Experimental Child Psychology* 48: 147-172.
- Berman, R.A. 1984. Syntactic components of the foreign language reading process. In J.C. Alderson & A.H. Urquhart (eds.). *Reading in a foreign language*, London: Longman.
- Bernhardt, E.B. 1983. Three approaches to reading comprehension in intermediate German. *Modern Language Journal* 67: 111-115.
- Bernhardt, E.B. 1991. *Reading development in a second language*. Norwood, NJ: Ablex.
- Birjandi, P., Alavi, S.M., & Salmani-Nodoushan, M.A. 2001. Text familiarity, reading tasks, and ESP test performance: a study on Iranian LEP and non-LEP university students. Unpublished Ph D dissertation, University of Tehran. Available online at: [www.geocities.com/nodoushan/HomePage](http://www.geocities.com/nodoushan/HomePage).html. Accessed: 9 April 2011.
- Block, E.L. 1992. See how they read: comprehension monitoring of L1 and L2 readers, *TESOL Quarterly* 26 (2): 319-343.
- 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.
- Brett, A., Rothlein, L. & Hurley, M. 1996. Vocabulary acquisition from listening to stories and explanations of target words. *Elementary School Journal*, 96 (4): 415-422.

- Bridgeman, B., Cline, F., & Hessinger, J. 2003. Effect of extra time on GRE Quantitative and verbal scores. *R&D Connections*, ETS; Princeton: NJ O8541-0001.
- Bridgeman, B., McBride, A., & Monaghan, W. 2004. Testing and time limits. *R&D Connections*, ETS: Princeton: NJ O8541-0001.
- Briggs, C. 2000. Interview. *Journal of Linguistic Anthropology*, 9 (1-2): 137-140.
- 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.html>
- Buck, G. 2001. *Assessing listening*. Cambridge. Cambridge University Press.
- Burgoyne, K., Whiteley, H.E. & Kelly, J.M. 2011. The role of background knowledge in text comprehension for children learning English as an additional language. *Journal of Research in Reading*, Article first published online: 17 MAY 2011 DOI: 10.1111/j.1467-9817.2011.01493.x.
- Canale, M. & Swain, M., 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1 (1): 1-47.
- Carrel, P. L. & Eisterhold, J. 1983. Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 1: 553-573.
- Carrel, P.L. 1987. Content and formal schemata in ESL reading. *TESOL Quarterly*, 19: 727-752.
- Carrel, P.L. 1992. Awareness of text structure: Effects on recall. *Language Learning*, 42: 1-20.
- 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 (2):165-174.

- 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.
- Chen, H.C. & Graves, M.F. 1995. Effects of previewing and providing background knowledge on Taiwanese college students' comprehension of American short stories. *TESOL Quarterly* 29: 663-686.
- Chen, H.C. 1998. Constraints of English proficiency on understanding English ambiguous sentences in Chinese EFL learners. *Modern Foreign Languages*, 2 (1-16).
- Chou, P. T. 2011. The effects of vocabulary knowledge and background knowledge on reading comprehension of Taiwanese EFL students. *Electronic Journal of Foreign Language Teaching*, 8 (1): 108-115
- Cizek, G.J. 2001. An overview of issues concerning cheating on large- scale tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.
- Clapham, C.M. 1996. *The development of IELTS: a study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.
- Cohen, A. & Upton, T. 2006. *Strategies in responding to the new TOEFL reading tasks*. Educational Testing Services
- 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.
- Cook, G. 1997. Key concepts in ELT: Schemas. *ELT Journal*, 51: 1- 86.
- Corbetta, P. 2004. *Social research theory, methods and techniques*. London: Sage Publications.
- Creswell, J. W. & Plano Clark, V. L. 2007. *Designing and conducting mixed methods research*, Thousand Oaks: Sage Publications.
- Creswell, J. W. 2003. *Research design: Qualitative, quantitative and mixed methods approaches* (2<sup>nd</sup> ed.), Thousand Oaks: Sage Publications.
- 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.

Crosson, A. C., Lesaux, N. & Martiniello, M. 2008. Factors that influence comprehension of connectives among language minority children from Spanish-speaking backgrounds. *Applied Psycholinguistics*, 29: 603-625

Daniels, M. 1994. The effect of sign language on hearing children's language development. *Communication Education*, 43 (4): 291-298.

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.

Denzin, N. K. 1989. *The research Act: A theoretical introduction to sociological methods*. New Jersey: Prentice Hall.

Dochy, F., Segers, M. & Buehl, M.M. 1999. The relation between assessment practices and outcomes of studies: the case of research on prior knowledge. *Review of Educational Research*, 69, pp. 145-186

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1995) Diagnostic oral reading tests. *The Reading Teacher*, 30 (4): 369-375.

Douglas, D. 2000. *Assessing language for specific purposes: Theory and Practice*. Cambridge: Cambridge University Press.

Driscoll, D.L., Appiah-Yeboah, A., Salib, P. & Rupert, D.J. 2007. Merging qualitative and quantitative data in mixed methods research: how to and why not. *Ecological and Environmental Anthropology*, 3 (1): 19-28.

Droop, M. & Verhoeven, L. 1998. Background knowledge, linguistic complexity, and second-language reading comprehension. *Journal of Literacy Research*: 30 (2): 253-271.

Du Bay, W.H. 2006. *Smart language: Readers, readability and the grading of text*. Costa Mesa: Impact Information.

Enright, M.K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P. and Schedl, M. 2000. *TOEFL® 2000 Reading Framework: A Working Paper*; Report Number: RM-00-04, TOEFL-MS-17.

- Erickson, M. & Molloy, J. 1983. ESP test development for engineering students. In J. Oller (ed.). *Issues in language testing research*. Rowley, MA: Newbury House.
- Faux, J. 2010. Pre-testing survey instruments. *Global Review of Accounting and Finance*. 1 (1): 100-111.
- Ferrara, S. 1997. Test security for high stakes performance assessments: consideration of ethics, validity, and data integrity, Paper presented at the annual meeting of the American Educational Research Association and Council on Measurement in Education, Chicago.
- Fletcher, J.M. 2006. Measuring reading comprehension. *Scientific Studies of Reading*, 10 (3): 323-330.
- Freebody, P. & Anderson, R.C. 1983. Effects of vocabulary difficulty, text cohesion, and schemata availability on reading comprehension. *Reading Research Quarterly* XVIII, 3: 277-294.
- Fulcher, G. 1997. Text difficulty and accessibility: Reading formulae and expert judgment. *System*, 25 (4): 497-513.
- Fulcher, G. 1999. Assessment in English for academic purposes: Putting content validity in its place, *Applied Linguistics*, 20 (2): 221-236.
- Glenn, C. 1978. The role of episodic structure and of story length in children's recall of simple stories. *Journal of Verbal Learning and Behavior*, 17 (4): 229-247.
- Goldman, S.R. 1997. Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23: 357-398.
- Gordon, J., Schumm, J., Coffland, C. & Doucette, M. (1992). Effects of inconsiderate vs considerate text on elementary students' vocabulary learning. *Reading Psychology*, 13 (2): 157-169.
- 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.
- Gray, D.E. 2004. *Doing research in the real world*. London: Sage Publications.

Gohmann, S.F. & Spector, L.C. 1989. Test scrambling and student performance. *The Journal of Economic Education*, 20, 3, pp. 235-238.

Gunning, T.G. 2006. Creating Reading Instruction for All Children. *Reading comprehension: theories and strategies*.  
<http://am3ng.blogspot.com/2009/04/reading-comprehension-theories-and.html>.  
Accessed March 2011

Haertel, E. 1999. Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice*, 18 (4): 5-9.

Hale, G.A. 1988. Student major field and text content: interactive effects on reading comprehension in the Test of English as a Foreign Language. *Language Testing*, 5 (1): 49-61.

Hammadou, J. 1991. Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *The Modern Language Journal*, 75: 27-38.

Hammadou, J. 2000. The impact of analogy and content knowledge on reading comprehension: What helps, what hurts. *The Modern Language Journal*, 75: 27-38.

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.

Holmes, J.E. & Smith, L.J. 2003. Student evaluations of faculty grading methods. *Journal of Education for Business*, 78 (6): 318-323.

[http://www.Utwente.Nl/cw/theorieenoverzicht/Theory % 20 clusters](http://www.Utwente.Nl/cw/theorieenoverzicht/Theory%20clusters)

<http://www.designedinstruction.com/learningleads/reading-vocabul>

Huesman, R.L. & Frisbie, D.A. 2000. The validity of ITBS reading comprehension test scores for learning disabled and non-learning disabled students under extended –time conditions. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

- Hymes, D. 1972. On communicative competence. In J.B. Pride & J. Holmes (eds.) *Sociolinguistics: Selected readings* (pp. 267-293). Harmondsworth, Middlesex: Penguin.
- Inbar-Lourie, O. 2008. Constructing a language assessment knowledge: a focus on language assessment course. *Language Testing*, 25: (4): 385-402.
- Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P. & Taylor, C. 2000. TOEFL 2000 *Framework: a working paper*. Educational Testing Service, Princeton.
- Jannuzi, C. 1997. Key concepts in FL literacy: Schema theory. *Literacy Across Cultures*, 1 / 2, Available online, <http://www2.aasa.ac.jp/-dedycus/LA>.
- Joshi, R. M. & Aaron, P.G. 2000. The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21: 85-97.
- Kajornboon, A.B. 2005. Using interviews as a research instrument. *Conducting interviews*, 29: 7-15.
- Kane, M. T., 2001. Current concerns in validity theory, *Journal of Educational Measurement*, 38, (4): 319-342.
- Kane, M. T., 2008. Terminology, Emphasis and utility in validation, *Educational Researcher*, 37: 276-282.
- 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.
- Keats, D.M. 1997. Interviewing for Clinical Research, in J. Keeves, (ed), *Educational Research, Methodology and Measurement: An International Book* (2nd ed), Permagon, Oxford, pp. 306 – 310.
- Kendeou, P. & van den Broek, P. 2005. The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, 97: 235-245.
- Kendeou, P. & van den Broek, P. 2007. The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory and Cognition*, 35 (7): 1567-1577.

- Kendeou, P. , Rapp, D.N., & van Den Broek, P. 2004. The influence of readers' prior knowledge on text comprehension and learning from text. In Nata (*ed.*) *Progress in education* (vol. 13, pp. 189-209). New York: Nova Science.
- Kintsch, W. 1988. The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review* 95 (2): 163-182
- 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
- Kitao, K.S. 2003. Textual Schemata and English Language Learning. *Cross Current Issues* 3: 147-155.
- Klein, M.L., Peterson, S., & Simington, L. 1991. *Teaching reading in elementary grades*. Needham Heights, Mass.: Allyn and Bacon.
- Kobayashi, M. 2004. Investigation of test method effects: text organization and response format: a response to Chen. *Language Testing* 21 (2): 235-244.
- Koh, M.Y. 1985. The role of prior knowledge in reading comprehension. *Journal of Reading in Foreign Language*, 3 (1): 375-380.
- Krekeler, C. 2006. Language for special academic purposes (LSAP) testing: the effect of background knowledge revisited. *Language Testing* 23 (1): 99-130.
- Kvale, D. 1996. *Interviews*. London: Sage Publications.
- Laberge, D., & Samuels, S. J. (1974). Towards a theory of automatic information processing in reading. *Cognitive Psychology*, (6): 293-323.
- Lane, S. 1999. Validity evidence for assessments, Reidy Interactive Lecture Series, National Center for the Improvement of Education Assessment.
- Leedy, P. & Ormrod, J. 2001. *Practical research: Planning and design* (7<sup>th</sup> ed.) Upper Saddle River, NJ: Merrill Prentice Hall. Thousand Oaks: Sage Publications.
- Lenney, E., Mitchell, L. & Browning, C. 1983. The effects of clear evaluation criteria on sex bias in judgments of performance. *Psychology of Women Quarterly*, 7, 4, pp. 313-328

- Levine, M.D. & Haus, G.J. 1985. The effect of background knowledge on the reading comprehension of second language learners. *Foreign Language Annals*, 18 (5): 391–397.
- Levorato, M.C., Roch M. & Nesi, B. 2007. A longitudinal study of idiom and text comprehension. *Journal of Child Language*. 34 (3): 473-494.
- Levorato, MC., Nesi, B. & Cacciari, C. 2004. Reading comprehension and understanding idiomatic expressions: a developmental study. *Brain Lang.* 91 (3): 303-314.
- Liu, F. 2011. A short analysis of the text variables affecting reading and testing reading. *Studies in Literature and Language* 2 (2): 44-49.
- Liu, O.L., Schedl, M., Malloy, J. & Kong, N. 2009. Does content knowledge affect TOEFL IBT TM reading performance? A confirmatory approach to differential item functioning, TOEFL IBT Research Report, Educational Testing Services, Princeton, New Jersey.
- Liu, Z. & Bever, T.G.. 2002. An experimental study of the function of syntactic analysis in reading comprehension. *Foreign Language Teaching and Research*, 34: 219-224.
- Lu, S. 1999. An investigation into EFL reading processes: Reading effectiveness, inference construction, metacognitive strategy. M.A. Dissertation, Zhejiang University, China.
- Mamba, K. 2007. Revisiting the English National Programme in connection with the English state examination, Licence thesis, English department, ISP/Kananga.
- Mandinach, E.B., Cahalan, C. & Camara, W.J. 2002. The impact of flagging on the admission process: Policies, practices, and implications (ETS RR 02-03, College Board Report 2002-20. New York and Princeton, NJ: The College Board and Educational Testing Service.
- Mandinach, E.B., Bridgeman, B., Cahalan-Laitusis, C. & Trapani, C. 2005. The impact of extended time on SAT test performance. College Board Research Report, 2005 ETS RR-05-20. New York and Princeton, NJ: The College Board and Educational Testing Service.

- Manxia, D. 2008. Content validity study on reading comprehension test of NMET. *CELEA Journal*, 31 (4):12-30.
- Marsland, N., Wilson, I., Abeyasekera, S., & Kleih, U. 2001. A methodological framework for combining qualitative and quantitative survey methods. An output from the DFID-funded Natural Resources Systems Programme.
- Martin-Chang, S.Y. & Gould, O.N. 2008. Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, 31: 273-284.
- Martohardjono, G., Otheguy, R., & Gabriele, A. 2005. The role of syntax in reading comprehension: a study of bilingual readers. In J. Cohen, K. McAlister and J. MacSwan (eds), *Proceedings of the 4<sup>th</sup> International Symposium on Bilingualism*, Somerville, MA: Cascadilla Press.
- McCann, T., & Clark, E. 2005. Using unstructured interviews with participants who have schizophrenia. *Nurse Researcher*, 13 (1): 7-18.
- McKeown, M.G., Beck, I.L., Sinatra, G.M., & Loxterman, J.A. 1992. The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly* 27, (1):79-93.
- McNamara, C. 1999. *General guidelines for conducting interviews*, Minnesota.
- McNamara, D. 2007. Reading comprehension strategies: theories, interventions and technologies. London: Routledge.
- Medo, M.A. & Ryder, R.J. 1993. The effect of vocabulary instruction in reader's ability to make causal connections. *Reading Research and Instructions*, 33 (2): 119-134.
- Mehrpour, S. & Riazi, A. 2006. The impact of text length on EFL students' reading comprehension. *Asian EFL Journal*, (6): 3-18.
- Messick, S. 1989. Validity. In R. L. Linn (ed.). *Educational measurement* (3<sup>rd</sup> ed. pp 13-103). New York: American Council on Education? Macmillan.
- Messick, S. 1992., *The interplay evidence and consequences in validation of performance*, (ETS RR-92-39). Princeton, NJ: Education Testing Service.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2): 13-23.

- Messick, S. 1996. Validity and washback in language testing. *Language Testing* 13 (3): 241-256.
- Moskal, B.M. 2000. Scoring rubrics: what, when, and how? *Practical Assessment, Research and Evaluation*, 7 (3): 123-131.
- Moss, P.A. 2007. Reconstructing validity. *Educational Researcher*, 36: 470-476.
- Nassaji, H. 2003. Higher-level and lower level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, 87: 261-276.
- Ngongo, D. 2007. *La recherche scientifiques en education: Paradigmes, methodes, techniques*. Kinshasa: Academia Bruylant.
- Norris, J. 2000. Purposeful language assessment. *English Teaching Forum*, 38 (1): 18-23.
- Nuttal, C. 1996. *Teaching reading skills in a foreign language*. Oxford: Heinemann English Language Teaching.
- O'Sullivan, B. 2000. Exploring gender and oral proficiency interview performance. *System*, 28 (3): 373-386.
- O'Sullivan, B. 2004. *Issues in testing business English: The BEC Revision Project*. Cambridge: Cambridge University Press/Cambridge ESOL
- O'Sullivan, B. 2005. Evidence-based validation: what makes a test "specific". Workshop-ALTE Cardiff.
- Oh, S.Y. 2000. Two types of input modifications in EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, 34: 4-12.
- Oller, J. 1979. *Language Tests in Schools*, Longman, London.
- Oller, J. 1995. Adding abstract to formal and content schemata: Results of recent work in Peircean semiotics. *Applied Linguistics*, 16 (3): 273-306
- Programme Commente D'anglais* (1988), DEPS, Kinshasa.
- 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.
- Patton, M. Q. 1990. *Qualitative evaluation and research methods* (2<sup>nd</sup> ed.), Newbury Park: Sage Publications.
- Peretz, A.S. & Shoham, M. 1990. Testing reading comprehension in LSP. *Reading in a foreign language* 7 (1): 447-455

- Perfetti, C. 1989. There are generalized abilities and one of them is reading. In L. Resnick (ed.), *Knowing, learning and instruction*. Hillsdale, NJ: Lawrence Erlbaum
- Perkins, K. & Jones, B. 1985. Measuring passage contribution in ESL reading comprehension. *TESOL Quarterly*, 19, (1): 137-153.
- Powers, D.E. 2010. The case for a comprehensive, four-skills assessment of English language proficiency. *R&D Connection*, 14, ETS Research and Development, Princeton.
- Programme National d'anglais* 1986. DEPS, Kinshasa
- Pulido, D. 2004. The effects of cultural familiarity on incidental vocabulary acquisition through reading. *The Reading Matrix an Online International Journal*, 4: 20-50.
- Punch, K.F. 1998. *Introduction to Social research: Quantitative and qualitative approaches*. Thousand Oaks: Sage Publications
- Pyrczak, F. & Axelrod, J. 1976. Determining the passage dependence of reading comprehension exercises: A call for replications. *Journal of Reading*, 19 (4): 279-283.
- Pyrczak, F. 1974. Passage-dependence of items designed to measure the ability to identify the main ideas of paragraphs: Implications for validity. *Educational and Psychological Measurement*, 34: 343-348.
- Qian, D.D. 1999. Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56 (2): 282-307.
- Qian, D.D. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52: 513-536.
- Rapport Academique de l'ISP Kananga, 2010
- Rayner, K. 1990. Comprehension process: an introduction. In D.A. Balota *et al.* (eds.)
- Read, J. 2000. *Assessing vocabulary*. Cambridge: Cambridge University Press.

Reddy, Y. M., 2007. Effect of rubrics on enhancement of student learning. *Educate*, 7: (1): 3-17.

Revue de l'Inspecteur de l'Enseignement, 2010, No1

Revue de l'Inspecteur de l'Enseignement, 1982, No 6.

Reynolds, R. 2002. Understanding the nature of reading comprehension: Basis research and instructional implications, Available online at [http://www.aprakaloss.se/reinolds\\_english.htm](http://www.aprakaloss.se/reinolds_english.htm)

Reynolds, R., Taylor, M.A., Steffensen, M.S., Shirey, L.L. & Anderson, R.C. 1982. Cultural schemata and reading comprehension. *Reading Research Quarterly*, 17 (3): 367-381

Ridgway, T. 1997. Thresholds of the background knowledge effect in foreign language reading. *Reading in a Foreign Language*, 3 (1): 375-380.

Rosenblatt, L.M. 1994. The transactional model of reading and writing. In Ruddell, R.B., Ruddell, M.R., & Singer, H., (eds.) (1994). *Theoretical models and processes of reading, 4th edition*. Newark, DE: International Reading Association.

Ruddell, R.B., Ruddell, M.R., & Singer, H., (eds.) (1994). *Theoretical models and processes of reading, 4th edition*. Newark, DE: International Reading Association.

Rumelhart, D.E. 1980. Schemata: the building blocks cognition. In R.J. Spiro et al. (eds), pp. 123-156

Rumelhart, D.E. 1985. Towards an interactive model of reading. In H. Singer and R.B. Ruddell (eds.), *Theoretical models and processes of reading*. Newark, Delaware: International Reading Association

Runyan, M.K. 1991. The effect of extra time on reading comprehension scores for university students with and without learning disabilities. *Journal of Learning Disabilities*, 24, 2, pp. 104-108.

- 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.
- Salager-Meyer, F. 1991. Reading expository prose at the post-secondary level: influence of textual variables on L2 reading comprehension (a genre-based approach). *Reading in a Foreign Language* 8 (1): 645-662
- Scarborough, H.S. 1991. Early syntactic development of dyslexic children. *Annals of Dyslexia* 41: 207-220.
- 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
- Shenton, A. K. 2004 The analysis of qualitative data in LIS research projects: A possible approach. *Education for Information* 22: 143-162.
- Shepard, L. 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.
- Shohamy, E., 1990. Language testing priorities: a different perspective. *Foreign Language Annals*, 23 (5): 385-394
- Sireci, S. 2007. On validity theory and test validation, *Educational Researcher*, 36 (8): 477-481.
- Stanovich, K.E. 1980. Toward an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16: 32-71
- Stein, N. L. & Glenn, C.G. 1979. An analysis of story comprehension in elementary school children. In R.O. Freedle (ed.) *New directions in discourse processing*. Norwood, NJ: Ablex.

Stein, N. L. & Policastro, M. 1984. The concept of story: A comparison between children's and teacher's viewpoints. In H. Mandl, N.L. Stein & T. Trabasso (eds.) *Learning and comprehension of text* (pp. 113-155). Hillsdale, NJ: Erlbaum

Stemler, S. 2001. An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). Retrieved October 26, 2011 from <http://PAREonline.net/getvn.asp?v=7&n=17> .

Tian, S. 2006. Passage dependence of reading comprehension items in the GEPT and TOEFL. *The Reading Matrix*, 6 (1): 66-84.

Trochim, W. E. 2006. Unit of analysis, *Research Methods Knowledge Base*, Third Edition.

Tuinman, J.J. 1974. Determining the passage dependency of comprehension questions in 5 major tests. *Reading Research Quarterly*, 9 (2): 206-223.

Urquhart, A.H. & Weir, C.J. 1998. *Reading in a second language: process, product and practice*. Harlow: Longman

Valette, R, 1977. *Modern language testing*. New York: Harcourt Brace Jovanovich.

Van Den Broek, P., Rapp, D.N. & Kendeou, P. 2005. Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 39: 299-316.

Van Dijk, T.A. 1985. Strategic discourse comprehension. In T. Ballmer (ed.), *Linguistic dynamics: Discourses, procedures, and evolution* (pp. 30-61). Berlin: Walter de Gruyter & Co.

Van Dijk, T.A. and Kintsch, W. 1983. *Strategies of discourse comprehension*. New York: Academic Press

Van Yperen, N. W. 2003. Task Interest and Actual Performance: The Moderating Effects of Assigned and Adopted Purpose Goals. *Journal of Personality and Social Psychology*, 85(6), pp. 1006-1015.

Weir, C. J. 1983. The Associated Examining Board's test of English for academic purposes: An exercise in content validation. In A. Hughes and D. Porter (eds.) *Current developments in language testing*. London: Academic Press, pp. 147-153

- Weir, C. J. 2005. *Language testing and validation: An evidence-based approach*. Palgrave, Macmillan.
- Widdowson, H.D. 1983. *Learning purpose and language*. Oxford: Oxford University Press.
- Williams, C. 2007. Research methods. *Journal of Business & Economic Research*, 5 (3): 65-71.
- Williams, R. & Dallas, D. 1984. Aspects of vocabulary in the readability of content area L2 educational textbooks: a case study. In J.C. Alderson and A.H. Urquhart (eds.). *Reading in a foreign language*. London: Longman
- Yano, Y., Long, M.L. & Ross, S. 1994. The effects of simplified and elaborated texts of foreign language reading comprehension. *Language Learning* 44 (2): 189-219.
- Yoshida, K. 2006. Theoretical frameworks of testing in SLA: processing perspectives and strategies in testing situations. *Shiken: JALT Testing and Evaluation Newsletter*, 10: 1
- 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

## APPENDIX 1

### QUESTIONNAIRE SUBMITTED TO HIGH SCHOOL FINALISTS BEFORE THEY PRESENT THE TEST

#### Introduction:

“Je m’appelle Godefroid Katalayi. J’enseigne l’anglais a l’ISP/Kananga il y a 17 ans. Je mene une etude sur les examens d’etat, plus precisement l’epreuve d’anglais. Mon objectif est d’ameliorer la qualite de l’evaluation des finalistes aux examens d’etat. J’ai besoin de votre contribution en repondant au questionnaire suivant. Votre identite ainsi que vos reponses seront confidentielles.”

**Translation:** “*My name is Godefroid Katalayi. I have been teaching English at Kananga Teachers Training College for 17 years. I am making an investigation on the state examinations, especially the English test. My aim is to enhance the quality of finalist students’ evaluation. I expect your contribution in answering this questionnaire. I declare that your identity as well as the answer provided will be kept confidential*”.

Generalites (*General details*)

Nom et Prenom (*Name + Surname*):.....

Sexe (*Gender*): 1. Masculin (*Male*)                      2. Feminin (*Female*)

Age (*age*): .....

Ecole (*School*):.....

Option (*Subject*):.....

Votre quantieme fois de presenter les examens d’Etat? (*Is it your first time to take the ESE?*).

1. Premiere fois (*first time*)
2. Deuxieme fois (*second time*)
3. Troisieme fois (*third time*)
4. Autre (*other*)

Depuis quand apprenez vous l’anglais? (Choisissez une option)

(*Since when have you been learning English? Select one option*):

1. 1ere secondaire (*1<sup>st</sup> high school grade*)
2. 2eme secondaire (*2<sup>nd</sup> high school grade*)
3. 3eme secondaire (*3<sup>rd</sup> high school grade*)
4. 4eme secondaire (*4<sup>th</sup> high school grade*)

5. Other (specify)

**Questionnaire (Questionnaire)**

Notez bien (*Notice*):

- A moins que ca soit dit examen d'état, toutes les propositions suivantes sont relatives a l'examen d'état d'anglais;

*(Unless explicitly stated SE (state examinations), all statements relate to the English state exam);*

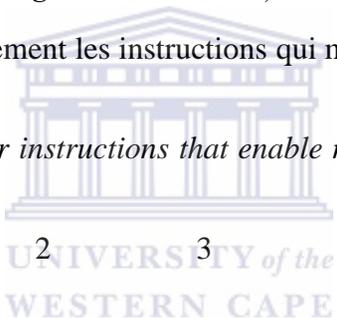
- Pour chacune de ces questions, encerclez le nombre qui REFLETE VOTRE POINT DE VUE sur cette echelle de cinq options ou 1=Tres d'accord; 2= D'accord; 3= Pas du tout d'accord; 4= Totalement pas d'accord; 5= Rien a dire

*(For each of the items below, circle the number that REFLECTS YOUR VIEWPOINT on a five -point scale where 1= Strongly agree; 2= agree; 3= disagree; 4=strongly disagree; 5= can't tell).*

1. Le test indique clairement les instructions qui me permettent de savoir ce que je dois faire

*(The test provides clear instructions that enable me to understand what I have to do).*

1                      2                      3                      4                      5



2. Pour chaque response correcte a la question, je sais que j'obtient 1 point, donc 1 %.

*(For each correct answer I provide to an item, I know that I get 1 point; thus 1 %).*

1                      2                      3                      4                      5

3. Les points que j'obtiendrai aux exetats seront combines avec les points des epreuves pratiques et dissertation francaise

*(Marks I get to the SE are combined with marks I get from the practical test and the French "Dissertation").*

1            2            3            4            5

4. Mes points aux exetats seront compares a mes points scolaires pour le resultat final

*(Scores I get to the SE will be moderated with my classroom scores in order to get the final score).*

1            2            3            4            5

5. J'ai quelque fois access aux ancient items

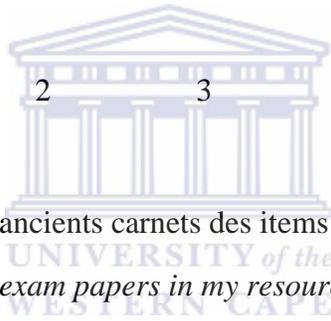
*(I sometime have access to some previous exam papers)*

1            2            3            4            5

6. Je possede quelques anciens carnets des items

*(I have some previous exam papers in my resource library)*

1            2            3            4            5



## APPENDIX 2

### QUESTIONNAIRE SUBMITTED TO HIGH SCHOOL FINALISTS AFTER THEY HAVE PRESENTED THE TEST

#### Introduction:

“Je m’appelle Godefroid Katalayi. J’enseigne l’anglais a l’ISP/ Kananga il y a 17 ans. Je mene une etude sur les examens d’etat, plus precisement l’epreuve d’anglais. Mon objectif est d’ameliorer la qualite de l’evaluation des finalistes aux examens d’etat. J’ai besoin de votre contribution en repondant au questionnaire suivant. Votre identite ainsi que vos reponses seront confidentielles.”

**Translation:** “*My name is Godefroid Katalayi. I have been teaching English at Kananga Teachers Training College for 17 years. I am making an investigation on the state examinations, especially the English test. My aim is to enhance the quality of finalist students’ evaluation. I expect your contribution in answering this questionnaire. I declare that your identity as well as the answers provided will be kept confidential*”.

#### Generalites (General details)

Nom et Prenom (*Name + Surname*):.....

Sexe (*Gender*): 1. Masculin (*Male*) 2. Feminin (*Female*)

Age (*age*): .....

Ecole (*School*):.....

Option (*Subject*):.....

Votre .... fois de presenter les examens d’Etat [Cochez la mention utile] (*Your ... time to take the ESE[Select the appropriate option]*).

1. Premiere fois (*first time*)
2. Deuxieme fois (*second time*)
3. Troisieme fois (*third time*)
4. Autre (*other*)

Depuis quand apprenez vous l’anglais? (Choisissez une option)

(*Since when have you been learning English?[ Select one option]*):

1. 1ere secondaire (*1<sup>st</sup> high school grade*)
2. 2eme secondaire (*2<sup>nd</sup> high school grade*)
3. 3eme secondaire (*3<sup>rd</sup> high school grade*)
4. 4eme secondaire (*4<sup>th</sup> high school grade*)

Other (specify)

**Notez bien (Notice):**

- A moins que ca soit dit examen d'état, toutes les propositions suivantes sont relatives a l'examen d'état d'anglais;

*(Unless explicitly stated SE (state examinations), all statements relate to the English state exam);*

- Pour chacune de ces questions, encerclez le nombre qui REFLETE VOTRE POINT DE VUE sur cette echelle de cinq options ou 1=Tres d'accord; 2= D'accord; 3= Pas du tout d'accord; 4= Totalement pas d'accord; 5= Rien a dire

*(For each of the items below, circle the number that REFLECTS YOUR VIEWPOINT on a five –point scale where 1= Strongly agree; 2= Agree; 3= Disagree; 4=Strongly disagree; 5= Can't tell).*

**Questionnaire (Questionnaire)**

1. Le test indique clairement les instructions qui me permettent de savoir ce que je dois faire

*(The test provides clear instructions that enable me to understand what I have to do).*

1                      2                      3                      4                      5

2. Pour chaque response correcte a la question, je sais que j'obtient 1 point, donc 1 %.

*(For each correct answer I provide to an item, I know that I get 1 point; thus 1 %).*

1                      2                      3                      4                      5

3. Les points que j'obtiendrai aux exetats seront combines avec les points des epreuves pratiques et dissertation francaise

*(Marks I get to the SE are combined with marks I get from the practical test and the French "Dissertation").*

1                    2                    3                    4                    5

4. Mes points aux exetats seront compares a mes points scolaires pour le resultat final

*(Scores I get to the SE will be moderated with my classroom scores in order to get the final score).*

1                    2                    3                    4                    5

5. Quatre heures (trois dans les options techniques) sont..... pour repondre a l'epreuve de langue (Anglais et Francais)

*(Four hours (three hours in technical sections) are ..... for taking the language test (French and English).*

1. Beaucoup trop (*Too much*)
2. Assez (*Enough*)
3. Insuffisant (*Not enough*)
4. Rien a dire (*Can't tell*)

6. Le texte etait..... a lire et comprendre

*(The test text was.....to read and understand).*

1. Tres difficile (*Very difficult*)
2. Difficile (*Difficult*)
3. Facile (*Easy* )
4. Tres facile (*Very easy*)
5. Rien a dire (*Can't tell* )

7. Le texte etait .....

*(The test text was.....)*

1. Tres long (*Very long*)
2. Long (*long*)
3. De longueur moyenne (*Of moderate length* )

4. Court (*Short*)

5. Rien a dire (*Can't tell*)

8. J'ai quelque fois access aux ancient items

*(I sometime have access to some ancient exam papers)*

1                      2                      3                      4                      5

9. Je possede quelques anciens carnets des items

*(I have some ancient exam papers in my resource library)*

1                      2                      3                      4                      5

10. J'ai eu access aux questions de l'examen avant l'epreuve

*(I could have access to the test content before I could take the test)*

1                      2                      3                      4                      5

11. Quelques finalistes avaient eu access aux questions avant l'epreuve

*(Some students have access to the test content before they take the test)*

1                      2                      3                      4                      5

### APPENDIX 3

#### QUESTIONNAIRE SUBMITTED TO LANGUAGE TEACHERS

##### **Introduction:**

My name is Godefroid Katalayi. I have been teaching English at Kananga Teachers Training College for 17 years. I am making an investigation on the state examinations, especially the English test. My aim is to enhance the quality of finalist students' evaluation. I expect your contribution in answering this questionnaire. I declare that your identity as well as the answer provided will be kept confidential).

Name and Surname:.....

School:.....

Qualification:

LA/L2.....G3.....D6.....

How long have you been teaching English in 6<sup>th</sup> grade? ..... Years

##### **Notice:**

- Unless explicitly stated SE (state examinations), all statements relate to the ESE (English state exam).

- For each of the items below, circle the number that REFLECTS YOUR VIEWPOINT on a five –point scale where 1= Strongly agree; 2= Agree; 3= Disagree; 4=Strongly disagree; 5= Can't tell.

1. The test provides clear instructions that enable candidates to understand what they have to do.

1                      2                      3                      4                      5

2. From the requirements of the tasks, candidates can use appropriate strategies to read the text and answer the test questions.

1                      2                      3                      4                      5

3. For each correct answer candidates provide to an item, I know that they get 1 point; thus 1 %.

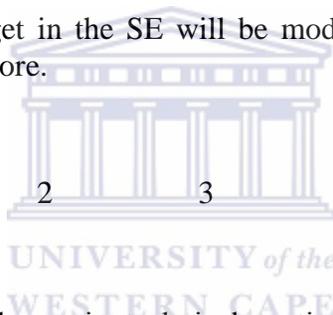
1                      2                      3                      4                      5

4. Marks candidates get in the SE are combined with marks they get from the practical test and the French “Dissertation”.

1                      2                      3                      4                      5

5. Scores candidates get in the SE will be moderated with classroom scores in order to get the final score.

1                      2                      3                      4                      5



6. Four hours (three hours in technical sections) are ..... for taking the language test (French and English).  
1. Too much                      2. Enough  
3. Not enough                      4. Can't tell

7. The test texts are generally .....to read and understand.

1. Very difficult    2. Difficult    3. Easy    4. Very easy    5. Can't tell

8. The test texts are generally .....

1. Very long    2. Long    3. Of moderate length    4. Short    5. Can't tell

9. Test items are grouped according to different skills they test

1                      2                      3                      4                      5

10. All candidates take the SE under the same conditions

1            2            3            4            5

11. I have access to previous exam papers

1            2            3            4            5

12. I have some previous exam papers in my resource library:

1            2            3            4            5

13. Some candidates have access to the test content before they take the test:

1            2            3            4            5





3. For each correct answer provided to an item, I know that it counts for 1 point; thus 1 %.

1                      2                      3                      4                      5

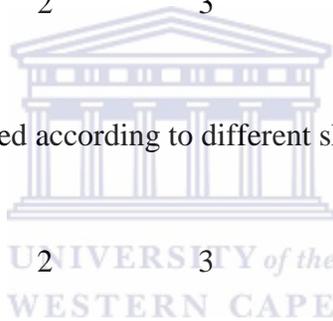
4. Marks candidates get to the SE are combined with marks they get from the practical test and the French "Dissertation".

1                      2                      3                      4                      5

5. Scores candidates get to the SE will be moderated with their classroom scores in order to get the final score.

1                      2                      3                      4                      5

6. Test items are grouped according to different skills they test



1                      2                      3                      4                      5

7. Circle the option of your choice. Four hours (three hours in technical sections) are ..... for taking the language test (French and English).

1. Too much              2. Enough              3. Not enough              4. Can't tell

8. Circle the option of your choice. The test texts are generally .....to read and understand.

1. Very difficult    2. Difficult    3. Easy    4. Very easy    5. Can't tell

9. Circle the option of your choice. The test texts are generally .....

1. Very long    2. Long    3. Of moderate length    4. Short    5. Can't tell

10. All candidates take the SE in the same conditions

1            2            3            4            5

11. Some candidates have access to the test content before they take the test

1            2            3            4            5



## APPENDIX 5

### QUESTIONNAIRE SUBMITTED TO OTHER STAKEHOLDERS

#### Introduction:

“Je m’appelle Godefroid Katalayi. J’enseigne l’anglais a l’ISP/ Kananga il y a 17 ans. Je mene une etude sur les examens d’etat, plus precisement l’epreuve d’anglais. Mon objectif est d’ameliorer la qualite de l’evaluation des finalistes aux examens d’etat. J’ai besoin de votre contribution en repondant au questionnaire suivant. Votre identite ainsi que vos reponses seront confidentielles.”

**Translation:** “*My name is Godefroid Katalayi. I have been teaching English at Kananga Teachers Training College for 17 years. I am making an investigation on the state examinations, especially the English test. My aim is to enhance the quality of finalist students’ evaluation. I expect your contribution in answering this questionnaire. I declare that your identity as well as the answer provided will be kept confidential*”.

#### Preliminaries:

-Nom +Postnom (Name +Surname):.....

-Profession (Occupation):.....

-Niveau d’etudes (facultative)/ Education level [optional]:.....

#### **Questionnaire (Questionnaire)**

Notez bien (Notice):

- A moins que ca soit dit examen d’etat, toutes les propositions suivantes sont relatives a l’examen d’etat d’anglais;

*(Unless explicitly stated SE (state examinations), all statements relate to the English state exam);*

- Pour chacune de ces questions, encerclez le nombre qui REFLETE VOTRE POINT DE VUE sur cette echelle de cinq options ou 1=Tres d’accord; 2= D’accord; 3= Pas du tout d’accord; 4= Totalement pas d’accord; 5= Rien a dire

*(For each of the items below, circle the number that REFLECTS YOUR VIEWPOINT on a five –point scale where 1= Strongly agree; 2= agree; 3= disagree; 4=strongly disagree; 5= can't tell).*

1. Le test indique clairement les instructions qui permettent au candidat de savoir ce qu'il doit faire

*(The test provides clear instructions that enable the candidate to understand what he/she has to do).*

1                      2                      3                      4                      5

2. Pour chaque response correcte a la question, chaque candidat obtient 1 point, donc 1 %.

*(For each correct answer I provide to an item, each candidate gets 1 point; thus 1 %).*

1                      2                      3                      4                      5



3. Les points que chaque candidat obtiendra aux exetats seront combines avec les points des epreuves pratiques et dissertation francaise

*(Marks each candidate gets to the SE are combined with marks he/she gets from the practical test and the French "Dissertation").*

1                      2                      3                      4                      5

4. Les points aux exetats seront compares aux points scolaires pour le resultat final

*(Scores obtained from the SE will be moderated with classroom scores in order to get the final score).*

1                      2                      3                      4                      5

5. Tous les candidats presentent les epreuves dans les memes conditions  
*(All candidates take the SE in the same conditions)*

1            2            3            4            5

6. Quelques candidats entre en possession de l'epreuve avant sa passation  
*(Some students have access to the test content before they take the test)*

1            2            3            4            5



## APPENDIX 6

### INTERVIEW WITH A., A HIGH SCHOOL FINALIST

**Interviewer:** Pouvez-vous me dire les différents supports que vous utilisez pour préparer l'épreuve d'anglais?

*Can you tell me the different supports that you use in the preparation of the English test?*

**A.:** Mes cahiers

*My class notes*

**I:** Seulement les cahiers?

*Only that?*

**A:** je n'ai pas des livres

*I don't have books*

**I:** Quest ce qui te manque encore?

*What else don't you have?*

**A:** Je n'ai pas de dictionnaire

*I don't have a dictionary*

**I:** Mais tu prépare les exéats avec les cahiers seulement?

*So you only use class notes for test preparation?*

**A:** Non, j'ai quelques anciens carnets des items

*No, I have some previous test papers*

**I:** Comment obtenez-vous ces carnets des items?

*How do you get them?*

**A:** On se débrouille!

*No, I try to fetch them*

**I:** Vous débrouillez comment? Vous achetez ça?

*How do you fetch them? Do you buy them?*



A: Bon il y a des bisseurs qui ont ça; peut être vous trouvez deux carnets d'une même série et il vous donne un; quelque fois on photocopie.

*Yes, there are those students who retake the test this year who have them; maybe he can have two identical test papers and he gives you one copy; or you photocopy yourself*

I: A quoi les anciens carnets sont importants pour vous?

*How are these previous test papers useful to you?*

A: On voit comment on pose les questions; on s'exerce en groupe pour répondre

*I see how questions are asked; then we do some exercises in group to answer them*

I: Les carnets ne vous servent pas à l'école?

*Aren't these papers useful to you in class sessions?*

A: À l'école on ne présente pas les interrogations à choix multiple; mais aux examens c'est le choix multiple

*In class we do not have multiple-choice tests; but at the state examination, we have multiple choice questions*

I: Mon cher Anaclet, revenons à notre petit test d'aujourd'hui.

*Anaclet, let's come back to today's test*

A: Oui

*Yes*

I: Comment tu as trouvé le texte?

*How did you find the text?*

A: Le texte était difficile; j'ai des problèmes en anglais

*The text was difficult; I have some problems with English*

I: As-tu lu tout le texte?

*Did you read all the text?*

A: Oui

*Yes*

I: Combien de fois?

*How many times?*

A: Deux fois; mais c'était toujours un peu difficile pour moi

*Twice; but it was still difficult for me*

I: Et Comment as-tu procede a repondre aux questions?

*How did you answer the test items?*

A: Eh, oui j'ai reflechis pour repondre a certaines questions

*Eh, yes, I could think to answer some questions*

I: A certaines, mais a d'autres?

*To some; but to others?*

A: A d'autres questions difficiles, j'ai seulement encercle la reponse

*To some difficult questions, I could just circle an option*

I: Pourquoi encercler la reponse sans lire les assertions?

*Why circle an option without reading the other options?*

A: Mais si je ne sais pas la reponse; que dois-je faire? Je ne dois pas laisser le vide

*But; I don't know the answer; what can I do? I cannot let a question unanswered*

I: Dites moi, il ya-t-il eu des questions auxquelles tu pouvais aussi voir facilement la reponse sans lire le texte?

*Tell me; did you find some questions to which you could provide answer without reading the text?*

A: Montre moi encore l'examen

*Show me the exam*

I: Regarde un peu

*Look*

A: Uhhh, voici cette question, je pouvais choisir 2 (assertion) a la question 4.

*Uhhh, look at this question, I could choose option 2 to question 4*

I: Pourquoi choisir 2?

*Why choose 2?*

A: Vous savez, nous savons tous. Nous parents nous dissent souvent ...1 zaire equivalait a 2 dollars et qu'ils pouvaient acheter beaucoup de nourriture avec

seulement 5e. Meme dans le cours d'histoire nous avons etudie aussi ca. Je ne pouvais pas perdre assez de temp a lire le text; la reponse etait la

*Uhhh, I could choose 2 [option] to question 4. You know, we all know. Our parents usually tell us... 1 zaire was 2 dollars and they could buy a lot of food with only 5 zaires. In history class we also studied. I could not spend time to read the text; the answer was there.*

[END]



## APPENDIX 7

### INTERVIEW WITH W., A HIGH SCHOOL FINALIST

**INTERVIEWER:** Comment avez-vous trouve le texte de l'examen?

*How did you find the text passage?*

**W.:** Tres difficile; je n'ai pas compris de quoi il etait question

*Very difficult; I couldn't understand what the text was about*

**I:** Qu'est qui a ete difficile?

*What was difficult?*

**W:** Ce texte parle meme de quoi?

*What is this text about?*

**I:** Vous n'avez pas compris de quoi le texte parle?

*So you didn't understand what the text was about?*

**W:** Regarde, ce n'est pas facile de comprendre ces choses: calipers,...

*Look; it is not easy to understand these things: calipers, ...*

**I:** Quoi encore?

*What else?*

**W:** Pliers eh...

*Pliers, eh...*

**I:** Encore un autre mot difficile pour toi?

*Another difficult word to you?*

**W:** loosening. Ces choses sont difficiles pour moi.

*Loosening. These words are difficult for me*

**I:** Quand tu lis le texte, ne peux tu pas comprendre la signification de ces mots a travers le context?

*When you read the text, can't you understand these words through context?*

**W:** Regarde ce texte, ca n'a pas de rapport avec ce que j'etudie.

*Look at this text, it does not relate to what I study*

I: Pourquoi tu dis ca?

*Why do you say so?*

W: Il n'ya pas des chose de mon domaine

*There is anything from my subject area*

I: Quoi par exemple?

*What for example?*

W: Ehh, tous les mots que j'ai cite la

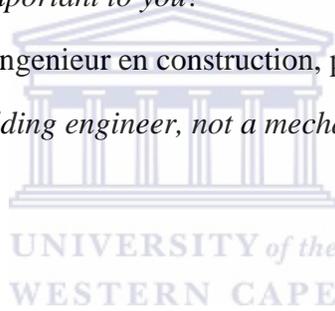
*Ehh, all these words that I have cited*

I: Ces mots ne sont pas importants pour toi?

*These words are not important to you?*

W: Je vais devenir un ingénieur en construction, pas un mecanicien.

*I'm going to be [a]building engineer, not a mechanic.*



## APPENDIX 8

### INTERVIEW WITH M., A HEAD INSPECTOR

**INTERVIEWER:** Pouvez-vous me citer certaines dispositions prise relatives a la securite des exetats?

*What are some of the precautions taken to ensure security of the state exam?*

M: Il ya des dispositions prevues tells que les malles des items ne doivent pas etre ouvertes avant les epreuves, il ya aussi le fait que le chef de centre ne doit pas entrer en contact avec le eleves et parents d'eleves avant la passation des epreuves

*There are some regulations for example: the item boxes and cases should not be open before the test starts; also the testing centre head inspector should not be in contact with students and parents until the test session ends*

I: Mais nous voyons après chaque epreuve, les carnets des items sont en circulation chez les finalistes

*But we notice that after each test session, many finalists still have test papers*

M: Oui nous constatons cela

*Yes, we also notice that*

I: Vous autorisez que les finalistes puissent sortir avec les carnets?

*Do you allow the candidates to go out with the test papers?*

M: Non, non, non personne ne doit avoir acces a ces carnets. Nous avons des instructions qui sont claires. Quand un candidat termine de presenter son examen, il remet la copie des reponses et le questionnaire. Mais il arrive surtout que dans beaucoup de centres des exetas que certains candidats s'arrangent avec des surveillants pour sortier avec ces carnets.

*No. No. No one should have access to the test papers. We have a regulation which is clear. When a candidate finishes writing his/her exam, he/she hands in both the answer sheet and the questionnaire papers. But in some testing centers, some candidates make arrangements with some proctors who let them go out with these archives.*

I: Comment mettre fin a ceci?

*How do you expect to stop this?*

M: Nous invitons le chef des differents centres des exetats a respecter scrupuleusement les instructions; a veiller que les surveillants ne laissent aucun candidat sortir avec les carnets.

*I invite all the responsible of testing centres to comply with the regulations and to ensure that proctors do not let any candidate leave the testing room with these test papers*

I: Pouvez vous dire que le test est administre dans les memes conditions a travers les differents centres ?

*Do you think that the test is administered under the same conditions?*

M: Non, je ne saurai pas dire cela

*I will not say so*

I: Si les conditions d'administration ne sont pas les memes, ne pensez vous pas que vous privilegiez certains candidats par rapport a d'autres?

*If administrative conditions are not the same, don't you think that you advantage some candidates on behalf of other candidates?*

M: Il est tres important et necessaire d'organiser l'epreuve dans les memes conditions. C'est ca etre juste. Les instructions sont claires en cette matiere. Mais sur terrain, les realites sont differentes. Par exemple, si en ville comme Kananga, il ya des sites pouvant arbriter facilement 2 a trois centres, ceci n'est pas le cas en milieu ruraux ou nous sommes souvent obliges d'utiliser meme des eglises pour organizer des centres

*It is a necessity, and it is desirable to organize the test under the same conditions. This is part of fairness. We have instructions for that; but the reality is quite different. For example, if in towns, like Kananga, we have venues that can accommodate two to three testing centers; this is not the case in rural areas where we are forced to use sometime churches to host testing venues*

[end]

## APPENDIX 9

### INTERVIEW WITH S., A LANGUAGE TEACHER

**INTERVIEWER:** Pouvez vous me dire les supports que vous utilisez pour préparer les finalistes aux exetats?

*Can you tell me the different supports that you use to prepare finalist students to the state exam?*

**S.:** Le grand support ce sont les anciens items

*The most important support are the previous test papers*

**I:** Pourquoi les anciens items?

*Why the previous test papers?*

**S:** C'est pour initier les eleves a s'abituier A a la forme de l'examen et aux types des questions qui sont posees

*This is to initiate the students to know how to answer this type of questions*

**I:** Comment obtenez vous ces carnets?

*How do you get them?*

**S:** Ca fait partie de ma bibliotheque

*This is part of my resource library*

**I:** Tu penses que tu ne peux pas preparer les eleves a l'epreuve sans necessairement utiliser ces items?

*Do you think you can't initiate the students to the test without these items?*

**S:** Comment puis-je initier les finalistes a reussir a l'epreuve d'anglais si je ne sais pas utiliser ces carnets comme support pedagogique? Je me base sur ces carnets et j'utilise quelques items pour l'evaluation scolaire

*How can I train students to be successful to the ESE if I can't use these papers as a teaching document? I base my preparation on the papers and I use some test items for classroom assessment*

## APPENDIX 10

### EXAMPLE OF AN EXPOSITORY TEST

Electricity is useful for many purposes. It is used for heating and lighting, and provides power for machines in houses and factories. Electricity is a form of energy. It can be produced from a number of sources. Generators and batteries are among these sources. Sometimes a small cell produces electricity of small voltage. Some of the sources of electricity can be taken and transported to fit the needs. The sun is considered as source of energy. With solar panels we can get electricity.

There are a number of dams which are built in the world to generate electricity. Some of the well known dams in Africa are: the Aswan Dam on the Nile and the Kariba Dam on the Zambezi. The Inga is on the river Congo. It is located in Bas-Congo region IN THE

Democratic Republic OF Congo. It is one of the most important dams in the world. It produces electricity which is used in local factories and some neighboring countries.

The movement of electricity is known as current. It flows from the source of electricity-place where it is created to the point of use and back to its source. The materials that carry electrical current are called conductors. Copper is used most often as an electric conductor. Conductors are shaped into tubes, bars, sheets and wires. Wires are used most commonly. Water and wet wood are also good conductors. Dry wood, rubber, glass and plastics are poor conductors. They are called insulators.

(Taken from the 2007 ESE session)

## APPENDIX 11

### EXAMPLE OF ARGUMENTATIVE TEST

Large forests are important to us in many ways. They give us wood for building and heating. They are a home for many kinds of plants and animals. For many city people, forest area place to go for vacation. There they can learn about nature, breathe fresh air, and sleep in a quiet place. There is more reason why forests are important for everyone. The leaves on trees help clean the air. Dirty air is a serious problem in many parts of the world. Without our forests, this problem might be even worse.

(Taken from the 2008 ESE session)





### Newsletter

Pour recevoir gratuitement  
notre newsletter hebdomadaire

[Inscrivez-vous ici](#)

### Rechercher



### Traduction



### L'INFO A LA SOURCE

ACCUEIL  
CONTACT  
CLIC-CLAC  
ANALYSES / OPINIONS  
TEMPS FORT  
AGENDA  
CULTURE & ARTS  
D'UN TRAIT  
ECHOS DE  
GROS PLAN  
LIENS  
MISE AU POINT  
MUSIQUE  
NEWS  
PEOPLE  
POLITIQUE  
SANTE  
SOCIETE  
SPORTS  
PUBLICITE

[ACCUEIL](#)

## SOCIETE

### "La baisse du niveau de l'enseignement en RDC est due à la modicité du budget y alloué"

Interview avec l'Inspecteur Général de l'EPSP Prosper Lituli Gay  
lundi 4 octobre 2004 Herman Bangi Bayo(AEM)



M. Prosper Lituli Gay, sur cette photo, inspectant l'expédition des épreuves vers les provinces.

**Le niveau de l'enseignement est à la baisse en R.d.C pour plusieurs raisons entre autres la carence des infrastructures scolaires, le manque des matériels didactiques et la non-motivation des enseignants. Et il s'avère de plus en plus vrai que les élèves ne maîtrisent pas la langue de l'enseignement qui est le français. Pour renverser la tendance, il faut que l'Etat mette les moyens et motive les enseignants, et enfin réforme l'actuel système éducatif pour le rendre compétitif.**



"Le bas niveau de notre enseignement est dû à des causes multiples. Il est tout d'abord lié à la déliquescence de la société congolaise en général. Dans une situation de crise socio-politique durable comme celle qu'a traversée notre pays, il va de soi que nous n'ayons pas un système éducatif performant", explique M. **Prosper Lituli Gay**, Inspecteur Général de l'enseignement primaire, secondaire et professionnel dans une interview à notre magazine.

**AFRIQU'ÉCHOS MAGAZINE : Vous venez d'être confirmé par arrêté ministériel comme Inspecteur Général de l'enseignement primaire et professionnel, quelles sont les tâches dévolues à l'inspection générale ?**

**Prosper Lituli Gay :** L'Inspection générale de l'EPSP est l'un de grands Services de l'Etat qui a une mission trilogique : contrôle classique, évaluation et formation. Concrètement, elle s'occupe du contrôle du système éducatif ; fait l'évaluation des connaissances des élèves et forme le personnel enseignant en cours d'emploi. Il y a peu de temps, le Ministère de tutelle vient de lui confier une mission supplétive, celle d'identifier les élèves ; d'imprimer et de délivrer des pièces scolaires au travers de la création d'un nouveau service dénommé SERNIE-ILPS

**A.E.M : \_Le niveau de l'enseignement est au rabais, à quoi attribuez-vous cette baisse ?**

**PLG :** Le bas niveau de notre enseignement est dû à des causes multiples. Il est tout d'abord lié à la déliquescence de la société congolaise en général. Dans une situation de crise socio-politique durable comme celle qu'a traversée notre pays, il va de soi que nous n'ayons pas un système éducatif performant. En outre, les crédits alloués au secteur de l'éducation sont insignifiants par rapport à la croissance exponentielle du nombre des élèves et par rapport à certains autres pays africains dont le pourcentage dépasse les 15% du budget national ; le 1% du budget alloué paraît comme une goutte d'eau dans un sol désertique. Et si, l'école n'a pas par bonheur perdu sa place dans la société ; par contre, la fonction d'enseignant a été fortement dévalorisée.

**A.E.M. Que préconisez-vous pour relever ce bas niveau de l'enseignement ?**

**PLG.** A mon sens, le financement de l'enseignement devrait permettre à équiper des écoles des moyens didactiques et d'autres infrastructures scolaires ; la révalorisation de la fonction enseignante aura des effets multiplicateurs sur tout le système éducatif. Il y a eu des Etats généraux de l'Education en 1996 et les différents séminaires et ateliers ; il en ressort que ce marasme est un problème conjoncturel et des moyens et non d'hommes. Il suffit de mettre des moyens conséquents et de motiver les enseignants, tout ira mieux. N'empêche, l'Etat doit également repenser les programmes et les modèles pour les adapter aux données du moment.

**A.E.M. Peut-on dire qu'aujourd'hui la RD Congo a un personnel enseignant suffisant et compétent ?**

**PLG.** Assurément, notre pays a fourni un grand effort de formation de cadres de haut niveau et dans des disciplines variées et aux compétences éprouvées. Mais les disparités entre les villes et les campagnes font qu'un grand nombre d'enseignants optent tout naturellement pour les milieux urbains pour des raisons que vous devinez facilement. Cette situation fait que l'école rurale se vide petit à petit de ses enseignants.

**A.E.M. On constate de plus en plus un taux élevé de déperdition scolaire, quelles mesures préconisez-vous pour inverser la tendance ?**

**PLG.** Comme je l'ai évoqué précédemment, la crise multidimensionnelle qui caractérise la société congolaise depuis plus de trois décennies n'est pas de nature à favoriser l'éducation des jeunes, surtout celle des jeunes filles. L'école s'est démocratisée sous d'autres cieux et il y a des lustres ; tandis que chez nous, l'instruction est devenue un luxe réservé à une bourgeoisie. L'Etat doit éviter de tuer des Mozart en favorisant une éducation de qualité pour tous et en garantissant un minimum vital digne et acceptable pour le commun de nos concitoyens.

**A.E.M. La non maîtrise du français qui est la langue de l'enseignement est aussi un de facteurs de la baisse du niveau de l'enseignement, que faire pour y remédier ?**

**PLG.** Cela est vrai et l'Inspection Générale de l'Enseignement s'y penche à travers des séances d'animation pédagogique sur ce sujet afin de réduire le taux d'échecs scolaires.

**A.E.M. Il vient d'être organisées en 2004 les premières épreuves des Examens d'Etat de la réunification de tous les territoires congolais, quelles sont les facilités dont vous avez bénéficié ?**

**PLG.** Votre question est lumineuse car il y a un mariage de raison entre l'Examen d'Etat et l'unité nationale. Pour rappel, ces épreuves existent depuis 1967 et ont résisté à toutes les turbulences que notre pays a connues. La tenue de ces épreuves à travers toute la République durant la balkanisation de la RD Congo suite à des rébellions à répétition est un symbole fort et le ferment de notre unité nationale. C'est l'unique maillon de l'administration qui couvrait tout le territoire national. Je profite de cette occasion pour rendre un hommage appuyé aux organismes du système des Nations Unies qui ont assuré le transport et la sécurité des épreuves. Avec la réunification des territoires, aucune récompense ni facilité n'a été accordée à l'IGE qui mériterait bien une médaille civique.

**A.E.M : Quelles sont des mesures prises pour éviter les fuites, la tricherie et la fraude ?**

**PLG.** Vous faites mention des fuites, tricherie et fraude lors des épreuves de l'Examen d'Etat ? Les mécanismes que nous avons mis en place sont consubstantiels à l'Examen d'Etat et l'IGE les affinent au fil des sessions. La sériation des items, le mixage inter optionnel dans les salles ainsi, que d'autres procédés en usage procèdent de la logique volontariste d'organiser une évaluation normative, objective garantissant les mêmes chances de réussite à tous les candidats. Comme il est naturel aux élèves les plus faibles à recourir à ces genres de pratiques anti-pédagogiques dues à une maîtrise insuffisante et un déficit des matières enseignées, nous tenons compte de tous ces enjeux car ces épreuves sanctionnent la fin des études secondaires du cycle long. Certains tricheurs organisés ont même développé la forme la plus astucieuse et la plus pernicieuse que représente la substitution des candidats et l'IGE s'y investit en sévissant par l'usage et le contrôle méticuleux des macarons.

**A.E.M : Votre mot de la fin.**

**PLG.** Je vous remercie pour nous avoir offert ces lignes afin de présenter au public l'IGE et comment elle s'emploie pour accomplir la lourde mission lui confiée par l'Etat. |Propos recueillis par Herman BANGI BAYO| Kinshasa-RD Congo

0 commentaires



Version à imprimer



envoyer par mail



Voir les commentaires/Réagir

**Dans la même rubrique**

- ▶ Angola : Le général Furtado aurait détourné 40 millions de dollars 10 07 2011 >>>
- ▶ Angola : Découverte d'un avion de combat enterré dans l'enceinte d'une école 4 07 2011 >>>
- ▶ Angola : Affrontements sanglants entre chercheurs de diamants et agents de sécurité à Lunda Norte 27 06 2011 >>>
- ▶ Assassinat de Chebeya en RDC : 4 policiers condamnés à mort 24 06 2011 >>>
- ▶ Procès Chebeya : Le verdict sera rendu aujourd'hui 23 06 2011 >>>
- ▶ RDC : Les assassinats et la maltraitance impunis des enfants dits « sorciers » évoqués au Parlement européen 19 06 2011 >>>
- ▶ L'Angola décide de s'attaquer à la prolifération et aux dérives des églises de réveil 19 06 2011 >>>
- ▶ Allemagne : Questions autour de la mort d'une Allemande d'origine nigérienne abattue par la police 19 06 2011 >>>
- ▶ Alain Marius Ngoya-Kessi : « Le taux de déforestation au Congo est de seulement 0,03 % contre 44 % en Amérique latine » 19 06 2011 >>>
- ▶ Conférence sur les enfants accusés de sorcellerie au Parlement européen 14 06 2011 >>>



UNIVERSITY *of the*  
WESTERN CAPE