DEVELOPMENT OF A LEARNING ANALYTICS APPROACH TO IDENTIFY AT-RISK STUDENTS IN HIGHER EDUCATION

Sonwabo Jongile

Student Number: 2113596

A full thesis submitted in fulfilment of the requirements for the

Magister Commercii Information Systems,

Faculty of Economic and Management Sciences,

Department of Information Systems,

University of the Western Cape

Supervisor: Dr Johan Breytenbach

Date: 30 November 2023

ABSTRACT

Learning Analytics (LA) has emerged as a study domain within higher education, combining elements of Business Intelligence (BI) and education-focused analytics. It implies principles and processes similar to BI in the business field. LA primarily focuses on analysing student-institution interactions, student success factors, and the effectiveness of teaching and learning approaches such as traditional face-to-face, online, and blended learning. Like in the business field, LA relies on quality data inputs, which vary in their accuracy and completeness.

Over the past two decades, higher education institutions (HEIs) have experienced significant changes related to the adoption of Information and Communications Technologies (ICTs). These changes aimed to improve operational efficiency, enhance management effectiveness, and increase competitiveness. Operational efficiency involved automating information-based processes, while management effectiveness included the implementation of Institutional Management Systems (IMS) such as Enterprise Resource Planning (ERP) and Student Information Systems (SIS). To improved competitiveness, HEIs implemented strategic information systems, shifted to online learning, and utilised blended learning practices through integrated Learning Management Systems (LMS) and Marks Administration (MAS).

In order to ensure seamless data flow and create a central data warehouse for LA purposes, Systems Integration (SI) became a key concern. SI connects various institutional data sources from multiple systems, activities, and resources to generate cohesive data for LA. This study focuses on investigating the critical success factors (CSF) that influence LA success at a South African HEI. The LA system in this study aims to provide comprehensive student-at-risk reports. The findings suggest that integrating data sources in higher education can enhance LA and enable HEIs to identify at-risk students more effectively.

KEY WORDS

Business intelligence

At-risk students

Institutional administrative systems

Learning Analytics

Learning Management System

Higher Education Institution

Online learning

Student performance

Systems integration

Student success

University retention

ABBREVIATIONS

A4L	-	Analytics for Learn
BI	-	Business Intelligence
CFA	-	Confirmatory Factor Analysis
CIET	-	Centre for Innovative Educational Technologies
CPUT	-	Cape Peninsula University of Technology
CRM	-	Customer Relations Management
CSF	-	Critical Success Factors
CTS	-	Computer and Telecommunications Services
DVC	-	Deputy Vice Chancellor
DWH	-	Data warehouse
EFA	-	Exploratory Factor Analysis
ELT	-	Extraction, Loading, Transforming
ERP	-	Enterprise Resource Planning
ETL	-	Extraction, Transforming, Loading
FRS	-	Functional Requirements Specification
GPA	-	Grade Point Average
HEIs	-	Higher Education Institutions
HTML	-	Hyper Text Mark-up Language
ICT	-	Information and Communications Technologies
IMS	-	Information Management Systems
I/S	-	Information Systems

ISPC	-	Institutional Strategic Planning Council
IT	-	Information Technology
KFC	-	Kentucky Fried Chicken
LA	-	Learning Analytics
LMS	-	Learning Management System
LTI	-	Learning Tool Interoperability
MAS	-	Marks Administration Systems
MOOCs	-	Massive Open Online Courses
NSFAS	-	National Student Financial Aid Scheme
OAAI	-	Open Academic Analytics Initiative
OBE	-	Outcome-Based Education
OLAP	-	Online Analytical Processing
PCSS	-	Psychological Counselling Service System
PDF	-	Portable Document Format
PoPI	-	Protection of Personal Information
PPDM	-	Privacy-Preserving Data Mining
SA	-	South Africa
SAP	-	System Analysis Program
SAS	-	Statistical Analysis System
SHEILA	-	Support Higher Education in Integrating Learning Analytics
SI	-	Systems Integration
SIS	-	Student Information Systems
UCDG	-	University Capacity Development Plan

UNESCO	-	United Nations Educational, Science and Cultural Organisation
WC	-	Western Cape
Wi-Fi	-	Wireless Fidelity
WIL	-	Work Integrated Learning
XML	-	Extensible Mark-up Language

DECLARATION

I hereby declare that the thesis entitled "*Development of a learning analytics approach to identify at-risk students in Higher Education*" is my own work. It has not been submitted, previously, for any degree or examination at any other university, and all the sources used, or quoted, have been indicated and acknowledged.

Sonwabo Jongile

Date: 30 November 2023

Signed: …Sjongile…………

DEDICATION

I would like to honour my late parents, Madoda and Nomboniso Jongile, by dedicating this thesis to them. They played a curtail role in shaping me into the person I am today- one who is strong, God-fearing, culturally aware, and independent.

I am eternally grateful to my father, Matshaya, Xesibe, Khandanyawana, and my mother Maradebe, Ndlebentle zombini, for the love and guidance they provided. Your memories will always be cherished in my heart. May your souls rest in peace.

I would also like to express my love and appreciation to my wife, Nomafu Ndinani Jongile, and my sons, Lithani Ntloso and Aluve Imolathile.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Qamata for granting me the gift of life, as well as for the strength and courage to successfully complete my research giving me life, strength, and courage to complete my research. I am also immensely thankful to my ancestors for their valuable guidance, protection, and the opportunity to exist. Without their blessings, none of this work would have been achievable.

Furthermore, I would like to extend my appreciation to Dr Johan Breytenbach, my Supervisor, for his unwavering guidance and support throughout this study. I will forever be indebted to you and your assistance.

I would be remiss not to acknowledge the unwavering support of my sons, Lithani and Aluve, and my siblings, Lindelwa, Linda, and Bongile Jongile, during the challenging times. Your compassionate love kept me motivated and inspired. My heartfelt gratitude also goes to my wife, Nomafu Ndinani Jongile, for her constant love, support, and understanding.

Above all, I am profoundly grateful to Qamata and my ancestors (OoMatshaya, Xesibe, Khandanyawana, Mkhuma; Rhadebe, Mthi mkhulu; Kabane, Mthi wembotyi; Jola, Jolinkomo, Mphankomo, and other unmentioned praises). During the completion of this thesis, I tragically lost both my parents simultaneously. It was through the mercies of Qamata and my ancestors that I managed to navigate through these trying times.

I would also like to acknowledge the support and encouragement from my colleagues at the Cape Peninsula University of Technology (CPUT), as well as the Information Systems department from the University of the Western Cape (UWC). Your unwavering support will forever remain etched in my heart.

May Qamata bless each and every one of you- from my ancestors to my sons, family, friends, colleagues, and all those who contributed to the successful completion of this research.

Warmest Regards,
Sonwabo Jongile

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

1.1 Introduction and background to the research problem

Over the past twenty years, most Higher Education Institutions (HEIs) have experienced significant changes in their use of Information Communication Technologies (ICTs) to enhance operational efficiency, improve management effectiveness, and increase competitiveness (Kelly et al., 2017). Operational efficiency improvements have focused on automating information-based processes (Sandberg et al., 2019). Management effectiveness enhancements have included the adoption of Institutional Management Systems (IMS) to meet information needs for decision-making and drive administrative reforms. This has involved implementing Enterprise Resource Planning (ERP) systems to manage various university functions, such as enrolment, student, student records, financial aid, finance, and human resources (Brainbridge et al., 2015). To improve competitiveness, HEIs have introduced strategic information systems, shifted towards online and blended learning modes, and utilised integrated Learning Management Systems (LMS) (Conijn et al., 2017) and an integrated Marks Administrative Systems (MAS) (Peppard & Ward, 2016).

As a result, the integration of various data sources, activities, hardware, software, and human resources through Systems Integration (SI) has become a crucial consideration for HEIs. This integration aims to create a cohesive flow of data and a central data warehouse for Learning Analytics (LA) purposes. LA involves measuring, collecting, analysing, and reporting data about learners and their contexts to enhance learning and optimise learning environments (Ferguson, 2012). Therefore, systems integration plays a vital role in collecting data from multiple unintegrated systems to improve LA applications. LA is a domain that intersect Business Intelligence (BI) and educational analytics and focuses on analysing student-institution interactions, student success factors, and the effectiveness of different teaching and learning approaches (Jayaprakash et al., 2014).

The success of BI and LA relies on the quality of the data they receive as inputs. Strategic Information Systems, like LA systems, rely on the data produced from student-institution interactions, which can vary in quality. However, if the institutional Information Systems and educational technologies generate functional log data that is not integrated, it becomes challenging. In such cases, LA is likely to fail (Pardo & Kloos, 2011).

Therefore, the integration of multiple institutional Information Systems becomes a crucial factor in determining the success of LA. Identifying the critical success factors (CSF) that can influence and improve the potential of LA is essential for improving the quality of education-based analytics (Lawson et al., 2016). This study aims to justify the capabilities of integrating institutional Information Systems by investigating the CSF that impact LA success at a South African HEI. Specifically, the study focuses on how integrated institutional Information Systems can enhance LA applications to identify students at-risk of academic failure or those facing obstacles to complete their studies.

1.2 Statement of the research problem

Despite the significant advances in BI and data analytics, their application in teaching and learning, specifically in higher education, has been largely overlooked. HEIs possess a wealth of student-generated data from various institutional Information Systems. However, HEIs lack the necessary knowledge and expertise to effectively utilise LA interventions in order to implement LA successfully. Using the analysis of students at-risk as an example, integrating HEI data sources into cohesive LA data repositories can improve the identification and support of students who may be at risk. This lack of integration is directly linked to the inability to identify at-risk students and hampers the decision-making process and potential intervention strategies within HEIs.

1.3 Primary research question

Based on the statement of the research problem, the primary research question was posed as:

What are the critical success factors of Information Systems integration necessary to facilitate LA at HEIs in SA?

In order to address the main research question, it is necessary to consider the following sub questions:

- What is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems in SA HEIs for LA adoption?

- Which data is necessary for conducting LA, and what are the resulting requirements for system integration?
- When examining students at-risk analysis as an instance of LA performed by HEIs, what data modelling choices are made accessible to HEIs following successful integration of BI and analytics systems?

1.4 Research aims and objectives

The primary aim of this study was to demonstrate how integrated institutional Information Systems can serve as reliable source systems. This was achieved by examining the CSF that influence a successful implementation of LA, specifically focusing on its application in identifying at-risk students. The study aimed to accomplish the following research objectives:

- Gain an understanding of the necessary capabilities of source systems integration to develop a LA model for analysing student at-risk data.
- Determine the optimal method for combining datasets (predictor variables) from multiple unintegrated source systems to inform LA, using the case study application of student at-risk data analysis.
- Propose critical success factors for implementing LA effectively, including recommendations for modelling the identification process of at-risk tertiary education students. Furthermore, highlight the potential benefits of using integrated source systems for timely and responsive interventions in LA.

1.5 Location of the study

The research was carried out on a specific case involving a Faculty at a university of technology in the Western Cape, South Africa. This chosen faculty serves as a representative example that sheds light on broader dynamics within the institution, which can be relevant to other faculties. Additionally, by focusing on a single faculty, the researcher ensured that the study was practical and could be effectively conducted with the available resources.

1.6 Alignment of the primary research question to research sub questions, method, and research objectives

The following table provides an overview of the alignment of the primary research question to the research sub questions, methods, and research objectives covered in this study.

Table 1.1: Alignment of the primary research question to the research sub questions, methods, and research objectives

Research question: What are the critical success factors of Information Systems integration necessary to facilitate LA at HEIs in SA?		
Research sub questions	**Method/s**	**Research subobjectives**
What is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems in South African higher education institutions (HEIs) for learning analytics (LA) adoption?	Literature Review Survey	Gain an understanding of the necessary capabilities of source systems integration to develop a LA model for analysing student at-risk data analysis.
Which data is necessary for conducting LA, and what are the resulting requirements for system integration?	Literature Review Survey	Determine the optimal method for combining datasets (predictor variables) from multiple unintegrated source systems to inform LA, using the case study application of student at-risk data analysis.
When examining students at-risk analysis as an instance of LA performed by HEIs, what data modelling choices are made accessible to HEIs following successful integration of BI and analytics systems?	Triangulation of integrated source system's data using the designed model	Propose critical success factors for implementing LA effectively, including recommendations for modelling the identification process of at-risk tertiary education students. Furthermore, highlight the potential benefits of using integrated source systems for timely and responsive interventions in LA.

1.7 Preliminary literature review

HEIs have accumulated a significant amount of student data in various systems such as admissions files, ERP, LMS, SIS, MAS, library records and other institutional Information Systems. As the education landscape has shifted from traditional paper-based methods to digital platforms and online learning, and with the implementation of Blended Learning Practices, HEIs are now leveraging this data to anticipate and address student needs for achieving academic success (Ibrahim et al., 2020). The objective is to utilise the data generated by different institutional Information Systems, which are dispersed throughout the institutions, in order to offer personalised interventions throughout entire student journey, from admissions to course progression and graduation (De Freitas et al., 2015).

The vast amount of digital data collected and stored by various institutional Information Systems creates a complicated contrast that puts pressure on HEIs to implement strategic information systems similar to LA systems. These systems should have the resources to consolidate the data into a central database for statistical analysis, generate reports or data visualisation, and identify patterns and trends (Daniel, 2015). The ultimate goal is to seamlessly integrate the different source systems and transition the data into a coherent data repository using LA technology. These LA systems are essential for connecting and analysing multiple data sources, as well as transforming the data to gain insights into interactions between student and institutions, factors that contribute to student success, and patterns of student behaviour. Additionally, they provide BI knowledge that aids in discovering new information and improving teaching and learning outcomes (Allen et al., 2017).

To gain a comprehensive understanding of the factors that influence the adoption of BI and analytics in businesses, a data warehouse, integrated source systems, and collection of data that is time-variant and non-volatile are required (Ain et al., 2019). Therefore, the adoption and success of LA systems in higher education can be a complex phenomenon that relies on integrated data from the source systems and the quality of the data received from those systems. This process also considers the underlying back-end systems, which may not have been suitable for LA applications. However, there is a lack of empirical research on the critical success factors for integrating Information Systems in higher education to facilitate LA. This preliminary literature review aims to explore and justify the capabilities of integrated institutional Information Systems by investigating (i) the current state of integration in the SA higher education systems, (ii) the relevant datasets generated from multiple source systems that contain predictor variables for designing a model to identify at-risk students, and (iii) the statistical models and theoretical knowledge required to build a LA model for analysing student at-risk data

1.7.1 The current state of integration in the SA higher education systems

Globally, the intersection of BI and analytics has led HEIs to seek strategic information systems that facilitate the effective extraction, transforming or loading (ETL/ELT), cleansing, and shaping of data from various educational source systems (Stodder & Matters, 2016). The same is true for SA HEIs. Whyte & Davies (2021) explained systems integration as a challenging and complex project whereby numerous interdependent systems with diverse components and knowledge are to be coordinated and adjusted to each other for a particular aim. For a long time, institutional Information Systems have supported SA HEIs in their diverse responsibilities. However, the implementation of strategic information systems such as LA requires integrating these institutional Information Systems to discover new interdependencies and prepare the data for analysis (Dietze et al., 2013). Placing systems integration, data warehousing, processing algorithms, data modelling, and analysis and visualisation techniques at a crucial level for HEIs that want to adopt and implement LA successfully (Papamitsiou & Economides, 2014).

While HEIs globally have successfully adopted and implemented LA, SA HEIs still face challenges related to disintegrated institutional Information Systems (Samara, 2015), aggregating educational data (Bock, 2014), and discovering relevant datasets from multiple data sources (Jayaprakash et al., 2014). Currently, SA HEIs who have been reckoned with LA systems have done so to collect and consolidate, drill, and analyse and report on the data generated from the different and dispersed institutional Information Systems, including:

- The ERP system: contains data on students' financial standing, Outcome-Based Education (OBE) background, and demographics.
- The LMS: includes data on student social integration through peer-reviews, discussion boards, blogs, group work challenges, and email communication.
- The MAS: provides data on student academic integration such as grade performance, as well as other institutional systems.

Managing the integration of institutional Information Systems in SA HEIs has been challenging due to their traditional use in separate silos (Chang et al., 2014). This has made it difficult for the systems to provide the necessary reliable, up-to-date, and secure data for analysis in the development of novel LA systems. To address this challenge, a strategic approach is needed to extract, transform, and load data from silo institutional Information Systems into a central data warehouse for analysis (Drake & Walz, 2022). The integration of these institutional Information Systems is crucial for the sustainability of Information Systems in higher education (Halili, 2019). The need for systems integration is evident in addressing the diverse challenges associated with LA in SA higher education. HEIs in SA face strategic challenges in terms of course and degree completion rates as well as overall student retention (Jayaprakash et al., 2014). According to the South African Department of Higher Education and Training's first annual statistical report to over twenty-three (23) public universities in the country, by the year 2013 the average graduation rate amongst Doctoral, Master's, Undergraduate and Diploma students stood at 13%, 21%, 15%, and 18% respectively (Statistics on Post-School Education and Training, 2015, p.17). The figures were relatively lower than that set more than a decade ago by the National Plan for Higher Education to the increase graduation rate from 15% to 20% in the long-term (Department of Education, 2001). Consequently, SA HEIs resolved to accelerate the integration of institutional Information Systems for LA purposes. The aims were to discover information, predict, and advise based on the student-institution interaction analysis, thus building the capacity of universities to respond in real-time, and to provide personalised interventions to the heterogeneous challenges faced by SA higher education.

Another report showed that 47.9% of SA university students do not complete their studies (South African Department of Higher Education and Training, 2015). There are several reasons for the low course and degree completion rates, and the overall university retention. Factors include high failure rates that may lead to academic exclusion (Akoojee & Nkomo, 2007), financial constraints where students do not have funding to see them through after enrolling for a course (Letseka & Maile, 2008), lack of first-year experience preparedness (De Klerk et al., 2017), students hopping from one course to another, and students not receiving enough support from their universities (Roger, 2003). The Council of Higher Education (2010) reported that South African students enter universities from positions of extreme inequality in terms of financial pressures, lack of academic preparedness -- both their social conditions and teaching and learning background. Financial pressures often lead students to either commit to a part-time job or to take a gap year to accumulate their study funds, while the social conditions of their environments can either discourage or encourage student commitment to graduate when the student has experienced poor teaching and learning or substantial financial problems (Scott et al., 2007). Although historical research findings explain that academic performance and the student level of commitment to graduate are moderate determinants of a student's decision to drop out, they are clear on the fact that financial hardship exerts a powerful influence (Bennett, 2003).

In business, organisations have deployed well-established processes that allow BI models to evaluate integrated data sources, and to identify patterns within the integrated data for decision-making (Chaudhuri et al., 2011). In HEIs, the complex nature of LA systems implementation and the role of institutional Information Systems integration, and how the integration affects education-based analytics remains unclear. Baskerville et al. (2018) pointed out that one of the challenges influencing resistance to integrate the institutional Information Systems is Privacy-Preserving Data Mining (PPDM), and that HEIs are of the view that the more the institutional Information Systems are integrated, the greater the need for self-protective cybersecurity countermeasures. PPDM has been an emerging research topic parallel to big-data mining, BI and analytics, and is now making its way into education-focused analytics.

Xu et al. (2014) described current studies of PPDM as those that "mainly focus on how to reduce the privacy risk brought by data analysis processes, while in fact, unwanted disclosure of sensitive information may also happen in the process of data collecting, data publishing, and information (i.e., the data mining results) delivering". Clearly, a reflection on the factors affecting the success of institutional Information Systems integration, sharing, and modelling of the student-at-risk identification process in education-based analytics is imperative. However, there is a need for HEIs to follow similar principles and processes to that of the well-established domain of BI and analytics in business, and to apply the principles and processes to the analysis of the student-at-risk in education-based analytics. If there is rigorous institutional Information Systems integrations configuration, the improved quality of the data produced from the integrated systems will enhance the competency of HEIs to provide personalised intervention that will support different students in different ways.

1.7.2 The relevant datasets generated from multiple source systems that contain predictor variables for designing a model to identify at-risk students

Datasets that contain predictor variables, which are created by institutional source systems, can be described as conceptual models that represent the structure of information in a problem domain in terms of entities and relationships (Lukyanenko et al., 2014). For instance, research conducted on course signal at Purdue highlighted that datasets such as demographic characteristics, past academic history, grades performance, and student interaction with the university LMS could be used to gauge student performance (Arnold & Pistilli, 2012). The underlying argument is that by identifying datasets generated from multiple integrated source systems, it is possible to design a model that can contribute to the process of identifying student at-risk in LA (Dietz-Uhler & Hurn, 2013).

The main question then becomes: what are these datasets? How effective and applicable are the relevant datasets mentioned in the literature when employed in different contexts? Furthermore, what is the significance of identifying a relevant dataset if the multiple institutional source systems are not integrated? What purpose does it serve to identify categories that influence students being at-risk if the relevant datasets are isolated? Several factors need consideration, such as the variety of interactions between students and institutions, the socio-economic and socio-technical aspects of LA in different countries, the diversity among institutions and their students, and the transformative potential of modern teaching and learning practices (Alyaseri et al., 2023).

1.7.3 The statistical models and theoretical knowledge required to build a LA model for analysing student at-risk data

According to Mikalef et al. (2019), BI allows for real-time insights and the ability to take action based on data analysis. The purpose of this subsection is to analyse education-based analytics using statistical and theoretical foundations found in the literature. The aim is to understand the underlying factors that contribute to students being at-risk by examining the datasets generated by source systems from both statistical and theoretical perspectives. While most on education-based analytics in LA focus primarily on statistics, this subsection aims to incorporate both statistical and theoretical viewpoints.

To establish a theoretical basis, the study initially chose the Vincent Tinto Longitudinal Model of Dropout (Tinto, 1975), which provides a theoretical synthesis of recent research on dropout rates in higher education. Additionally, the study selected the Survival Analysis Model by John Grant (as mentioned in Rupert & Miller, 1998) from a statistical perspective. Finally, the Cox's Proportional-Hazards Model (Cox, 1972) was chosen to further ground the datasets from a statistical standpoint.

The exploration of each model was divided into two phases. The first phase provided a contextual description of the models, while the second phase discussed the significance of the models within the context of this study's objectives. Based on the descriptions and significance of these models, the study identified a strong and adaptable model that best suited the higher education student context in SA. These models incorporated dimensions that thoroughly explained the reasons why students are at-risk.

1.7.3.1 Vincent Tinto's Longitudinal Model of Dropout

Contextual description of the model: In Vincent Tinto's (1975) study, the longitudinal model of dropout is defined as a process where a person's experiences in higher education, as measured by their integration into academic and social systems, continuously modify their commitments and goals, ultimately leading to their persistence or dropout. The theory takes into account the interaction between factors such as student family background, individual attributes, and pre-university schooling, which directly and indirectly impact student performance. Yet, Tinto (1975) emphasises that while these factors can predict student failure, it is more important to understand the reasons why become at-risk rather than identifying which students are at-risk.

Additionally, Tinto's (1975) findings suggest that not only are background characteristics and pre-university schooling significant, but also the student goal commitments, the institution's commitment, and the educational expectations delivered by the institution. Thus, it is crucial for HEIs to recognise that students choose to attend a particular university based on their own learning experience and expectations (Mupinga et al., 2006). These factors form the foundation of Tinto's Longitudinal Model of Dropout, which considers family background, individual attributes, and pre-university schooling as influences on student commitments and decisions to persist or dropout. Overall, the decision to persist or dropout is shaped by both academic and social dimensions, as well as the students' expectations and experiences from the university.

Significance of the model within the context of this study: Tinto's Longitudinal Model of Dropout presents a comprehensive framework that offers HEIs a holistic view of the dropout process. This framework considers various factors such as student educational expectations, goal commitment to complete their studies, and differing dropout behaviours, while also taking into account the institutional aspect. By utilising Tinto's framework, HEIs can approach the inquiry into at-risk students from a conceptual perspective rather than solely relying on searching for at-risk datasets. This means considering the various reasons why students may be at-risk, such as their academic performance, social behaviour, and background, through datasets that reflect these aspects. One advantage of adopting Tinto's Longitudinal Model of Dropout is the ability to collect variables related to student grade performance from integrated institutional Information Systems, such as the MAS. Additionally, information on student learning activities and their social interactions can be gathered and consolidated from the LMS. This integration allows for a comprehensive student-at-risk identification process, as Tinto's framework has the flexibility to outline the datasets necessary for successful identification.

In summary, Tinto's Longitudinal Model of Dropout provides HEIs with a comprehensive framework for understanding and addressing the dropout process. By considering multiple factors and utilising integrated Information Systems, this framework has the potential to effectively identify and support students at-risk of dropping out.

1.7.3.2 The Survival Analysis Model

Contextual description of the model: According to Smith (2001), the Survival Analysis Model is a statistical approach used to study the time between entering into observation and a subsequent event, such death. The model was originally used to investigate mortality and morbidity, and gained momentum when John Graunt published a table on human survival processes in 1662 (Liu, 2012). Almost twenty years ago, Lu (2002) argued that conventional statistical methods like logistic regression and decision tree could predict customer churn to some extent. However, these methods were unable to determine exactly when a customer will churn or how long they could be retained.

Therefore, the Survival Analysis Model was applied to predict customer churn and improve customer retention. In the context of higher education, churn refers to datasets that help to identify student at-risk based on various factors, such as academic failure, pregnancy, off-campus living, poor performance, and even academic exclusions imposed by the university.

Significance of the model within the context of this study: The Survival Analysis Model can be utilised in this study to assess the progression of students throughout their academic journey, from beginning to completion, with the intention of identifying students who are at-risk of academic failure and those who face obstacles that may prevent them from finishing their studies. This model allows for the identification of the precise point in time at which students are most likely to fail or drop out. For instance, previous research by Murtaugh at al. (1999) found that dropout rates increased with students' age, but decreased with higher levels of pre-university education and higher GPAs during the first quarter of university. Additionally, literature indicates that on-campus students generally have lower drop-out rates compared to off-campus students, including international students.

Therefore, by applying the Survival Analysis Model, it becomes possible to predict which students are most at-risk of failing or dropping out and when they are likely to do so. However, it is important to note that this model primarily focuses on identifying the conditions under which dropout occur and when they are expected to happen, without considering the underlying reasons that lead to failure or dropout. Moreover, it fails to account for the fact that students' circumstances are not static and can change at any given moment, Consequently, HEIs may lack a comprehensive understanding of students who are at-risk and the process of dropout, as well as the ability to provide immediate support and early interventions.

Therefore, the utilisation of the Survival Analysis Model as a main framework in this study may present challenging due to the aforementioned limitations, considerations, and the model's predictive capabilities.

1.7.3.3 The Cox proportional-hazards model

Contextual description of the model: The Cox Proportional-Hazards Model, first proposed by Cox in 1972, is a widely used regression model in medical research for examining the relationship between patient survival time and datasets (Cox, 1972). According to Imbayarwo-Chikosi (2015), this model serves as a standard analysis for competing risks data by modelling cause-specific hazard functions based on a proportional hazards assumption. One of the key advantage of the Cox Hazards Model is its ability to incorporate survival data, duration, and response to estimate the expected time to failure (van Os et al., 2022). For instance, in the context of course completion, the observation of students' progress can be considered survival data, while the duration represents the time until course completion. Consequently, predicting the response will depend on historical predictions of withdrawal times made previously (Chimka et al., 2007).

Significance of the model within the context of this study: In this study, the Cox Propositional-Hazards Model is used to illustrate an example. The example involves students who choose certain courses that typically require a first language of English in high school, but there are also students with English as an additional language who are interested in the same courses. The proficiency in English can be seen as survival data, with the duration representing the time it takes for the additional language students to become proficient. The response in this case is influenced by historical data on the time to withdrawal for students with English as an additional language who had registered for similar courses. This highlights the importance of universities acknowledging the diversity in student populations and placing emphasis on integrating academic and language skills (Harris & Ashton, 2011).

Another example that is different from the previous one involves students from low socio-economic background who are first in their family to attend university. Delving (2010) argues that universities should strive to provide a successful experience for all students, including those from low socio-economic backgrounds who may be less familiar with tertiary education. Similar to the previous example, the observation of the low socio-economic background students can be seen as survival data, with the duration representing the time until they are offered in-course support tailored to their needs. The response here is influenced by historic data on the time to withdrawal for students from low socio-economic backgrounds.

When HEIs are interested in examining the relationship between the survival time for first-year students and the factors that contribute to identifying students at-risk, the Cox Propositional-Hazard Model can be helpful. However, applying this model to the current study presents a challenge due to the presence of censored data. For example, there may be differences in the observation time for students with English as an additional language compared to those who had English as a first language. This limited observation time means that student data may be collected only until the completion of the first year, rather than the completion of the entire course (Lindsey, 2000). This study aims to use a suitable model that allows for interventions and close monitoring of student at-risk, from their first year to their completion and graduation.

1.7.4 Justifying and selecting the model that underpins the datasets

The statistical and theoretical models mentioned previously have different approaches and capabilities for identifying and mapping the datasets related to identifying student at-risk. This includes determining the appropriate variables to select when integrating the source system for LA purposes. Each model has a strong potential, flexibility, and inclusiveness in identifying student at-risk of failing and those who face challenging circumstances that jeopardise their ability to complete their studies, with the goal of retaining these students.

The results of this study indicate that Tinto's Longitudinal Model of Dropout (1975) offers a more comprehensive theoretical perspective on the process of identifying students at-risk compared to the statistical models, which mainly focus on the conditions in which risk factors occurs. Additionally, the study suggests that the identification of students at-risk should be viewed as a longitudinal process when modelling the risk dimension. Furthermore, both the Survival Analysis and the Cox Propositional-Hazard Models use censored student data, which often lead to incomplete information about the students (Tinto, 2006). Therefore, the study recommends using Tinto's Longitudinal Model of Dropout (1975) due to its strong, flexible, and inclusive theoretical standpoint and its ability to guide the investigation of students at-risk.

1.7.5 Conclusion: Preliminary literature review

The initial findings of this preliminary literature review indicate that a crucial initial step in any LA initiative is to integrate the separate information Systems used by different institutions. While access to higher education has improved over time, there is still a significant risk of dropout for both students at-risk and the overall student population (Chen, 2012). A study conducted in baccalaureate institutions in the United States found that despite spending over $9 billion on first-year students, 30% of them still dropped out (Aulck et al., 2016). Therefore, in order to develop and improve LA approaches, it is important to thoroughly understand the reasons why students are at-risk and how to provide personalised interventions.

The introduction of LA in higher education has empowered institutions to address the issues of students at-risk and dropout rates. This study suggests that insights gained from analysing student-produced data patterns, generated from integrated institutional Information Systems should not be disconnected from theoretical reasoning. Therefore, solutions to challenges faced by at-risk students can be based on well-established theoretical frameworks, such Tinto's Longitudinal Model of Dropout. This approach allows for a comprehensive assessment of the underlying concepts in the analysis of student-produced datasets and patterns, as well as the effects of integrated institutional Information Systems data. Consequently, HEIs can select and successfully implement personalised interventions for students identified as at-risk.

1.8 Research design and methodology

The purpose of this study was to investigate the CSF that contribute to the successful integration of Information Systems in institutions and their impact on LA at HEIs. The findings of this study present CSF that can be helpful to HEIs in identifying students who are at-risk. It is crucial to use an appropriate research methodology to ensure the accuracy and validity of the study (Quintão et al., 2020). In this section, the researcher outlines research design and methodology, including the following subsections: (i) research design; (ii) unit of analysis; (iii) instrument development; (iv) data sources and sampling; (v) research methods; (vi) data analysis.

1.8.1 Research design

According to Bordens and Abbott (2002), research design serves as a tool for researchers to make decisions regarding the design and implementation of their study, and how these decisions will impact data collection, analysis, and interpretation. In this particular study, the goal is investigate the CSF of integrating institutional Information Systems to facilitate LA, with the ultimate aim of developing model to identify at-risk students in HEIs. The study is based on empirical evidence.

The research design consisted of two dimensions. The first dimension focused on addressing the research questions, while the second dimension focused on the case study itself. In the first dimension, three phases were conducted:

- The first phase involved a review of the mechanisms required to integrate multiple institutional Information Systems for LA adoption.
- The second phase involved a review of the literature on the data needed for LA and the corresponding requirements for systems integration.
- The third phase explored the application of educational-based analytics theory to establish a theoretical basis for the integrated datasets. Although various statistical models were discussed, there was no indication of their actual implementation. These statistical models would potentially be performed within an analytical system, with only the outcomes of the logical statements presented in this study. Therefore, no statistical analysis was carried out.

The three phases of the first dimension resulted in a Functional Requirements Specification (FRS) checklist of CSF necessary for implementing a successful LA project in higher education. The study then transitioned to the second dimension, the case study. Within this dimension, the study provided a narrative of a specific university in SA that had recently adopted LA. The results of the case study were then compared to the FRS checklist developed in the first dimension. This comparison allowed for an understanding of whether the observation (FRS checklist compared to the case study narrative) supported or rejected the hypothesis.

1.8.2 Unit of analysis

The researcher employed an exploratory case-study methodology, following Becker's (1970) definition of a case study as an empirical investigation that delves into a single case to gain an understanding of a phenomenon. In this case, a South African university of technology in the Western Cape was studied to explore the CSF of integrating institutional Information Systems to support LA. The unit of analysis for this research was the review of literature that informed the development of the CSF checklist. Additionally, secondary data from the Faculty of Engineering and Built Environments (FEBE) were examined to determine if the factors influencing student risks were predictive of at-risk students. FEBE was chosen to ensure that the study is logically feasible and manageable within the variable resources.

1.8.3 Instrument development

The instrument development for this study was carried out in two dimensions. The first dimension involved a comprehensive review of literature to determine the best mechanisms for integration multiple institutional Information Systems and identifying relevant datasets. Furthermore, a theoretical model was applied to categorise the factors that contribute to students being at risk.

The second dimension focused on comparing the outcomes of the first dimensions with the results of a case study to evaluate the effectiveness of the model in identifying at-risk tertiary students. involved reporting on the results from the contrast between outcomes from the first dimension and the footprint of the case study, and testing to see if the designed model was successful in identifying at-risk tertiary students. Student records from integrated source systems were used for this analysis, with the data anonymised and pressed through an Online Analytical Processing (OLAP) System.

1.8.4 Data source and sampling

According to Babbie & Mouton (2005), sampling involves selecting a small but representative portion of data that can answer research questions. In this study, a purposive sampling method was employed, intentionally selecting a sample that could provide unique information not attainable through other means (Teddlie & Yu, 2007). The purposive sample consisted of a range of 200 to 300 anonymous rows of student records.

1.8.5 Research Methods

The study employed a single case study to gather and analyse the data. Data was collected using a qualitative method, which involves direct fieldwork observations, in-depth interviews, and analysis of written documents. This study did not involve any statistical analysis.

To support the capabilities of institutional Information Systems, the study utilised a qualitative review of literature. This review helped in understanding the mechanisms for integrating multiple institutional Information Systems and identifying the necessary datasets for identifying at-risk students. The process involved qualitative observation of project conceptualisation and technical documentation.

Furthermore, the study solely relied on the outcomes of a query conducted using Learning Analytic System and secondary data containing student information. The purpose was to assess whether the designed model could successfully identify at-risk students.

1.9 Ethical considerations

Having read and understood the University of the Western Cape research ethics code of conduct, the nature of this study involves the acquisition of information gathered from the institutional information systems by the selected Faculty departments of a single university in the Western Cape, South Africa. The study obtained permission to research the Institutional Ethics Committee of the identified university prior to research. The well-being of the students took precedence over the anticipated knowledge benefits. Students have the right to refuse prior collection of their data through the institutional information systems. The selected university data and material brought into being is to be handled confidentially and professionally. Students have the right to:

- privacy,
- remain anonymous,
- respected confidentiality,

- no release of information inside and outside the university,
- fair and accurate evidence; and
- unbiased attitude.

It should be confirmed that the researcher adhered and adheres to the above. All task intentions and processes of analyses were transparent and sufficiently outlined to substantiate appropriation and expectations.

1.10 Assumptions and limitations

Assumptions: This study was conducted under the assumption that the selected university institutional Information Systems database was successfully integrated and that the university has instigated a Learning Analytic System.

Limitations: A limitation in this research was that it is a single case study of a single academic university and the researcher understands that the results of the model designed and tested in this study cannot be generalised across other academic universities. Moreover, the findings of this study were based on records analysis from secondary data, which may be open to measurement error, missing values, and unpredictable calculations.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

There has been a growing interest in LA for its potential to benefit students and improve teaching in higher education (Baker & Inventando, 2014). One important area of focus in LA is identifying at-risk students and those facing obstacles to completing their studies, as this can help improve student retention and success (Dietz-Uhler & Hurn, 2013).

Toetenel and Rienties (2016) note that HEIs have been adapting their practices through LA to better support student learning and retention in the diverse educational landscape, which includes traditional face-to-face, flexible, remote, online, and blended learning. This shift is driven by the increasing availability of digital data generated from institutional Information Systems and advancements in BI and analytics.

However, there is concern that HEIs are struggling to effectively implement LA interventions based on student data (Guitart & Conesa, 2016). The literature suggest that challenges include disintegration of Institutional Information Systems, managing and securing student data (Gursoy et al., 2016), lack of explicit theorisation of learning outcomes (Wilson et al., 2017), and issues with digital education governance (Williamson, 2016).

This gap between interest and successful application of LA has led to a need for a framework that can guide HEIs in adopting LA effectively (Kika et al., 2017). This literature review aims to address this gap between interest and successful application by examining the approach developed by McKinsey and Company (Barton & Court, 2012), which highlights CSF for integrating Information Systems to facilitate LA in HEIs. The approach focuses on three themes: data, model, and transformation.

Table 2.1 summarises this approach. While HEIs have opportunities to collect student data from various institutional Information Systems, the systematic adoption of LA is often lacking, as there is a need for a roadmap that incorporates lessons learned and best practices from BI and analytics (Siemens & Baker, 2012).

Table 2.1: *An approach for systematic implementation of LA in higher education institutions*

DATA	MODEL	TRANSFORMATION
Development of principles for creative data sourcing and data integration. Increasing awareness of data limitations. Securing necessary information technology support.	Following question-driven approaches to the applications of machine learning. Informing the use of machine learning by educational research and practice.	Development of Institutional policy and strategy for LA. Establishing effective leadership models to drive and oversee the implantation. Adopting principles for privacy protection and ethical use of analytics. Implementation of LA tools catering for the primary stakeholders. Development of analytics-informed decision-making culture.

2.2 Data: what is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems in SA HEIs for LA adoption?

Many previous studies in the field of Information Systems (I/S) have focused on the adoption, implementation, integration, and use of information Systems in higher education (Islam, 2011). These studies have highlighted the significant investments made by universities to implement institutional Information Systems in order to streamline administrative processes, enhance teaching and learning, and ultimately improve the quality of education (Deng & Tavares, 2013). Information Systems in this context refer to a combination of interconnected components that collect, process, store, and distribute information to facilitate decision making, control, analyses, and visualisation within an organisation (Laudon & Laudon, 2012).

Over the past decade, there has been a rapid integration of Information Systems and educational technologies higher education, driven by managerial and pedagogical objectives, as well as the need for greater flexibility in content delivery and student engagement with course material (Macfadyen & Dawson, 2010). As a result of these efforts, significant amounts of data have been generated from multiple independent institutional Information Systems, serving as valuable resources for studying student-institution interactions.

2.2.1 Development of principles for creative data sourcing and data integration

Identifying the appropriate source of data from various Information Systems in higher education is a crucial aspect of any data warehouse program. Data sourcing involves systematically extracting data institutional Information Systems, profiling their properties, and creating relevant datasets for data integration (Hosseinpoor et al., 2018). This process insures that the uniform access to data is provided from multiple source.

Data integration, as defined by Zipkin et al. (2021), is a statistical approach that incorporates multiple data sources into an analytical framework. It aims to provide consistent access to data generated from the various source systems enlisted during data sourcing. According to Chaki (2015), information systems integration is the foundation of any analytics initiative and should be considered when advising on integration approaches for HEIs.

Chaki (2015) outlines five (5) key drivers to consider when determining an integration approach:

i. The nature of extraction process between source systems and consuming systems (push/pull).
ii. The type of connectors required for pulling data from source systems.
iii. The choice between using a data integration engine or a database engine for data transformations.
iv. The required outbound extract formats for consuming applications.
v. The need to address data security and comply with any country-specific data regulations during the integration process.

2.2.1.1 The nature of extraction process between source systems and consuming systems (push/pull)

In their study, Biplob et al. (2018) defined extraction as the process of retrieving data from a source system. The source system can generate either push or pull extraction logic, which is used database implementations that support the storage and analysis of historic data. This suggests that universities have the responsibility to choose an extraction method that aligns with their institutional settings. Consuming systems, as described by Guerrero (2021), are transactional systems that utilise relevant datasets from source systems to address specific design challenges.

Chaki (2015) states that if a university opts for a push-based extraction, the generated file from the source system is transferred to a secure area for the data integration process to verify its status. This is where the source extract file undergoes preparation before the extraction process begins. Even if universities choose a pull-based extraction mechanism from the source system, an extraction logic is still required for the integration process. Access to source system tables is necessary to run the query before sending it for extraction processing. Typically, the extraction process is considered the most time-consuming aspect of conceptual design, whether it follows the extraction, transformation, and loading (ETL) approach, or the extraction, loading, and transformation (ELT) approach into the LA system or consuming system (Prakash & Rangdale, 2017).

Therefore, universities must decide between push and pull extraction methods from source systems. It should be noted that if the extraction is pull-based (requiring access to source system tables), the university must determine the conditions for granting access to these tables. This means that universities can either allow direct access to their institutional source system tables or create a replica source database accessible to the integration process instead of the original tables, often in real-time (Chaki, 2015). However, implementing the extraction process between university source systems and consuming systems can introduce complexity and may require significant modifications. To address this challenge, universities should consider centralising their source systems using connectors that facilitate pulling data from source systems and a clearly defined extraction logic.

2.2.1.2 The type of connectors required for pulling data from source systems

According to Ebert et al. (2017), connectors are reusable components that can connect data between different source systems using definable extraction logic. Even though many modern institutional source systems are cloud-based, they still need to exchange data with other institutional Information Systems, like university legacy systems (Wortmann & Fluchter, 2015). This raises the question of what kind of connectors are necessary to pull data from source system tables. The specific type of connectors required depends on the nature of the institutional source systems. For institution that anticipate using SQL servers, DB2, Oracle, or XML for database access, Chaki (2015) suggests the use of Open Database Connectivity (ODBC) connectors, which can establish intricate connections from source systems data tables. On the other hand, institutions that anticipate using packaged applications such as ERP, SAP, Snowflake, or Tableau will find that application-specific connectors that utilise existing system reports as data sources are essential to the integration process.

2.2.1.3 The choice between using a data integration engine or a database engine for transformations

In the previous section, we discussed two sets of techniques for transformation raw data from the source systems into a data warehouse or a target database to achieve a specific objective: extract, transform, and load (ETL) and extract, load, and transform (ELT). The decision that HEIs need to make at this stage is whether to load data into the target system before or after transformation. Whether using ETL or ELT, the transformation of source data involves the following three steps:

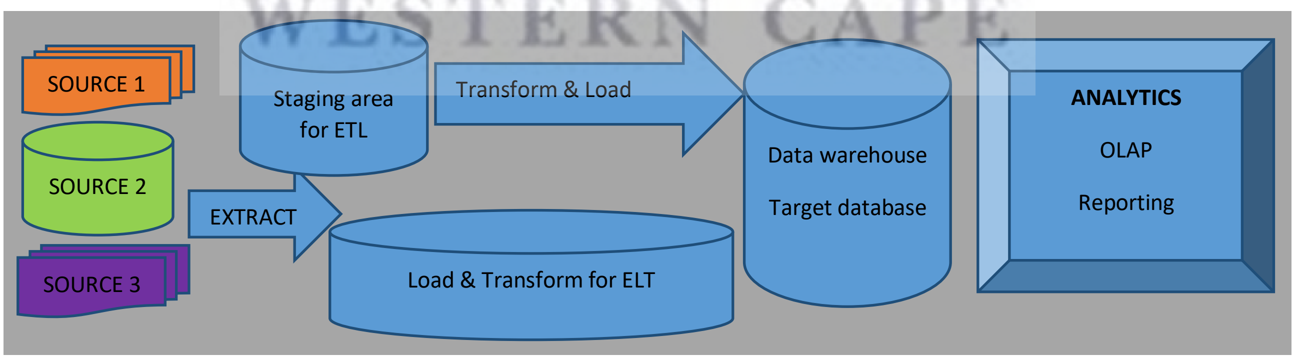


Figure 2.1: ETL/ELT approaches

Extract: This involves pulling data from different source systems and moving it to the target area. In ETL, the data is transformed according to the university rules before being loaded into the target database (Azaiez & Akaichi, 2018). In ELT, the data is loaded into target databases like Teradata, and transformation happens using fast data engines (Moly et al., 2019).

Transform: This step involves cleaning the data and applying transformation rules. HEIs can use this opportunity to ensure that the data is not corrupted during extraction or the subsequent ETL or ELT stages (Homayouni, 2018). ELT has an advantage here as transformation happens inside the database, eliminating the need to send transformed data over the network (Chaki, 2015).

Load: This is the process of loading extracted data into the data warehouse or target database in a format suitable for BI ready analysis.

Data warehouse technologies serve central repositories of integrated data, allowing universities to access and query data from source systems for analysis and reporting (Brum et al., 2019). On the other hand, a target database is the central database used in the ETL/ELT approach. BI involves technology-driven processes and strategies for collecting, integrating, analysing, and presenting business information to enhance decision-making (Garani et al., 2019). The goal of using a data integration engine or database engine for data transformation is to enable data integration for Online Analytical Processing (OLAP). Data warehouse systems connect multiple data sources and provide a central point for cleansing, transforming, and shaping data before implementing self-service models. It is important to note that BI data is stored in the data warehouse, and OLAP is one of the components of a BI solution. OLAP is cloud-hosted business intelligence and analytics platform that provides instant insights, interactive visualisations, and business intelligence know-how to users (Lachev & Prince, 2018).

2.2.1.4 The required outbound extract formats for consuming applications

In the overall integration process, automating the outbound extract formats required for consuming applications is crucial. This is especially important when a HEI has bilateral agreements with external parties such as other local and international tertiary institutions, financial institutions, or exchange partners. To ensure seamless data integration, it is essential to plan and define the outbound extract formats, like CSV, XML, Excel, HTML and PDF, along with the interface agreement (Chaki, 2015). This planning also includes determining the specific scope of extraction formats to be generated, sent, and shared with external parties. It is crucial to identify and distinguish these extraction formats during the integration design stage to prevent any potential data breaches during and after the integration process.

2.2.1.5 The need to address data security and comply with an any county-specific data regulations during the integration process

Education-based analytics have introduced significant data security risks in relation to BI and data warehouse technologies. When integrating data, it has been made clear that source systems share extract files that are either transformed and loaded, or loaded and transformed into a format suitable for analysis using database engines of the data warehouse or a target database. In order to ensure data security these extract files need to be protected using secure file transfer methods, both during the extraction process and in file landing zone (Chaki, 2015).

Therefore, it is recommended that HEIs consider data security requirements throughout the integration process and engage in activities such as the development of information security policies, information architecture, IT infrastructure, and compliance training to ensure the integrity of information security (Soomro et al., 2016). By doing so, data breaches during the integration process can be eliminated. Additionally, the study provides a detailed description of the integration approaches used to integrate multiple institutional Information Systems as source systems for LA purposes. The study then proceeds to examine the limitations of data during the data analysis stage.

2.2.2 Increasing awareness of data limitations

The successful implementation of well-established BI and analytics in the business sector has demonstrated the potential for applying education-focused analytics in higher education. However, the data integration strategy in higher education settings differ from that of BI and analytics in business. The existing literature on information systems integration and data exchange in BI and analytics in business primarily focuses on operational systems such as Customer Relations Management (CRM), trading partners, and suppliers (Pokhriyal & Jacques, 2017). On the other hand, HEIs often rely on multiple unintegrated institutional Information Systems as source systems. These institutional Information Systems include Learning Management Systems (LMS), admission files, Library records, Marks Administration Systems (MAS); university services usage, Social Media (SM), Student Success System (similar to

Customer Relations Management in business), Student Information Systems (SIS), and Enterprise Resource Planning (ERP) system. Each of these systems represent different aspects of student-institution interaction and engagement. Although HEIs have numerous institutional Information Systems representing student-institution interaction, this study focuses on four (4) key systems (LMS, MAS, SIS, and ERP) to demonstrate the integration process for educational analytics purposes.

Learning Management System (LMS): The LMS is defined by Gautreau (2011) as "a self-contained webpage with embedded instructional tools that permit faculty to organise academic content and engage students in their learning." LMSs such as Blackboard, Moodle, Canvas, and iSpring Learn are widely used internet technologies that support the shift from traditional (face-to-face) to online learning, as well as the adoption of blended (hybrid) teaching and learning practices (Fathema et al., 2015). An average LMS typically includes a content area for instructors to upload relevant files and multimedia resources, assessment tools for group or individual assignments, collaborative tools such as discussion boards and chat for real-time communication, course management tools for marks management and task information, and a retention centre to identify at-risk students. LMSs collect functional data that describe student learning activities and online/blended learning behaviours, which can be analysed to identify students at-risk of academic difficulties.

Marks Administration System (MAS): The Cape Peninsula University of Technology's 2018 general handbook described MAS as a system that allows staff to record, evaluate, and upload student marks, as well as generate class lists (CPUT, 2018). However, Fakude et al. (2014) argue that the increasing enrolment targets in South African universities have led to human errors in using MAS due to its time-consuming and tedious nature. Despite this, a survey conducted by a university of technology found that 85% of respondents agreed with the accuracy of information generated by MAS, 90% agreed with its usability, and 88% indicated that MAS was always accessible. Additionally, 90.5% stated that they always received the necessary information from the system and were satisfied with it (Irakoze, 2015). Based on these findings, the researcher affirms that MAS, along with the grade centre within the LMS, are reliable institutional Information Systems that can be utilised to extract student performance data during the student-at-risk data analysis process.

Student Information System (SIS): The SIS serves as a user-friendly interface for creating, maintaining, and managing accurate academic records of students, including course details, course progress, completed and upcoming semesters, curriculum details, batch executive details, and placement details, years, faculty information, and other resources relevant for student reporting purposes (Bharamagoudar et al., 2013). By using SIS data, HEIs can generate reports on the faculty in which a student is enrolled, the courses they have registered for, the exams they will undertake, and certifications, along with other necessary documents, such as proof of employment for part-time students. SISs are typically web-based and integrated into a university's website, allowing students to update their information as needed, leading to more efficient student record management. To ensure security, students are encouraged to access the system using their student login credentials, which enable the system to retrieve their data.

Enterprise Resource Planning (ERP) system: The ERP system, as described by Bidgoli (2004), automates the management of internal and external information in areas such as finance, human resources, sales, service information, and student information management. ERP systems facilitate the flow of information between various functional areas within the university and its external relationships. These systems integrate data generated across campus and standardise processes for capturing, processing, and disseminating the data (Grant et al., 2006). As a result, universities must either conform their processes entirely to ERP system or customise the system to align with their existing procedure. This presents a challenge as the pedagogic and socio-economic forces driving the incorporation of Information Systems and internet technologies into teaching and learning require a flexible approach to data management and delivery (Yoloye, 2015). This includes online applications, registration, content delivery, communication, synchronous and asynchronous learning, and pedagogic enhancements, all of which necessitate a dynamic structure and management of data.

Educational technology is driving various teaching and methods, including chalk-and-talk, blended, flexible, online, and remote approaches. Diana Laurillard (2013) suggests utilising educational technologies similar to the LMS to provide diverse learning opportunities for students. These methods have been employed by HEIs for nearly two decades, integrating traditional and flexible learning experiences (Garrisin & Kanuka, 2004). They have proven effective in in accommodating diverse students with different challenges (Alammary et al, 2014).

Consequently, a significant amount of student learning data has been generated, enabling the use of LA to measure variations in academic performance and success. This section aims to shift the focus from integrated source systems to contemporary literature, exploring how different datasets can impact the process of identifying at-risk students. It is important to note that the relevant datasets for identifying at-risk students encamps both academic elements and personal dynamics.

The study examines eight (8) datasets within the SA higher education context. The first set focuses on academic failure risk factors, such as students who have not accessed the institutional LMS for a specific period, those with minimal activity in the LMS, missed deadlines, and average grade performance.

The second set considers conditions that may threaten students' ability to complete their studies, including linguistic barriers, conflicting peer-group interactions based on ethnic and cultural values, on-campus and off-campus challenges (transportation, time and cost), and pregnancy.

2.2.2.1 Students who have not accessed the institutional LMS for a specific period

In an effect to improve student retention, universities utilise educational data analytics to identify and support students who have not utilised the institution's LMS for a specific period of time. This support aims to increase their access to the LMS and enhance their engagement with their studies (Williams et al., 2018). However, students in developing countries like SA face limited internet access due to socio-economic conditions. As a result, they often rely on the university's internet as well as other external locations with free Wi-Fi, such as shopping malls, popular family restaurants such as MacDonald's and Kentucky Fried Chicken (KFC), and local internet cafes (Dillahunt et al., 2014). This reliant on external Wi-Fi make it less likely that students will frequently access the institutional LMS and other online educational technologies.

It should be noted that SA students have developed various coping and support mechanisms to overcome these challenges. For instance, more privileged students who can frequently access the LMS do so on their smartphones, downloading the necessary materials and sharing them with their peers who do not have the same level of access. This may create a misconception that these privileged are not at-risk, in contrast to the underprivileged students from low socio-economic backgrounds who either lack internet access or do not own smart devices. Consequently, the chosen OLAP system may flag the underprivileged students as being at-risk, even though they may be actively completing and submitting their tasks on time without accessing the LMS.

2.2.2.2 Students with minimal activity in the LMS

Research indicates that the frequency of students clicks in the institutional LMS and their interaction with course content, as well as the time spent in online courses, can predict variations in final grade performance (Smith et al., 2012). However, Tempelaar et al. (2015) argue that the number of clicks made by eight hundred and seventy-three (873) students in two (2) blended courses could only explain a marginal four (4) percent of the variance in final exam performance. Furthermore, Bailey et al. (2015) found that out of fifty-one thousand three hundred and six (51, 306) students enrolled in their Massive Open Online Courses (MOOCs), twenty-seven thousand six hundred and seventy-nine (27, 679) students demonstrated above average click activity, but only seven hundred and ninety-five (795) completed the course.

Based on these findings, one could argue that there is a minimal correlation between the number of clicks in the LMS and student performance in an exams or their overall final grade. In developing countries like SA, ICT competencies are a scarce (Kirilidog et al, 2018), underprivileged students may access their online courses on campus or from locations with free Internet access. These students may click on various course content areas as they acquaint themselves with the LMS. This behaviour leads the LMS to record their clicks as they navigate the system. the number of clicks accomplished by the students trying to find their way through the system.

Consequently, analytic system removes these students from the at-risk list and marks them as active participants. Meanwhile, students who are familiar with the LMS log in, access the necessary course content, click on relevant materials, and then log out. Due to lack of computer literacy and Internet access, the majority of students may have a high chance of being flagged as at-risk due to their fewer clicks compared to average number of clicks in the class. In addition, if lecturers promise to load content at specific times but fail to do so, students will repeatedly log in and review the course pages to check for the promised material. This behaviour leads to inaccurate records and compromises the effectiveness and reliability of using the ‘least number of clicks’ as a predictor variable.

2.2.2.3 Missed deadlines

According to Falkner's (2012), knowing when students initially submit their assignments and their current course level can be a reliable indicator of their performance. However, it is important to consider that SA students often submit their work at the last minute. This increases the likelihood of unsuccessful submissions due to technical and administrative issues. Therefore, HEIs have a responsibility to offer significant technical and administrative support before the submission deadline if they identify "missed deadlines" as a risk behaviour. Without this support, the prediction of at-risk behaviour may not be accurate (Judd et al., 2010).

In some cases, students in the same course program may receive multiple assignments from different subjects that are due around the same time. This can be overwhelming for students and results in missed deadlines or low-quality work, leading to poor grades. Therefore, to accurately assess the effectiveness and relevance of the "students who missed deadline" variable, a comprehensive evaluation of students' progress in the course and the support provided by faculties and departments is necessary.

Additionally, the shift from traditional to online learning and the increase in Blended learning practices in countries like SA have been experimental for both lecturers and students. Challenges such as limited access to computers and smartphones, a lack of computer literacy when it comes to typing assignments, the distribution of multiple subject assignments within a course, are significant obstacles faces by SA HEIs. All these factors undermine the effectiveness and relevance of the "student- who-missed-deadline" variable.

2.2.2.4 Average grade performance

In order for educational analytics research to effectively predict grade performance, it is crucial to consider the specific context in which online learning and blended practices are used. This information must be taken into account before merging log-data to create a generalised model (Gasevic et al., 2016). Previous studies by Falakmsir and Jafar (2010) and Macfadyen and Dawson (2012) have shown that analysing different types of log-data, such as discussion forum activity, tests, and assignments, can provide valuable insights into students' final grades.

However, a survey conducted by Unwin et al. (2010) revealed that many lecturers in African countries have limited knowledge and interest in using LMSs to predict students' grades. This highlights the importance of increasing the use of LMS in higher education before implementing predictive analysis. Venter et al. (2012) suggest that initiatives should be taken to enhance the perceived usefulness and attractiveness of institutional LMS before examining predictor variables related to student performance.

Another complication is that HEIs often focus on students who received less than fifty percent (50%) grade mark. This can overlook interventions that may be needed for students who initially perform well but experience a decline in their grades. For example, a student may start a course with an average mark of eighty percent (80%) but see their average drop to sixty percent (60%). It is important to give equal attention and interventions to these students as well.

2.2.2.5 Linguistic barriers

South African HEIs are increasingly popular among both local and international students and lecturers. However, they face challenges such as limited English proficiency and cultural diversity (Ralarala, Pineteh, & Mchiza, 2016). This is a challenge for SA students as well. Most universities in SA are recognised as English speaking, but many students in Africa and from other countries speak English as a second or third language (Ralarala et al., 2016).

To determine the impact of language barriers on academic success, it is important to consider social exclusivity and cultural integration (Cummin, 2008). Adaptive learning approaches have been developed to address the diverse English proficiency levels among students. By identifying and supporting students facing language barriers, universities can help them feel accepted and included, and improve their chances of completing their studies successfully (Becker et al., 2017).

Adaptive learning is a pedagogical approach that supports self-regulated learning and can be applied in both formal and informal teaching and learning settings (Dabbagh & Kitsantas, 2012). For example, universities can provide recorded lectures, offer academic literacy development programmes, create multilingual glossaries, and offer other supportive resources. By doing so, SA universities can identify students with language barriers early on and provide the necessary support to make language barriers a more effective and manageable variable in the academic context.

2.2.2.6 Conflicting peer-group interactions based on ethnic and cultural values

Another variable that literature suggests can indicate risk is peer-group interactions. Tinto (1976) explains that these interactions result in collective affiliations, which are considered important social rewards that influence an individual's evaluation of the costs and benefits of attending university and shape their educational and institutional commitments. Lecturers gather data from records of social interaction practices, such as discussion forums, consultations, email correspondence, and peer and self-assessments.

However, countries similar to SA face various social risk factors, including conflicting ethnic and cultural values or traumatic peer exchanges and social interactions (King, 2004). This means some students may choose to submit group assignments individually or interact with peers of certain gender in discussion forums. As a result, their involvement in peer-group interactions is limited. To avoid misinterpreting this lack of group work as a risk indicator, it may be possible to consider the student's progress in other academic activities, even those meant to be completed as a group.

Furthermore, research shows that students may hold prejudice against gay and lesbian peers and lecturers, which may explain why some students avoid working with such peers or consulting with such lecturers (Ewing et al., 2003). This research suggest that peer-group interactions and consultation logs may not be reliable predictors in diverse cultural contexts, especially those similar to SA with a multitude of cultures and languages.

2.2.2.7 On-campus and off-campus challenges (transportation, time, and cost)

The literature suggests that the issue of students living on-campus versus off-campus is an additional variable that can indicate risk. However, a study by Frazier (2009) found no significant differences in grades between these two groups of students. Theis support the idea that the educational context, particularly teaching and learning practices, is changing. Therefore, it is important to consider the specific nature and context of a university before applying the predictor variable based on on-campus versus off-campus living arrangements.

For example, with current teaching and learning practises such as blended learning and adaptive learning, students who live sixty (60) kilometres away from campus may not worry about missing the first period. Lecturers can use flexible content delivery methods like recording lectures and creating podcasts. These approaches not only help students who have missed classes due to distance or language berries, but also assist lecturers who are not native English speakers. This highlights the importance of innovative approaches like blended learning, flexible learning, and adaptive learning, which diminish the predictive power of variables such as on-campus or off-campus residence and English proficiency (Ng et al., 2017).

Furthermore, Muslim et al. (2012) found that on-campus students face challenges related to their living environments that can impact their well-being. For example, a student living on campus who is hungry may not perform as well as a well-fed student living off-campus. These are important considerations that SA higher education practitioners should take into account when assessing the effectiveness and applicability of on-campus and off-campus student variables in identifying students who may face obstacles to academic success and completion.

2.2.2.8 Pregnancy

In traditional teaching and learning methods, unplanned pregnancies created various challenges such as disruptions to academic programs and financial burdens on institution, families, and individuals (Naidoo & Kasiram, 2014). However, the rise of online adaptive learning and blended teaching and learning approaches has reduced the impact of pregnancy on students' ability to complete their studies and graduate. Nonetheless, SA students and their institutions still face difficulties in implementing these innovative teaching practices, making it hard for students to fully rely on blended learning. This challenge is particularly significant for students with Internet access or necessary technology, as well as for lecturers who are resistant to changing their teaching methods.

For instance, even if both the lecturer and the pregnant student can easily adapt to online and blended learning, the student still requires a stress-free environment and ample rest during pregnancy. Moreover, after giving birth, the student must juggle attending classes, taking care of the baby, and helping with household chores. This can create significant stress in terms of time, cost, and pressure. Furthermore, it is crucial for HEIS to be aware of the risk factors that may hinder students from completing their studies and graduating.

Therese risk factors, not covered in this research, exist outside student's academic and social trajectory yet may still be connected to their personal situations. Examples chronic illness, troubled household dynamics, and bereavement, amongst other categories. While acknowledging these "unknowns," this study focuses on academic elements and personal dynamics of students that can be traced back that can be traced through institutional Information Systems like LMS and other sources.

These datasets provide universities with the necessary information to identify predictor variables for LA purposes. Additionally. The study considers the context, effectiveness, and transferability of the identified predictor variables, aiming to understand the environment in which they are applied. The table below represents the relevant datasets identified from literature and mapped against the source systems. These datasets aim to provide insights that can inform the design of different LA interventions.

Table 2.2: Mapping out the relevant datasets against the institutional Information Systems

Relevant datasets (predictor variables)	**Institutional Information Systems**
Minimal activity (those who have not accessed the institutional LMS over specified period).	Learning Management System (LMS)
Tool use frequency (those with least number of clicks in the LMS)	Learning Management System (LMS)
Student participation in discussion forums	Learning Management System (LMS)
Low assignment grades (those who missed submission deadline)	Learning Management System (LMS)
Students performance on activities that affect their final grades	Marks Administrative System (MAS)
Time spent on activity (pace)	Learning Management System (LMS)
Average grade mark performance	Marks Administrative System (MAS)
Student background	Student Information System (SIS)
Race	Student Information System (SIS)
Absenteeism (absent without leave)	Learning Management System (LMS)
Gender	Student Information System (SIS)
Demographics	Enterprise Resource Planning (ERP)

Black male	Student Information System (SIS)
Students on Financial Aid	Enterprise Resource Planning (ERP)
Comes from average secondary schools	Student Information System (SIS)
Poor grades	Enterprise Resource Planning (ERP)
Not proficient in English	Student Information System (SIS)
Academic and epistemological obstructions	Psychological Counselling Service System (PCSS)
Intrusive Advising	Psychological Counselling Service System (PCSS)
First generation in their family to be at university (lack of parental engagement)	Enterprise Resource Planning (ERP)
English as a second or third language	Enterprise Resource Planning (ERP)
Students from average secondary schools	Enterprise Resource Planning (ERP)
Poor grades	Enterprise Resource Planning (ERP)
Necessary employment	Enterprise Resource Planning (ERP)
Residence	Enterprise Resource Planning (ERP)
Transportation	Enterprise Resource Planning (ERP)

2.2.3 Securing necessary information technology support

The adoption and implementation of LA require the active participation and support of multidisciplinary teams and stakeholders, including senior leaders, information technology units, teaching and learning units, faculty representatives, students, at-risk officers, ethics committees, and legal departments. Their involvement is crucial, as highlighted by Tsai et al. (2018). Macfadyen et al. (2014) conducted a study that confirmed the complexity of higher education and developed a policy and planning framework to guide multidisciplinary teams in integrating and optimizing their institutional source systems with LA.

Senior leaders are instrumental in driving and ensuring adherence to the institutional policy for the LA project. Meanwhile, the IT department is responsible for developing the data sourcing process based on input from other stakeholders such as at-risk officers, faculty, and students. Lecturers play an active role in the teaching process, and students are active participants in their own learning (Gasevic et al., 2015). Therefore, ethical and legal considerations must be addressed upfront, including obtaining informed consent from students to use their data, involving ethics committees, and having legal departments evaluate policy frameworks before deciding on question-driven approaches or using machine learning based on educational research and practice.

2.3 Model: which data is necessary for conducting LA, and what are the resulting requirements for systems integration?

Once the process of data sourcing and data integration have been put into effect and established within an institution, the next step in the LA implementation journey typically involves designing and implementing LA models based on specific demands. These models are designed to address challenges related to identifying at-risk-students or determining factors that contribute to student success. For the purpose of this study, the researcher focus on the case study subject area of "at-risk identification." The term "at-risk" came into widespread use in the 1980s and refers to students who are at risk of not succeeding or graduating (Narrz'ello, 2014). This issue of at-risk students, defined as those who are in danger of dropping out of university due to academic failure or personal problems, is a significant problem on a global scale in higher education (McMillan & Reed, 1994).

HEIs around the world have made significant investments in educational technologies and institutional Information Systems, with the belief that these technologies will help address issues such as at-risk students (Kozma & Croninger, 1992). However, for an equal number of years that the term at-risk has been in use, educators and policymakers have been searching for strategies to address this issue. They have explored various solutions, including the use of new analytic technologies as strategic information systems (Darling-Hammond et al., 2014). Recently, the analysis conducted for the 2018 New Media Consortium (NMC) Horizon report reveals that analytics technologies are crucial components of the digital economy and are driving advancements in information systems and educational technologies in higher education (Becker et al., 2018).

2.3.1 Following question-driven approaches in the application of machine learning

Machine learning, analytics, and interpretations have a wide range of applications in the business sector, and these concepts can also be applied in the field of education-based analytics. Goyal & Vohra (2012) defined machine learning as a way for universities to analyse and interpret meaningful knowledge from large datasets extracted from institutional source systems and other educational technologies. The goal is to identify patterns and provide instant interventions and support. By integrating data from various sources, universities can explore different machine learning techniques to examine the generated datasets.

This process involves discovering patterns help analyse the data in the database. Bhargava et al. (2013) explain that machine learning techniques can be used to reason about educational data and find patterns and consistency within the data sets. In addition to analysing the integrated institutional data sources, universities should also focus on pedagogic strategies that prioritise student-centred instruction and provide multiple pathways for learning. This ensures that the extracted and loaded student learning pathway records in the target database are meaningful of high quality, thus enabling analysis and reporting using OLAP systems like Power BI and Pyramid Analytics.

Power BI already include data mining techniques such as cluster analysis, decision tree, The data, factor analysis, and regression analysis. These techniques help extract meaningful insights institutional data sources. The researcher informs the reader about the broad nature of these data mining techniques to foster an understanding of their application in higher education and the key concepts involved.

2.3.1.1 Cluster analysis

Clustering analysis, which divides objects into homogeneous groups for comparison, is commonly used in LA to assess the significance of variances in student data (Scott & Knott, 1974). For instance, it can be used by HEIs to group students based on similar behaviour, such as students staying off-campus. This allows for more targeted interventions and support.

In addition, clustering analysis can be utilised by lecturers offering early morning classes to identify students who face distance challenges. By analysing student data, the lecturers can send personalised podcasts of missed or late-arriving class sessions to help these students catch up. Various clustering algorithms, such as the connectivity, centroid, distribution, destiny, and subspace models are available for this purpose (Goyal & Vohra, 2012).

2.3.1.2 Decision tree

A decision tree is a decision support system that uses a tree-like graph technique to classify and predict categorical variables. It enables the creation of accurate predictive models, allowing users to extract meaningful information (Yu et al., 2010). In the context of educational data, a decision tree can be applied to estimate student grade performance based on their learning activity levels in the integrated LMS. It can also analyse admission criteria for mainstream courses by modelling the overall grade performance of students in their first year.

One limitation of decision is the need to classify numerical attributes already in a specific order to determine where to split a node, which can be time and memory consuming (Ben-Haim & Tom-Tov, 2010). Therefore, it may not be ideal for categorical variables with diverse levels. In such cases, factor analysis, an alternative approach designed for similar situations, may be more suitable.

2.3.1.3 Factor analysis

Factor analysis is a statistical approach that helps researchers simplify and understand the relationships between different variables (Yong & Pearce, 2013). For example, it can be used to assess the performance of candidates in an entrance test and determine if the test accurately measures what is necessary for admission. This technique allows for the development of theoretical constructs (Williams et al., 2010), and can improve the understanding of how variables are related. To make factor analysis more accessible, newer methods like exploratory factor analysis and confirmatory factor analysis have been developed, aiming to demystify the process (Comrey & Lee, 2013).

2.3.1.4 Regression analysis

Regression analysis is a commonly used statistical technique that allows researchers to examine relationships between variables (Chaubey & Bhattacharya, 2015). In simple terms, regression analysis collects data from students, creates a model based on that data, and then uses regression to assess how well the model fits the data. One common application of regression analysis in higher education is predicting the performance of at-risk students or those facing challenges that may hinder their success. For example, to predict the grades of students with below-average marks, researchers look at variables such as tool use frequency in the LMS, participation in discussion forums, and time spent on activities. They also consider independent variables such as lack of subject interest, poor relationship with lecturers, and ineffective teaching strategies. There are various regression techniques available in the literature, including linear regression, stepwise linear regression, and multiple linear regression.

Linear regression analysis involves a single regressor *(x)* that is related to a response variable *(y)* through an unknown intercept (β_o), slope (β^1), and a random error (ε) (Montgomery et al., 2012). The equation of linear regression is: $y = \beta_o + \beta^1 x + \varepsilon$. Stepwise regression, on the other hand, is a method used in educational research is used to select relevant variables and evaluate their importance (Thompson, 1995). Multiple linear regression is employed when the dependent variable is analysed in relation to various independent variables of interest (Cohen et al., 1983).

In higher education, linear regression analysis is applied to predict student performance, such as their academic achievements. Stepwise linear regression help predict the time students spent on the LMS (pace). Multiple linear regression is used to identify multiple variables that can predict students' success in university or their first-year grade point average. To aid in these analyses, machine learning techniques such as cluster analysis, decision tree, and factor analysis, as well as regression analysis on historic educational datasets, are commonly used. These datasets are extracted, transformed, and loaded into a target database for analysis and reporting using analytical systems such as Power BI and Pyramid Analytic systems.

2.3.2 Informing the use of machine learning by educational research and practice

Most HEIs in Africa, including SA, have responded to the need for equity by increasing enrolment. However, they have not adequately addressed the issue of ensuring equity in student success (Mohamedbhai, 2014). This has led to the #FeesMustFall movement in SA, which focuses on issues such as food and housing for enrolled students, curriculum transformation, financial barriers, and academic exclusion (Dominguez-Whitehead, 2017). Despite these efforts, it is still unclear how successful students from low socio-economic backgrounds are graduating. In 2015, undergraduate success rates were higher for students studying in person compared to those in distance programs. Additionally, White and Indian/Asian students had higher success rates than other groups, while African students had the lowest success rates (Department of Higher Education and Training, 2017).

In 2017, President Jacob Zuma announced fee-free education for poor and working-class students starting in 2018. While attention was given to the socio-economic impact of free education on access, there were limited measures in place to ensure student retention and success. As a result, there was a need to develop LA approaches to identify students at-risk and minimise waste of public funds on students who do not graduate (Chui et al., 2018). The University Capacity Development Plan (UCDP) called on universities to develop strategies to analyse student data and create early warning systems for at-risk students (Ministerial Statement, 2018-2020).

However, there is scarcity of flexible predictor variables and transferable models in educational settings (Muthukrishnan et al., 2017), Therefore, it is important to prioritise high-quality research on LA models and their predictor variables. While the importance of predictor variables for at-risk students has been recognised globally, there is limited research on their effectiveness and transferability across different contexts (DeZure et al., 2012). The following section provides a conceptual perspective on the effectiveness and transferability of predictor variables in successfully identifying at-risk students in different academic settings.

2.3.2.1 The need for LA to be informed by educational research

The existing literature on LA is a compilation of educational data studies that provide clear datasets for identifying students who are at-risk. The aim of educational data analytics in higher education is to enhance student success by integrating datasets from multiple source systems into LA systems for analysis and evaluation. However, there is a lack of a guiding model for this research (Mattingly et al., 2012). Therefore, the first step in implementing educational research to increase student retention, improve student success, and reduce accountability for HEIs is to understand, based on student-institution interactions, the various reasons why students become at-risk.

Subsequently, inquiries about the datasets that contribute to the effective identification of at-risk students can be informed by theories. Instead of solely relying on literature searches, selecting datasets should involve a profound understanding and though examination of how these datasets impact the process of identifying at-risk students. For instance, it is important to consider at what level (institutional, faculty, departmental, program, course, or individual) these datasets have an effect on students and in what way. Furthermore, it is crucial to determine if the selected datasets have the same predictive capacity when applied in different educational contexts. The literature presents numerous datasets that play a significant role in the identification process of at-risk students. To ensure accuracy and objectivity, this study categorises multiple datasets into three (3) groups: (i) those with below-average grades; (ii) those who have been academically excluded or suspended; and (iii) those from low socio-economic backgrounds (Suh et al., 2007).

Students with below-average grades

Research on factors contributing to below average grades in students has revealed several common trends. Failures and the need to repeat failed modules, returning to education after a gap year, and extended completion times are all consequences faced by students with below average grades (Asikhia, 2010; de Valero, 2001; Mugali et al., 2017; Peterson & Barrett, 1987). Previous studies have attempted to identify the key factors that play a role in these lower grades, with findings suggesting that factors such as a lack of subject interest, poor teaching strategies, negative relationship with lecturers, unfavourable learning environment, too much socialisation, and the necessity of part-time jobs all contribute to lower grades (Wadesango & Machingambi, 2011).

However, when conducting LA research, it is important to create models based on individual courses before merging the data to develop a generalised model for predicting grade performance. This is because the adoption and application of LA, online learning, and blended practices can vary significantly between courses (Gasevic et al., 2016). The shift from traditional to online learning, as well as the increase in blended learning practices, has fundamentally changed the higher education landscapes (Chaubey & Bhattacharya, 2015). Blended learning, defined as the integration of digital technologies to enhance both student and lecturer experiences, has become a prominent approach (Doolan & Guiza, 2015).

The predictive power of analytics related to students below average marks in the context of student at-risk has been a source of frustration. Various predictors, such as students' engagement with the LMS, assignment grades, activity levels, pace, and time spent on each activity, have been examined (Smith et al., 2012). Studies similar to Falakmsir and Jafar (2010) have shown that student participation in discussion forums is a particularly strong predictor of their grades. Moving forward, it is crucial for LA research to focus on analysing patterns of student engagement, understanding the reasons behind their engagement, understanding the reasons behind their engagement or disengagement, and exploring their participation in the LMS (Macfadyen & Dawson, 2012).

By harnessing data from institutional Information Systems, specifically the MAS used by academic departments (Bytheway, 2000), in conjunction with the integrated LMS to track student attendance (Ezen-Can et al., 2015), group participation, and consulting with lecturers, it becomes possible to identify students at-risk with below average marks.

Students who have been academically excluded or suspended

Academic exclusion or suspension is when a student is not allowed to attend university for a certain period of time. Despite being used as a warning by universities, there is limited evidence that excluding or suspending misbehaving and poor-performing students actually improves their behaviour (Losen & Skiba, 2010). In fact, being excluded or suspended can cause emotional distress for these students and have negative consequences for both the universities and the communities they come from. A study by Noltemeyer and Ward (2015) also found that there is an inverse relationship between academic exclusion or suspension and student outcomes, meaning that students are more at-risk of not succeeding or graduating when they are suspended.

This suggests that academically excluded or suspended students often do not return to university once their suspension has been lifted. To really understand the connection between academic exclusion or suspension and poor academic performance, engagement, and attendance, it is important to consider socio-economic disparities among students, such as their such as their backgrounds, demographics, race, and gender, which can contribute to their behaviour (Toldson et al., 2015). For example, Rooney's (2015) research findings has shown that black males who are on financial aid, come from average secondary schools with poor grades, and have limited English proficiency are more likely to be excluded compared to their white female peers who are ineligible for financial aid and proficient in English.

These findings highlight the need for universities to provide supportive interventions such as psychological counselling services to address the underlying reasons for academic struggles and misbehaviour. Intrusive advising is often recommended for at-risk students, especially those from minority or disadvantage backgrounds, to help them identify and develop strategies to succeed academically (Walter, 1998). By using personalised approaches and intervention offers that acknowledge the students' efforts, universities can support at-risk students and prevent the need for exclusion or suspension (Fowler & Boylan, 2010). This way, student perception of university responses such as academic exclusion or suspension, can be guided and supported by consolidating results into conceptual models aimed at identifying students at-risk and offering a personalised relationship with such students through offers of intervention that acknowledge the student's own efforts (Candela et al., 2015).

To better understand and address the risk of academically exclusion or suspension, universities should collect data from various source systems, such as counselling service, financial status, high school and university performance, demographics, participation and engagement, and progress reports. This data can be sued to create conceptual models and analytics that help identify students at-risk and provide early intervention.

<u>Students from low socio-economic background</u>

Devlin (2008) found that the graduation rate of students from low socio-economic backgrounds in higher education has remained at approximately 15 per cent for over fifteen (15) years in Australia. Similarly, in SA, Letseka and Maile (2008) reported that out of 120 000 students who enrolled in higher education in year 2000, 36 000 (30%) dropped out in their first year and an additional 24 000 (20%) dropped out in their second and third years. Furthermore, 70% of these dropouts were from low economic background families. To address this issue, some countries, including SA, have implemented subsidised or fee-free education to increase access for students from low socio-economic backgrounds.

Even in developed countries like the United States, the graduation rates of students from low socio-economic backgrounds have not caught up with those from higher socio-economic backgrounds, despite an increase in enrolment numbers (Castleman & Long, 2013). Dietrichson et al. (2017) highlighted the significance of socio-economic status as a predictor of educational achievement and argue that it is possible to improve the educational outcomes of students from low socio-economic backgrounds. For instance, some Australian regional universities use a *Student Readiness Questionnaire* to identify at-risk students before they begin their studies, allowing the university to provide early support and allocate resources effectively (Purnell et al., 2010).

A report from UNESCO has proposed that an inclusive report has suggested that an inclusive higher education system should provide opportunities for exceptional grades, access, a positive atmosphere without exclusion or suspension, participation, multilingualism, and success regardless of socio-economic background (Unesco, 1994). Devlin (2010) argued in a conference proceedings that as the number of enrolled students increases, universities should strive for a successful experience for all students, particularly those from low socio-economic backgrounds who will now be studying alongside conventional students. However, Okioga (2013) noted in more recent studies that the parents of "conventional" students are more engaged in their children's education and development compared to students from a low socio-economic backgrounds.

Typically, students from low socio-economic backgrounds are the first in their families to attend university, resulting in a lack of parental engagement. Therefore, it is crucial to have a comprehensive understanding of how students from low socio-economic backgrounds are recognised (or misrecognised), how they access education, and their experiences (Fataar, 2018). As a result, this study suggests that HEIs should develop capabilities similar to psychological counselling services and intrusive advising to identify students from low socio-economic backgrounds and allocate resources to support them in enhancing their educational achievements.

The table below illustrates the three data categories emphasised in the literature that can inform the use of machine learning in educational research and practice: students with below-average grades, students who have been academically excluded or suspended, and students from low socio-economic backgrounds.

Table 2.3: Data categories informing the use of machine learning by educational research and practice

Data categories based on student academic elements and student personal dynamics
Below average marks   • Lack of subject interest

• Poor teaching strategies by lecturers • Poor relationship with lecturers • Unfavourable learning environment • Too much socialisation • Necessary part-time jobs
Exclusion and suspension • Below-average academic performance • Poor academic engagement • Misbehaviour
Students from low socio-economic background • *Poor academic achievement* • *Lack of parental academic support* • *Financial constraints*

2.3.2.2 The need for theory-informed use of LA

Despite the fact that most LA approaches rely on data analysis and analytical skills, there is a growing interest in applying theoretical concepts to guide the identification of at-risk students. This can involve using machine learning techniques to analyse datasets and conceptual understanding of the at-risk process. For example, a study on technology use by underserved students found that academic factors such as lack of support and learning English, as well as personal dynamics such as employment and pregnancy, contribute to the increase in at-risk students (Zielezinski & Darling-Hammond, 2016).

To support this study, the literature was reviewed and three potential statistical models or theories were identified to analyse the datasets. Vercellis (2011) defined a statistical model as a set of mathematical models and analysis methodologies that use available data to make informed decisions. In this study, the statistical models aim to identify the categories that contribute to students being at-risk. The Survival Analysis Model and the Longitudinal Model of Dropout are two potential frameworks that can provide insights on the reasons and consequences of students being at-risk.

The Survival Analysis Model

According to Ameri et al. (2016), the Survival Analysis framework is a set of statistical methods used to predict student dropouts before graduation. The framework focuses on estimating the time it will take a student to drop out. The results of the study show that the framework is successful in accurately predicting when a student will dropout. The framework uses various datasets, such as demographics, family background, financial information, high school records, and university enrolment data.

However, it is important to note that the framework assumes that the students' conditions will remain constant until they dropout, which may not always be the case. The main objective of the study is to identify at-risk students and provide interventions to prevent dropout. The framework considers multiple variables that influence at-risk behaviour.

One challenge with the Survival Analysis Model is that it primarily focus on time-varying factors and does not comprehensively address multiple risk factors for students and the dropout process as a whole. Additionally, the framework predicts future events based on students' current identity and conditions, which may raise ethical concerns. Prinsloo and Slade (2013) argue that students' identity and conditions are context-dependent and can change over time, potentially invalidating the predictions made by the framework.

Considering these challenges to the model's predictive power and ethical considerations, using the Survival Analysis Model as the base framework for this study may be challenging.

The Longitudinal Model of Dropout

Vincent Tinto's (1975) study defined the longitudinal model of university dropout as an ongoing interaction between the individual and the academic and social systems of the college that constantly shape their goals and commitments, ultimately leading to their persistence or different forms of dropout. Tinto argued that while understanding the factors that predict academic failure or threatens students' ability to complete their studies is important, it does not provide insights into how these factors contribute to the dropout process. Tinto attributed the lack of understanding about the university dropout to a focus on identifying predictor variables without a conceptual framework to guide the inquiry (Pascarella & Terenzini, 1980).

While the field LA has traditionally relied on data-driven approaches rather than theoretical reasoning, this study adopts a theoretical perspective to ground LA. In addition to deriving empirical approaches from the literature review, this study utilises Vincent Tinto's (1975) Longitudinal Model of Dropout theory as a framework to uncover the theoretical aspects of the analysed datasets. According to Tinto (1975), the longitudinal model of dropout involves the continual interactions between the individuals and the college's academic and social systems, leading to modifications in their goals and commitments that result in persistence or various form of dropout.

Pitre (1990) provided a definition of theory as a coherent explanation or description of observed or experienced phenomena. This study selected Vincent Tinto's (1975) Longitudinal Model of Dropout as a foundation to understand how integrated source systems' datasets contribute to the identification of at-risk students. Tinto's theoretical framework was used to gain knowledge on the factors influencing students' risk of dropping out. His model also serves as a guide for future studies on education-based data modelling using datasets from HEIs.

Vincent Tinto's theoretical model draws form Durkheim's theory of suicide in sociology and concepts from the field of educational economics, specifically cost-benefit analysis (Tinto, 1975). Tinto's model has remained a fundamental reference for understanding at-risk students, providing insights into the interaction process that lead to different students in higher education to engage in risky behaviours (Tinto, 2017). While Tinto's (1975) Longitudinal Model of Dropout focuses on the interaction between student attributes and the influences and expectations from academic and social systems in university, Spady (1970) argues that the at-risk process should also consider these academic and social systems (Spady, 1970).

Both Johnson's (1965) exploration of Durkheim's theory of suicide and the application of cost-benefit analysis in education economics (Koch et al, 2015) have influenced Tinto's model. These frameworks are employed in the institutional model of dropout. Stengel (1964) defines suicide as personal unhappiness and the perception that integration into society cannot rectify this condition. Similarly, Shi et al. (2015) links students at-risk with low academic performance, high opportunity cost, and low socio-economic status, which are factors that can be explained using cost-benefit analysis.

The figure 2.2 below depicts Tinto's (1975) conceptual schema for identifying at-risk students who are likely to drop out of university:

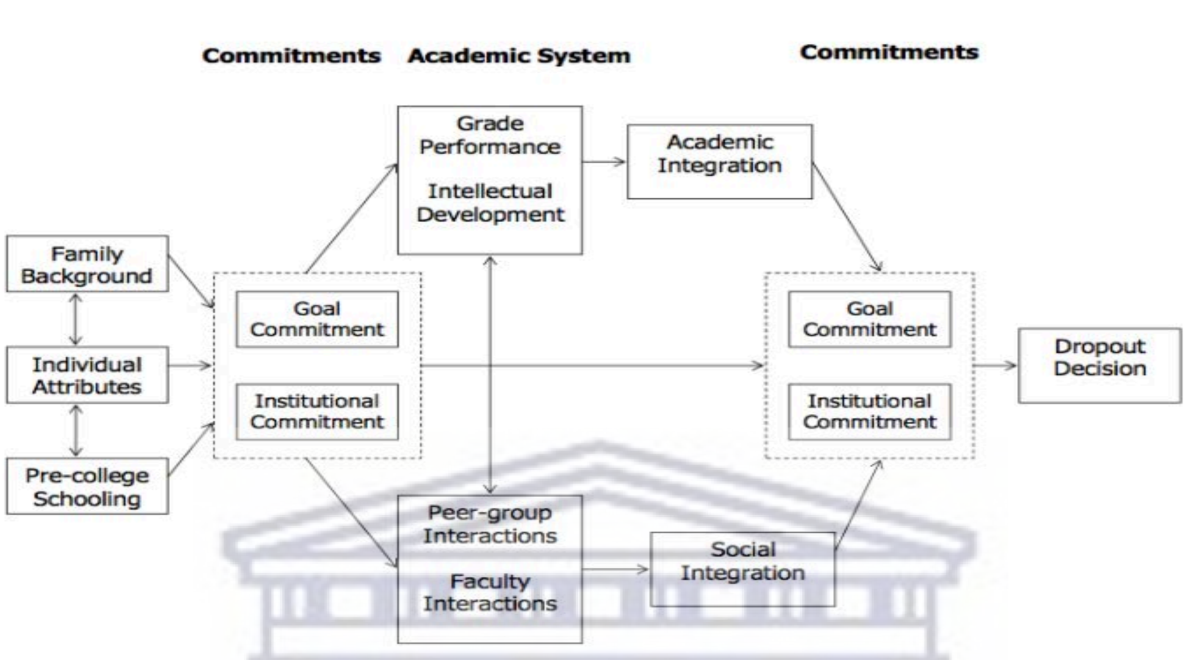


Figure 2.2: Tinto's (1975) conceptual schema for dropout from university

The conceptual framework of the student at-risk process, as presented by Pascarella and Chapman (1993), offer valuable insights into factors that influence student persistence in completing their studies. These factors include the student's attributes, family background, pre-university schooling, and interactions with the academic and social systems of the university. The likelihood of students becoming at-risk increases when there is a lack of integration between the academic and social systems, which then impacts the student's decision to commit or discontinue their studies. It is important to consider the educational history and expectations that a student brings to university, including their family background, individual attributes, and pre-university experiences .

In a study conducted by Welsh et al. (2001), the relationship between the academic and the social systems of the university was examined. The findings of this study revealed a direct influence of academic achievement on social competence, and a reciprocal relationship between social competence and academic achievement. The literature on Tinto's Longitudinal Model of Dropout has consistently validated the model and its potential implications, with few negative criticism. However, some research has uncovered slight variations when the datasets are disaggregated from the framework. For instance, Terenzini and Pascarella (1980) highlighted the importance of student informal contact with faculty and background characteristics in relation to persistence and student-university interactions. They concluded that Tinto's model remains conceptually useful for future studies or practical actions. When the sample was disaggregated by gender, Terenzini and Pascarella (1983) found that persistent decisions were more influenced by the social systems for females, whereas for males, the academic system played a more significant role. On the other hand, Bean (1985) argued that informal faculty contacts were less influential than group work and student peer assessments in terms of socialisation and persistence.

When considering suicide as a result of unhappiness and social exclusion, it is possible to draw a connection between the causes of suicide and the risk of dropping out of university. This can be attributed to factors such as low academic performance, a lower socio-economic status, and a lack of integration between the academic and social aspects of university life. It is important to note that Tinto's model emphasises the significance of both social and academic integration in determining a student's decision to persist at university. Additionally, the background characteristics of students, including their family background, individual attributes, and pre-schooling, directly impact on how they interact with the academic and social systems of the university. Within this context, student peers and informal faculty contacts are essential for social integration, while grade performance, personal development, and academic self-esteem are crucial for academic integration.

Over time, Tinto's theory has successfully identified various processes and correlations that classify students at-risk of dropping out from higher education (Duarte, Ramos-Pires, & Goncalves, 2013). Consequently, numerous studies recommend applying this theory in experiments related to this topic (Kember, 1995). Thus, the findings presented in this study are derived from Vincent Tinto's (1975) Longitudinal Model of Dropout, which provides a conceptual framework for understanding the longitudinal process of student-institution interaction. This framework maps the conditions that contribute to a student's risk of dropping out and facilitate real-time interventions for analysis (Conijn et al., 2017).

Understanding the process of university dropout is a critical task that requires immediate attention, even before examining how certain factors contribute to dropout rates (Tinto, 1975). Tinto's framework provides guidance for studying the dropout process by considering different predictor variables. This framework explains the interaction between students and universities, and identifies the factors that influence various forms of dropout behaviour (Tinto, 2017). Tinto's model has become a fundamental resource for understanding the dynamics that lead different students to dropout. For instance, students from average pre-university schools who have English as a second or third language in an English-speaking university may be at-risk of dropping out due to social exclusion and below average grades. According to Tinto’s theory, teaching these students in English hinders their assimilation and acquisition of content, which is more significant that solely relying on predictor variables such as grades. Focusing on the “why” question, rather than relying on datasets, is crucial for intervention strategies. By understanding the different forms of dropout early on, universities can identify at-risk students and provide early support to encourage them to reconsider their decision and complete their studies.

Students from low socio-economic backgrounds are more likely to be at-risk of expulsion or suspension due to poor academic performance or misconduct. Once expelled, these students may never return to higher education or may seek alternative options upon the completion of their expulsion period (Hemphill et al., 2014). Following Tinto’s theory, this behaviour reflects a low commitment to graduation and a lack of commitment to both the academic and social systems. It could also suggest that even if students have a high commitment to the social system, their low goal commitments impede their success. Furthermore, students from low socio-economic backgrounds may struggle to integrate socially with their peers, leading to academic exclusion or expulsion.

In all these scenarios, Tinto’s model emphasizes that student goals and institutional commitment are the most significant predictors of dropout intentions. Recent studies that examined Tinto’s model not only highlight social risk factors similar to faculty and departmental visits, but also emphasise the importance of consultations with the institutional Centre for Innovative Educational Technologies (CIET) in promoting student success (Braxton, 2019). As a result, Tinto’s model enhances our understanding of the dropout process and has meaningful implications for developing LA approaches to identify at-risk students in higher education (Fortin et al., 2013).

The potential of education-based analytics to improve intervention for at-risk students in higher education has gained attention. However, there is a significant need for a theoretical framework to guide research on predicting at-risk students, as much of the current work in this area lacks a solid theoretical basis. Theoretical models, such as Tinto’s Longitudinal Model, can provide valuable insights into university dropout process and help institutions identify at-risk students (Nandeshwar et al., 2011) . This study emphasises the importance of Tinto’s longitudinal model as theoretical framework before seeking predictor variables for future predictive modelling studies.

There is a substantial amount of literature on at-risk datasets that can utilised to identify students at-risk of failing or facing challenges that hinder their ability to complete their studies. However, without a thorough understanding of the dropout process, this data lacks meaning. By gaining a comprehensive understanding of the at-risk process, relevant datasets associated with different risk behaviours can be extracted from integrated source systems. This will enable the creation of a model based on the guidelines of a BI framework.

2.4 Transformation (the institutional priorities in the adoption of LA)

The focus of transformation is on addressing implementation priorities at an institutional level by considering the comprehensive dynamics in LA. The comprehensive dynamics include various aspects such as building institutional policy and strategy, establishing effective leadership models, defining principles for privacy protection and ethical use of analytics, implementing LA tools for primary stakeholders, and fostering an analytics-informed decision-making culture.

2.4.1 Building institutional policy and strategy for LA

Recently, there has been a trend towards the development of a large-scale LA policy and strategy through a European research project team known as Support Higher Education in Integrating Learning Analytics (universally referred to as a SHEILA framework). This framework was derived from interviews with seventy-eight (78) senior staff members from fifty-one (51) European HEIs across sixteen (16) countries (Tsai Y. , Moreno-Marcos, Jivet, Scheffel, & Tammets, 2018). The SHEILA framework suggests six (6) dimensions for building institutional policy and strategy, including:

(i) mapping the political context
(ii) identifying key stakeholders
(iii) identifying desired behaviour changes
(iv) developing an engagement strategy
(v) analysing internal capacity for change
(vi) establishing monitoring and learning frameworks.

However, there has been a lack of guidance and alignment between the SHEILA framework dimensions and the successful implementation of LA encountered by LA practitioners. Therefore, studies similar to Broos et al. (2020) have explored principles or factors that can guide the alignment between the dimensions of the SHEILA framework and successful LA implementation. The coordination model suggests that LA policy building initiatives should coincide with a systematic implementation of LA, with a coordinated effort spread over time.

Additionally, regular evaluation of LA adoption is recommended to ensure alignment with institutional policy and strategy (Tsai et al., 2021). Once the institutional policy and strategy for LA are established and coordinated, the next step is to establish leadership models to drive the implementation.

2.4.2 Establishing effective leadership models to drive and oversee the implementation

The implementations of LA in educational institutions involves various aspects, such as integrating multiple Information Systems, designing models, mining data, and presenting information to different stakeholders (Wise & Vytasek, 2017). This technical component plays a crucial role in driving the implementation process. However, it is equally important to involve stakeholders in establishing effective leadership models for adopting LA. To achieve this, it is necessary to understand the perspectives of faculty, lecturers, and students regarding the institutional-wide adoption of LA (Herodotou et al., 2020). This involves faculty representatives understanding and accepting LA (Rienties et al., 2018), motivating lecturers to support LA, and ensuring students are aware and accepting of the changes (Song & Kong, 2017). By considering the perspective of all stakeholders, the institution can determine if it is ready to fully embrace and implement LA.

2.4.3 Defining principles for privacy protection and ethical use of analytics

As Information Systems and internet technologies become more integrated into higher education, it is important for ethical guidelines to be developed and followed in regards to the ownership and protection of student produced data (Ferguson, 2012). Collecting and analysing student data should be done with a critical perspective, taking into account the and analysis methods used (Slade & Prinsloo, 2013). Ensuring informed consent and upholding data privacy and protection are key factors in the classification and management of this data.

2.4.3.1 The analysis, informed consent, and the location of student produced data

The analysis of student-produced data by HEIs poses a threat to student privacy. Sharing misleading information could compromise a student's identity (van der Bank, 2012). To address this, universities and students should establish a mutually agreed upon method for obtaining informed consent that allows the universities to use the student-produced data. This method should also allow students access to their data and provide clear, valid reasons for the collection and analysis. Additionally, students should have an option to have their analytical record cleared upon leaving or completing their studies.

The clearance of student analytical records is based on the understanding that a student's identity and circumstances are temporary and context-dependent. For example, if a student is identified as at-risk based on their lack of engagement in a discussion forum, real-time intervention may be necessary to help them improve. If support interventions are successful and the student become more engaged and committed to their studies, their socio-economic status may change. Therefore, the information collected to predict and understand their learning needs and performance should be transparent and impermanent.

2.4.3.2 Privacy and management of student produced data

In a study by Pardo & Siemens (2014), it has been proposed that the stakeholders in higher education have an obligation to discuss the four fundamental principles to privacy: transparency, student control over data, security, and accountability. This is important to ensure compliance with current laws, regulations, and societal requirements. Section 14 of the Constitution of the Republic of South Africa (1996) protects the right to privacy, stating that individuals have the right to privacy and should not have their person, home, property, possessions, or communications violated. Consequently, universities must not compromise the right to privacy in relation to student data collected from any interaction with the university, including applications to study, registration, learning activities, grades, and performance.

The Protection of Personal Information (PoPI) Act (2013) has the main objective of ensuring that all South Africans institutions handle personal information in a responsible manner. This includes collecting, processing, storing and sharing personal information from another entity. The Act (2013) holds institutions accountable and ensures they will be held responsible if they abuse or compromise personal information in any way. Ideally, students should give informed consent, understanding the purpose of data collection and that its use in LA will be limited to understanding and predicting their personal learning needs and performance.

2.4.4 Implementation of LA tools catering the primary stakeholders

According to Gasevic et al. (2015), analysing data produced by students when interacting with information systems and the internet technologies has the potential to advance our understanding of learning science. Over time, universities have shifted from teacher-centred to student-centred approaches to learning, and now, utilising LA can enhance learning environments through data-informed decision-making (McCabe & O'Connor, 2014). In student-centred approaches, students take on more responsibility for their own learning, which heavily relies on their individual attributes.

Both theoretical and empirical evidence in the learning sciences confirm that a student's attitudes, character, dispositions, skills, and values are a complex blend that influence their deep engagement in the learning process (Shum & Crick, 2012). Unfortunately, when these attributes are not acknowledged or accommodated, a student's ability to complete studies and graduate may be hindered. Therefore, it is essential for lecturers, management, and education practitioners to leverage LA in conjunction with the learning sciences and pedagogic practices. Thus approach would accommodate student attributes, assess their critical thinking, resilience, and social skills, and provide valuable feedback.

2.4.5 Development of analytics-informed decision-making culture

One of the challenges that institutions face as they shift from traditional educational environments to online and blended learning practices has been how to create active and interactive learning environments developed by an analytics-informed decision making culture and related to student needs. Essentially, online and blended learning environments are intended to individualise education. However, while online and blended learning practices attract a diverse range of students, Vanslambrouck et al. (2018) note lecturers' lack of comprehension of analytics-informed, decision-making culture, which in turn leaves them in the dark when it comes to anticipating students' individual needs. In contrast, the development of analytics-informed decision-making culture related to students' individual needs could be adapted as a pedagogic strategy that creates student-centred instructions that provide students with multiple learning pathways (Munene, Darby, & Doherty, 2015).

In fact, an effective online and blended learning environment offers a more inclusive and equitable learning experience for all students and in many ways provides students with greater educational access (Cunningham, 2014). Supporting research suggests that student confidence, enjoyment, and motivational strategies are essential for student learning success beyond assessments and grade performance (Ferguson, 2012). In addition, course redesign for modern pedagogic requirements, assessment activities, and feedback seems to authenticate the online formative assessments in higher education (Gikandi et al., 2011). In turn, Lynam & Cachia (2018) argue that assessments and feedback in higher education remain the area of concern for students for the reason that very little research has been investigated to consider students' experience of assessments.

For the LMS to integrate collaborative and interactive learning activities, institutional and sociocultural commitment from all stakeholders is necessary (Dias & Diniz, 2013). Such collaborative participation and commitment permits personalised feedback from the lecturers to their students using the LMS-developed analytic models to improve and optimise their learning experience. This means that the perspectives of a student-centred environment will be more important than those of the lecturers and the institution. In turn, this allows institution to provide greater levels of flexibility and choice to all students.

2.5 Conclusion

The study conducted a literature review to explore the CSF of Information Systems integration needed to support LA at HEIs in SA. The review employed two structuring principles:

- The conceptual framework proposed by McKinsey and Company (Barton & Court, 2012), which identifies the mechanisms required to integrate multiple institutional source systems and the necessary data for LA, along with the ensuing system integration requirements.
- Tinto's (1975) Longitudinal Model of Dropout, used to provide a theoretical basis for the datasets analysed in the study.

Firstly, the literature review introduced McKinsey and Company's framework (Barton & Court, 2012), highlighting the importance of source extraction mechanisms for integrating institutional

Information Systems. This step was found to be crucial for all LA initiatives. Once the source extraction mechanisms were defined, and the source systems were integrated into a coherent data repository (such as a data warehouse or target system), HEIs were deemed closer to their goal of utilising data analysis for decision-making.

The table represented below summarises the datasets generated from the integrated institutional source systems. These include categories that influence the factors leading to students being at risk.

Table 2.4: Summary of the datasets, the source systems in which the datasets are derived from, and examples of the reasons why students come to be at-risk.

Datasets (predictor variables)	**Information Systems (source systems)**
Data Categories Based on Student Academic Elements and Personal Dynamics	
Below average marks Lack of subject interest| Poor teaching strategies by lecturers| Poor relationship with lecturers| Unfavourable learning environment| Too much socialisation| Necessary part-time jobs	
Minimal activity (those who have not accessed the LMS over specified period).	Learning Management System (LMS)
Tool use frequency (those with least number of clicks in the LMS)	Learning Management System (LMS)
Student participation in discussion forums	Learning Management System (LMS)
Low assignment grades	Learning Management System (LMS)
Students performance in activities that affect their final grades	Marks Administrative System (MAS)
Time spent on activity (pace)	Learning Management System (LMS)
Average grade mark performance	Marks Administrative System (MAS)
Exclusion and suspension Below average academic performance| Poor academic engagement| Misbehaviour	
Student background	Student Information System (SIS)
Race	Student Information System (SIS)
Absenteeism (absent without leave)	Learning Management System (LMS)
Gender	Student Information System (SIS)

Demographics	Enterprise Resource Planning (ERP)
Black male	Student Information System (SIS)
Students on Financial Aid	Enterprise Resource Planning (ERP)
Comes from average secondary schools	Student Information System (SIS)
Poor grades	Enterprise Resource Planning (ERP)
Not proficient in English	Student Information System (SIS)
Academic and epistemological challenges	Psychological Counselling Service System (PCSS)
Intrusive Advising	Psychological Counselling Service System (PCSS)
Students from low socio-economic background Poor academic achievement| Lack of parental academic support| Financial constraints	
First generation in their family to be at university	Enterprise Resource Planning (ERP)
English as a second or third language	Enterprise Resource Planning (ERP)
Students from average secondary schools	Enterprise Resource Planning (ERP)
Poor grades	Enterprise Resource Planning (ERP)
Necessary employment	Enterprise Resource Planning (ERP)
Other	
Residence	Enterprise Resource Planning (ERP)
Transportation	Enterprise Resource Planning (ERP)

The review progressed to the second argument, which focused on profiling the relevant datasets that were extracted, transformed, and loaded from the source systems into the data warehouse of the institution. During this argument, the study examined the effectiveness and applicability of datasets developed in a different context. These datasets included information on at-risk students who were likely to fail academically, such as those who had not accessed the institutional LMS within a specified period, those who had spent the least amount of time on course content, those who had missed deadlines, and those with below average grade. Additionally, the study also considered students who faced various challenges that hindered their ability to complete their studies, such as language barriers, conflicts with their peer group due to ethnic and cultural differences, students living on-campus versus off-campus, transportation issues, time constraints, costs, and pregnancy.

It is important to note that the term “different” here refers to datasets created for developed countries but adopted by developing countries, with different context and challenges. For example, students in SA may face security concerns that prevent them from using their electronic devices in public transportation, whereas this may not be the case in the USA. This study argued that by considering both the first exploration and the second argument of the literature review, the findings can inform the design of a model that can help HEIs identify at-risk students more effectively.

Furthermore, the study delved into the third debate discussed in the second part of the literature review, which focused on the theoretical capabilities needed for building an LA model, and identified Tinto’s Longitudinal Model of Dropout (1975) as a preferred theory that underpins the datasets discovered from the source systems.

The results of the literature review provided insights into some CSF that are essential for the implementation of any LA initiative in HEIs. The following table presents a checklist of CSF that HEIs should consider when implement their LA initiatives.

Table 2.5: Critical Success Factor (CSF) checklist acquired from the review of literature

Functional Requirement Specification (FRS) checklist of Critical Success Factors (CSF) required.
Data *What is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems for LA adoption?* ☐ Has the university built a data informed culture in decision making for LA based on a predefined design challenge, for example, the student at-risk identification? ☐ Has the university established an effective source systems integration approach, and understand the limitations in which the identified source systems data may have to offer? ☐ Has the university secured the necessary information technology support and other stakeholders similar to the institutional ethics and legal department?

Model
Which data is necessary for conducting LA, and what are the resulting requirements for systems integration? - Are the datasets generated from the integrated information systems used to inform model design primarily for the design challenge? - Is the model design informed by educational research and practice? - Has the university reviewed the efficacy and transferability of datasets that are developed in foreign context? - Has the university avoided prioritising question-driven approaches to the application of LA (data driven) and rather designed models informed by educational theories capable to account for contextual factors? - Did the university avoid making use of LMS vendors or external data specialist to create its dashboards and started developing its own dashboards, using its own data specialists?
Transformation
What are the institutional priorities in the implementation of LA? - Has the Executive Management been involved in the creating of the institutional policy and strategy for LA, in order to drive and oversee the implementation? - Has the university considered the legal and ethical implications on the use of student data?

Concluding this literature review, the analysis conducted here represents a ground-breaking approach to design. By examining the CSFs identified in Chapter Four and mapping them against the LA project (case study) of a specific university in SA, the study tested the effectiveness of a model in identifying at-risk students. The results obtained from aligning the CSF with the project groundwork, as well as testing the model's ability to improve LA output, provided insights into the hypothesis testing of the study. This study suggested that if HEIs can gain an understanding of the crucial factors for integrating information systems to facilitate education-based analytics, they can enhance their analytics reporting and improve LA outcomes.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

In Chapter Two (2) of the literature review, the first dimension of the research design was examined. This first dimension focused on the research questions by (i) reviewing literature on the state of source extraction mechanisms that are crucial for integrating multiple institutional source systems for LA adoption, (ii) reviewing literature on source systems for relevant datasets to inform the model design in the student at-risk identification process, and (iii) reviewing a theory used in education-based analytics to ground the datasets from the integrated source system. The literature review resulted in a FRS checklist of CSF needed to implement LA in higher education.

To ensure an effective research methodology, the choice of data collection techniques, research design construction, and analysis process were aligned with the central research question. The central research question focused on the critical success factors of Information Systems integrations necessary to facilitate LA at HEIs in SA.

In this chapter, the researcher adopted a research methodology construction based on the theoretical concept of the research 'onion' proposed by Saunders et al. (2009) and revised in 2019 (Saunders et al., 2019). While research on source systems integration is not new, the integration of source systems for education-based analytics is a relatively new field for scientific studies (Viberg et al., 2018). Therefore, it is important to investigate the CSF that influence successful implementation of LA and establish theoretical frameworks.

Figure 3.1 illustrates the six (6) layers of the research 'onion'. The researcher started peeling these layers from the outer layers towards the core, explaining and justifying the significance of each peel in relation to the research questions (Sahay, 2016). The outer layer began with delineating the research philosophy, followed by choosing approaches to theory development, methodological choices, strategy(ies), and defining time horizons. These layers ultimately led to the core of the onion, which focuses on the research design, techniques, and procedures for data collection and analysis (Abdelhakim, 2021).

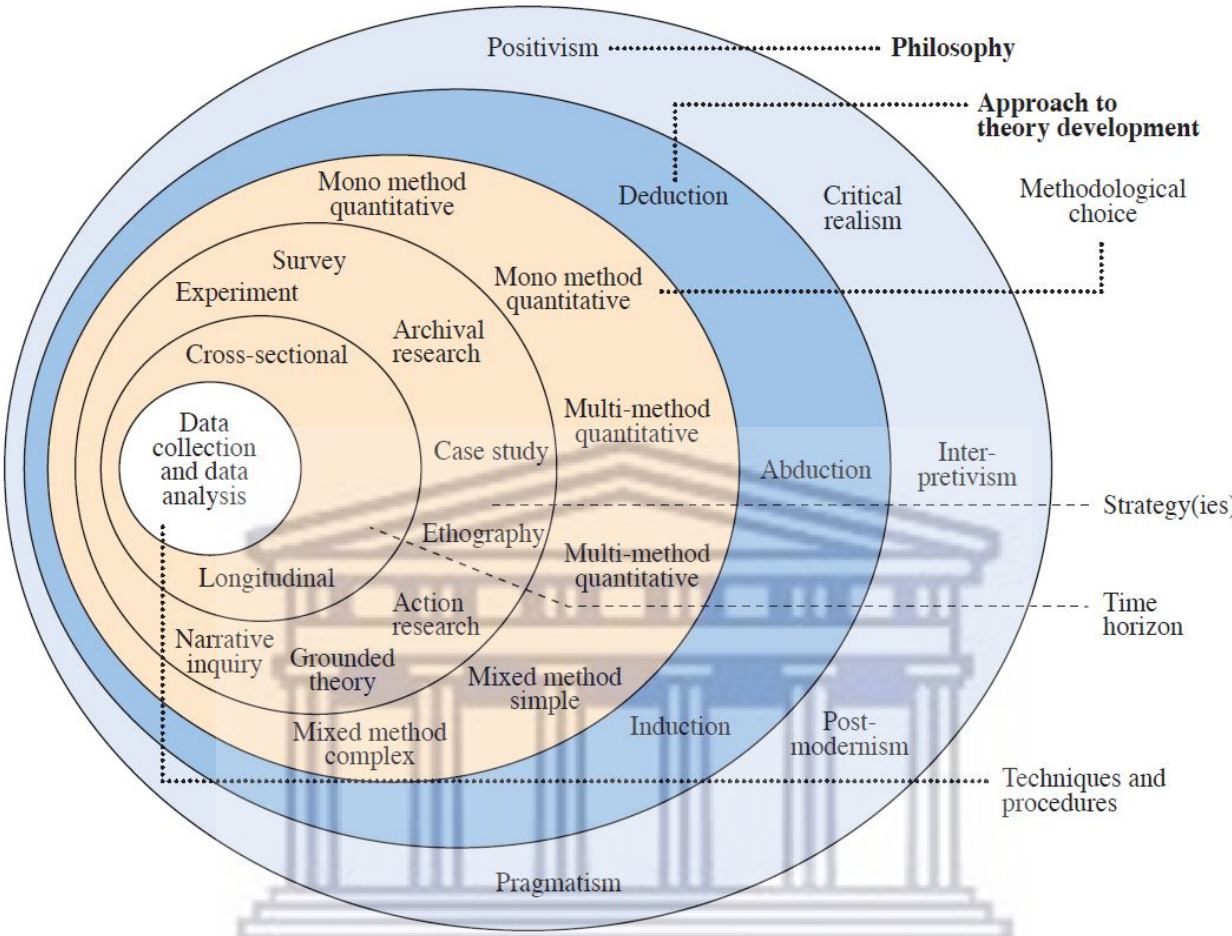


Figure 3.1: The research onion by Saunders et al. (2019)

3.2 Research philosophy

To begin exploring the research onion and its layers, the researcher must first consider different research paradigms such as positivism, critical realism, interpretivism, postmodernism, and pragmatism. These paradigms will inform the choice of research paradigm for this study.

Positivism: According to Marsonet (2019), positivism views knowledge as being based on empirical evidence and immediate observations. It involves articulating hypotheses and using observation to gain insights and understanding of behaviour.

Critical realism: As defined by Wynn & Williams (2012), offers a holistic approach to investigating phenomena by acknowledging the influence of external factors such as technology, environment, and social factors on behaviour and outcomes.

Interpretivism: Adopts the perspective that our understanding of reality comes from subjective experiences and social constructions (Walsham, 1995). This view recognises that a single phenomenon can have many interpretations, in contrast to positivism which focuses on knowledge derived from observation.

Post-modernism: According to Yin (2018), post-modernism is a complex field that challenges the principles of modernism and philosophy. It embraces ways of knowing and rejects the focus on a single approach to knowledge.

Pragmatism: As defined by Kaushik and Walsh (2019), embraces a variety of research methods and encourages researchers to use the approach that works best for a particular research problem.

In this study, the observable phenomenon involves the integration of Information Systems to facilitate LA in HEIs in SA. Additionally, data is required for LA to profile source systems and inform the design of a model for identifying at-risk students.

Based on these observable phenomena, the research is conducted as a qualitative study in the positivism position. However, it also involves the use of theory to unveil the profiled datasets and test the efficacy of the designed model in identifying at-risk students. This inclusion of multiple methods and the empirical testing of propositions aligns the study with a qualitative research approach. Williamson and Johanson (2017) defines "propositions" as broad statements drawn from theory for comparison with empirical evidence. As the study progresses, the researcher moves away from the limitations of positivism towards a post-positivism paradigm.

Post-positivism: is defined as a rich paradigm that combines elements of positivism and interpretivism, considering the experiences of the majority and incorporating histological, comparative, and phenomenological analysis (Panhwar et al., 2017).

3.3 Approach to theory development

Results obtained after peeling back the first layer of the research onion provided support for the identification of the philosophical paradigm employed in this study. This has led to the next layer of the research onion, which involves the approach to theory development. Saunders et al. (2009)

suggest that it is useful to align the approach to theory development with the chosen research philosophy. Given the research problem being addressed and the adoption of post-positivism philosophy, Mitchell & Jolley (2012) describes the research approach as a logical method of using impartial observation to formulate and test a phenomenon. Peeling back this this second layer of the research onion reveals three (3) main approaches to theory development: deductive, abductive, and inductive.

Deduction & Induction: are two contrasting approaches to theory development. Duductive reasoning involves moving from an existing theory and formulating a hypothesis to test this theory, before observing the data. On the other hand, induction works by observing specific data and using it to build a theory (Saunders et al., 2009).

Abduction: To address weaknesses found in both deductive and inductive approaches, the abduction approach is employed. This involves using incomplete observations to make logical inferences and compare different theories (Biskupek, 2019)

For this particular study, the selected approach to theory development is deductive reasoning. Gray (2009) explains that in deductive research logic, the process begins with an existing theory, followed by the formulation of a hypothesis based on this theory, collecting data to test the hypothesis, and finally analysing the results to either accept or reject the hypothesis.

The existing theory: The current theory suggest that HEI's possess valuable student-generated data but struggle to extract and integrate information from multiple systems. This prevents them from identifying students who are at-risk of failing or facing conditions that hinder their ability to complete their studies. As a result, early interventions cannot be effectively implemented using LA. This inability to identify at-risk students is directly linked to the lack of integration of HEI data sources into cohesive data repositories, which inhibits decision-making.

The formulated hypothesis: Based on this theory, the formulated hypothesis states that If HEIs can develop a comprehensive understanding of the CSF of Information Systems integration required to facilitate LA, they can enhance their analytics capabilities and improve LA outcomes.

The collection of data to test the hypothesis: To test this hypothesis, data was collected in two ways. Firstly, a meta-analysis of literature on the CSF of Information Systems integration necessary for education-based analytics were conducted. Additionally, relevant working documents, operations, and reports from a case study of an implemented LA project were gathered.

The second part of the data collection involved obtaining integrated student data from the systems described in the case study. This data was used to test whether the model that was designed, which is made possible through systems integration, could successfully identify at-risk students.

Analyse the results: To analyse the results, the study compared the Functional Requirements Specifications (FRS) checklist of Critical Success Factors (CSF) from the literature review in Chapter 2 against the information provided in the case study narration in Chapter 4. By assessing whether all the boxes are checked, the study either support or reject the hypothesis.

Table 3.1: Critical Success Factors (CSF) acquired from literature to be mapped against the case study groundwork

Functional Requirement Specification (FRS) checklist of Critical Success Factors (CSF) required.
Data *What is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems in SA HEIs for capabilities for LA adoption?* ☐ Has the university built a data informed culture in decision making for LA based on a predefined design challenge, for example, the student at-risk identification? ☐ Has the university established an effective source systems integration approach, and understand the limitations in which the identified source systems data may have to offer? ☐ Has the university secured the necessary information technology support and other stakeholders similar to the institutional ethics and legal department?

Model
Which data is necessary for conducting LA, and what are the resulting requirements for system integration? - Are the datasets generated from the integrated information systems used to inform model design primarily for the design challenge? - Is the model design informed by educational research and practice? - Has the university reviewed the efficacy and transferability of datasets that are developed in foreign context? - Has the university avoided prioritising question-driven approaches to the application of LA (data driven) and rather designed models informed by educational theories capable to account for contextual factors? - Did the university avoid making use of LMS vendors or external data specialist to create its dashboards and started developing its own dashboards, using its own data specialists?
Transformation
What were the institutional priorities in the implementation of LA? - Has the Executive Management been involved in the creating of the institutional policy and strategy for LA, in order to drive and oversee the implementation? - Has the university considered the legal and ethical implications on the use of student data?

In addition, out of the 200 rows of student secondary data, was the designed model able to identify students at-risk?

3.4 Methodological choices

The study explores different layers of the research onion, starting with a post-positivist approach and deductive reasoning, leading to theory development. The third layer of the research onion involves methodological choices, including options such as mono method quantitative, mono method qualitative, multi-method quantitative, multi-method qualitative, mixed method simple, and mixed method complex for conducting research.

Quantitative method: Aliaga and Gunderson (2002) defined a quantitative method as collecting numeric data and analysing it using mathematically based methods.

Qualitative method: Silverman (2016) defined a qualitative method as a theoretically driven social practices and experiences of how social phenomena are constituted in their natural setting in real time. Qualitative approaches have their philosophical roots in naturalistic investigations (Newman & Ridenour, 1998).

Mono method and methods: Choice of researchers to use a single data collection technique and corresponding analysis procedure such as questionnaires in quantitative design and in-depth interviews in qualitative design (Saunders & Tosey, 2012/2013).

Multiple methods: Choice of researchers to use more than one data collection technique with corresponding analysis procedure such as questionnaires and observations in quantitative design and interviews and diary accounts.

Mixed methods: A combinative research design that mixes both quantitative and qualitative data collection techniques, methods, approaches and corresponding analysis procedure into a single study (Onwegbuzie & Leech, 2005). In mixed methods, the researcher can either use qualitative analysis technique followed by a quantitative analysis technique, referred to a mixed method simple. On the other hand, a researcher can use a qualitative analysis technique to analyse quantitative data qualitatively, also referred to as mixed method complex (Felizer, 2010).

Although there are debates about the compatibility of quantitative and qualitative methods, mixed methods research has gained popularity in contemporary research (Johnson & Onwuegbuzie, 2004). Some argue that it fits well within post-positivist epistemology, while others suggest that pragmatism is the best paradigm for mixed methods (Kock, Gallivan et al., 2008). However, post-positivism is suggestive of mixed methods research and that pragmatism as a paradigm has no exclusive rights over mixed methods (Gallivan, 1997). As a result, post-positivism was chosen as the philosophical premise for this study in conjunction with the qualitative method.

3.5 Strategy(ies)

By peeling back the fourth layer of the research onion, the researcher was able to determine the appropriate research strategy to address the research question. The research question focused on the CSF of Information Systems integration for facilitating LA at HEIs in SA. Additionally, the study aimed to report the outcomes of a model designed to identify at-risk students using predictor variables from integrated Information Systems.

Various research strategies can be implemented in a study, including experiments, surveys, archival research, case studies, ethnography, action research, grounded theory, and narrative inquiry (Saunders et al., 2019).

An experiment strategy could be adopted to explore the relationship between two variables and inform decision-making (Barton, 2010). This strategy involves controlling one variable while measuring another to establish a cause-and-effect relationship.

A survey strategy, on the other hand, would be suitable when standardised information is needed for quantitative research related to the question (Pinsonneault & Kraemer, 1993).

Archival research strategy is useful when pre-existing data collected prior to the research is required. This includes sources such as consensus data, patent office records, credit history, and educational records (Das et al., 2018).

For a deeper understanding of a specific problem, a case study strategy can been adopted. This strategy involves methods such as participant observation, in-debt interviews, and longitudinal studies, emphasising qualitative analysis (Gable, 1994).

Ethnographic research, stemming from social and cultural anthropology, involves immersing oneself in the lives of the people being studied and spending considerable time in the field (Myers, 1999).

Action research strategy allows the researcher to generate knowledge and gain a better understanding of complex problems simultaneously (Baskenville, 1996).

Grounded theory strategy involves developing a theory grounded in empirical observation in the field of information systems research (Wiesche et al., 2017). It follows an inductive approach to theory development.

Lastly, narrative inquiry involves the documentation and analysis of contextually bound stories to gain rich and in-depth understandings (Tan & Hunter, 2003). This strategy focusses on telling compelling stories within a specific domain of discourse.

Therefore, it is crucial for a researcher to have a clear understanding of the study's nature before implementing various strategies and methods. Consequently, in order to evaluate the potential strategies suitable for this study, the researcher initially analysed the nature of the research. The study itself is exploratory since it delves into the CSF that contribute to the effective implementation of LA.

Specifically, it focuses on creating a LA approach that not only identifies at-risk students but also provides real-time personalised interventions. According to Robson (2002), an exploratory study involves investigating a phenomenon that has not been thoroughly explored in order to gain understanding, insights, and assess the phenomenon from a fresh perspective. At this point, the study has progressed to the stage where the research questions will be addressed using the nature of the study, the position of post-positivism, a deductive approach to theory development, and the qualitative method as the chosen methodology for conducting a case study.

Conducting a case study

According to Yin (2011), a case study strategy is a crucial aspect of social science that involves examining a specific contemporary phenomenon within its real-life context. This research design utilised a holistic case study methodology that focused on a single SA university of technology, specifically the faculty of Engineering & Built Environment. It should be noted that the researcher conducting this holistic case study is currently employed at this university, hence why it was chosen as the case. Case studies are valuable for exploring theories, particularly when there is ambiguity between the phenomenon being studied and its context (Yin, 1981).

Referring back to Ferguson's (2012) definition of LA as the measurement, collection, analysis and reporting of data about learners and their contexts to optimise learning and the learning environment, the case study strategy employed in this study aligns with Ferguson's definition of LA. The goal of this study was to gain comprehensive insights and understanding of the CSF associated with the integration of Information Systems to enable a successful implementation of LA and the specific context of this university of technology (Morris & Wood, 1991).

3.6 Time horizon

Unveiling the strategies uncovers the fifth layer of the research onion, referred to as the time horizon. In their 2008 publication, Kosow and Gabner (2008) described this layer as a collection of chronological horizons with varying breadth. Within this layer, two approaches are examined: the cross-sectional and the longitudinal method. The cross-sectional method focuses on a specific phenomenon or phenomena during a brief period of time (Saunders et al., 2009). On the other hand, the longitudinal study involves individuals or events over an extended period to analyse change and development (Adams & Schvaneveldt, 1991).

The present investigation not only review literature on the CSF of Information Systems integration for facilitating LA in higher education, including case study materials, but also tested the impact of the developed model in identifying at-risk students using anonymous secondary data from a single university of technology over the course of a year. Thus, it can be categorised as a longitudinal case study.

3.7 Data collection and data analysis

Following the research onion layers step-by-step, the sixth layer is the core of the research onion belonging to techniques and procedures- data collection and analysis. Moving the research design towards the practicalities of data collection and analysis. Bryman (2012) explain that data collection and analysis depend on the methodological approach adopted by a study. This process contributes significantly to the overall reliability and validity of a study (Saunders et al., 2009).

3.7.1 Reliability and validity

Reliability refers to consistency in findings, should the study's data collection technique and analysis procedures be subjected to varied measures and observations (Golafshani, 2003). While validity is concerned with the effectiveness of the finding (Kirk et al., 1986). It should be noted that this qualitative method study focused on the collection and analysis of literature primarily to explore CSF that contribute to the integration of source systems for LA purposes. In addition to testing the influence of the designed model in identifying student-at-risk. Also, the secondary data collection and analysis may be open to measurement error, missing values, and unpredictable calculations. However, such challenges were not significant enough to impend the reliability and validity of the research.

3.7.2 Generalisability

This qualitative method research adopted a case-study strategy to go about answering the research question from a single university of technology in SA. Therefore, the researcher understands that the research results from data collection and analysis are not generalisable to other research settings. Even more so, given that the study adopted the case study research strategy.

3.7.3 Type of data collected

The study only used secondary data. There are two types of data collection in qualitative method: the primary data and secondary data. Primary data can be derived from first-hand sources similar to interview data, surveys, census, or even text being analysed (Flick, 2011). While secondary data can be explained as information that has already been processed, which can be derived from the work or opinions of other researchers (Newman, 1998).

The study made use of secondary data collected from the integrated institutional source systems such as the LMS, ERP, MAS, and the SIS, stored in the target system. The study made use of an OLAP system to analyse the effect of student-at-risk identification process from the multiple-source secondary data.

3.7.4 Assumptions

This study was conducted under the assumption that the selected university of technology information systems are successfully integrated and that the university implemented an OLAP system to analyse the data.

3.7.5 Identification of the research population

To experiment the influence of CSF that contributed to the successful design and development of the student at-risk identification process, by means of testing the designed model, the study needed a target population. The target population remained as the rows of student secondary data collected from the integrated institutional source systems of a single university of technology in SA.

3.7.6 Sample

The researcher had to decide between using probability sampling plans, where every member of the population has an equal chance of being selected, and non-probability sampling plans, where the chances of being selected are not equal (Daniel, 2011). As a result, the researcher chose to use a non-probability sampling plan and selected a purposive sample of at least 200 rows of student secondary data from the Faculty of Engineering and Built Environments. A purposive sample is one that is selected by the researcher based on who they believe is most appropriate for the study Schutt (2015). It is important to give primary consideration to the sample, as suggested by Som (1995), especially when making inferences from secondary data analysis.

3.8 Conclusion

The study peeled the six layers of the research onion as follows:

Table 3.2: Research Methodology Summary

Research onion layers	**Research approach**
Research Philosophy	Post-positivism
Approach to theory development	Deductive approach
Methodological Choices	Qualitative method
Strategy	Case study
Time Horizon	Longitudinal
Techniques and Procedures	Data collection: secondary data. Data collection: rows of student data from the institutional source systems. Data analysis: OLAP system. Data analysis: Designed model datasets.

3.9 Ethics of research design

Having read and understood the University of the Western Cape ethics code of conduct, the nature of this study involves the acquisition of secondary data collected from the institutional information systems. Specifically, secondary data of a single Faculty from a single university of technology in the WC, SA. The study obtained permission to research the Institutional Ethics Committee of the identified university prior research. The well-being of the students took precedence over the expected benefits to knowledge. Students are allowed to exercise their right to refuse prior collection of their data as primary data, as well as consent to the usage of their data as secondary data from the institutional information systems.

The depth of confidentiality with regards to the selected university data and material brought into being is to be handled professionally. Students have the right to:

- privacy,
- remain anonymous,
- respected confidentiality,
- no release of information inside and outside the university,
- fair and accurate evidence; and
- Unbiased attitude.

It should be confirmed that the researcher adheres to the above listed. All task intentions and processes of analyses were transparent and sufficiently outlined to substantiate appropriation and expectations.

CHAPTER FOUR
PRESENTATION AND DISCUSSION OF FINDINGS

4.1 Introduction

The research design in this thesis had two dimensions. The first dimension was focused on addressing the research questions, while the second dimension centred on the case study of the project. In the first dimension, the research questions identified in Chapter 2 of the literature review were examined. The literature review chapter utilised McKinsey and Company's (Barton & Court, 2012) suggested framework for the systematic implementation of LA systems to accomplish the following:

a) Review the literature on source extraction mechanisms that are necessary for integrating multiple institutional source systems for LA adoption.
b) Review the literature on the data requirements for LA and the resulting integration needs for systems.

The literature review also included an examination of Vincent Tinto's (1975) Longitudinal Model of Dropout as a theory-based conceptual framework applied in education-based analytics to ground the datasets sourced from integrated systems on theoretical reasoning. The purpose of this research was to validate the importance of source systems integration by investigating the CSF of Information Systems integration required for successful implementation of LA.

These three phases of the first dimension (literature review) resulted in a FRS checklist of CSF needed for implementing a LA project. Subsequently, in this chapter, the study shifted to the second dimension – the case study, which involved a narrative description of the CPUT Analytics for Learn (A4L) project. The study compared the findings from the A4L project's narrative with the FRS checklist of CSF acquired from the first dimension. The observation from this comparison provided insights into whether hypothesis was supported or rejected.

4.2 The CPUT Analytics for Learn (A4L) case study overview

The A4L project, funded by the University Capacity Development Grant (UCDG)), was undertaken by the Centre for Innovative Educational Technology (CIET) at the Cape Peninsula University of Technology (CPUT). The purpose of the project was to utilise data from various institutional Information Systems to improve student retention. As CPUT collaborated with Blackboard Solution for the institutional Learning Management System (LMS), they sought guidance from Blackboard on how to best utilise the funding. Blackboard recommended two approaches:

- Learning Analytics Data Strategy: To assess readiness and plan for deployment in order to understand the future state.
- Expanding and/or customising the current A4L warehouse to include data from additional source systems or expand data from the Student Information Systems.

The execution of the A4L project at CPUT was aligned with the university's vision for the years 2021-2030, known as the One Smart CPUT futuristic vision. This vision was developed during the Institutional Strategic Planning Council (ISPC) meeting held on November 27 and 28, 2017. This vision aimed to position CPUT as a leader in technology education and innovation in Africa. Within the scope of the One Smart CPUT vision, the A4L project focused on enhancing smart teaching, learning, and learning environments throughout the institution. The project aimed to utilise learner information more effectively, hence the reference to the A4L project. The high-level goals and objectives of the A4L project included:

Table 4.1: High-level goals and objectives of the A4L project

Institutional Goals	**Objectives**
To increase adoption of the institutional Learning Management System (LMS), Blackboard Learn.	✓ Understand the relationship between activity and success. ✓ Understand the relationship between other variables and success. ✓ Understand the adoption of Blackboard Learn. ✓ Provide evidence for teaching excellence. ✓ Benchmark the best practice. ✓ Promote use of assessing tools in Blackboard Learn. ✓ Quality assurance during accreditation.
To improve retention and throughput of at-risk students.	✓ Identify at-risk students. ✓ Identify at-risk subjects.

	✓ Quantify risk. ✓ Identify students with no activity. ✓ SIS Mismatch

The A4L project timeline followed the waterfall methodology of seven (7) phases:

- Planning (kick-off, project definition, technical preparation)
- Installation and configuration (historical load)
- Orientation
- Configuration, customisation, and data review
- Deployment preparation
- Report-writing training and development
- Project close

During the project kick-off and definition (refer to Appendix A), the staffing model (role and responsibility) proposed for the deployment of the A4L project included the following:

Table 4.2: Staffing model for the deployment of the A4L Project

Role	**Responsibility**
✓ Deputy Vice Chancellor (DVC): Academic ✓ Quality Management Department	Management Oversight and Project Owner: For general oversights
✓ CIET Director ✓ CIET Educational Technologist	Project Manager and Primary Contact: Project Management team for project scope, time, and financial resources
✓ CIET Instructional Designers and a Materials Developer ✓ Deans ✓ Faculty Teaching and Learning Coordinators ✓ Faculty IT Coordinators ✓ Retention Officers ✓ Selected lecturers (CIET Champions)	Functional Consultants for required gatherings, end-user training, and setup of reports

✓ Head of Departments (HODs) ✓ Students	
✓ CIET Technician with SQL Expertise ✓ Blackboard Solution Administrator(s) ✓ Registrar's Office ✓ Management Information Services ✓ Computer and Telecommunications Services (CTS)	Technical Consultant: Responsible for the IT infrastructure and to provide technical implementation and configuration of baseline Blackboard Analytics for Learn product, data sourcing and data integration

The narrative of the A4L project (case study) question structure will be derived from McKinsey and Company's theoretical framework (Barton & Court, 2012), which was adopted in Chapter 2 of the literature review. The theoretical framework consist of three elements: data, model, and transformation. The following questions have been derived from this framework:

- **Data:** What is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems for LA adoption?
- **Model:** Which data is necessary for conducting LA, and what are the resulting systems requirements for system integration?
- **Transformation:** What were the institutional priorities in the adoption of LA?

4.3 Data: What is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems for LA adoption?

In terms of the project plans, installation, and configuration, the A4L project placed a significant emphasis on data integration. After consulting with Blackboard Solution and the CPUT Technical Consultants, it was determined that customising the existing A4L data warehouse with additional source systems would be the most suitable approach (refer to Appendix B). This approach differed from traditional analytics, which primarily focus on using data to improve services and business practices by reporting on past events (Gagliardi, Parnell, & Carpenter-Hubin, 2018). Instead, the A4L project focused on utilising student learning activity data to provide analytics-informed interventions and enhance student success (Sclater et al., 2017).

During the technical preparations, Blackboard Solution and other CPUT Technical Consultants creatively developed principles for data sourcing and integration. The team began by mapping institutional requirements to identify the information systems used for data sourcing and relevant datasets. The integrated Learning Management System (LMS), Enterprise Resource Planning (ERP) system, Student Information System (SIS), and Marks Administration Systems (MAS) were identified as the relevant institutional Information Systems for data sourcing.

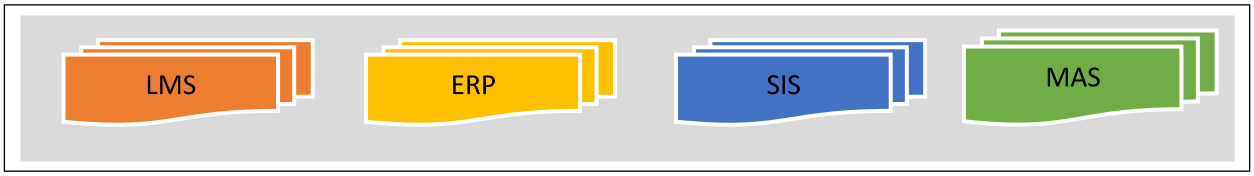


Figure 4.1: The data sourcing systems mapped for integration

The decision to source data from these identified systems was based in the following reasons:

- The integrated LMS directly sources student learning activities and course usage data.
- The ERP and SIS systems provide data on student demographics, payroll, human resources, and finances.
- The MAS sources student grade performance data.

These systems were chosen due to their ability to source their data tables and provide extract files for data integration.

The importance of data integration in the project was highlighted by Chaki (2015), who outlined the following key factors to consider when determining a data integration process:

(i) The nature of extraction process between source systems and consuming systems (push/pull).

(ii) The type of connecters required for pulling data from source systems.

(iii) The choice between using a data integration engine or a database engine for data transformations.

(iv) The required outbound extract formats for consuming applications.

(v) The need to address data security and comply with any country-specific data regulations during the integration process.

4.3.1 The nature of extraction process between source systems and consuming systems (push/pull)

The university opted for the push extraction logic to be used in the integration process. This logic involved running a predefined cycle every four hours, along with extracting student activity data once a day at midnight using SQL Agent jobs. These jobs were responsible for extracting data from the Statistical Analysis System (SAS) through a linked server connection and from Blackboard Learn through a web service extract.

The extracted files were then transferred to a secure area and prepared for the extraction process. The CPUT Network manager , Business Administration Systems (BAS) from Computer and Telecommunications Services (CTS), and the CIET IT Technician with SQL expertise worked together with the assigned Blackboard Solution Consultant.

4.3.2 The type of connectors required for pulling data from source systems

To establish connection between the extracted data from various sources, connectors were utilised as part of the initial integration process. Due to the fact that CPUT source systems are cloud-based packaged applications, the university utilised application-specific connectors to access existing system reports and the Ms OLAP server as data sources. Initially, Originally, Power BI from Microsoft was used as an OLAP, but later the university transitioned to Pyramid BI Software.

Furthermore, SQL servers were provided for offline analytics, with Microsoft SQL being used for local data and Snowflake for LMS data access. These servers were connected to the source system through Open Database Connectivity (ODBC) connectors, which enabled the creation complex connections.

The figure below is indicative of the connectors pulling from source systems:

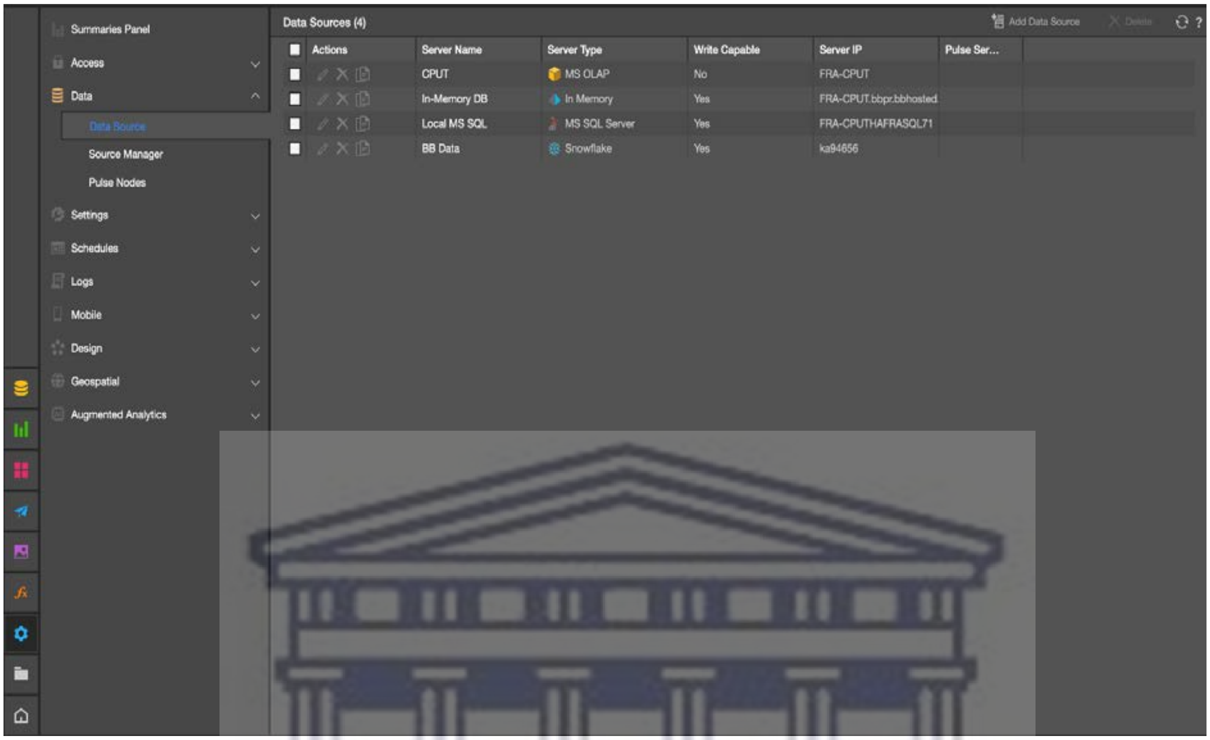


Figure 4.2: The data warehouse for the A4L project

4.3.3 Leverage data integration engine for transformations of source data or use database engine for transformations.

The UCDG funds played a key role in facilitating data integration in the A4L project. The university employed the Extract, Transform and Load (ETL) technique to retrieve raw data from the source systems cleanse it, ensuring that no data was corrupted during the extraction phase. The figure shown below illustrates the ETL approach used in the A4L project during integration:

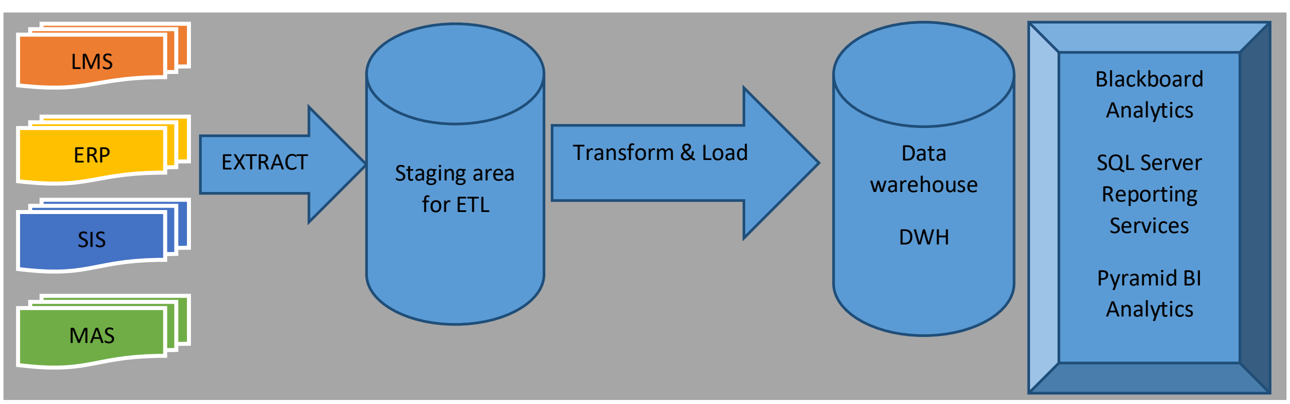


Figure 4.3: CPUT A4L ETL approach

Once the data was cleansed, it was sent over the network and loaded into the A4L data warehouse in a format ready for analysis. The database engines of the data warehouse were responsible for this loading process. The figure depicted below shows the whiteboard schema created when mapping the relevant sources of data for the integration process:

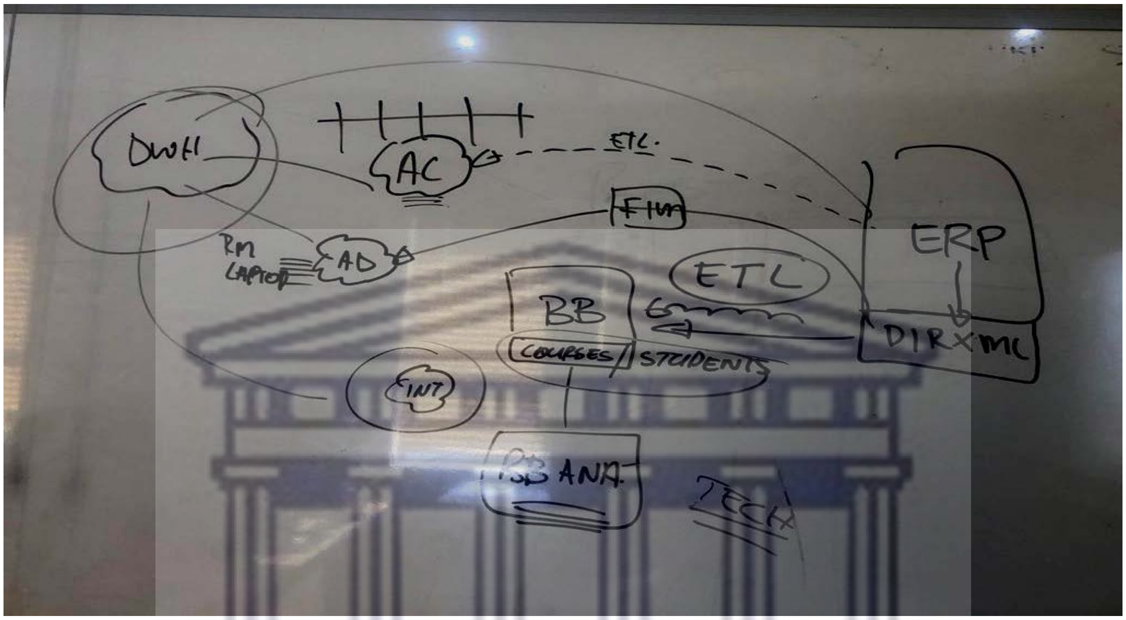


Figure 4.4: Whiteboard schema on systems integration capability

At this stage, the A4L project timeline has completed several key milestones, including the definition, kick-start, installation and configurations, data sourcing, and data integration. Moving forward, the project has now entered the preparation phase for deployment (refer to Appendix C for details). The deployment plans began with pilot preparations, where we pilot the analysis-ready data from the data warehouse into various delivery phases. These phases involved generating reports through different interfaces for various stakeholders and use cases, such as:

- Lecturers and students using Blackboard Analytics Integrated Reports
- Students supporting utilising SQL Reporting Services (SSRS)
- Power users accessing Pyramid BI Analytics reports, including dashboards and publications for leadership, as well as alerts for student support.

Pilot Blackboard Analytics Integrated Reports

During the piloted delivery phase, the focus was on course level analytics for functional consultants, including course lecturers (CIET Champions), CIET Instructional Designers, Educational Technologists, and Faculty IT Coordinators. These consultants had the ability to generate various course reports using the SSRS and/or Pyramid BI Analytics systems. The reports were then integrated into the LMS interface to facilitate access, communication, and immediate interventions between lecturers and students.

The figure below illustrates the course analytics and how they can be accessed through the integrated LMS:

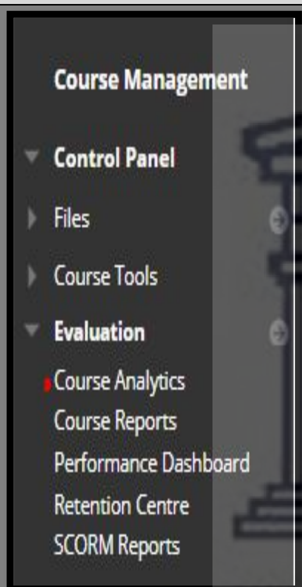
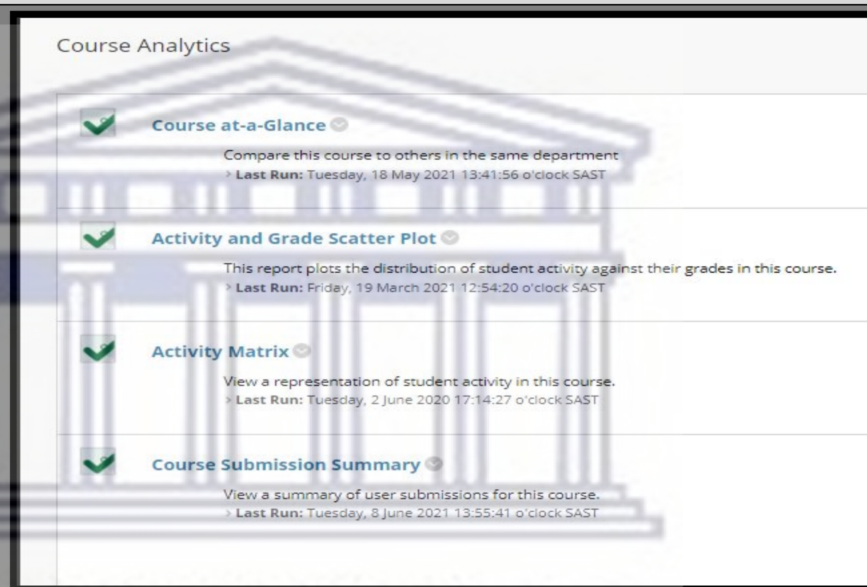
Course reports	Course report types from the integrated LMS
	

Figure 4.5: First pilot focused on using course analytics from the integrated LMS Analytics.

The course analytics designs aimed to provide lecturers and other consultants with different types of reports. These reports allowed CPUT lecturers and functional consultants to compare student performances across different modules within the same department, analyse the distribution of student activity and grades, visualise student activity in the course, and view a summary of student submissions.

Additionally, lecturers could assess student performance using data from the integrated LMS, grade centre, and the MAS grades journey, utilising various tools provided by the LMS such as content area, assignment and test tools, discussion forums, and announcements. The Retention Centre within the integrated LMS complemented the course analytics by identifying students at-risk and determining the reasons behind their risk status.

The table below displays a list of students at-risk, including the reasons why the students came to be at risk using four (4) customisable pre-defined variables: missed deadline, grades, course activity, and course access:

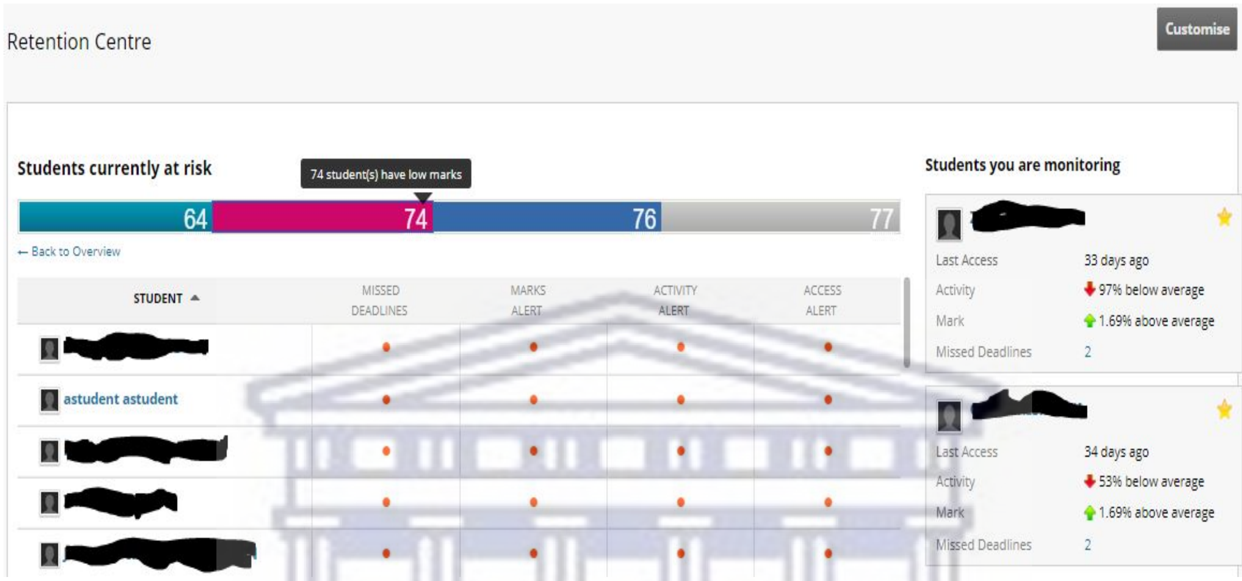
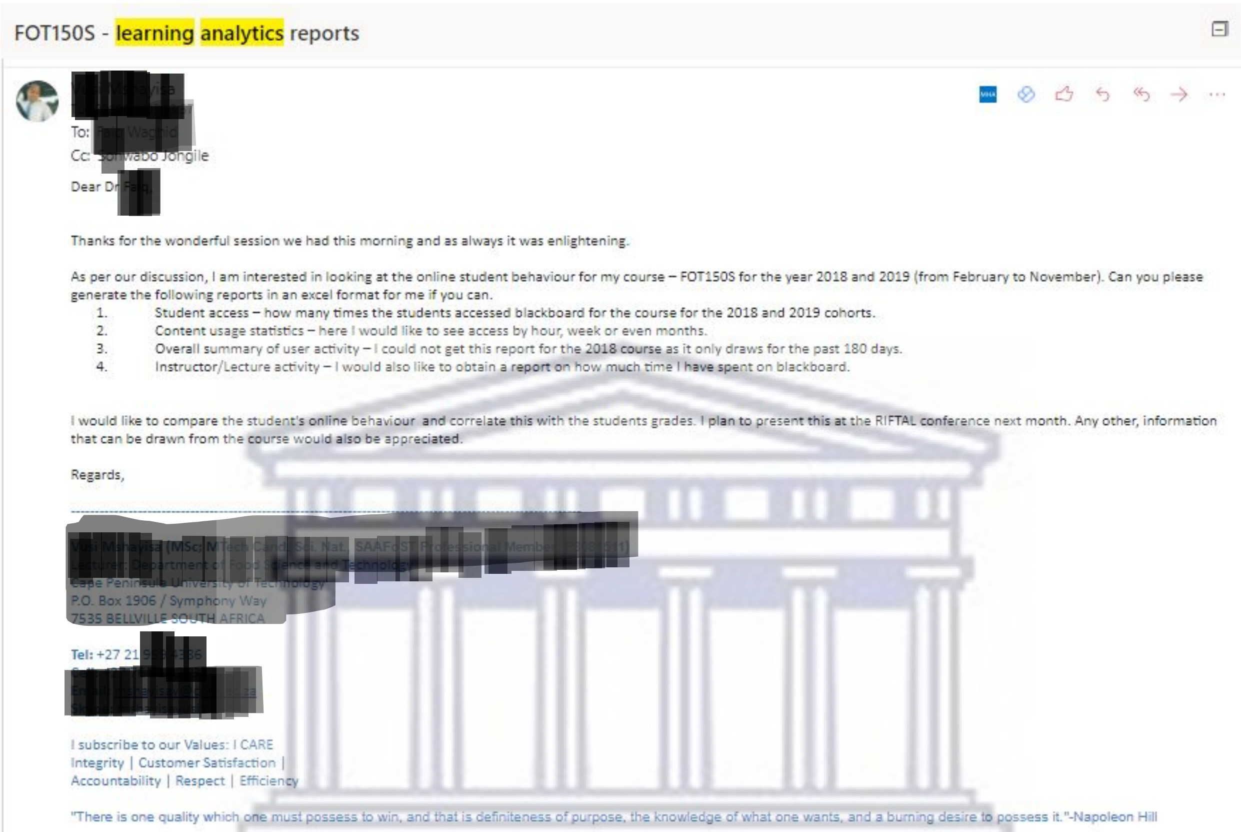


Figure 4.6: The Retention Centre from the integrated LMS Analytics

In order to access the analytics reports, lecturers, CIET Instructional Designers, and IT Coordinator needed the appropriate role privileges within the LMS. These privileges enabled instant alerts and communication channels (email or instant messaging) between lecturers and students. For instance, lecturers could develop intervention frameworks and provide support to students who had not accessed the LMS for a specified number of days. The Retention Centre would generate a list of at-risk students and send email notifications to alert lecturers.

Similarly, if students performed poorly in a task, tools such as "adaptive release" could be used to give them a second chance. The roles and privileges within the LMS ensured compartmentalised access to analytics, reports, and dashboards for lecturers and students, reducing the need for additional licenses for functional consultants to access SSRS and Pyramid BI Analytics. This setup allowed the analytics to be pushed into the institutional LMS, where lecturers and students were already located, eliminating the need for direct access to SQL and Pyramid. Feedback from lecturers was collected after each training session to evaluate the effectiveness of the setup.

The figure below illustrates the feedback provided by one of the lecturers who is also serves as a departmental champion:



FOT150S - learning analytics reports

[redacted]
To: [redacted]
Cc: [redacted] Jongile

Dear Dr [redacted]

Thanks for the wonderful session we had this morning and as always it was enlightening.

As per our discussion, I am interested in looking at the online student behaviour for my course – FOT150S for the year 2018 and 2019 (from February to November). Can you please generate the following reports in an excel format for me if you can.

1. Student access – how many times the students accessed blackboard for the course for the 2018 and 2019 cohorts.
2. Content usage statistics – here I would like to see access by hour, week or even months.
3. Overall summary of user activity – I could not get this report for the 2018 course as it only draws for the past 180 days.
4. Instructor/Lecture activity – I would also like to obtain a report on how much time I have spent on blackboard.

I would like to compare the student's online behaviour and correlate this with the students grades. I plan to present this at the RIFTAL conference next month. Any other, information that can be drawn from the course would also be appreciated.

Regards,

[redacted]
P.O. Box 1906 / Symphony Way
7535 BELLVILLE SOUTH AFRICA

Tel: +27 21 [redacted]

[redacted]

I subscribe to our Values: I CARE
Integrity | Customer Satisfaction |
Accountability | Respect | Efficiency

"There is one quality which one must possess to win, and that is definiteness of purpose, the knowledge of what one wants, and a burning desire to possess it."-Napoleon Hill

Figure 4.7: Feedback from a CPUT Lecturer and CIET champion

SQL Server Reporting Services (SSRS)

The initial focus of the delivery phase was on the SQL Server Reporting Services (SSRS), which contained a collection of example reports for A4L product.

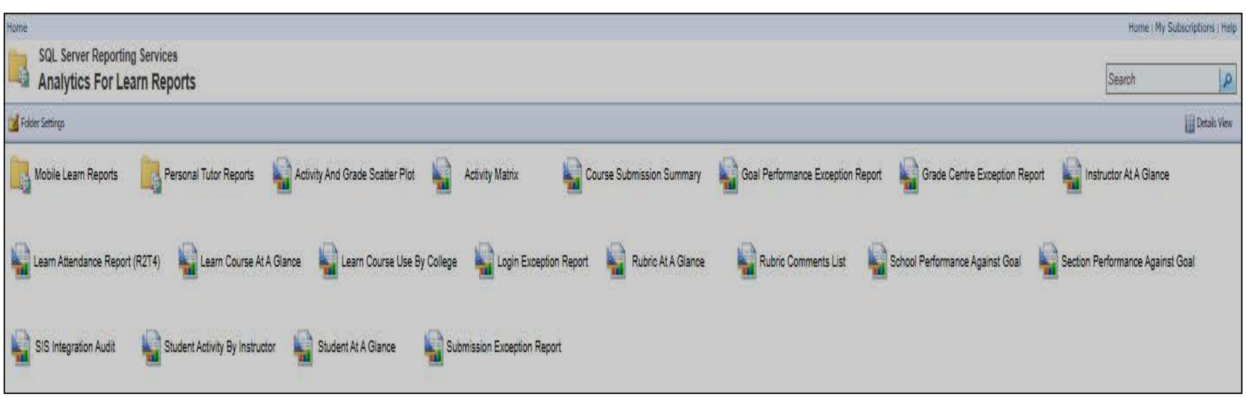


Figure 4.8: Second pilot focused on using SQL Server Reporting Services (SSRS)

To meet the institution's requirements, the CIET IT Technician with SQL expertise had to develop dashboard reports using the analytics reports generated from the SSRS. These reports specifically focused on identifying student at-risk of failing academically and those facing challenges that could hinder their ability to complete their studies. The IT Developer with SQL expertise and the assigned Blackboard Data Specialist collaborated to develop the necessary analytics and integrate them into the LMS analytics. These analytics were then presented to course lecturers for analysis, serving as the first pilot for the project.

During the implementation, the CIET team was fortunate to have the expertise of an IT Technician with SQL skills. Working together with the CPUT Network Manager from CTS and the assigned Blackboard Data Specialist, they were able to fulfil data requests from various stakeholders, including lecturers, Deans, the IT Coordinator, and Retention Officers. These stakeholders, who reached out to CIET Instructional Designers and Educational Technologists, were able to access standardised reports based on their roles within the LMS and SSRS systems.

Unfortunately, the CIET IT Technician with SQL expertise resigned shortly after the A4L project was implementated. Tas a result, the institution had to rely on the assigned Blackboard Data Specialist for report generation. This necessitated training the other A4L stakeholders on how to create dashboards.

Pyramid Analytics

The third pilot focused on using Pyramid BI Analytics to provide comprehensive analytics specifically for senior stakeholders, including DVC Academic and Deans, as well as Systems Administrators such as the IT Coordinators and CIET Functional Consultants. This system allowed these stakeholders to generate new reports, edit existing reports, dashboards, and publications at an institutional level.

To access the Pyramid BI Analytics, users would need to log in through a browser, as shown in the screenshot below:

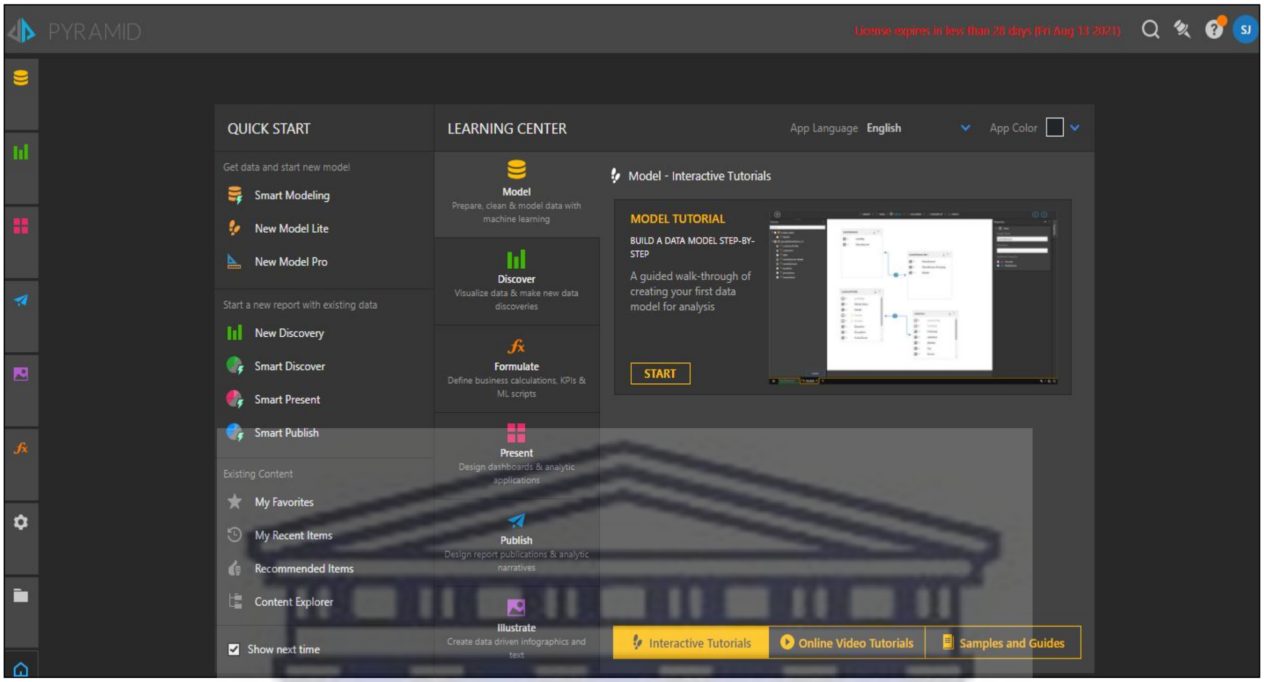


Figure 4.9: Third pilot focused on using analytics from the Pyramid BI Analytics

The purpose of conducting the Pyramid BI Analytics pilot was to generate reports and analytics at the Programme, Faculty, and Institutional levels. Its primary focus was on extracting information from data source systems beyond the scope of student learning activities, such as the adoption of the integrated LMS, instructor and student tool usage, time spent on task, student residence and transportation data, and the secondary schools attended by students.

This pilot outperformed the integrated LMS Course Analytics and SQL analytics as it utilised additional datasets obtained from integrating various source systems, including Applications to Admissions files, the Registration System, including the Library record system (student mode of access to the library system). This allowed for more powerful and higher quality institutional analytics.

4.3.4 The required outbound extract formats for consuming applications.

Due to sensitive nature of student data and the importance of data security, the university did not prioritise outbound data extracts. While the university had relationships with external parties such as Learning Tools Interoperability (LTI) and Work-Integrated Learning (WIL), the management of student data and compliance requirements took precedence during the integration design process.

4.3.5 The need to address data security and comply with any country-specific data regulations during the integration process.

Despite the A4L project's focus on identifying at-risk students, the development principles for obtaining data remained consistent. This included integrating data sources for Course Level Analytics, as well as leveraging more advanced analytics offered by Pyramid Analytics. The project received IT support from the Centre for Innovative Educational Technologies (CIET), which was overseen by the Deputy Vice Chancellor (DVC) Teaching and Learning. The success of the project relied on the active involvement of various stakeholders, including:

- Project Manager from CIET
- Functional Staff such as Instructional Designers, Educational Technologists, Materials Developers, Faculty IT Coordinators, and Curriculum Design Representatives who are activists of digital curricula to support seamless teaching and learning
- Academic representatives (lecturers -- also known as CIET champions who continuously attend and implement technology-enhanced teaching and learning approaches and strategies from CIET), Course Coordinators, Executive, and Student Support Units
- Computer and Telecommunications Services (CTS) such as the Network Manager for data sourcing and data integration, data warehouse, security and other IT related support
- An in-house (within the CIET team) IT Technician with database and SQL expertise working collaboratively with CTS on how the institutional IT status quo and practices can best support the adoption and implementation of LA
- The Quality Management department for general oversights, to ensure that the project fulfils all the quality requirements
- The CPUT Legal Department for all legal related matters and compliances such as PoPI Act compliance on the use of student data
- The ethics specialist invited to start the conversation on ethical considerations when using student data.

Unfortunately, the resignation of the CIET IT Technician with SQL expertise created challenges for the project team and stakeholders, as their contribution was critical to the data sourcing and integration process. However, despite this setback, various factors essential for integrating multiple institutional source systems for LA adoption were still in place. These included:

- the groundwork and creativity in data collection,
- the integration process of source data tables,
- understanding data limitations related to source systems integration,
- involvement of IT and other stakeholders,
- data quality and governance; and
- the different approaches for LA implementation across CPUT.

To assess the state of source extraction mechanisms and their ability to integrate multiple institutional Information Systems for LA adoption, study compared the checklist of CSF derived from literature review with the findings of the case study. The goal was to determine if the results supported or rejected the hypothesis:

Table 4.3: Data element- results from mapping CSF against A4L project groundwork

Critical Success Factor	**Results**
Has the university built a data informed culture in decision making for LA based on a predefined design challenge, for example, the student at-risk identification?	✓ The data **supports** the hypothesis
Has the university established an effective source systems integration approach, and does it understand the limitations of the identified source systems data?	✓ The data **supports** the hypothesis
Has the university secured the necessary information technology support and other stakeholders similar to the institutional ethics and legal department?	✓ The data **supports** the hypothesis

The results of the first part of the theoretical framework demonstrated that the information gathered from the case study supported the hypothesis presented in the literature review on all three aspects.

4.4 Model: Which data is necessary for conducting LA, and what are the resulting requirements for systems integration?

The objective of this question was to identify patterns in educational data by creating models specifically designed to process the datasets obtained from the source systems. Additionally, the aim was to address the institutional need to identify students who may be at risk. Since most of the source systems at CPUT use in-memory databases and ODBC connectors, it was decided to temporarily outsource the development of models to Blackboard Solution.

Based on the findings in this section, a proposal was made to Blackboard Solution to assign a Data Specialist to provide training on how to develop models using the SQL Server Reporting Services and the Pyramid BI Analytics OLAPs. The ultimate goal was to fully transfer the responsibility of model development to the CPUT stakeholders, particularly the CIET team.

However, a complete transfer of model development became unfeasible, primarily due to the mentioned resignation and the fact that the train-the-trainer sessions did not resonate with any of the project stakeholders, even after a replacement Data Specialist was appointed. The original Data Specialist lacked training skills, patience, and an understanding of the fact that none of the project stakeholders possessed data analysis skills.

Consequently, the report-writing training and development schedules with the Data Specialist were unsuccessful (refer to Appendix D). Blackboard Solution assigned a new Data Specialist to the university and a new contract was established. The renewable annual contract for the Blackboard Data Specialist proved to be crucial for the CIET team as none of them, nor the Functional and Technical Consultants, had specialised expertise in constructing and deploying dashboards and reports. The plan was to gather information from the faculties on the datasets they believed should be included in the design model for identifying student at-risk.

The CIET team and the Faculty IT Coordinators were instructed to conduct roadshows on the datasets. They intended to gather information from faculty members about datasets that could inform the design of a model to identify at-risk students in their specific faculty context. Unfortunately, the roadshows were unsuccessful as most faculty representatives did not participate. The few representatives who did participate provided responses that were mostly based on their own knowledge and internet searches, lacking educational research.

Additionally, there was a lack of theory-informed analytics to guide the model design for at-risk students. To overcome this issue, the Blackboard Analytics Data Specialist suggested using common datasets from universities within and outside of SA, which they had previously supported. However, this approach did not account for the specific context of CPUT and how well these models could be implemented. As a result, the project took a question-driven approach to machine learning application. Fortunately, the researcher of this thesis, who was a member of the CIET team, suggested reviewing at-risk datasets recommended in the literature on education-based analytics.

They decided to adopt Tinto's (1975) Longitudinal Model of Dropout theory as the basis for the datasets, in order to personalise the intervention and support for at-risk students. The introduction of Tinto's theoretical model into the A4L project aimed to connect theoretical arguments with the testing of the designed model for identifying at-risk students. The researcher proposed an integrated approach, using Tinto's conceptual schema as a knowledge provider of the student at-risk identification process. This involved mapping the datasets with the institutional source systems in the case study.

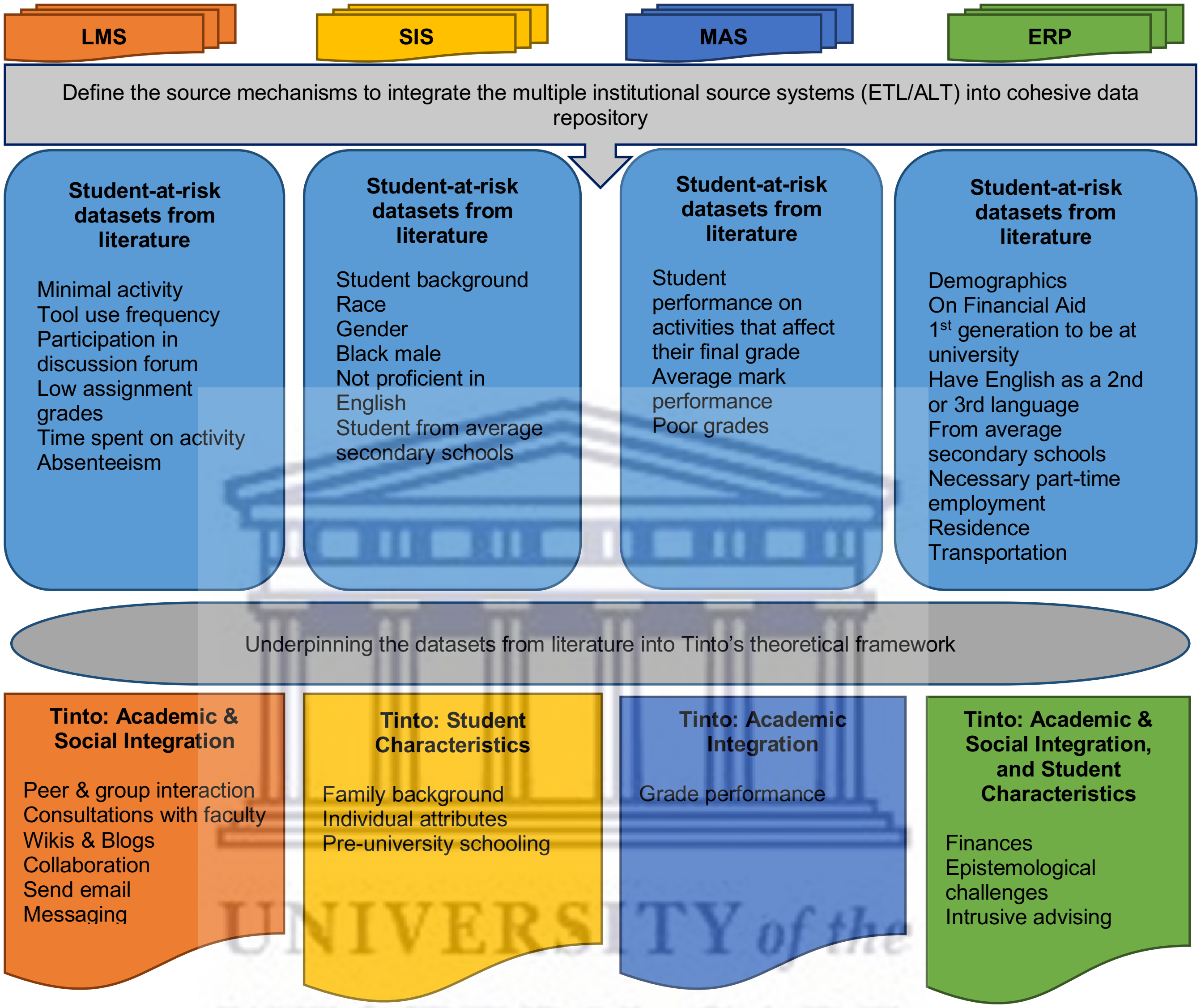


Figure 4.10: Underpinning datasets into theory

Based on Tinto's (1975) Longitudinal Model of Dropout, important sources of information that can shed light on the interactions leading to academic and social risk factors for students include:

- Student characteristics, such as age, ethnicity, residency, demographics, and previous schooling.
- Academic integration, including grade performance in both continuous and final assessments, personal development, and study patterns.
- Social integration, such as student interaction with the university's LMS, peers, faculty, and formal and informal visits.

By going beyond literature searches and Internet sources, the A4L project team gained a deeper understanding why students become at risk. This understanding allowed for the implementation of personalised interventions based on the specific reasons for student vulnerability. For example, an LA system could flag the "mark-below-average percentage" dataset generated from the MAS or the LMS grade centre column. However, this approach may lack a conceptual framework that explains why some students are consistently underperforming.

This data-driven analysis categorised all students as at-risk. Incorporating theoretical reasoning, knowledge, and understanding of factors contributing to "average performance," such as personal circumstances like family loss, students from average secondary schools struggling to catch up, or non-native English speakers, enables lecturers to respond with real-time personalised interventions and empathy towards students. By utilising datasets from source systems grounded in Tinto's (1975) theoretical model, the university can identify risk patterns in student behaviour outlined in Tinto's theory and provide immediate support and interventions long before they could escalate to student failure.

The table below presents a list of datasets recommended by the Blackboard Data Specialist, project stakeholders, faculty representatives from the CIET roadshow for data models, and the researcher's findings from the literature review:

Table 4.4: Designed model for the identification of at-risk students

Datasets	**Delineation**
Last Access	Last day a student accessed the LMS
Gender	Male/Female
No. of Days in Course	Number of days registered in course
Access in First Two Weeks	Early preparations Students who accessed the LMS before the commencement date
Tool Clicks	Number of clicks per tool in the LMS
Click Quartile	Number of clicks per tool in a quarter
No. of Forum Posts	Number of student posts in forums
No. of Submission	Number of student submissions
No. Content Accesses	Number of content accesses
No. Course Accesses	Average accesses per course

Course Access Quartile	Average accesses per course in a quarter
Assessment Accesses	Number of assessments accessed online
No. Item Accesses	Number of items accessed within a course
Accessed Prior to Term	Students accessing the LMS before the start of term
No. Late Assignment	Assignments submitted after the due date

The datasets mentioned previously were compared to integrated data sources to determine if the data could be extracted. The data was then transformed and loaded using both the SQL report and Pyramid Analytics as the institutional OLAP. This analysis focused on 200 student rows of data from a single faculty, based on the model design presented in Table 4.4:

RISK CATEGORY	LAST ACCESSED	GENDER	NO. of DAYS IN COURSE	ACCESSED IN FIRST 2 WEEKS	TOOL CLICKS	CLICK QUARTILE	NO. FORUM POSTS	NO. SUBMISSIONS	NO. CONTENT ACCESSES	NO. COURSE ACCESSES	COURSE ACCESS QUARTILE	ASSESSMENT ACCESSES	NO. ITEM ACCESSES	ACCESSED PRIOR TO TERM	NO. LATE ASSIGNMENT
Low Grades & Not Engaged	More than a week ago	M	232	N	67	3rd	0	4	35	21	4	9	44	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	20	4th	0	0	40	10	4	5	45	N	0
Low Grades & Not Engaged	3 days ago	F	232	N	81	2nd	0	8	45	31	4	14	59	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	37	3rd	0	2	15	19	3	3	18	N	0
Low Grades & Not Engaged	7 days ago	F	232	N	95	2nd	0	9	52	25	3	11	63	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	109	2nd	0	8	79	35	2	13	92	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	73	2nd	0	4	38	17	4	7	45	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	118	2nd	0	5	67	24	3	11	78	N	0
Low Grades & Not Engaged	3 days ago	M	232	N	19	4th	0	4	10	18	4	7	17	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	90	2nd	0	4	32	23	3	11	43	N	0
Low Grades & Not Engaged	Yesterday	M	232	N	149	2nd	0	0	62	28	3	8	70	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	78	2nd	0	5	38	17	4	6	44	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	96	2nd	0	0	30	3	4	0	30	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	59	3rd	0	8	40	22	3	11	51	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	48	3rd	0	1	17	10	4	5	22	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	59	3rd	0	1	16	13	4	6	22	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	0	4th	0	0	0	0	4	0	0	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	93	2nd	1	9	29	28	3	16	45	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	7	4th	0	1	1	6	4	2	3	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	24	4th	0	1	14	8	4	4	18	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	49	3rd	0	9	33	22	3	13	46	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	82	2nd	0	2	13	19	4	14	27	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	165	1st	1	11	33	46	2	19	52	N	0
Low Grades & Not Engaged	5 days ago	M	232	N	77	2nd	0	5	30	13	4	13	43	N	0
Low Grades & Not Engaged	More than a week ago	F	232	N	38	3rd	0	0	3	32	3	1	4	N	0
Low Grades & Not Engaged	More than a week ago	M	232	N	118	2nd	0	0	51	10	4	2	53	N	0

	RISK CATEGORY	LAST ACCESSED	GENDER	NO. of DAYS IN COURSE	ACCESSED IN FIRST 2 WEEKS	TOOL CLICKS	CLICK QUARTILE	NO. FORUM POSTS	NO. SUBMISSIONS	NO. CONTENT ACCESSES	NO. COURSE ACCESSES	COURSE ACCESS QUARTILE	ASSESSMENT ACCESSES	NO. ITEM ACCESSES	ACCESSED PRIOR TO TERM	NO. LATE ASSIGNMENT
	Low Grades & Not Engaged	More than a week ago	M	232	N	1	4th	0	0	0	1	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	F	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	F	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	12	4th	0	0	3	2	4	1	4	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	25	4th	0	1	14	4	4	3	17	N	0
	Low Grades & Not Engaged	More than a week ago	F	232	N	28	4th	0	2	15	22	3	8	23	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	2	4th	0	0	2	1	4	0	2	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	150	1st	0	4	54	21	4	13	67	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	46	3rd	0	10	23	8	4	6	29	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	12	4th	0	0	0	6	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	94	2nd	0	3	9	32	3	9	18	N	0
	Low Grades & Not Engaged	Yesterday	M	232	N	57	3rd	0	2	17	14	4	2	19	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	5	4th	0	0	5	5	4	1	6	N	0
	Low Grades & Not Engaged	More than a week ago	F	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	9	4th	0	0	0	2	4	1	1	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	1	4th	0	0	0	1	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	2	4th	0	0	0	5	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	F	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	F	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	11	4th	0	0	3	1	4	0	3	N	0
	Low Grades & Not Engaged	5 days ago	M	232	N	76	2nd	0	1	11	20	3	4	15	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	16	4th	0	0	4	11	4	1	5	N	0
	Low Grades & Not Engaged	More than a week ago	M	232	N	0	4th	0	0	0	0	4	0	0	N	0
	Low Grades & Engaged	2 days ago	M	232	N	112	2nd	0	7	58	27	2	12	70	N	0
	Low Grades & Engaged	2 days ago	M	232	N	51	3rd	0	2	45	28	1	17	62	N	0
	Low Grades & Engaged	Yesterday	F	232	N	262	1st	0	8	138	40	2	13	151	N	0
	Low Grades & Engaged	Yesterday	F	232	N	314	1st	0	10	85	82	1	21	106	N	0
	Low Grades & Engaged	Yesterday	F	232	N	378	1st	0	8	161	101	1	17	178	N	0
	Low Grades & Engaged	2 days ago	M	232	N	168	1st	0	8	85	39	2	18	103	N	0
	Low Grades & Engaged	More than a week ago	F	232	N	420	1st	1	15	149	133	1	39	188	N	0
	Low Grades & Engaged	Yesterday	M	232	N	80	2nd	0	3	81	20	3	8	89	N	0

RISK CATEGORY	LAST ACCESSED	GENDER	NO. of DAYS IN COURSE	ACCESSED IN FIRST 2 WEEKS	TOOL CLICKS	CLICK QUARTILE	NO. FORUM POSTS	NO. SUBMISSIONS	NO. CONTENT ACCESSES	NO. COURSE ACCESSES	COURSE ACCESS QUARTILE	ASSESSMENT ACCESSES	NO. ITEM ACCESSES	ACCESSED PRIOR TO TERM	NO. LATE ASSIGNMENT
Low Grades & Engaged	3 days ago	F	232	N	280	1st	0	4	37	84	1	10	47	N	0
Low Grades & Engaged	3 days ago	F	232	N	97	2nd	0	7	79	67	1	46	125	N	0
Low Grades & Engaged	2 days ago	M	232	N	330	1st	0	20	106	55	1	35	141	N	0
Low Grades & Engaged	Yesterday	M	232	N	244	1st	1	15	48	51	2	9	57	N	0
Low Grades & Engaged	Yesterday	F	232	N	265	1st	0	12	122	110	1	46	168	N	0
Low Grades & Engaged	2 days ago	M	232	N	390	1st	0	5	217	103	1	28	245	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	66	3rd	0	6	33	26	3	20	53	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	95	2nd	0	2	48	25	3	7	55	N	0
High Grades & Engaged	More than a week ago	F	232	N	117	2nd	0	7	53	34	2	16	69	N	0
High Grades & Engaged	2 days ago	M	232	N	174	1st	0	6	88	62	1	19	107	N	0
High Grades & Engaged	More than a week ago	M	232	N	190	1st	0	8	107	49	1	27	134	N	0
High Grades & Not Engaged	3 days ago	M	232	N	89	2nd	0	7	54	42	1	26	80	N	0
High Grades & Not Engaged	3 days ago	M	232	N	62	3rd	0	7	29	20	4	9	38	N	0
High Grades & Engaged	2 days ago	F	232	N	115	2nd	0	8	61	31	2	13	74	N	0
High Grades & Not Engaged	2 days ago	M	232	N	18	4th	0	1	13	17	3	3	16	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	29	4th	0	4	34	14	3	10	44	N	0
High Grades & Engaged	More than a week ago	M	232	N	58	3rd	0	5	54	26	1	14	68	N	0
High Grades & Engaged	More than a week ago	F	232	N	49	3rd	0	2	44	26	2	14	58	N	0
High Grades & Engaged	2 days ago	M	232	N	134	2nd	2	7	90	48	1	34	124	N	0
High Grades & Engaged	2 days ago	M	232	N	54	3rd	0	5	44	26	2	19	63	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	18	4th	0	2	25	19	2	4	29	N	0
High Grades & Engaged	Yesterday	M	232	Y	281	1st	0	17	210	146	1	64	274	N	0
High Grades & Not Engaged	2 days ago	M	232	N	105	2nd	0	8	99	52	3	28	127	N	0
High Grades & Engaged	Yesterday	F	232	N	232	1st	0	11	175	107	2	32	207	N	0
High Grades & Not Engaged	3 days ago	F	232	N	152	1st	0	10	115	58	2	30	145	N	0
High Grades & Not Engaged	3 days ago	F	232	N	53	3rd	0	10	53	38	3	19	72	N	0
High Grades & Not Engaged	Yesterday	M	232	N	33	3rd	0	6	30	28	4	12	42	N	0
High Grades & Engaged	Yesterday	M	232	N	170	1st	0	11	118	61	2	25	143	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	12	4th	0	1	10	11	4	5	15	N	0
High Grades & Not Engaged	Yesterday	F	232	N	68	3rd	0	11	31	30	4	26	57	N	0
High Grades & Engaged	Yesterday	F	232	N	400	1st	1	17	271	159	1	56	327	N	0
High Grades & Engaged	2 days ago	M	232	N	138	2nd	0	19	509	57	2	54	563	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	21	4th	0	4	38	12	4	3	41	N	0
High Grades & Engaged	2 days ago	F	232	N	319	1st	0	18	162	127	1	48	210	N	0

RISK CATEGORY	LAST ACCESSED	GENDER	NO. of DAYS IN COURSE	ACCESSED IN FIRST 2 WEEKS	TOOL CLICKS	CLICK QUARTILE	NO. FORUM POSTS	NO. SUBMISSIONS	NO. CONTENT ACCESSES	NO. COURSE ACCESSES	COURSE ACCESS QUARTILE	ASSESSMENT ACCESSES	NO. ITEM ACCESSES	ACCESSED PRIOR TO TERM	NO. LATE ASSIGNMENT
High Grades & Not Engaged	Yesterday	M	232	N	127	2nd	0	9	94	25	4	25	119	N	0
High Grades & Not Engaged	3 days ago	F	232	N	102	2nd	0	12	59	50	3	33	92	N	0
High Grades & Engaged	3 days ago	F	232	N	91	2nd	0	8	113	47	3	40	153	N	0
High Grades & Not Engaged	Yesterday	M	232	N	98	2nd	0	9	41	38	3	19	60	N	0
High Grades & Not Engaged	Yesterday	M	232	N	70	3rd	0	8	26	38	4	19	45	N	0
High Grades & Not Engaged	2 days ago	M	232	N	107	2nd	0	18	102	51	3	33	135	N	0
High Grades & Engaged	Yesterday	F	232	N	564	1st	0	10	349	124	1	42	391	N	0
High Grades & Engaged	2 days ago	M	232	N	138	2nd	0	8	141	71	2	45	186	N	0
High Grades & Not Engaged	2 days ago	F	232	N	71	3rd	0	11	63	40	3	44	107	N	0
High Grades & Engaged	Yesterday	F	232	N	367	1st	0	16	215	92	2	50	265	N	0
High Grades & Engaged	Yesterday	M	232	N	169	1st	0	21	81	112	1	72	153	N	0
High Grades & Engaged	2 days ago	M	232	N	124	2nd	0	13	161	63	2	32	193	N	0
High Grades & Engaged	Yesterday	F	232	N	559	1st	0	15	318	149	1	41	359	N	0
High Grades & Engaged	2 days ago	M	232	N	408	1st	0	16	233	124	1	42	275	N	0
High Grades & Not Engaged	Yesterday	M	232	N	72	3rd	0	16	86	60	2	34	120	N	0
High Grades & Engaged	Yesterday	F	232	N	267	1st	0	9	181	132	1	41	222	N	0
High Grades & Not Engaged	2 days ago	M	232	N	107	2nd	0	13	72	55	3	34	106	N	0
High Grades & Engaged	Yesterday	F	232	N	607	1st	0	22	195	178	1	62	257	N	0
High Grades & Engaged	2 days ago	M	232	N	195	1st	0	7	93	44	2	16	109	N	0
High Grades & Engaged	More than a week ago	F	232	N	298	1st	0	8	130	70	1	13	143	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	72	3rd	0	5	28	19	3	12	40	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	161	1st	0	8	69	18	4	11	80	N	0
High Grades & Engaged	Yesterday	M	232	N	332	1st	0	6	181	91	1	12	193	N	0
High Grades & Engaged	2 days ago	M	232	N	176	1st	0	9	90	38	2	14	104	N	0
High Grades & Not Engaged	2 days ago	M	232	N	122	2nd	0	5	72	17	4	6	78	N	0
High Grades & Engaged	Yesterday	M	232	N	179	1st	0	6	98	60	1	12	110	N	0
High Grades & Engaged	More than a week ago	F	232	Y	196	1st	0	7	108	45	1	13	121	N	0
High Grades & Engaged	2 days ago	M	232	N	296	1st	0	8	164	79	1	11	175	N	0
High Grades & Engaged	Yesterday	F	232	N	165	1st	0	7	82	40	2	15	97	N	0
High Grades & Engaged	3 days ago	M	232	N	222	1st	0	9	97	42	2	10	107	N	0
High Grades & Engaged	Yesterday	F	232	N	307	1st	0	9	185	63	1	16	201	N	0
High Grades & Engaged	Yesterday	M	232	N	137	2nd	0	7	68	30	2	11	79	N	0
High Grades & Engaged	3 days ago	M	232	N	249	1st	0	8	70	40	2	12	82	N	0
High Grades & Not Engaged	Yesterday	M	232	N	51	3rd	0	2	22	12	4	3	25	N	0

RISK CATEGORY	LAST ACCESSED	GENDER	NO. of DAYS IN COURSE	ACCESSED IN FIRST 2 WEEKS	TOOL CLICKS	CLICK QUARTILE	NO. FORUM POSTS	NO. SUBMISSIONS	NO. CONTENT ACCESSES	NO. COURSE ACCESSES	COURSE ACCESS QUARTILE	ASSESSMENT ACCESSES	NO. ITEM ACCESSES	ACCESSED PRIOR TO TERM	NO. LATE ASSIGNMENT
High Grades & Engaged	More than a week ago	F	232	N	196	1st	0	8	91	64	1	13	104	N	0
High Grades & Not Engaged	Yesterday	M	232	N	65	3rd	0	7	33	27	3	9	42	N	0
High Grades & Not Engaged	4 days ago	M	232	N	78	2nd	0	5	41	25	3	16	57	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	130	2nd	0	3	36	28	3	12	48	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	250	1st	0	4	91	28	3	6	97	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	203	1st	2	10	76	32	3	13	89	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	160	1st	1	13	48	34	3	18	66	N	0
High Grades & Engaged	More than a week ago	F	232	N	162	1st	0	10	57	33	3	24	81	N	0
High Grades & Engaged	Yesterday	M	232	N	209	1st	0	12	75	47	2	24	99	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	129	2nd	0	7	45	38	3	26	71	N	0
High Grades & Engaged	3 days ago	F	232	N	597	1st	1	17	226	75	1	42	268	N	0
High Grades & Engaged	More than a week ago	M	232	N	447	1st	17	26	93	49	2	18	111	N	0
High Grades & Engaged	More than a week ago	F	232	N	218	1st	0	7	103	41	2	20	123	N	0
High Grades & Engaged	More than a week ago	M	232	N	151	1st	1	14	44	51	1	30	74	N	0
High Grades & Engaged	2 days ago	M	232	N	177	1st	1	9	95	42	2	17	112	N	0
High Grades & Engaged	More than a week ago	F	232	N	261	1st	1	10	97	36	3	18	115	N	0
High Grades & Engaged	3 days ago	F	232	N	546	1st	2	16	235	124	1	30	265	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	93	2nd	3	12	26	28	3	15	41	N	0
High Grades & Engaged	Yesterday	F	232	N	969	1st	4	16	232	177	1	120	352	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	82	2nd	1	11	31	27	3	19	50	N	0
High Grades & Engaged	More than a week ago	F	232	N	408	1st	3	18	53	59	1	33	86	N	0
High Grades & Engaged	3 days ago	M	232	N	225	1st	1	13	106	62	1	21	127	N	0
High Grades & Not Engaged	4 days ago	F	232	N	151	1st	1	9	33	43	2	15	48	N	0
High Grades & Engaged	More than a week ago	M	232	N	250	1st	1	14	110	64	1	27	137	N	0
High Grades & Engaged	More than a week ago	M	232	N	286	1st	0	10	78	65	1	32	110	N	0
High Grades & Engaged	5 days ago	M	232	N	519	1st	1	14	251	83	1	38	289	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	31	4th	0	6	15	26	3	21	36	N	0
High Grades & Engaged	More than a week ago	F	232	N	314	1st	0	9	76	82	1	41	117	N	0
High Grades & Engaged	More than a week ago	M	232	N	222	1st	1	16	56	47	2	31	87	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	110	2nd	0	10	34	49	2	28	62	N	0
High Grades & Engaged	More than a week ago	M	232	N	496	1st	0	12	118	47	2	31	149	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	128	2nd	0	11	73	41	2	19	92	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	52	3rd	1	11	13	41	2	35	48	N	0
High Grades & Engaged	3 days ago	M	232	N	342	1st	1	10	73	46	2	25	98	N	0

RISK CATEGORY	LAST ACCESSED	GENDER	NO. of DAYS IN COURSE	ACCESSED IN FIRST 2 WEEKS	TOOL CLICKS	CLICK QUARTILE	NO. FORUM POSTS	NO. SUBMISSIONS	NO. CONTENT ACCESSES	NO. COURSE ACCESSES	COURSE ACCESS QUARTILE	ASSESSMENT ACCESSES	NO. ITEM ACCESSES	ACCESSED PRIOR TO TERM	NO. LATE ASSIGNMENT
High Grades & Engaged	More than a week ago	M	232	N	226	1st	0	9	75	50	2	20	95	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	177	1st	1	12	80	41	2	17	97	N	0
High Grades & Engaged	More than a week ago	F	232	N	278	1st	0	10	86	57	1	19	105	N	0
High Grades & Engaged	More than a week ago	M	232	N	633	1st	1	11	269	61	1	20	289	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	183	1st	1	14	38	38	3	24	62	N	0
High Grades & Engaged	More than a week ago	M	232	N	210	1st	0	10	72	46	2	18	90	N	0
High Grades & Not Engaged	3 days ago	M	232	N	170	1st	1	13	40	42	2	18	58	N	0
High Grades & Engaged	More than a week ago	M	232	N	273	1st	9	18	88	50	2	22	110	N	0
High Grades & Engaged	2 days ago	F	232	N	384	1st	0	15	239	100	1	30	269	N	0
High Grades & Engaged	4 days ago	M	232	N	214	1st	1	9	66	49	2	27	93	N	0
High Grades & Engaged	Yesterday	M	232	N	400	1st	1	13	78	65	1	29	107	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	56	3rd	1	9	21	17	4	12	33	N	0
High Grades & Engaged	More than a week ago	M	232	N	186	1st	1	10	52	44	2	22	74	N	0
High Grades & Engaged	More than a week ago	F	232	N	231	1st	1	16	71	64	1	21	92	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	159	1st	0	4	48	32	3	12	60	N	0
High Grades & Engaged	6 days ago	F	232	N	207	1st	1	14	55	64	1	15	70	N	0
High Grades & Engaged	More than a week ago	M	232	N	287	1st	2	10	112	38	3	27	139	N	0
High Grades & Not Engaged	2 days ago	F	232	N	60	3rd	1	10	18	28	3	16	34	N	0
High Grades & Engaged	More than a week ago	M	232	N	213	1st	1	13	84	41	2	23	107	N	0
High Grades & Engaged	More than a week ago	F	232	N	323	1st	1	12	114	57	1	32	146	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	5	4th	0	5	3	8	4	5	8	N	0
High Grades & Not Engaged	5 days ago	F	232	N	54	3rd	1	10	22	23	3	20	42	N	0
High Grades & Engaged	2 days ago	M	232	N	317	1st	1	14	127	54	1	27	154	N	0
High Grades & Engaged	More than a week ago	M	232	N	202	1st	1	13	70	28	3	22	92	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	41	3rd	0	9	16	39	2	17	33	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	64	3rd	0	1	14	20	4	5	19	N	0
High Grades & Engaged	More than a week ago	M	232	N	190	1st	1	10	58	51	1	28	86	N	0
High Grades & Engaged	More than a week ago	M	232	N	375	1st	1	14	82	76	1	44	126	N	0
High Grades & Engaged	3 days ago	M	232	N	121	2nd	1	15	28	39	2	19	47	N	0
High Grades & Not Engaged	3 days ago	M	232	N	115	2nd	0	5	41	25	3	14	55	N	0
High Grades & Engaged	More than a week ago	F	232	N	317	1st	1	14	111	38	3	17	128	N	0
High Grades & Engaged	Yesterday	M	232	N	361	1st	3	14	118	51	1	30	148	N	0
High Grades & Not Engaged	More than a week ago	F	232	N	51	3rd	0	2	21	22	3	7	28	N	0
High Grades & Not Engaged	More than a week ago	M	232	N	125	2nd	0	9	61	45	2	28	89	N	0

Figure 4.11: Student-at-risk identification using the designed model

The figure in the analysis displayed student at-risk alerts divided into four (4) quadrants, colour-coded like traffic lights:

- Students who are Not Engaged with Low Grades (highlighted in red),
- Students who are Engaged with Low Grades (highlighted in amber),
- Students who are Not Engaged with High Grades (highlighted in green),
- Students who are Engaged with High Grades (highlighted in green).

The findings in Table 4.4 showed that out of 200 students rows analysed:

- 57 students (28.5%) were Not Engaged with Low Grades (highlighted in red).
- 14 students (7%) were Engaged with Low Grades (highlighted in amber).
- 129 students (64.5%) were both Not Engaged with High Grades and Engaged with High Grades.

The designed model successfully identified at-risk students. However, the results were considered inadequate because the model did not provide insights into why Not Engaged students were at-risk and how the university could offer customised support. To effectively support at-risk students, different approaches based on their specific reasons for being Not Engaged should be implemented.

Additionally, there was contradictory evidence indicating "Engaged" students with Low Grades highlighted in green. In SA HEIs, assessment relies on Continuous Assessment (which includes attendance, participation in tasks, quizzes, assignments, and discussions) and Final Assessments (exams). Students with Low Grades should not be disregarded when it comes to active engagement and participation, as they significantly contribute to the overall grade. If SA universities introduce alternative assessment types, such as portfolios of evidence or open book exams with higher-order questions, the level of "Engagement" will become an even greater risk factor than "Low Grades".

Therefore, the answer to the question of what data is required for LA and the resulting systems integration requirements is that was datasets should be designed based on a clear, data-informed decision-making culture. This will ensure that the designed datasets can be effectively mapped against the integrated source system.

Table 4.5: Model element- results from mapping the CSF against the A4L project groundwork

Critical Success Factors	**Results**
Are the datasets generated from the integrated information systems used to inform model design primarily for the design challenge?	✓ The data **supports** the hypothesis.
Is the model design informed by theory, educational research, and practice?	✓ The data **supports** the hypothesis on the basis that the researcher made the recommendation.
Has the university reviewed the efficacy and transferability of datasets that are developed in foreign context?	✓ The data **rejects** the hypothesis.
Has the university avoided prioritising question-driven approaches to the application of LA (data driven), and rather designed models informed by educational theories able to account for contextual factors?	✓ The data **supports** the hypothesis.
Did the university avoid using LMS vendors or external data specialist to create its dashboards, and did it start developing its own dashboards, using its own data specialists?	✓ The data **rejects** the hypothesis.

Most important was a full understanding of what can be achieved with datasets and model designs underpinned by theory-informed frameworks. Such understanding supports the argument that lack of theoretically informed analytics can lead to meaningless datasets and model designs that are able to identify students at-risk, but cannot distinguish why the list of at-risk students came to be at-risk. Understanding why the at-risk students come to be at risk adds value to our body of knowledge when universities invest in providing immediate intervention and support measures that can prevent students from failing or dropping out.

4.5 Transformation: What were the institutional priorities in the adoption of LA?

The purpose of this question was to understand the process of transitioning from technical to social systems in the A4L case study. So far, the case study has primarily focused on the technical systems required for data sourcing and integration to create a learning analytics model for student at-risk analytics. However, moving forward, the case study will explore the social aspects prioritised during the A4L project implementation and at CPUT.

The documents and reports that related to the A4L project do not provide any evidence of initiatives to develop a LA policy or strategies to align with the implementation of LA at CPUT. Despite the CIET team being responsible for implementing the A4L project based on the institutional vision 2030 in the area of smart teaching and learning, the project lacked a systematic approach to transitioning. There was no coordinated effort to establish a LA policy that aligns with the overall strategy. Consequently, the project lacked effective leadership models to drive the implementation to its full potential.

The project leaders had divided visions, resulting in sporadic implementation. Each management unit focused on its own deliverables, leading to scattered project objectives, and a bottom-up rather than a top-to-bottom enterprise. This highlighted the need for improvement in project management.

Although the CIET team was dedicated to the development of the A4L project, certain transformation aspects, such as defining principles for privacy and ethical use of student data, had to be considered. Similar to the process of data sourcing, integration, and model design, the CIET team organised an information sharing session:

Data analytics strategy workshop 6-9 April 2020

Dear Colleagues,
Higher education institutions are continually being challenged to put in place pre-emptive measures as a means to maximise student success. The ministerial statement on the University Capacity Development Plan (UCDP), states that all universities should employ the use of student data analytics as a means to potentially analyse, plan and predict key risk factors within student populations, and subsequently devise interventions to negate these risk factors.
Towards the realization of these goals you are invited by the Centre for Innovative Educational Technology (CIET), to attend a workshop series. Our Learner Management System (LMS) partner, Blackboard, is at the cutting edge of using predictive information, from a range of sources, as to develop appropriate reporting systems evaluating student at risk. Over the course of 4 days this workshop series will facilitate the exploration, cross skilling, and high level analysis on how the institution can leverage analytics across the LMS and other data elements to provide insight, feasibility, provide analytics, segregate insights and become empowered and competent across the analytics platform to make better decisions and leverage the data for practical and profound use.

The Analytics Data Strategy workshop series is designed to help institutions understand the data that is collected by their institutional, technology investments, and how this data can be leveraged both to improve teaching and learning and support strategic efforts to address core institutional challenges such as enrollment, progression, retention, graduation, course design, adoption and key risk reporting.

The Workshop Goals include the following:

- Analytics overview and strategic use – the broader perspective
- Provide planning around analytics platform usage
- Provide insight into how LMS data can be leveraged to optimize teaching and learning
- Provide data streams to support decision making processes
- Data model overview
- Understand analytics platforms and when to leverage and cross apply different solutions
- Become adept and competent in analytics discovery and construction of reporting metrics
- Extensibility of the platform through integration and modelling insights for feasibility analysis

The workshop will be hosted by CIET at:
E- Learning building – 6 April 2020
Granger Bay Hotel School – 7, 8, 9 April 2020

Could you kindly send two people from your faculty/Unit who your want be trained to assist the faculty/Unit in the training, mining and using of data. Your response is appreciated. We have only space for 25 people hence the request for two representatives.
kind regards
[redacted]

Associate Professor [redacted] (PhD)
[redacted] for Innovative Educational Technology
[redacted] University of Technology,

Figure 4.12: Invite to the data analytics strategy workshops

This session involved:

- The A4L project stakeholders from CPUT as a task team that governed the project.
- The legal department to address the legal implications of the student data use (specifically the PoPI Act).
- An ethics expert for insights on ethical considerations.
- LA representatives from other SA universities that had implemented similar projects.

The purpose of the information sharing session was for the A4L task team to develop guidelines on how to handle the student data extracted from source systems for analysis. In order to answer the final question about the institutional priorities in LA implementation, the study created Table 4.6, which reflects a checklist of critical success factors identified though a literature review.

Table 4.6: Transformation element- results from mapping the CSF against the A4L project groundwork

Critical Success Factors	Results
Has the Executive Management been involved in creating the institutional policy and strategy for LA, in order to drive and oversee the implementation?	✓ The data **rejects** the hypothesis.
Has the university considered the legal and ethical implications on the use of student data?	✓ The data **supports** the hypothesis

In essence, this study has presented and discussed findings of critical success factors based on the theoretical framework consisting of three elements – data, model, and transformation with reference to the literature review and the A4L project as a single case study. The study mapped the critical success factors against the groundwork of the A4L case study to observe whether the results in each element (data, model, transformation) supported or rejected the hypothesis. The primary research question of this study was: **What are the critical success factors of Information Systems integrations required to facilitate LA at HEIs in SA?**

Results from aligning the critical success factors from literature with the groundwork of the A4L project as a case study provided some understanding to the hypothesis: **if HEIs can build an understanding of the critical success factors of Information Systems integration required to facilitate education-based analytics, then HEIs can run better analytics reports and improve LA output**. The majority of the data observed from the case study support the hypothesis in all three elements. What was more interesting in the observation was that, even though some of the data from the model and transformation elements rejected the hypothesis, positive results from the data element became strong enough to answer the primary research question and to support the hypothesis that HEIs can run better analytics reports when institutional source data is integrated. Therefore, building an understanding of source system integration capabilities required to build a LA model for analytics related to student-at-risk was found to be the most profound success factor that universities must enforce, for the reason that it improves analytics. One of the benefits of improved analytics that was highlighted by the case study was an improved ability for CPUT to identify at-risk students.

The next chapter (Chapter Five) concludes the study, discusses its limitations and makes recommendations for universities who wish to get started with systems integration for LA purposes, and for future research.

CHAPTER FIVE
RECOMMENDATIONS

5.1 Summary of the research

5.1.1 Overview

Higher Education Institutions (HEIs) from developed and developing countries such as South Africa (SA) have implemented Learning Analytics (LA) to, amongst other things, identify students at-risk of failing and those who face conditions that threaten their potential to complete their studies and to apply early interventions. Chapter One observed a challenge to the ability for HEIs to identify successfully implement LA for these purposes, for this inability was directly linked to higher education data sources not being integrated into coherent data repositories that facilitate decision making. Thus, there was a need to explore the critical success factors of Information Systems integration necessary to facilitate LA at these HEIs.

Chapter Two showed why it was necessary to optimise critical success factors by exploring an approach to the systematic adoption of LA in HEIs that had been developed and used by McKinsey and Company (Barton & Court, 2012). That approach provided a theoretical framework consisting of three elements -- data, model, and transformation -- on which the rest of the study was built. The theoretical underpinnings that informed the critical success factors of Information Systems integration included the review of literature on the state of source extraction mechanisms that are essential to integrate the multiple source systems for LA adoption. This second chapter not only explored the above-mentioned technical systems required during the process of embedding LA. It also examined the legal, ethical and leadership matters by exploring the adoption principles for privacy protection and ethical use of analytics, and the leadership models that could drive and oversee the implementation. In addition, the chapter reviewed literature concerned with profiling the source systems for the relevant datasets that could be used to inform a model design for at-risk students. Subsequently, an educational theory was adopted to ground the relevant datasets that influence why students come to be at risk using Tinto's (1975) Longitudinal Model of Dropout. Results from the review of literature offered the study critical success factors from each of the three elements mentioned above.

Chapter Three identified the empirical research design and methodology used to collect and analyse the data using the research onion devised by Saunders and others (2009). The researcher peeled away the layers of the research onion using the philosophy of post-positivism by means of a deductive approach. In addition, the study adopted a single case study strategy on the implementation of LA in a HEI in SA, namely the Cape Peninsula University of Technology (CPUT). The chapter concluded with a non-probability sampling plan. It used a purposive sample of at-least 200 rows of student secondary data from the Faculty of Engineering and Built Environments (FEBE) to test the designed model's capacity to identify students at-risk.

Chapter Four highlighted the CPUT case study, in which the questions were structured according to the theoretical framework adopted in Chapter Two from McKinsey and Company (Barton & Court, 2012). Results from the three elements -- data, model, and transformation -- supported the hypothesis, with minor critical success factors from the model and transformation element that rejected the hypothesis.

The study concluded that building a data-informed culture in decision-making for LA (from the data element), particularly if it led to systems integration, is a far-reaching step universities should take, for the simple reason that it can lead to improved analytics.

5.1.2 Research problem and questions

HEIs have substantive student-produced data at hand, yet are unable to extract datasets from multiple Information Systems that would enable them to identify the student-at-risk of failing and those who face conditions that threaten their potential to complete their studies, and so apply early interventions from LA. This inability is directly linked to HEI data sources that are not integrated into coherent data repositories that facilitate decision-making. The study had the following research question:

What are the critical success factors of Information Systems integrations necessary to facilitate LA at HEIs in SA?

Furthermore, to answer the research question, the following sub-questions had to be addressed:

- Data- What is the state of source extraction mechanisms that are crucial for integrating multiple institutional source systems for LA adoption?
- Model- Which data is necessary for conducting LA, and what are the resulting requirements for systems integration?
- When examining student at-risk analysis as an instance of LA performed by HEIs, what data modelling choices are made accessible to HEIs following successful integration of BI and analytics systems?

5.1.3 Aim and objectives of the study

The main aim of this study was to justify the capabilities of integrated institutional Information Systems as source systems by investigating the critical success factors that influence the successful implementation of LA using the application in the analytics related to students-at-risk as a distinct example that may improve the ability for HEIs to identify student-at-risk. In order to achieve this aim, the study had the following objectives:

- To gain an understanding of the necessary capabilities of source systems integration to develop a LA model for analysing student at-risk data.
- To determine the optimal option for combining dataset (predictor variables) from multiple unintegrated source systems to inform LA, using the case study application of student at-risk data analysis.
- To propose critical success factors for implementing LA effectively, including recommendations for modelling the identification process of at-risk tertiary education students. Furthermore, highlight the potential benefits of using integrated source systems for timely and responsive interventions in LA.

5.1.4 Literature study

Chapter Two presented the literature review of this study. The questions were derived from the McKinsey and Company (Barton & Court, 2012) theoretical framework consisting of the three elements:

- **Data:** What is the state of source extraction mechanisms that are crucial for integrating multiple institutional source systems for LA adoption?
- **Model:** Which data is necessary for conducting LA, and what are the resulting requirements for systems integration?

- **Transformation:** What were the institutional priorities in the adoption of LA?

The literature review also included Tinto's (1975) Longitudinal Model of Dropout to provide a theory informed conceptual framework applied in education-based analytics to ground the datasets acquired from the integrated source systems on theoretical reasoning.

Chapter One contains a preliminary literature review, in which the focus was on:

(i) substantive digital data collected and stored in institutional Information Systems

(ii) the current state of institutional Information Systems' integration in SA HEIs

(iii) the relevant datasets generated from multiple source systems to inform the model design of student at-risk

(iv) the statistical and theoretical models required to ground the acquired datasets into a statistical or theoretical point of view, when building a LA model.

5.1.5 Empirical research design

A qualitative research design (single method) was used to facilitate an in-depth study within the context of Information Systems integration from LA implementation at the Cape Peninsula University of Technology, the empirical setting of the case study. Results from the case study groundwork were mapped against the critical success factors derived from the theoretical framework that consist of elements such as data, model, and transformation derived from the literature review. Also, 200 rows of secondary data were used to gauge if the designed model had the capacity to identify students at-risk. Tinto's model was used to ground on theoretical reasoning the datasets designed to identify the at-risk student.

5.1.6 Analysis and results

The qualitative method analysis involved an extensive literature review by McKinsey and Company (Barton & Court, 2012) that consists of three elements -- data, model, and transformation. The data element focused on the source extraction mechanisms that are essential to integrate the multiple institutional source systems for LA implementation. The model element focused on the data required for LA by mapping out the source systems for relevant datasets which were used to inform the model design for the student at-risk identification process. The transformation element focused on the institutional priorities in the adoption of LA.

The study expanded on the literature review by grounding the datasets designed to identify at-risk students in theoretical reasoning. The aim was to explore the datasets to identify the reasons behind why students might become at-risk, in order to design differing interventions appropriate to the differing reasons. Critical success factors derived from the literature review were mapped against the groundwork from the case study, in order to analyse which of the critical success factors from the three elements the data might reject or support.

5.2 Conclusions

Based on the review of literature and in-depth qualitative analysis of sections from the systematic adoption of LA, the study not only identified the critical success factors of Information Systems integration required to facilitate LA, but it also explored social systems such as the development of institutional policy and strategy for LA, and the effective leadership development models to drive and oversee the implementation. The results of the integrated source system led to the design of a model fitted to identifying the at-risk students.

5.2.1 The research objectives in light of the findings

Objective 1: *To gain understanding of the necessary capabilities of source systems integration to develop a LA model for analysing student at-risk data.*

Chapter Two presented a theoretical framework, pointing out that universities have over the years invested considerable capital in a number of institutional Information Systems for different purposes. These institutional Information Systems have collected a substantive amount of data. Hence, it helps when universities are aware of the data afforded by the institutional information systems so that they may build an understanding of source systems integration requirements that are creative enough to enable them to address institutional dynamics such as the identification of at-risk students. Thus, it was necessary to consider the five key drivers afforded by Chaki (2015) when determining an integration approach. These key drivers include the following: (i) nature of extraction process between source systems and consuming systems (push/pull); (ii) type of connectors required for pulling data from source systems; (iii) the choice between using a data integration engine or a database engine for data transformations; (iv) the required outbound extract formats for consuming applications; and (v) the need to address data security and comply with any country-specific regulations during the integration process.

Chapter Two also highlighted how the process of finding relevant sources of data, based on the question to be addressed, should be informed by principles and theories established in educational research and practice. Chapter Two also shifted the understanding of datasets from not just searches on the internet towards more recent understanding and application of datasets found in contemporary literature on educational analytics. This included an in-depth interrogation on how an integration approach, a narration of datasets in educational analytics informed by a theory, and the limitations thereof that come with data integration, can affect the quality of results when addressing the student at-risk identification question.

Finally, the case study presented in Chapter Four indicated the involvement of internal stakeholders such as the IT department, functional staff, academic representatives, IT technicians with database expertise, the quality assurance department, and the Legal Department. The external stakeholders included a data specialist with specialised expertise, and if possible, outsourced from the Data Analytics consulting organisation appointed by the university, an ethics expert, and users from neighbouring universities that had already implemented LA.

Objective 2: *To determine the optimal method for combining datasets (predictor variables) from multiple unintegrated source systems to inform LA, using the case study application of student at-risk data analysis.*

Both Chapters Two and Four responded to the research objective. The theoretical framework presented in Chapter Two considered two approaches to adopt learning analytics. The first approach is data-driven and follows question-driven analytics to the application of machine learning where a university has neither internal capacity nor a well-defined challenge to address through LA. Such an approach has no distinct source systems to map the data from, as there is no distinct challenge to address through LA. Thus, should the design challenge be for at-risk students, results may reflect a list of students at-risk, but without any justifications of how the students come to be at-risk in order for the support system to provide differing interventions.

The second approach informs the use of machine learning via educational research and practice where a university has internal capacity and a well-defined challenge to address through LA. The model design to address the challenge is informed by an educational theory, as illustrated in Chapter Two. Results from this approach can differentiate between the reasons why students come to be at-risk, and to whom differing interventions can be applied. Unlike simple data-informed analytics, it does not put all the students in “one box”.

The second phase of Chapter Four of the case study highlighted four categories or reasons behind why a student may become at-risk: low engagement with low grades; high engagement and low grades; low engagement with high grades; and high engagement with high grades. Demonstrating that the designed model acquired from the data required from the integrated Information Systems were able to identify students-at-risk.

Objective 3: *To propose critical success factors for implementing LA effectively, including recommendations for modelling the identification process of at-risk tertiary education students. Furthermore, highlights the potential benefits of using integrated source systems for timely and responsive interventions in LA..*

Finally, the theoretical framework presented in Chapter Two transitioned from the technical systems of LA adoption such as systems integration, datasets of the source systems, and towards more socially oriented systems such as institutional policy and strategy, and effective leadership models to drive and oversee the implementation, to develop an analytics-informed decision-making culture.

Results from the above objectives made it possible for the researcher to identify and analyse the results of critical success factors needed to answer the primary research question: **What are the critical success factors of Information Systems integrations necessary to facilitate LA at HEIs in SA?**

5.2.2 Development of a LA approach conclusion

This research identified and analysed the results of critical success factors of Information Systems integration necessary to facilitate LA at HEIs in SA as follows:

Functional Requirement Specification (FRS) checklist of Critical Success Factors (CSF) required.
Data What is the current state of source extraction mechanisms that are crucial for integrating multiple institutional source systems for LA adoption? ☐ Has the university built a data informed culture in decision making for LA based on a predefined design challenge, for example, the student at-risk identification? ☐ Has the university established an effective source systems integration approach, and understood the limitations which the identified source systems data may have to offer? ☐ Has the university secured the necessary information technology support and other stakeholders similar to the institutional ethics and legal department?
Model Which data is necessary for conducting LA, and what are the resulting requirements for systems integration? ☐ Are the datasets generated from the integrated information systems used to inform model design primarily for the design challenge? ☐ Is the model design informed by educational research and practice? ☐ Has the university reviewed the efficacy and transferability of datasets that are developed in a foreign context? ☐ Has the university avoided prioritising question-driven approaches to the application of LA (data driven) and rather designed models informed by educational theories able to account for contextual factors? ☐ Did the university avoid making use of LMS vendors or external data specialists to create its dashboards and started developing its own dashboards, using its own data specialists?
Transformation What were the institutional priorities in the implementation of LA? ☐ Has the Executive Management been involved in creating of the institutional policy and strategy for LA, in order to drive and oversee the implementation? ☐ Has the university considered the legal and ethical implications of the use of student data?

The study mapped the identified critical success factors against the groundwork of the A4L case study at CPUT for each element -- data, model, and transformation.

5.2.3 Contribution of the study

The data element -- the identification of data source systems and the integration of data sources -- represents the most critical factor of Information Systems integration, and contributes to the successful implementation of LA. The key drivers when determining an integration approach, the theoretical framework that consists of the three elements – data, model, and transformation, including a theory underpinning the identified datasets -- are a guide for HEIs that wish to develop a LA approach for a predefined challenge such as the student at-risk. Informed by educational theory and practice, institutions that have already implemented their LA initiatives can review their project scope based on the findings of the study, starting from a data integration approach to the model design. Such practice will enable universities to address predefined challenges such as the student at-risk identification.

The model design also emphasised the importance of a predefined institutional challenge to address through the use of LA in order to understand what can be achieved with datasets extracted from the integrated source systems. The key drivers when determining an integrated approach, the critical factors of Information Systems integration carried out to facilitate the successful implementation of LA in HEIs, contribute to the emerging body of knowledge about LA implementation in a developing country. The model design highlighted that LA implementation in developing countries should go beyond data-informed analytics. Finding relevant sources of data, determining an integrated approach to source the data into a data warehouse, and contextualising the datasets extracted from the integrated source systems using a theory informed framework can allow HEIs to design more meaningful educational analytics. Furthermore, the review of literature regarding the key drivers when determining an integration approach, including the LA implementation approach, provide a comprehensive interpretation of evidence in LA implementation.

5.3 Recommendations

5.3.1 The resulting systems integration requirements

For quality analytics and informed interventions, the quality assurance process should be inclusive in the review of the LA implementation project. In the long run, should the institution decide to expand its analytics, acquiring additional source systems beyond the current design challenge for integration will be an advantage. More powerful and meaningful educational analytics that may be required in future so that universities may provide better support and real time interventions. For this to apply, universities should support the pull nature of extraction between source systems and consuming systems. As a security measure, the study also recommends universities should replicate the source database. Nor should they allow external consulting organisations direct access to their institutional databases and source systems tables. Simply put, they should retain the integrity of their student data.

The extraction process may be lengthy and complex, for the reason that some of the data extracted from the source systems may be too large or else may contain irrelevant data. It should be noted that since institutional needs currently being addressed through the use of analytics may change, the integration approach and mapping of source systems should always be of better quality than that of the initial challenge which the project was developed to address. Moreover, universities should acknowledge that some of the source systems (especially the legacy information systems) were not developed for Online-Analytical-Processing (OLAP) technologies (Chatti, Dyckhoff, Schroeder, & Thus, 2012). Nevertheless, they hold some of the most valuable institutional data that may improve institutional analytics.

5.3.2 The data required for LA

Special care is vital when designing a model with data extracted from the source system. Missing datasets may compromise the predictive power of a model. In addition, models designed in one faculty may not provide similar results when applied in a different faculty. Therefore, when designing or deciding on a model, a small survey to determine student context is important.

5.3.3 The theoretical framework required to ground the LA datasets

Different theoretical frameworks for different LA needs should be always applied. Should the institutional design challenge for LA be about at-risk students or the LMS adoption, a theoretical framework grounding the datasets for at-risk students should be applied. The majority of analytics in HEIs are data driven (Sclater et al., 2016), for instance:

- the use of BI tactics such as factor analysis and logistic regression to produce appropriate course signals at Purdue University in Indiana (Arnold et al., 2012),
- the Open Academic Analytics Initiative (OAAI) led by Marist College in New York to transfer predictive models to other institutions based on data from Marist College, building on Purdue's approach (Miteva et al., 2017); and
- the use of diagnostic tools to analyse social networks at the University of Wollongong in Australia (Dawson et al., 2011).

While theoretical frameworks such as Tinto's have over the years offered evidence about the heterogeneity of student-at-risk identification, scientific research in LA has hardly dealt with this concern (Janosz et al., 2000). Instead, research has been presenting the degree to which the datasets relate to the at-risk behaviour with marginal knowledge and in-depth understanding of the authentic student-at-risk process. This explains the special relevance of Tinto's theory for uncovering and examining the profiled datasets. Respective universities should build a typology of risk factors based on students' academic and social experiences, while simultaneously creating awareness of how the student-at-risk datasets that influence diverse students in dissimilar ways at different times affect the authenticity of the student-at-risk identification process.

In a study conducted by Ishitani (2008) on the timing of the student-at-risk of dropping out over a five-year period, findings showed that the effects of the datasets that show how students come to be at risk change over time. Based on these findings, an assertive deduction is that student conditions are not permanent, and that their risk conditions are not alike. For example, ten (10) students may be at risk due to poor marks, but the conditions that led each of the students to be at risk may vary, hence the need for scientific research to adopt theories similar to Tinto's (1975) Longitudinal Model of Dropout that ground the student-at-risk process. Once the overall process of why students come to be at risk has been understood, then universities will have the capacity to provide differential interventions using datasets from the integrated institutional information systems data showing that different students are affected in different ways at different times.

5.4 Recommendations for further study

The study recommends further research on the impact of legacy systems that are complex to integrate or use as source systems. Legacy systems have valuable data that can be used to help improve the quality of analysis in institutions, but they were not developed for learning analytics.

In addition, certain risk conditions are external to the students' academic and social systems within the university have a direct impact on students' academic performance. These include death in the family, chronic illness, troubled household, and low socio-economic status. Further research on how universities can secure institutional Information Systems integration with external data sources that hold records of such risk conditions is necessary. Examples of these data sources include the Health Department for when a student or a next of kin has been admitted, or the Department of Labour for when a student or a guardian has attained employment or is unemployed. The same data is available to institutional bursary proprietors such as the National Student Financial Aid Scheme (NSFAS) using the Labour Department's data when it needs to confirm if the parents or guardians of students are employed.

To end, further research on the use of diverse educational theories for LA implementation is necessary, particularly where different theories are being used to ground LA datasets into theoretical reasoning.

5.5 Limitations

A limitation of this research is that it relies on a single case study of a single academic university, and the researcher understands that the results of the model designed and tested in this study cannot be generalised for other academic universities. Moreover, the findings of this study were based on records analysis from secondary data, which may be open to measurement error, missing values, and unpredictable calculations.

REFERENCES

Abdelhakim, A. (2021). Adopted Research Designs by Tourism and Hospitality Postgraduates in the Light of Research Onion . International Journal of Tourism and Hospitality Management , 98-124.

Adams, G., & Schvaneveldt, J. (1991). Understanding Research Methods. New York: Longman.

Ain, N., Vaia, G., DeLone, W., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success–A systematic literature review. Decision Support Systems, 113113.

Akoojee, S., & Nkomo, M. (2007). Access and quality in South African higher education: the twin challenges of transformation. South African Journal of Higher Education, 21(3), 385-399.

Alammary, A., Sheard, J., & Carbone, A. (2014). Blended learning in higher education. Australasian Journal of Educational Technology , 30(4), 440-454.

Aliaga, M., & Gunderson, B. (2002). Interactive Statistics . Sage .

Allen, J., Cavanagh, T., Gunkel, M., & Whitmer, J. (2017). 7 Things you should Know About Developments in Learning Analytics . EDUCAUSE Learning Initiatives .

Allen, J., Cavanagh, T., Whitmer, J., & Gunkel, M. (2017, July 11). 7 Things You Should Know About Developements in Learning Analytics. Retrieved September 2, 2019, from EDUCAUSE: https://library.educause.edu/resources/2017/7/7-things-you-should-know-about-developments-in-learning-analytics

Alyaseri, N., Salman, M., Maseer, R., Hussein, E., Subhi, K., Alwan, S., & Abed, R. (2023). Exploring the Modeling of Socio-Technical Systems in the Fields of Sport, Engineering and Economics. Revista iberoamericana de psicología del ejercicio y el deporte, 338-341.

Ameri, S., Fard, M., Chinnam, R., & Reddy, C. (2016). Survival analysis based framework for early prediction of student dropout . USA: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management .

Arndt, T., & Guercio, A. (2016). A Formalism for Plan- A Big Data Personal Learning Assistant for University Students. Journal of e-Learning and Knowledge Society, 12(2), 13-25.

Arnold, K., & Pistilli, M. (2012). Course Signal at Purdue: Using Learning Analytics to Increase Student Success. Canada: ACM International Conference Proceeding Series.

Arnold, K., & Pistilli, M. (2012). Course signals at Purdue: using learning analytics to increase student success. Canada: Learning Analytics and Knowledge .

Asikhia, O. (2010). Student and Teachers' Perspective of the Causes of Poor Academic Performance in Ogun State Secondary Schools: Implications for National Development . European Journal of Social Sciences, 13(2), 229-242.

Aulck, L., Nishant, V., Joshua, W., & Jevin, B. (2016). Predicting Student Dropout in Higher Education. eprint arXiv:1606.06364, 1(4), 16-20.

Azaiez, N., & Akaichi, J. (2018). Integrating Trajectory Data in the Warehousing Chain: A New Way to Handle the Trajectory ELT Process. In G. De Pietro, L. Gallo, & R. J. Howlett (Eds.), Intelligent Interactive Multimedia Systems and Services (pp. 353-361). Australia: Springer .

Babbie, E., & Mouton, J. (2005). The Practice of Social Research. Cape Town: Oxford University Press Southern Africa.

Baker, R. S., & Inventando, P. S. (2014). Educational data mining and learning analytics. In J. Larusson, & B. White (Eds.), Learning Analytics: From Research to Practice (pp. 61-75). New York : Springer.

Barton, D., & Court, D. (2012). Making advanced analytics work for you . Harvard Business Review , 90(10), 78-83.

Barton, R. (2010). Simulation experiment design. 2010 Winter Simulation Conference (pp. 75-86). Pennsylvania: IEEE.

Baskenville, R. (1996). A Critical Perspective on Action Research as a Method for Information Systems Research . Journal of Information Technology, 235-246.

Baskerville, R., Rowe, F., & Wolff, F. (2018). Integration of Information Systems and Cybersecurity Countermeasures: An Exposure to Risk Perspective. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 49(1), 33-52.

Bean, J., & Metzner, B. (1985). A Conceptual Model of Nontraditional Undergraduate Student Attrition. Review of Educational Research, 55(4).

Becker, H. (1970). Sociological work: Method and substance. Chicago: Aldine Publishing.

Becker, S., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V., & Pomerantz, J. (2018). The NMC Horizon Report: 2018 Higher Education Edition. Louisville: EDUCAUSE Publications.

Becker, S., Cummins, M., Davis, A., Freeman, A., Hall, C., & Ananthanarayanan, V. (2017). NMC horizon report: 2017 Higher Education Edition. Texas: The New Media Consortium.

Ben-Haim, Y., & Tom-Tov, E. (2010). A Streaming Parallel Decision Tree Algorithm. Journal of Machine Learning Research, 11(1), 849-872.

Bennett, R. (2003). Determinants of Undergraduate Students Drop Out Rates in a University Business Studies Department. Journal of Future and Higher Education, 27(2), 123-141.

Bharamagoudar, S., Geeta, R., & Totad, S. (2013). Web Based Student Information Management System. International Journal of Advanced Research in Computer Engineering , 2(6), 2342- 2348.

Bhargava, N., Sharma, G., Bhargava, R., & Mathuria, M. (2013). Decision Tree Analysis on J48 Algorithm for Data Mining . International Journal of Advanced Research in Computer Science and Software Engineering , 3(6), 1114-1119.

Bidgoli, H. (2004). The Internet Encyclopaedia (Volume 3 ed., Vol. 3). California: John Wiley & Sons .

Biplob, M., Sheraji, G., & Khan, S. (2018). Comparison of different extraction transformation and loading tools for data warehousing. International Conference on Innovations in Science, Engineering and Technology (ICISET), 262-267.

Biskupek, A. (2019). Research on the Impact of Project Team on Selected Areas of Project Management. Trends Economics and Management , 29-46.

Bock, D. (2014). Multilevel Analysis of Educational Data. London: Elsevier.

Bordens, K. S., & Abbott, B. B. (2002). Research designs and methods: A process approach (5th ed.). New York: McGraw-Hill.

Brainbridge, J., Melitski, J., Zahradnik, A., Lauria, E. J., & Baron, J. (2015). Using Learning Analytics to Predict At-Risk Students in Online Graduate Public Affairs and Administration Education. Journal of Public Affairs Education, 21(2), 247-262.

Braxton, J. (2019). Rethinking the Causes and Cures of Student Attrition by Vincent Tinto . Journal of Collage Student Development , 60(1), 129-134.

Broos, T., Hilliger, I., Perez-Sanagustin, M., Htun, N., Millecamp, M., Pesantez-Cabrera, P., . . . De Laet, T. (2020). Coordinating learning analytics policymaking and implementation at scale. British Journal of Educational Technology , 51(4), 938-954.

Brum, L., Lampert, V., & Camargo, S. (2019). Business intelligence and data warehouse in agrarian sector: a bibliometric study. Embrapa Pecuaria Sul-Artigo em periodico indexado (ALICE), 353-368.

Bryman, A. (2012). Social research methods . Oxford: Oxford University Press.

Bytheway, A. (2000, December 11). An analysis of the current and future deployment of information systems and technology at the university of the western cape. Retrieved 03 25, 2019, from https://repository.uwc.ac.za/handle/10566/1319

Candela, L., Gutierrez, A., & Keating, S. (2015). What predicts nurse faculty members' intent to stay in the academic organisation? A structural equation model of national survey of nursing faculty . Nurse education today, 35(4), 580-589.

Castleman, B., & Long, B. (2013, December). Looking Beyond Enrolment: The Causal Effect of Need-Based Grants on Collage Access, Persistence and Graduation. NBER Working Paper No. 19306, 1-42. Cambridge: National Centre for Postsecondary Research .

Chaki, S. (2015). Enterprise Information Management in Practice . New York: Aspress.

Chang, S., Chang, I., & Wang, T. (2014). Information systems integration after merger and acquisition. Industrial Management & Data Systems, 114(1), 37-52.

Chatterjee, S., & Hadi, A. (2015). Regression analysis by example (5th ed.). Canada: John Wiley & Sons.

Chatti, M., Dyckhoff, A., Schroeder, U., & Thus, H. (2012). A Reference Model for Learning Analytics . International Journal of Technology Enhanced Learning (IJTEL) , 4(5-6), 318-331.

Chaubey, A., & Bhattacharya, B. (2015). Learning Management Systems in Higher Education. International Journal of Science Technology & Engineering, 2(3), 158-162.

Chaubey, A., & Bhattacharya, B. (2015). Learning Management Systems in Higher Education . International Journal of Science Technoogy & Engineering, 2(3), 158-162.

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. Communications of the ACM, 54(8), 88-98.

Chen, R. (2012). Institutional Characteristics and Collage Student Dropout Risk: A Multilevel Event History Analysis. Research in Higher Education, 53(5), 487-505.

Chen, R. (2012). Institutional Characteristics and College Student Dropout Risks: A Multilevel Event History Analysis. Research in Higher Education, 53(5), 487-505.

Chimka, J., Reed-Rhoads, T., & Barker, K. (2007). Proportional hazards models of graduation. Journal of College Student Retention: Research, Theory & Practice, 221-232.

Chui, K., Fung, D., Lytras, M., & Lam, T. (2018, June 27). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. In Press, pp. 1-7.

Cohen, P., Stephen, W. S., & Aiken, L. (1983). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd ed.). New York: Psychology Press.

Comrey, A., & Lee, H. (2013). A first course in factor analysis (2nd ed.). New York: Psychology Press.

Conijn, R., & van Zaanen, M. (2017). Trends in student behavior in online courses. Proceedings of the 3rd International Conference on Higher Education Advances, (pp. 649-657). Spain.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. IEEE Transactions on Learning Technologies, 10(1), 17-29.

Council on Higher Education. (2010). Access and throughput in South African Higher Education: Three case studies. Pretoria: Council on Higher Education.

Cox, D. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society , 34(2), 187-220.

CPUT, C. P. (2018, January 3). General Handbook: Academic and Student Rules and Regulations. Retrieved November 27, 2018, from https://www.cput.ac.za/storage/study_at_cput/Student_Rules%20_Regulations_2018.pdf

Cummin, J. (2008). BICS AND CALP: Empirical and theoretical status of the distinction (2nd ed.). (B. Street, & N. Hornberger, Eds.) New York: Encyclopaedia of Language and Education.

Cunningham, U. (2014). Teaching the disembodied: othering and activity systems in blended syncronous learning situation. The International Review of Research in Open and Distance Learning , 15(6), 33-51.

Dabbagh, N., & Kitsantas, A. (2012). Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. The Internet and Higher Education, 15(1), 3-8.

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. British journal of educational technology, 46(5), 904-920.

Daniel, J. (2011). Sampling Essentials: Practical Guidelines for Making Sampling Choices. USA: Sage Publications.

Darling-Hammond, L., Zielezinski, M., & Goldman, S. (2014). Using technology to support at-risk students' learning . California: Allience for Excellent Education .

Das, R., Jain, K., & Mishra, S. (2018). Archival research: A neglected method in organisation studies. Benchmarking: An International Journal.

Daud, A., Aljohani, N., Abbasi, R., Lytras, M., Abbas, F., & Alowibdi, J. (2017). Predicting student performance using advanced learning analytics. Australia: International World Wide Web Commette .

Dawson, S., Macfadyen, L., Lockyer, L., & Mazzochi-Jones, D. (2011). Using social network metrics to assess the effectiveness of broad based admission practices. Australasian Journal of Educational Technology, 27(1), 16-27.

De Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., & Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. British Journal of Educational Technology, 46(6), 1175-1188.

De Klerk, D., Spark, L., Jones, A., & Maleswena, T. (2017). Paving the Road to Success: Reflecting critically on year one of an undergraduate student support programme at a large South African university. Journal of Student Affairs in Africa, 5(1), 1-13.

de Valero, Y. (2001). Departmental factors affecting time-to-degree and completion rates of doctoral students at One Land-Grant Research Institution. The Journal of Higher Education, 72(3), 341-364.

DeLone, W., & McLean, E. (1992). Information Systems Success: The Quest for the Dependent Variable. Information Systems Research, 3(1), 60-95.

Delvin, M. (2010). Non-traditional achievement: Theory, policy and practice in Australian higher education. 13th Pacific Rim First Year in Higher Education Conference. Brisbane.

Deng, L., & Tavares, N. (2013). From Moodle to Facebook: exploring students' motivation and experiences in online communities. Computers & Education, 68(1), 167-176.

Department of Education. (2001). Higher Education Institutional Plans: An Overview of the First Planning Phase. Pretoria: The Republic of South African Department of Education .

Department of Higher Education and Training. (2017). Statistics on Post-School Education and Training in South Africa. Pretoria: Department of Higher Education and Training.

Devlin, M. (2008). Research challenges inheriting in determining improvement in university teaching . Issues in Educational Research, 18(1), 12-25.

Devlin, M. (2010). Non-traditional achievement: Theory, policy and practice in Australian higher education . Brisbane: 13th Pacific Rim First Year in Higher Education Conference.

DeZure, M., Van Note Chism, N., Deane Sorcinelli, M., Cheong, G., Ellozy, A., Holley, M., & Atrushi, D. (2012). Building international faculty-development collaborations: The evolving role of American teaching centres . Change: The Magazine of Higher Learning , 44(3), 23-33.

Dias, S., & Diniz, J. (2013). Towards an Enhanced Learning Management System for Blended Learning in Higher Education Incorporating Distinct Learners' Profiles. Journal of Educational Technology & Society , 17, 307-319.

Dietrichson, J., Bog, M., Filges, T., & Jorgensen, A.-M. (2017). Academic Interventions for Elementary and Middle School Students with Low Socioeconomic Status: A Systematic Review and Meta-Analysis. Review of Educational Research, 87(2), 243-282.

Dietz-Uhler, B., & Hurn, J. (2013). Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective. Journal of Interactive Online Learning, 12(1), 17-26.

Dietze, S., Sanchez-Alonso, S., Ebner, H., Qing Yu, H., Giordano, D., Merenzi, I., & Pereira-Nunes, B. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches. Program, 47(1), 60-91.

Dillahunt, T., Wang, B., Teasley, & S. (2014). Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. The International Review of Research in Open and Distributed Learning, 15(5), 178-196.

Dominguez-Whitehead, Y. (2017). Food and housing challenges: (Re)framing exclusion in higher education. Journal of Education (University of KwaZulu-Natal), 1(68), 149-169.

Doolan, M., & Guiza, M. (2015). Towards a Novel Methodology for Adopting Blended Collaborative Learning Solutions. Nassau: Bahamas: 10th International Conference on e-Learning: ICEL 2015.

Drake, B., & Walz, A. (2022). Evolving business intelligence and data analytics in higher education. New Directions for Institutional Research, 2018(178), 39-52.

Duarte, R., Ramos-Pires, A., & Goncalves, H. (2013). Identifying at-risk students in higher education. Total Quality Management & Business Excellence, 25(7-8), 1-19.

Duarte, R., Ramos-Pires, A., & Goncalves, H. (2013). Identifying at-risk students in higher education . Total Quality Management & Business, 25(7-8), 1-19.

Ebert, N., Weber, K., & Koruna, S. (2017). Integration platform as a service. Business & Information Systems Engineering , 59(5), 375-379.

Ewing, V., Stukas Jr, A., & Sheehan, E. (2003). Student prejudice against gay male and lesbian lecturers. The Journal of Social Psychology, 143(5), 569-79.

Ezen-Can, A., Boyer, K., Kellogg, S., & Booth, S. (2015). Unsupervised modelling for understanding MOOC discussion forums: a learning analytic approach. New York: Proceeding of the fifth international conference on learning analytics and knowledge: ACM.

Fakude, L., Roux, L., Daniels, F., & Scheepers, N. (2014). Reflections of nursing students, lecturers and clinical supervisors in the Western Cape large classes. South African Journal of Higher Education, 28(6), 1762-1775.

Falakmasir, M., & Habib, J. (2010). Using educational data mining methods to study the impact of virtual classroom in e-learning . USA: Educational Data Mining .

Falkner, N., & Falkner, K. (2012). A fast measure for identifying at-risk students in computer science. New Zealand: Proceedings of the ninth annual international conference on International computing education research .

Fataar, A. (2018). Placing Students at the Centre of the Decolonizing Education Imperative: Engaging the (Mis)Recognition Struggle of Students at the Post-apartheid University. A Journal of the American Educational Studies Association, 54(4), 1-14.

Fathema, N., Shannon, D., & Ross, M. (2015). Expanding The Technology Acceptance Model (TAM) to Examine Faculty Use of Learning Management Systems (LMSs) in Higher Education Institutions. Merlot Journal of Online Learning and Teaching, 11(2), 210-232.

Felizer, Y. (2010). Doing mixed methods research pragmatically: Implication for the discovery of pragmatism as a research paradigm. Journal of mixed methods research , 6-16.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning , 4(5-6), 304-317.

Ferguson, R. (2012). Learning Analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6), 304-317.

Flick, U. (2011). Introducing research methodology: A beginner's guide to doing a research project. London: Sage.

Fortin, L., Marcotte, D., Diallo, T., Potvin, P., & Royer, E. (2013). A multidimensional model of school dropout from an 8-year longitudinal study in a general high school population. European journal of psychology of education, 28(2), 563-583.

Fournier, H., Kop, R., & Sitlia, H. (2011). The value of learning analytics to networked learning on a personal learning environment . Canada: ACM: Digital Libraries.

Fowler, P., & Boylan, H. (2010). Increasing Student Success and Retention: A Multidimensional Approach. Journal of Developmental Education, 34(2), 2-10.

Frazier, W. (2009). A study of themed residential learning communities at a Midwest four-year university: North Dakota State University. North Dekota: Proquest.

Gable, G. (1994). Integrating case study and survey research methods: an example in information systems. European journal of information systems, 112-126.

Gagliardi, J., Parnell, A., & Carpenter-Hubin, J. (2018). The analytics revolution in higher education. Change: The Magazine of Higher Learning , 50(2), 22-29.

Gallivan, M. (1997). Value in Triangulation: A Comparison of Two Approaches for Combining Qualitative and Quantitative Methods . Information Systems and Qualitative Research, 417-443.

Garani, G., Chernov, A., & Savvas, I. (2019). A data Warehouse Approach for Business Intelligence. In 2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE) (pp. 70-75). IEEE.

Garrisin, D., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. The Internet and Higher Education, 7(2), 95-105.

Gasevic, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. TechTrends, 59(1), 64-71.

Gasevic, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. Internet and Higher Education, 28(1), 68-84.

Gautreau, C. (2011). Motivational Factors Affecting the Integration of a Learning Management System by Faculty. The Journal of Educators Online, 8(1), 1-25.

Gikandi, J., Morrow, D., & Davisa, N. (2011). Online formative assessment in higher education . Computers & Education , 57(4), 2333-2351.

Gioioa, D., & Pitre, E. (1990). Multiparadigm perspectives on theory building. Academy of Management Review, 15(4), 584-602.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. The qualitative report, 597-607.

Goyal, M., & Vohra, R. (2012). Application of Data Mining in Higher Education . International Journal of Computer Science Issues, 9(2), 113-120.

Grant, D., Hall, R., Wailes, N., & Wright, C. (2006). The false promise of technological determinism: the case of enterprise resource planning systems. New Technology, Work and Employment , 21(1), 2-15.

Gray, D. (2009). Doing Research in the Real World (2nd ed.). SAGE Publications.

Guerrero, S. (2021). Consuming Data in Fiori Applications . In S. Guerrero, Custom Fiori Applications in SAP HANA (pp. 37-80). Apress, Berkeley, CA.

Guitart, I., & Conesa, J. (2016). Creating University Analytical Information Systems: A Grand Challenge for Information Systems Research . In Formative Assessment, Learning Data Analytics and Gamification (pp. 167-186). Barcelona : Academic Press.

Gursoy, M., Inan, A., Nergiz, M., & Saygin, Y. (2016). Privacy-preserving learning analytics: challenges and techniques. IEEE Transactions on Learning technologies, 10(1), 68-81.

Halili, S. (2019). Technological advancements in education 4.0. The Online Journal of Distance Education and e-Learning, 7(1), 63-69.

Harris, A., & Ashton, J. (2011). Integrating academic and language skills within management units. In Developing student skills for the next decade, 1-9.

Hemphill, S., Plenty, S., & Herrenkohl, T. (2014). Student and school factors associated with school suspension: A multilevel analysis of students in Victoria, Australia and Washington State, United States. . Children and youth services review, 36(1), 187-194.

Herodotou, C., Rienties, B., Hlosta, M., Boroowa, A., Mangafa, C., & Zdrahal, Z. (2020). The scalable implementation of predictive learning analytics at a distance learning university: Insights from a longitudinal case study. The Internet and Higher Education, 45, 100725.

Homayouni, H. (2018). Testing Extract-Transform-Load Process in Data Warehouse Systems. (pp. 158-161). Memphis, USA: 2018 IEEE international symposium on software reliability engineering workshops (ISSREW).

Hosseinpoor, A., Nambiar, D., Suparmi, E., & Kusumawardani, N. (2018). Data source mapping: an essential step for health inequality monitoring . Global Health Action, 13-19.

Ibrahim, F., Susanto, H., Haghi, P., & Setiana, D. (2020). Shifting paradigm of education landscape in time of the COVID-19 pandemic: Revealing of a digital education management information system. Applied System Innovation, 3(4), 1-21.

Imbayarwo-Chikosi, V. (2015, November 27). Genetic Evaluation of Functional Longevity in South African Holstein Cattle Using a Proportional Hazards Model. Retrieved from PhD dissertation, Stellenbosch University : https://scholar.sun.ac.za/items/b7466a61-d730-4a75-af33-c3709e1fe4c5

Irakoze, J. M. (2015). The Effects of Information Systems on End-Users and the Organisational Processes at University of Technology. Cape Town: Cape Peninsula University of Technology .

Ishitani, T. (2008). How do transfers survive after “transfer shock”? A longitudinal study of transfer student departure at a four-year institution. Research in Higher Education, 403-419.

Islam, A. (2011). The determinants of the post-adoption satisfaction of educators with an e-learning system. Journal of Information Systems Education , 22(4), 319-331.

Janosz, M., Le Blanc, M., Boulerice, B., & Tremblay, R. (2000). Predicting Different Types of School Dropouts: A Typological Approach With Two Longitudinal Samples. Journal of Educational Psychology , 92(1), 171-190.

Jayaprakash, S., Moody, E., Lauría, E., Regan, J., & Baron, J. (2014). Early alert of academically at-risk students: An open source analytics initiative. Journal of Learning Analytics, 1(1), 6-47.

Johnson, B. (1965). Durkheim's one cause of suicide. American Sociological Review, 875-886.

Johnson, B., & Onwuegbuzie, A. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. American Educational Research Association, 14-26.

Judd, T., Kennedy, G., & Cropper, S. (2010). Using wikis for collaborative learning through contribution. Australasian Journal of Educational Technology, 26(3), 341-354.

Kaushik, V., & Walsh, C. (2019). Pragmatism as a research paradigm and its implication for social work research . Social Science, 255.

Kelly, N., Montenegro, M., Gonzalez, C., Clasing, P., Sandoval, A., Jara, M., . . . Alarcon, R. (2017). Combining event- and variable- centred approaches to institution- facing learning analytics at unit of study level. The International Journal of Information and Learning Technology, 34(1), 63-78.

Kember, D. (1995). Open Learning Courses for Adults: A Model of Student Progress . New Jersey: Educational Technology Publications.

Kika, C., Duan, Y. D., Cao, & Guangming. (2017). The Use and Critical Success Factors of Learning Analytics: An Organisational Absorptive Capacity Analysis. Association for Information Systems.

King, N. (2004). Advising underprepared students. NACADA: Summer Institute on Advising.

Kirilidog, M., van der Vyver, C., Zeeman, M., & Coetzee, W. (2018). Unfulfilled need: reasons for insufficient ICT skills in South Africa. Information Development, 34(1), 5-19.

Kirk, J., Miller, M., & Miller, M. (1986). Reliability and validity in qualitative research. Sage.

Koch, A., Nafziger, J., & Nielsen, H. (2015). Behavioural economics of education. Journal of Economic Behaviour & Organization , 115, 3-17.

Kock, N., Gallivan, M., & DeLuca, D. (2008). Furthering information systems action research: a post-positivist synthesis for four dialectics. Journal of the Association for Information Systems, 4.

Kosow, H., & Gabner, R. (2008). Methods of Future and Scenario Analysis: Overview, assessment, and selection criteria. Germany: Studies/Deutsches Institut für Entwicklungspolitik.

Kozma, R., & Croninger, R. (1992). Technology and the fate of at-risk students. Education and Urban Society , 4(24), 440-453.

Lachev, T., & Prince, E. (2018). Applied Microsoft Power BI: Bring your data to life (3rd ed.). Atlanta: Prologika Press.

Laudon, K., & Laudon, J. (2012). Management Information Systems (12th ed.). USA: Pearson Prentice Hall.

Laurillard, D. (2013). Teaching as a design science: Building pedagogical patterns for learning and technology. New York: Routledge.

Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of 'at-risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. Educational Technology Research and Development , 64(5), 957-968.

Letseka, M., & Maile, S. (2008). High university drop-out rates: A threat to South Africa's future. Pretoria: Human Science Research Council.

Lindsey, J. (2000). Applying generalized linear models. New York: Springer Science & Business Media.

Liu, X. (2012). Survival analysis: Models and Applications. United Kingdom: John Wiley & Sons.

Losen, D., & Skiba, R. (2010). Suspended Education: Urban Middle School in Crisis. California: Southern Poverty Law Centre.

Lu, J. (2002). Predicting Customer Churn in the Telecommunications Industry- An Application of Survival Analysis Modelling Using SAS. SAS User Group International (SUGI27) Online Proceedings, 114-27.

Lukyanenko, R., Parsons, J., & Wiersma, Y. (2014). The impact of conceptual modelling on dataset completeness: A field experiment. Thirty Fifth International Conference on Information Systems, 1, pp. 1-19. Auckland.

Lynam, S., & Cachia, M. (2018). Student's perspective of the role of assessment at higher education. Assessment & Evaluation in Higher Education , 43(2), 223-234.

Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education, 54(2), 588-599.

Macfadyen, L., & Dawson, S. (2012). Numbers are not enough: Why e-learning analytics failed to inform an institutional strategic plan. Journal of Educational Technology & Society, 15(3), 149-163.

Macfadyen, L., Dawson, S., Pardo, A., & Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. Research & Practice in Assessment , 9, 17-28.

Marsonet, M. (2019). Philosophy and logical positivism. Academics International Scientific Journal, 10(19), 32-36.

Mattingly, K., Rice, M., & Berge, Z. (2012). Learning analytics as a tool for closing the assessment loop in higher education . Knowledge Management & E-Learning: An International Journal , 4(3), 236-247.

McCabe, A., & O'Connor, U. (2014). Student-centred learning: the role and responsibility of the lecturer . Teaching in Higher Education , 19(4), 350-359.

McMillan, J., & Reed, D. (1994). At-risk Students and Resiliency: Factors Contributing to Academic Success . Taylor & Francis , 67(3), 137-140.

Melnokovas, A. (2018). Towards an Explicit Research Methodology: Adapting Research Onion Model for Futures Studies. Journal of Futures Studies, 28-44.

Mikalef, P., Pappas, I., Krogstie, J., & Pavlou, P. (2019). Big data and business analytics: A research agenda for realizing business value. Information and Management, 54(1), 1-35.

Ministerial Statement, o. t. (2018-2020). Transforming teaching, learning, researching and leading towards enhanced quality, success and equity in universities. Republic of South Africa: Department of Higher Education and Training: Republic of South of South Africa.

Mitchell, M., & Jolley, J. (2012). Research Design Explained. Cengage Learning .

Miteva, D., Stefanov, K., & Stefanova, E. (2017). e-Analytics for e-Learning. International Journal of Human Capital and Information Technology Professionals (IJHCITP), 8(4), 1-13.

Mohamedbhai, G. (2014). Massification higher education institutions in Africa. International Journal of African Higher Education, 1(1), 59-83.

Moly, M., Roy, O., & Hossain, M. (2019). An Advanced ETL Technique for Error Free Data in Data Warehousing Environment. International Journal of Scientific Research & Engineering Trends , 5(2), 2395-566X.

Montgomery, D., Peck, E., & Vining, G. (2012). Introduction to linear regression analysis (4th ed.). New Jersey : John Wiley & Sons.

Morris, T., & Wood, S. (1991). Testing the survey method: continuity and change in British industrial relations. Work Employment and Society, 5(2), 259-282.

Mugali, J., Chate, S., Patil, N., Tekkalaki, B., Patil, S., Pattanashetty, N., & Pattanashetti, M. (2017). Study of Somatic Complain amongst Children with Poor Academic Performance. The International Journal of Indian Psychology , 4(2), 5-10.

Munene, I., Darby, F., & Doherty, J. (2015). Blended for student engagement and retention: The case of cinema and visual culture and healthy lifestyle studies. In J. Keengwe, & J. Agamba (Eds.), Models for Improving and Optimizing Online and Blended Learning in Higher Education (pp. 129-146). USA: IGI Global .

Mupinga, D. M., Nora, R. T., & Yaw, D. C. (2006). The Learning Styles, Expectations, and Needs of Online Students . Collage of Teaching, 54(1), 185-189.

Murtaugh, P., Burns, L., & Schuster, J. (1999). Predicting the retention of university students. Research in Higher Education, 40(3), 355-371.

Muslim, M., Karim, H., & Abdullah. (2012). Satisfaction of Students' Living Environment between On-Campus and Off-Campus Settings: A Conceptual Overview. Procedia-Social and Behavioral Sciences, 68(1), 601-614.

Muthukrishnan, S., Govindasamy, M., & Mustapha, M. (2017). Systematic mapping review on student's performance analysis using big data predictive model. Journal of Fundamental and Applied Sciences, 9(4), 730-758.

Myers, M. (1999). Investigating information systems with ethnographic research. Communications of the Association for Information Systems, 23.

Naidoo, P., & Kasiram, M. (2014). Exploring unplanned pregnancy amongst university students. Sicial Work/Maatskaplike Werk, 42(3), 341-354.

Nandeshwar, A., Menzies, T., & Nelson, A. (2011). Learning patterns of university student retention . Expert Systems with Applications, 38(12), 14984- 14996.

Narrz'ello, G. (2014). At-risk students . In D. Levinson, P. Cookson, & A. Sadovnik (Eds.), Education and sociology: An encyclopaedia (p. 49). New York: Routledge.

Newman, I. (1998). Qualitative-quantitative research methodology: Exploring the interactive continuum. Carbondale: Southern Illinois University .

Newman, I., & Ridenour, C. (1998). Qualitative-Quantitative Research Methodology: Exploring the Interactive Continuum. Educational Leadership Faculty Publications, 122.

Ng, R., Lam, R., Ng, K., & Lai, I. (2017). Identifying the Needs of Flexible and Technology Enhanced Learning in Vocational and Professional Education and Training's (VPET) Workplaces. In New Ecology for Education- Communication X Learning (pp. 107-117). Singapore: Springer.

Noltemeyer, A., & Ward, R. (2015). Relationship Between School Suspension and Student Outcomes: A Meta-Analysis . School Psychology Review, 44(2), 224-240.

Okioga, C. (2013). The Impact of Students' Socio-economic Background on Academic Performance in Universities, A Case of Students in Kisii University Collage. American International Journal of Social Science , 2(2), 39-46.

Onwegbuzie, A., & Leech, N. (2005). On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. International journal of social research methodology , 375-387.

Panhwar, A., Ansari, S., & Shah, A. (2017). Post-positivism: An effective paradigm for social and educational research. International Research Journal of Arts & Humanities (IRJAH), 45.

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Journal of Educational Technology & Society , 17(4), 49-64.

Pardo, A., & Kloos, C. D. (2011). Stepping out of the box. Towards analytics outside the Learning Management System. Spain: Semantic Scholar .

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology , 45(3), 438-450.

Pascarella, E., & Chapman, D. (1983). A Multi-institutional, Path Analytic Validation of Tinto's Model of College Withdrawal. American Educational Research Journal, 20(1), 87-102.

Pascarella, E., & Terenzini, P. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model . The journal of higher education , 51(1), 60-75.

Peppard, J., & Ward, J. (2016). The strategic management of information systems: Building a digital strategy. John Wiley & Sons.

Peterson, C., & Barrett, L. (1987). Explanatory style and academic performance among university freshman. Journal of personality and social psychology , 53(3), 603.

Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: an assessment . Journal of management information systems , 75-105.

Pokhriyal, N., & Jacques, D. (2017). Combining disparate data sources for improved poverty prediction and mapping. Proceedings of the National Academy of Science , 114(46), 9783-9792.

PoPI Act. (2013). Protection of Personal Information Act. Republic of South Africa: Constitution of the Republic of South Africa.

Prakash, G., & Rangdale, S. (2017). ETL Data conversion: Extraction, transformation, and loading Data conversion . International Journal of Engineering and Computer Science , 6(10), 22545-22550.

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. Belgium: Proceeding of the 3rd international conference on learning analytics and knowledge.

Purnell, K., McCarthy, R., & McLeod, M. (2010). Student Success at University: Using early profiling and interventions to support learning. Studies in Learning, Evaluation, Innovation and Development, 7(3), 77-86.

Quintão, C., Andrade, P., & Almeida, F. (2020). How to Improve the Validity and Reliability of a Case Study Approach. Journal of Interdisciplinary Studies in Education, 264-275.

Ralarala, M., Pineteh, E., & Mchiza, Z. (2016). A case study on the language and social-cultural challenges experienced by international students studying at Cape Peninsula University of technology. South African Journal of Higher Education , 30(4), 231-255.

Rienties, B., Herodotou, C., Olney, T., Schencks, M., & Boroowa, A. (2018). Making sense of learning analytics dashboards: A technology acceptance perspective of 95 teachers. International Review of Research in Open and Distributed Learning, 19(5).

Robson, C. (2002). Real World Research. Blackwell : Oxford.

Robson, C., & McCartan, K. (2016). Real world research. John Wiley & Sons.

Roger, B. (2003). Determinants of Undergraduate Student Drop Out Rates in a University Business Studies Department. Journal of Future and Higher Education, 27(2), 123-144.

Rooney, C. (2015). Using survival analysis to identify the determinants of academic exclusion and graduation in three faculties at UCT. Doctoral dissertation, University of Cape Town .

Rupert, G., & Miller, J. (1998). Survival Analysis (Library ed.). New York: John Wiley & Sons.

Sahay, A. (2016). Peeling Saunder's Research Onion. Research Gate, 1-5.

Samara, T. (2015). ERP and information systems: Integration or disintegration (5th ed.). London: John Wiley & Sons.

Sandberg, J., Holmström, J., & Lyytinen, K. (2019). Digital transformation of ABB through platforms: the emergence of hybrid architecture in process automation. Digitalization Cases: How Organizations Rethink Their Business for the Digital Age, 7(2), 273-291.

Saunders, M., & Tosey, P. (2012/2013). Research Design. Academia .

Saunders, M., Lewis, P., & Thornhill, A. (2009). Research Methods for Business Students. England: Pearson Education .

Saunders, M., Lewis, P., & Thornhill, A. (2019). Research Methods for Business Students. New York: Pearson .

Schutt, R. (2015). Sampling: In Investigating the Social World. Singapore : Sage.

Sclater, N., Peasgood, A., & J, M. (2017). Learning analytics in higher education. London: Jisc, 8, 176.

Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning Analytics in Higher Education: A review of UK and international practice . London: Jisc.

Scott, A., & Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. Biometrics, 30(3), 507-512.

Scott, I., Yeld, N., & Hendry, J. (2007). Higher Education Monitor 6: a case for improving teaching and learning in South African higher education. Pretoria: Council of Higher Education.

Shi, Y., Zhang, L., Ma, Y., & Yi, H. (2015). Dropping Out of Rural China's Secondary Schools: A Mixed-methods Analysis. The China Quarterly , 1048-1069.

Shum, S., & Crick, R. (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. ACM (pp. 92-101). Canada: 2nd international conference on learning analytics and knowledge.

Siemens, G. (2012). Learning analytics: envisioning a research decipline and a domain of practice. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 4-8). ACM.

Siemens, G. (2012). Learning analytics: envisioning a research discipline and a domain of practice. In Proceedings of the 2nd international conference on learning analytics and knowledge , 4-8.

Siemens, G., & Baker, R. S. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. Canada: ACM Digital Library.

Silverman, D. (2016). Qualitative research. Sage .

Slade, S., & Prinsloo, P. (2013). Learning Analytics: ethical issues and dilemmas . American Behavioural Scientist, 57(10), 1509-1523.

Smith, T., & Smith, B. (2001). Survival analysis and the application of Cox's proportional hazards modelling using SAS. In Proceedings of the 26th Annual SAS Users Group International Conference, NC: SAS Institute Inc, 244-246.

Smith, V., Lange, A., & Huston, D. (2012). Predictive modelling to forecast student outcomes and drive effective interventions in online community college courses. Journal of Asynchronous Learning Networks, 16(3), 51-61.

Som, R. (1995). Practical Sampling Techniques. CRC Press.

Song, Y., & Kong, S. (2017). Investigating students' acceptance of a statistics learning platform using technology acceptance model. Journal of Educational Computing Research, 55(6), 865-897.

Soomro, Z., Shah, M., & Ahmed, J. (2016). Information security management needs more holistic approach: A literature review. International Journal of Information Management, 36(2), 215-225.

South African Department of Higher Education and Training. (2015). Statistics on Post-School Education and Training in South Africa: 2013. Pretoria: Department of Higher Education and Training.

Spady, W. (1970). Dropout from Higher Education: An Interdisciplinary Review and Synthesis. Interchange, 1(1), 64-85.

Stengel, E. (1964). Suicide and attempted suicide . Penguin Books.

Stodder, D., & Matters, W. (2016). Improving data preparation for business analytics. Transforming Data With Intelligence, 1(1), 41.

Suh, S., Suh, J., & Houston, I. (2007). Predictors of Categorical At-Risk High School Dropout. Journal of Counselling & Development, 85(2), 196-203.

Tan, F., & Hunter, M. (2003). Using Narrative Inquiry in a Study of Information Systems Professionals. 36th Annual Hawaii International Conference on System Science (p. 7). Hawaii: IEEE.

Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A typology with Examples. Journal of Mixed Methods, 1(1), 77-100.

Terenzini, P., & Pascarella, E. (1980). Toward the validation of Tinto's model of college student attrition: A review of recent studies. Research in Higher Education, 271-282.

Thakar, P. (2015). Performance analysis and prediction in educational data mining: A research travelogue . International Journal of Computer Application , 110(15), 60-68.

Thompson, B. (1995). Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guideline Editorial . Educational and Psychological Measurement, 5(4), 525-534.

Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. American Educational Research Association, 45(1), 89-125.

Tinto, V. (2006). Research and practice of student retention: what next? Journal of College Student Retention: Research, 1-19.

Tinto, V. (2017). Reflections on student persistence. Student Success, 8(2), 1-8.

Tinto, V. (2017). Through the Eyes of Students. The Journal of College Student Retention: Research, Theory & Practice (CSR), 19(3), 254-269.

Toetenel, L., & Rienties, B. (2016). Analysing 157 learning design using learning analytic approaches as means to evaluate the impact of pedagogical decision-making . British Journal of Educational Technology , 5(47), 981-992.

Toldson, I., McGee, T., & Lemmons, B. (2015). Reducing suspensions by improving academic engagement amongst school-age Black males. In D. Losen (Ed.), Closing the school discipline gap: Equitable remedies for excessive exclusion (pp. 107-117). New York: Teachers Collage Press.

Tsai, Y., Kovanovic, V., & Gasevic, D. (2021). Connecting the dots: An exploratory study on learning analytics adoption factors, experience, and priorities. The Internet and Higher Education, 50, 100794.

Tsai, Y., Moreno-Marcos, P., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gasevic, D. (2018). The SHEILA Framework: Informing Institutional Strategies and Policy Processes of Learning Analytics. Journal of Learning Analytics , 5(3), 5-20.

Unesco. (1994). The Salamanca Statement and Framework for action on special needs education: adopted by the World Conference on Special Needs Education. Spain : UNESCO.

Unwin, T., Kleessen, B., Hollow, D., William, J., & Oloo, L. (2010). Digital learning management systems in Africa. Open Learning: The Journal of Open, Distance and e-Learning , 25(1), 5-23.

van der Bank, C. (2012). The right to privacy- South African and Comparative Perspectives. European Journal of Business and Social Science, 1(6), 77-86.

van Os, H., Kanning, J., Wermer, M., Chavannes, N., Numans, M., Ruigrok, Y., & Groenwold, R. (2022). Developing Clinical Prediction Models Using Primary Care Electronic Health Record Data: The Impact of Data Preparation Choices on Model Performance. Frontiers in Epidemiology, 10(4), 871630.

Vanslambrouck, S., Zhu, C., Lombaerts, K., Philipsen, B., & Tondeur, J. (2018). Students' motivation and subjective task value of participation in online and blended learning environments. The Internet and Higher Education, 36, 33-40.

Venter, P., van Rensburg, J., & Davis, A. (2012). Drivers of learning management system use in a South African open distance learning institution. Australasian Journal of Educational Technology , 28(2), 183-198.

Vercellis, C. (2011). Business intelligence: data mining and optimization for decision making. United Kingdom: John Wiley & Sons.

Viberg, O., Hatakka, M., Balter, O., & Mavroudi, A. (2018). The Current Landscape of Learning Analytics in Higher Education. Computers in Human Behaviour, 98-110.

Wadesango, N., & Machingambi, S. (2011). Causes and Structural Effects of Student Absenteeism: A Case of Three South African Universities. Journal of Social Sciences, 2(26), 89-97.

Walsham, G. (1995). The Emergence of Interpretivism in IS Research. Information Systems Research , 376-394.

Walter, E. (1998). Intrusive advising of refreshment in academic difficulty. NACADA Journal, 8(2), 27-33.

Wang, F., & Hannafin, M. (2005). Design-based research and technology-enhanced learning . Educational Technology Resources and Development , 53(4), 5-23.

Welsh, M., Parke, R., Wildaman, K., & O'Neil, R. (2001). Linkages Between Children's Social and Academic Competence: A Longitudinal Analysis. Journal of School Psychology , 463-482.

Whyte, J., & Davies, A. (2021). Reframing Systems Integration: A Process Perspective on Projects. Project Management Journal, 52(3), 237-249.

Wiesche, M., Jurisch, M., Yetton, P., & Krcmar, H. (2017). Grounded theory methodology in information system research . MIS quarterly , 685-701.

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices . eJournal of Emergency Primary Health Care, 8(3), 1-13.

Williams, T., Morton, J., Kilgour, P., & Northcote, M. (2018). How engaged are our students? Using Analytics to Identify Students At-risk. (pp. 122- 128). Amsterdam, Netherlands : Association for the Advancement of Computing in Education (AACE).

Williamson, B. (2016). Digital education governance: data visualization, predictive analytics, and ‘real-time’ policy instruments. Journal of Education Policy , 31(2), 123-141.

Williamson, K., & Johanson, G. (2017). Research Methods: Information, systems, and contexts . Chandos Publishing.

Wilson, A., Watson, C., Thompson, T., Drew, T., & Doyle, S. (2017). Learning analytics: Challenges and limitations. Teaching in Higher Education , 22(8), 991-1007.

Wise, A., & Vytasek, J. (2017). Learning analytics implementation design. Handbook of learning analytics, 1, 151-160.

Wortmann, F., & Fluchter, K. (2015). Internet of Things: Technology and Value Added. Bus Inf Syst Eng, 57(3), 221-224.

Wynn, D., & Williams, C. (2012). Principles for Conducting Critical Realist Case Study Research in Information Systems. MIS/OM/DS Faculty Publications, 787-810.

Xu, L., Jiang, C., Wang, J., Yuan, J., & Ren, Y. (2014). Information security in big data: privacy and data mining. Ieee Access, 2(1), 1149-1176.

Yin, J. (2018). Beyond postmodernism: A non-western perspective on identity. Journal of Multicultural Discourses, 193-219.

Yin, R. (1981). Applications of case study research (Vol. 3). Sage.

Yin, R. (1981). The case study as a serious research strategy. Knowledge, 3(1), 97-114.

Yin, R. (1981). The Case Study Crisis: Some Answers. Admin Science Quarterly , 58-65.

Yoloye, E. (2015). New technologies for teaching and learning: Challenges for higher education learning institutions in developing countries. In C. Nwokeafor (Ed.), Information Communication Technology (ICT) Integration to Educational Curricula: A New Direction for Africa (Vol. 3, pp. 250- 260). United States of America: University Press of America.

Yong, A., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. Tutorials in quantitative methods for psychology , 9(2), 79-94.

Yu, Z., Haghighat, F., Fung, B., & Yoshino, H. (2010). A decision tree method for building energy demand modelling. Energy and Buildings , 42(10), 1637-1646.

Zielezinski, M., & Darling-Hammond, L. (2016). Promising practices: A literature review of technology use by underserved students. California: Stanford Centre for Opportunity Policy in Education .

Zipkin, E., Zylstra, E., Wright, A., Saunders, S., Finley, A. D., & Tingley, M. (2021). Addressing data integration challenges to link ecological processes across scales. Frontiers in Ecology and the Environment , 30-38.

APPENDICES

Appendix A: Analytics for Learn (A4L) Project Kick-off and Definition

1. Introduction

The Project Definition document will provide a plan for the CPUT Analytics for Learn Project. This document is the result of a full day workshop of discussion that is typically spread out over multiple remote sessions to allow for complete participation by the appropriate client staff. The document defines the goals, objectives and scope of the project to implement Blackboard Analytics for Learn Data Warehouse and Business Intelligence products. Additionally, the Project Definition will serve as an agreement between the following parties: Project Sponsor, Project Manager, Project Team, and other personnel associated with and/or affected by the project.

Version 2017-04-25

The Project Plan defines the following:

- Project purpose
- Business and project goals and objectives
- Scope of Deployment
- Roles and responsibilities
- Project Workplan

2. Project Approach

The project implementation approach will follow a waterfall methodology, although some phases overlap:

Phase I: Project Planning (Kickoff)

Phase II: Project Planning (Functional Project Definition)

Project Planning (Technical Preparation)

Phase III:	Installation and Configuration
Phase IV:	Orientation
Phase V:	Training
Phase VI:	Deployment
Phase VII:	Close

3. Goals and Objectives

3.1. Business Goals and Objectives

The following high-level goals were identified. Note that it may not be possible to cover all objectives in the implementation of Analytics for Learn, but the deployment model can be reused for later phases of internal deployment (i.e. post-implementation).

1. Adoption of Blackboard Learn
2. Retention and throughput of at-risk students

3.2. Project Goals and Objectives

Goals and objectives of the project implementation are defined as:

Ensure that end users have input into the design process.

- Accomplish project business goals and objectives within defined budget and time parameters.
- Minimise impact to standard business operations within the affected units.
- Identify Pilot group
- Deploy key reports
- Investigate and identify scope for emerging use cases such as major-level reporting
- Success criteria:
    - All baseline reports are available.
    - Defined custom reports are available.
    - Selected report customisations (within budgeted hours) are delivered.
    - Client has sufficient knowledge of the software and data model to enable future client-generated reports.
    - Client has sufficient knowledge of the data warehouse to ingest data into the existing institutional data warehouse for use alongside other data.

4. Scope

4.1. Scope Definition

The project deliverables shall include:

- Dimensional Data Models
    - A functioning relational data warehouse and nightly ETL process
- OLAP Analytics Data Models
    - A functioning OLAP database and nightly ETL process
- Library of Operational (SSRS - 25 Reports) and Analytic (Pyramid - 62 Reports) Reports
- Dashboard Templates (Pyramid - 4 Dashboards)
- Training and Knowledge Transfer for Functional and Technical staff on the maintenance and use of the application and reporting tools.

4.2. Items Beyond Scope

- Development of intervention frameworks.
- Training on development of SSRS reporting and related Microsoft SQL server products.
- Training of staff outside of the client implementation team, including:
  - End-user documentation and translation.
  - Training of students, instructors, department chairs, deans etc., on the use of reports.

4.3. Configuration

The following configuration changes were decided during this phase, and will be used during installation to tailor the solution to meet your needs:

- Term hierarchy – CPUT has a mixture of term types for different educational areas. The majority follow a semester-based structure but there are some courses granulated to the academic year and some following a seasonal structure (e.g. Winter, Summer courses).
- Grading Schema
  - Percentage grades are used in almost all cases..
- Institutional hierarchy
  - CPUT follows the expected hierarchy of Institution > College (Faculty/School) > Department > Subject (Programme) > Course Number > Course Section.

5. Deployment Planning

During the project definition phase we discussed the various considerations of deployment. Based on these discussions, we recommend the following client-side actions:

- The development of an intervention framework defining:
  - Responsibility for the intervention (who runs the report)
  - Timescale (when to run the report)
  - Action (what to do with the information)
  - Support (what the central support unit should do to enable this intervention, i.e. automated alerts, documentation, reminders)
- Develop Communication and Training Plan to support those interventions.

6. Stakeholders & Use Cases

6.1. Stakeholders

During the project definition workshops, the following stakeholders were identified for the implementation phase of deployment.

Stakeholder Group	Priority
Faculty T&L Coordinators	1
Retention Officers	1

IT Coordinators	2
Executive Management	3
Deans	3
HODs	3
Lecturers	3
Students	3
Quality Management Department	4
Faculty T&L Coordinators	1
Retention Officers	1

6.2. Use Cases

The following use cases were identified and are detailed here for clarity. In some cases, similar use cases have been merged to illustrate how they map across stakeholders. In addition, we have identified baseline content which aligns with each use case. Please see the accompanying **Report List** spreadsheet for details and screenshots of reporting content.

No.	Area	Use case(s)	Sample Content
1	Student Retention	Identification of at-risk students	Student at-a-glance (I, R), Course at-a-glance (I, R), Exception reports (R), Course Access Exception Report (P)
2	Promote Learn Adoption	Set targets for adoption	Sections Using Learn Courses (P), Course Activity Trend by Instruction Method (P)
2	Student Retention	Identify at-risk subjects	Top 50 Courses: Avg Interactions (P), Custom reports (P)
3	Student Retention	Quantify Risk	Student Performance Correlation Reports (P)
4	Promote Learn Adoption	Understand current levels of adoption of Bb Learn	Learn Course Use by College (R), Sections Using Learn Courses (P)
4	Student Success	Understand relationship between activity & success	Student Performance Correlation Reports (P)
4	Student Success	Understand relationship between other variables and success	Student Performance Correlation Reports (P), Custom Reports (P)

5	Promote Learn Adoption	Benchmarking for best practice	Learn Course Use by College (R), Sections Using Learn Courses (P)
5	Promote Learn Adoption	Provide evidence for teaching excellence	Top X Courses/Instructors (P)
5	Promote Learn Adoption	Quality assurance during accreditation	
5	Promote Learn Adoption	Promote use of assessment tools in Learn	Grade Center Effect on Activity (P)
6	Administration	SIS Mismatch	Custom Reports (P)

Sample Content Key: P – Pyramid, R – Reporting Services, I – Integrated

6.1. Use Case Matrix

Use cases can often be mapped to multiple stakeholders, often with different granularity. The following matrix (overleaf) maps each use case to each stakeholder. These have been confirmed during the Report Writing on-site.

	Use Case	Identify and support at-risk students	Identify & mitigating at-risk subjects	Quantify risk	Investigate relationship between activity & success	Investigate relationship between other variables and success	Increase adoption of Bb Learn	Provide evidence for teaching excellence	Benchmarking for best practice	Increase use of assessment tools in Learn	Quality assurance during accreditation	Students with No Activity	SIS Mismatch
Stakeholder	**Priority**	**1**	**2**	**3**	**4**	**4**	**4**	**5**	**5**	**5**	**5**	**6**	**6**
CIET	**1**			3	4	4	4	5	5	5			6
Faculty T&L Coordinators	**1**		2					5					
Retention Officers	**1**	1	2									6	
IT Coordinators	**2**		4				8		10				
Executive Management	**3**		6				12		15	15	15		
Deans	**3**		6				12		15	15	15		
HODs	**3**	3	6				12		15	15	15		
Lecturers	**3**	3	6					15			15	18	18
Students	**3**	3											
Quality Management Department	**4**		8					20	20		20		

Appendix B: Plans to Customise the Currently Installed A4L Data Warehouse

Cape Peninsula University of Technology

Data Project Leveraging USDG Funds

Service Overview

Summary

Cape Peninsula University of Technology (CPUT) have received UCDG funds to work on a project to develop capacity in supporting at-risk students. The first block of funds must be ringfenced before end November 2019.

CPUT has expressed particular interest in exploring available data across the institution in the context of driving student retention and progression. To this end, CPUT has liaised with Blackboard (due to current partnership) to recommend the best use of this funding. Following consultation, we recommend the following two-phased approach:

- Workshop: Learning Analytics Data Strategy
  - This service is a gap analysis that also looks at readiness and deployment planning to better understand a future state.

Depending on the outcome of this workshop, one of the following options could be the follow-up phase:

- Expanding and/or customizing the currently installed A4L warehouse. This may involve adding data from additional source systems, expanding data from the student information system or taking advantage of new features in the Pyramid BI software to add value.
  OR
- Continue working with A4L as is but allow for additional functional supporting services (like writing new reports, additional training etc.)

The idea is to give CPUT maximum flexibility on where to apply their funding that would best suit their data needs. Hence, this second phase would be delivered as a timebox of hours.

Blackboard

Below please find a draft statement of work to further explain how Blackboard consulting could work with CPUT. *Please note that this is not a formal contract but rather a proposal.*

1 Scope of Services

1.1 Learning Analytics Data Strategy

1.1.1 Scope

This service that will identify requirements around Learning Analytics with a view to provide recommendations around data, reports and implementation to achieve the desired future state. Key considerations explored in the engagement include:

- What is the institution's short-term learning data needs and longer-term data aspirations to provide key insights regarding the student experience?
- What is the current gap between available data and desired data and information to identify at-risk students?
- How does institution convert learning data into actionable knowledge?

Data Analytics readiness assessment

- Blackboard Consultant will review existing documentation around Learning Analytics at the institution and usage, including university strategies and goals, operational plans, existing learning analytics use cases, usage and systems maintenance of the current warehouse information. This review will inform the onsite activities and recommendations report.

Onsite workshop (2 Days):

- Blackboard Consultant will conduct an opening forum with key stakeholders to confirm workshop parameters, discuss preliminary findings, etc., conduct focus group sessions and stakeholder interviews, Facilitate a preliminary qualitative analysis of information gathered, Facilitate a closing planning session with key stakeholders to define and prioritise a set of action steps to identify the At Risk Students.

Deployment planning

- Following the onsite, the Blackboard consultants will analyze all the data and information collected throughout the engagement. The findings will be reviewed, interpreted and documented. Following which the Blackboard consultants will provide a recommendation report as to required use cases, reports and data gaps related to at risk student population. The report will include Definition of the Learning Analytics requirements which will prioritise the recommended initiatives based on complexity of the initiative, the estimated effort of the initiative, the implementation timeframe of the initiative, and the strategic importance of the initiative. This report will represent the final deliverable of the engagement

1.1.2 Deliverables

- Requirements and Recommendations Report

1.1.3 Customer Assumptions/Requirements

- Work to be carried out on-site and remote
- The customer is responsible for staffing resources on the project that have the necessary functional and technical knowledge to execute required tasks, please refer to section (4) for a

list that identifies possible client-side participants for this engagement.

1.2 Report customization & extended support timebox

1.2.1 Scope

Blackboard Consulting will provide technical and functional support to enable the client to apply additional scope within a limit of 60 consulting hours, which can include the following (after scoping and estimation is performed on the set of requests):

- Customizations to the A4L warehouse
- Additional functional supporting services, for A4L

To provide flexible access to the consulting support required, this will be provided in the form of a 80 hours Timebox. This service is designed to complement the client's own team.

At initiation, the Blackboard Project Manager will work with the client Project Lead to build out a workplan identifying work packages and agreed deliverables for the timebox effort. The "Timebox" will run no longer than 12 months from the contract date.

1.2.1.1 Artifacts/ Deliverables

- Project Workbook
- Recordings of Remote sessions

1.2.1.2 Customer Assumptions/Requirements

- The consultant's time must be scheduled in advance
- All work to be performed remotely
- In addressing questions, some may require researched; and will thus require a turnaround time.
- Hours spent researching questions will be included in the timebox available.
- Client and Blackboard will report on and manage resource consumption as appropriate, through the project management structure
- Any days not consumed by 12 months from the contract date will be assumed to be no longer required and are non-refundable

2 Resource Requirements

Blackboard proposes the following projected staffing model to deliver this project.

Role	Activities and Responsibilities
Management Oversight	Responsible for general oversight, serves as client escalation point, additional subject matter expert coordination
Project Manager	Responsible for management of project tasks, schedule, and resources
Functional Consultant	Provides requirements gathering, end user training, and setup of reports
Technical Consultant	Provides technical implementation and configuration of baseline Blackboard Analytics for Learn product

3 Customer Responsibilities

Blackboard Consulting's approach assumes active participation from the customer team. The customer is responsible for staffing resources on the project that have the necessary functional and technical knowledge to execute required tasks. The list below identifies possible client-side participants for this engagement.

Customer Role	Involvement
Project Owner/Executive Sponsor	The Project Owner provides strategic direction and executive sponsorship of the engagement.
Project Manager/Primary Contact	Responsible for management of customer project tasks, schedules, and resources
Academic Computing/Education Technology Director	Person responsible for the education technology infrastructure of the institution.
Blackboard Solution Administrator(s)	Individual(s) responsible for the configuration and administration of the component systems that comprise the Blackboard solution.

4 Fees, Expenses, and Terms

4.1 Firm-Fixed Price Services

The costs for additional services to be provided on a Firm-Fixed Price basis are detailed below:

Service Name	Product Code	Term of Service	Fees
Analytics Data Strategy *Once-off fee	AN-LANA-FFP	Upon delivery	$20,000
Cross-functional Services Timebox *Once-off fee	AS-ICS-HRS-FFP	Upon Delivery	$14,550
ICM (integration customization maintenance) *Annually recurring fee.	AS-ICMCUSDV	Upon Activation	$3,450
		***Total**	$38,000
		ZAR ($1=R14.85)	R564,300
		VAT 15%	R84,645
		FINAL TOTAL	R648,945

***Please note:** CPUT has historically paid Blackboard directly and in USD. The ZAR value indicated here is for information only.

4.2 Travel Expenses related to Consulting and Training Services

Travel costs are included in the Service Pricing Section. Blackboard Consulting will make reasonable efforts to manage travel costs without compromising project objectives.

5 Project Timeline

The project plan will be drafted, agreed to, and tracked with the Customer during or after the planning phase. Timing and dependencies are identified as outputs from planning sessions and a formal plan will be drafted and tracked in partnership with the Customer's project lead.

Appendix C: Implementation Planning Phase

Implementation Planning Decision Template

Please use the tables in this document to record decisions made within or following Project Definition Implementation Planning workshop. These do not need to be completed in full immediately but please try to record decisions as soon as they are made so that none are forgotten. Please send this to your Functional Consultant whenever a change is made.

Institutional Goals

Please define and prioritise high-level institutional goals which Analytics for Learn is expected to support.

Goal	Priority
Retention and throughput of students	1
Adoption of BB Learn	2

Stakeholders

Use the table below to define and prioritise stakeholder groups at your institution

Stakeholder Group	Description / Examples	Priority
Faculty T&L Coordinators		1
Retention Officers		1
IT Coordinators		2
Executive Management		3
Deans		3
HODs		3
Lecturers		3
Students		3
Quality Management Department		4

Use Cases

Use Case Title	Description / Details	Aligned Stakeholder(s)	Priority
Identify at-risk students			1
Identify at-risk subjects			2
Quantify risk			3
Understand relationship between activity & success			4
Understand relationship between other variables and success			4

Blackboard

Understand adoption of Bb Learn			4
Provide evidence for teaching excellence			5
Benchmarking for best practice			5
Promote use of assessment tools in Learn			5
Quality assurance during accreditation			5
Students with No Activity			
SIS Mismatch			
Reporting on library usage	For consideration post-deployment		-

Analytics for Learn – Action Framework

The following table is designed to help you map stakeholders to use cases and turn insight from data into action. Note that each stakeholder is likely to have several use cases, and a single use case may require action by several stakeholders. This will in turn result in several actions using different report content.

Stakeholder	Use Case	Content (Report, Dashboard, Alert, Publication) – Complete after Orientation	Timing	Trigger	Action	Support
Instructor	*Identifying students who have low activity*	*Course at-a-glance*	*End of Week 1 of teaching*	*Students who have not accessed the course*	*Email students with low interactions*	*L&T to remind instructor; Screencast of integrated reports training.*
Student Advisor	*Monitoring students' progress*	*Advisor at-a-glance*	*Prior to meeting with student*		*Discuss progress with student*	*Student Support office to email advisors*
Student Advisor	*Review Report to identify students who have not logged on*	*Login Exception Report*	*End of Week 1 of teaching*		*Email student with no/low site accesses*	*L&T to send reminder email with draft text*
Department Administrator	*Review report to identify student who are not submitting*	*Submission Exception Report*	*Following first assessment submission date or Discussion board requirements*		*Refer students who have low/no submissions to Student Advisor*	*L&T to provide access and training.*
Dean	*Understanding current practice within Learn*	*Dean Dashboard*	*Termly/Annually*		*Develop targets for Learn use and*	*L&T to develop appropriate reports*

					Staff Development priorities	
Retention Officer	Identifying students with low or no activity	Login Exception Report	End of week 2 of teaching	Students who have not accessed one or more courses in the last 2 weeks.	Contact Instructor to advise action	Provide access to report. Provide training PDF.
IT Coordinators	Identify courses with large files	Top 50 courses – file size	Monthly	Courses with an average file size > 100mb	Guide instructor on reducing file size.	Supporting documentation

Communication Plan

Effective communication is the key to success for any institutional change project. For Analytics for Learn, each stakeholder group should be considered for both the type of information required and the method of delivery. Note that communication will likely need to be widespread and take into consideration roles which are not included in the action framework. The following communication types will be required to rollout to the various stakeholders

- Strategic direction Communication – Interested in how this project will contribute to the University's plan
- Technical – Interested in the technical changes, how they occur, how they fit into existing systems and architecture
- Training – Interested in how to use the product, interpret data and drive actions
- Adoption – Requiring information on what is available and when to use the system

Interested Party	Information Required	Information Provider	Frequency of Communication	Method of Communication
Instructors	*Training; Adoption*	*eLearning Team*	*At key points of semesters*	*Workshops & Seminars; Email Updates*
Students	*Training; Adoption*	*Instructors*	*At induction to University/Blackboard*	*Email announcements*

Training Plan

The primary goal of training will be to inform relevant users on how to operationally use the Blackboard Analytics platform as well as drive learning and teaching practices based on the data users see. The training intends to:

- Teach instructors and facilitators (of various technical capabilities) how to access and use the integrated reports
- Make instructors and facilitators aware of the various data sources and how to interpret reports
- Make instructors and facilitators aware of intervention strategies associate with Analytics results
- Make students aware of the integrated reports and how to interpret them
- Teach Pyramid users how to access and use the application and gain value from the reports.

There are three key workshops that should be run to support the various users:

Topic	Type	Primary Stakeholders	Offered by
Blackboard Analytics Training: Integrated Reports	*Workshop (Online/Face-to-face)*	*Instructors*	*L&T*
Intro to Student Analytics	*Documentation / Video*	*Students*	*L&T*
Introduction to Pyramid Reports	*Face-to-Face Workshop*	*Pyramid Users*	*L&T*

Blackboard

Appendix D: Report Writing Training and Development Schedule

Analytics for Pyramid & Deployment Agenda

Date of Service: 19, 20 and 21 April 2017, CIET, room 2.17, Cape Town

Objective:

The goal of pyramid and deployment session is twofold. The session will provide detailed training on the use of the pyramid application so that functional staff have the ability to build and deploy their own reports. The first two days are highly interactive with practice and report building activities tailored to the institution. The final part of this session is designed to work through deployment planning initiatives

Recommended Audience:

Project Manager (PM), Functional Staff (Instructional Technology/Curriculum Design Representative), Bb Learn Support Staff.

Other Stakeholders (Optional): Academic Representatives, Advisors, Student Support, Executive

Facilities and Access:

- Meeting room with projector
- Computer Training Room/Class Room/Lab with projector (computers will require Silverlight plugin)
- Access to network
- **MS Silverlight plugin** inside **Internet Explorer** will need to be installed on computers

USER ACCOUNTS MUST BE CREATED PRIOR TO THE SESSION

1. **CIET functional/support staff for Learn and Analytics (LMS & ETU)**
2. **6 x IT coordinators**
3. **QMD (Luclaire Airey)**
4. **MIS (David Bleazard)**
5. **Registrar's office (Happy Mantshi/Kuselwa Marala)**
6. **A4L technical support- CIET technical (Cecilia), BAS (Ruqeyah Waja)**
7. **PM – project Management team- Izak, Sakkie and Cecilia**

Blackboard

Day 1 An introduction to the pyramid application.		
Time	**Agenda**	**Participants**
9am – 10am	**Setup** • Confirm network and user access	PM Only
10am – 12pm	**Pyramid Introduction** • Overview of Pyramid BioXL and Basic Navigation • Organizing Personal Content	PM, Functional Staff, Bb Learn Staff, Stakeholders
1pm – 4pm	**Pyramid Continued** • Using and Modifying Delivered Reports • Basic Report Formatting • Changing Data Layout • Building a New Report	PM, Functional Staff, Bb Learn Staff, Stakeholders

Blackboard

Day 2 Day two will continue with the pyramid tool with a focus on advance reporting options and dashboard development finished with a review of the admin and management side.		
Time	**Agenda**	**Participants**
9am – 12pm	**Advanced Report Options** • Review of Report Building • Creating new Measures • Creating new Members and Sets • Parameters • Creating Cascading filters • Changing Data Sources	PM, Functional Staff, Bb Learn Staff, Stakeholders
1pm – 2pm	**Report building Activities** • Tailored activities to build client reports	PM, Functional Staff, Bb Learn Staff, Stakeholders
2pm – 3pm	**Pyramid Dashboards** • Overview of Pyramid bioPoint Dashboards • Dashboard Demonstration • Basic Dashboard Design o Build desired report objects o Create dashboard pages o Creating Slicer Interactions o Create Global Slicers o Dashboard Building Activities	PM, Functional Staff, Bb Learn Staff, Stakeholders
2pm – 4pm	**Pyramid Admin & Security** • Pyramid Technical Administration • Security (Users, Roles, Feature Access Profiles), • Color Themes and Branding	PM, A4L Administrators

Blackboard

Day 3

This session aims to define key stakeholders and the adoption strategy around Analytics for Learn. We will discuss reporting requirements and the interventions and strategies arising from the reports. This is not a technical session; it will focus on how end users might use the reports and what support will be available to them. The outcome of this session will lead to the development of a Training, Communication and Adoption plan which should guide the institution on how to deploy Analytics for Learn.

Time	Agenda	Participants
9am – 10am	Detailed discussion on Use Cases - specific ways & reports that Blackboard Analytics will be used and identification of key stakeholders.	PM, Functional Staff
10am – 11am	Discussion on the client's intervention and analytics usage strategy. Based on the client's internal policies and procedures, define what 'actions' are needed when to drive usage and take action with the data. The adoption plan is focused on the practicalities of using Analytics and support requirements.	PM, Functional Staff
11am - 12pm	Discussion on how Analytics will be rolled out. Defining required training and communications in the lead up to deployment and in an ongoing capacity throughout teaching periods to support end users.	PM, Functional Staff
1pm – 3pm	Preparing for A4L (Administration Activities) • Monitoring and Troubleshooting • Support • Setting up User Access	PM, A4L Administrators
3pm – 4pm	Debrief and Next Steps	PM Only