UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Evaluating technology-based interventions to enhance the learning of first year Statistics students

Fransonet Reyneke

Submitted in fulfilment of the requirements for the degree

PhD in Science and Mathematical Education

In the Faculty of Natural & Agricultural Science

University of Pretoria

Pretoria

Supervisors: Dr L Fletcher and Prof AF Harding

@2021

ABSTRACT

This retrospective study focuses on the unique contribution of the various technology-based interventions to improve the learning experience and performance of first-year statistics students. Eight consecutive cohorts from 2011 to 2018 were divided into five intervention cohorts for comparison of performance measured by the examination mark, of which four included the initial introduction of a technology intervention and one formed the base year. The other three cohorts were non-initial technology intervention years.

Departing from a traditional teaching model in 2011 the technology-based intervention strategies implemented were firstly the Aplia™ interactive online homework system in 2012 and then, in addition, the flipped classroom in 2013. The flipped classroom as pedagogical model, grounded in constructivism, substituted the traditional teaching model. The impact of the flipped classroom versus the traditional model was measured and the role of the online homework system in both teaching models. QT-clickers were subsequently introduced in 2015 to enable active and cooperative learning for face-to-face engagement inside the classroom.
The pedagogical influence of the QT-clicker and the effect of partial grade crediting were investigated. Due to students' different personality types and learning styles it was decided in 2018 to finally implement and evaluate peer learning activities using South African content and lecturer-switching within the flipped classroom.

To make the consecutive cohorts comparable, it was decided to divide the students in each cohort into two samples; the prerequisite sample consisting of only new students with at least 60% for Grade 12 mathematics and an AP Score of at least 26, while the remaining students formed the non-prerequisite sample. The non-prerequisite sample was further divided into new and repeat students.

The results from the various graphs, ANOVA and chi-square tests confirmed that the differentiator between the technology-based interventions is pedagogy. The flipped classroom model cohorts outperformed the traditional model cohorts. The uniqueness of this study is that it illustrates the benefits of the online homework system, QT-clickers, and peer learning activities through the lens of a flipped classroom. The positioning of the online homework system as a self-paced preparation tool concentrates on a student-centred environment, with self-directed learning and self-knowledge, helped to control cognitive overload. This study

further demonstrates the pedagogical influence of the QT-clicker and the effect of partial grade crediting in a large, flipped classroom setting. Most of the new students in both prerequisite and non-prerequisite samples benefited from the flipped classroom and QT-clicker usage, with the stronger students' final marks being moved into higher percentage brackets. Surprisingly, most of the repeat students also gained from the flipped classroom and peer learning activities with localised content and lecturer-switching, which indicate their need for alternative learning activities and styles.

Several general linear models (GLM) and a machine learning algorithm, XGBOOST (XGB), were fitted to the data to investigate how the technology-based interventions are related to the students' examination performance. The outcome of the GLM models and XGB indicates that the association of higher examination marks hold for the flipped classroom cohorts, i.e., 2013, 2015 and 2018. A qualitative component of the study reports on the student voice testifying to positive experience.

From this retrospective research study the following papers originated:

Fransonet Reyneke, Lizelle Fletcher, & Ansie Harding. (2018). The effect of technology-based interventions on the performance of first year university statistics students. *African Journal of Research in Mathematics, Science and Technology Education*, *22*(2) 231–242. https://doi.org/10.1080/18117295.2018.1477557

Fransonet Reyneke, Lizelle Fletcher, & Ansie Harding. (2021). Enhancing a flipped statistics first year course by using QT-Clickers, *Journal of Statistics and Data Science Education, 29*(1), 71-83. https://doi.org/10.1080/10691898.2021.1895694

Keywords: QT-clicker; flipped classroom; formative assessment; summative assessment; self-paced polling; partial grading; online homework;constructivism; peer learning

ACKNOWLEDGEMENTS

I would like to thank the following people for their assistance during the completion of this dissertation. This study would not have been possible without them:

- My Creator, without whom none of this would have been possible.
- Dr Lizelle Fletcher and Prof Ansie Harding, whose insights and support were invaluable. Whenever I wanted to give up, they inspired me to continue. This thesis matured greatly under their guidance. Thank you for the time you've spent with me and always motivating me to improve my research to the next level.
- My husband, Hennie, who was my pillar of strength throughout the whole process. Thank you for always supporting me with love.
- My family who are so dear to me for their continued support throughout and for always believing in me.
- My friends, who inspired me to rise every time you fall and to live a vision that supersedes every challenge that comes your way.

ACRONYMS AND ABBREVIATIONS

ANOVA	ANalysis Of Variance
AP Score	Advancement Placement Score
BIRAP	Bureau for Institutional Research and Planning
BYOD	Bring Your Own Device
CAF	Clicker Assessment and Feedback
FLN	Flipped Learning Network
GAISE	Guidelines for Assessment and Instruction in Statistics Education
GLM	General Linear Model
GPA	Grade Point Average
HG	Higher Grade
HIM	High Impact Module
JITT	Just In Time Teaching
LBL	Learn Before Lecture
MCQ	Multiple Choice Question
MOOC	Massive Open Online Course
OBE	Outcome-Based Education
RMSE	Root Mean Square Error
RSS	Residual Sum of Squares
SG	Standard Grade
SHAP	SHaply Additive exPlanations
SPSS	Statistical Package for Social Sciences
STK110	Statistics110
STK120	Statistics120
TSS	Total Sum of Squares
XGBOOST (XGB)	eXtreme Gradient Boosting

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
OVERVIEW OF THE STUDY

1.1 INTRODUCTION

This thesis documents a retrospective, cohort study which evaluated the technology-based interventions implemented. Interventions included online homework, flipped learning embedded with online homework and the use of QT-clickers, underpinned by constructivism, to alter the teaching and learning model and ultimately to increase the pass rates of first-year statistics students.

1.2 BACKGROUND

1.2.1 HISTORICAL PERSPECTIVES UP TO 2011

First-year statistics as presented by the Department of Statistics at the University of Pretoria (UP) consists of two modules, namely Statistics 110 (STK110), which is offered in the first semester, and Statistics 120 (STK120), which runs during the second semester. This thesis only focuses on STK110, the first semester module. Only a brief description has been added since a more comprehensive overview of the fail/repeat problem in coherence with section 1.2.1 can be found in Annexures 2 and 3.

At the end of the first semester in 2004, it was identified that the 2004 pass rate was particularly low (see Figure A3-1 in Annexure 3). The challenges faced with first-year students' pass rates are not unique to UP, nor to South Africa (Kantanis, 2000; Pather et al., 2017). Like many other tertiary institutions all over the world, UP has a diverse intake of students, many of whom are mathematically ill-prepared for university (Engelbrecht & Harding, 2008; Engelbrecht et al., 2010; Hourigan & O'Donoghue, 2007; Hoyles et al., 2001; Luk, 2005). We are, however, of the opinion that the diversity in South Africa is more extreme and the disparity in the mathematics standard larger than in many other countries, due to the historic legacy of the country (Fedderke et al., 2000).

The low pass rates called for a revision of the Grade 12 mathematics entrance criteria for the first-year statistics students. A study was conducted to investigate the 2003 Grade 12 learners' mathematics marks versus their 2004 Statistics 110 marks. (A detailed discussion is available in Annexure 2.) From the study it was clear that performance in STK110 is correlated with the mathematics results, and higher Grade 12 mathematics marks as a prerequisite for STK110 was implemented in 2005. The higher Grade 12 mathematics prerequisites had an immediate and positive effect on the pass rate of the new students of 2005. However, the pass rate of the repeat students dropped even further. This prompted a second intervention, namely the introduction of compulsory homework assignments in 2006. At first, these compulsory assignments helped to improve the pass rate of new as well as repeat students; however, the effect was, unfortunately, temporary. The flaw in the system was that the assignment mark weighed relatively heavily and ultimately made up 10% of the final mark. We realised that students could copy their fellow classmates' assignments which inflated their marks to give a false representation of how well they were doing. As a result, this caused a problem by allowing students to proceed to the second year who should not have been allowed to do so. Consequently, it set them up for failure in their second year. The assignments were therefore replaced in 2007 by class tests and the upward trend was reversed.

The Outcome-Based Education (OBE) approach, referred to as Curriculum 2005, was an initiative of the post-apartheid government to equalise school education for all South Africans (Jansen, 1998). The first group of students who had full exposure to OBE over their entire school career entered the university in 2009. (Context of OBE is given in Annexure A4.) Since 1998, statistics formed a much larger part of the new OBE curriculum compared to the old curriculum (North & Zewotir, 2006; Wessels, 2008; Zewotir & North, 2011). Therefore, it was expected that the 2009 cohort should perform better in the first semester in statistics, since they were already familiar with many of the concepts that are taught in STK110. Figure A3-1 in Annexure 3 shows that this expectation was not met, as many students either dropped out or failed STK110. The subsequent disparity between students' knowledge base and what is expected from them at university inevitably impacted negatively on our first-year students' performance (Engelbrecht et al., 2010).

A logit model was fitted to the STK110 data to explore the relationship between statistics achievement and Grade 12 mathematics marks in 2009 and in 2010 (see Annexure 5). It can be deduced from the logit model that the probability to fail STK110 in 2009 was higher compared to 2010 across all the Grade 12 mathematics percentage categories. The reason could have been the inflated mathematics marks of the 2008 Grade 12 learners; therefore many students had been wrongly admitted to STK110 in 2009.

During 2009, STK110 was also identified as a High Impact Module (HIM) by the Faculty of Natural and Agricultural Sciences in which STK110 resides. An HIM has many students, in the order of 1000+ and an initial low pass rate of below 80%. STK110 had a pass rate of below 60% and 3091 students when identified as an HIM. The faculty provided more funding for additional tutoring support. Regular meetings were held to monitor the performance of all HIMs. The additional support had a positive impact on progression but was not the only intervention to improve students' pass rates.

Up until the end of 2011, the low pass rates had not improved dramatically despite continuing efforts, e.g. implementing new prerequisites and class tests on the part of the Department of Statistics to address the problem. Pass rates remained between 59% and 68% and therefore continued to be a matter of concern. The resulting large proportion of annual repeat students in the class was also problematic, because if students fail the module once, they tend to fall into a failing cycle. Therefore, repeat students should be identified and warrant further investigation.

Apart from mathematically ill-prepared students, the statistical community likewise realised that there should be a change. Since the late 1980s there has been a universal call for change in the pedagogy and content of statistics at tertiary level. The quest for change culminated in the "reform movement in statistical education" to improve the content and pedagogy of introductory statistics courses (Cobb, 1993; Cox, 1997; Garfield, 1995; Garfield et al., 2002; Moore, 1997). Moore (1997), in *New Pedagogy and New Content,* emphasised the need for change: "less theory, more data"; "less lecture, more active learning"; "less calculation, more technology". Cobb (1993) reports on the implementation of statistical laboratories where students work with their own data sets, e.g. pulse rates or traffic counts, and on simulation-based teaching. Moore (1997) gives guidelines for teaching introductory statistics to students who should construct their own learning. They also suggest incorporating technology into teaching and learning. Nolan and Lang (2007) introduce visualisation and simulations into their revised introductory statistics course, to highlight the importance of statistical thinking and literacy.

The recommendations of the GAISE (Guidelines for Assessment and Instruction in Statistics Education) report (2005, revised 2016) by the American Statistical Association (2016) are the teaching of statistical thinking, conceptual understanding, real data, active learning, technology and assessments. With the advent of educational technology, there was a call for incorporating technology into the teaching and learning of statistics and for a more "constructivist"[1] approach (Tishkovskaya & Lancaster, 2012).

Technology has broadened the scope of graphical and visualisation techniques to enable students to use case studies to explore and analyse data and concentrate on interpretation rather than laborious computations (Nolan & Speed, 2000). The technology revolution changed what and how we teach (Chance et al., 2007; Cobb, 2007; Garfield, 1995; Garfield & Ben-Zvi, 2007; Xu et al., 2014). The movement towards further change has continued since then, taking shape as innovative teaching and learning possibilities emerged, which included online homework systems and the notion of a flipped classroom, amongst others (Garfield & Ben-Zvi, 2007; Tishkovskaya & Lancaster, 2012). The movement towards further change should also incorporate data science into the statistics curriculum to equip students to meditate with data (Hardin et al., 2015).

1.2.2 INTERVENTIONS 2011 – 2018

The questions under consideration are what the effect and contribution of the three technology-based interventions would be and how to quantify the success of the interventions, discussed subsequently.

The first intervention was the implementation of an online homework system in collaboration with the traditional teaching model. Ungraded pen-and-paper homework was substituted by online interactive homework, which was automatically graded, and immediate detailed feedback provided. After the statistical concepts were discussed in class, online post-class assignments were done. Students could make use of up to three attempts, to learn from mistakes and improve their understanding of course concepts.

[1] To be explained in more detail in Chapter 2

The second intervention was the change in pedagogy, namely the flipped classroom model as substitute for the traditional teaching model. Students had to prepare before they attended classes by pre-reading of the textbook and class notes and attempting the online pre-class assignments. Difficult concepts were revisited during class time, but more time was spent on problem-based exercises. Lectures gradually took the form of tutorial classes. As from 2015, not only were pre-class assignments used for preparation, but also post-class assignments for consolidation.

The third intervention was the implementation of QT-clickers with full featured keypads. With the flipped classroom, there was a need for active learning in large classrooms and QT-clickers were the obvious solution. Clicker questions were initially used to evaluate student preparation and to assess the understanding as well as application of concepts in class, with the advantage of instant and anonymous feedback. The value of QT-clickers for formative assessment created awareness of its use for summative assessment. Multiple choice examination papers were substituted by written examination papers, where final answers had to be submitted via QT-clickers.

The fourth intervention extended the third intervention in 2018 with problem-based, peer learning activities using South African content and lecturer-switching.

1.3 RESEARCH AIMS

The research aims were as follows:

- To investigate what the impact of the four technology-based interventions were with respect to student learning and performance.
    - The first aim of the study was to measure the impact of the flipped classroom model versus the traditional classroom model, as well as the role that online homework plays in both models.
    - The second aim was to investigate the use of QT-clickers in a large first-year flipped statistics module along two lines: the pedagogical influence of the QT-clicker and the effect of partial grade crediting.
    - The third aim was to evaluate problem-based peer learning activities with localised content and lecturer-switching, within a flipped classroom environment and QT-clickers to help students’ understanding of difficult statistical concepts.

- To formulate recommendations that would be of value to other South African universities and, more generally, to teachers venturing into technology-enhanced teaching in statistics.

1.4 RESEARCH QUESTIONS

Four technology-based interventions were introduced and the effect regarding the learning process of each intervention is evaluated:

- Research question 1: How do the technology-based interventions (online homework and/or flipped learning and/or QT-clickers and/or peer learning activities) impact on first-year statistics students' learning and pass rates?
  - Research question 1a: Do the five intervention cohorts have different mean final marks?

    $H_0: \mu_{2011} = \mu_{2012} = \mu_{2013} = \mu_{2015} = \mu_{2018}$

    $H_a: Not\ all\ population\ means\ are\ equal$

  - Research question 1b: Do the five intervention cohorts have different pass rates?

    $H_0: \pi_{2011} = \pi_{2012} = \pi_{2013} = \pi_{2015} = \pi_{2018}$

    $H_a: Not\ all\ population\ proportions\ are\ equal$

  - Research question 1c: Is there an association between the five intervention cohorts and the final marks distribution?

    $H_0: There\ is\ an\ association\ between\ the\ final\ marks\ distribution\ and\ cohorts$

    $H_a: There\ is\ no\ association\ between\ the\ final\ marks\ distribution\ and\ cohorts$

  - Research question 1d: Is there an association between QT-clicker use in 2017 versus 2014 cohorts and the examination marks?

$H_0: There\ is\ an\ association\ between\ clicker\ use\ and\ examination\ marks$

$H_a: There\ is\ no\ association\ between\ clicker\ use\ and\ examination\ marks$

  - Research question 1e: Is there an association between the five intervention cohorts and the examination marks?

$H_0: There\ is\ an\ association\ between\ the\ cohorts\ and\ examination\ marks$

$H_a: There\ is\ no\ association\ between\ the\ cohorts\ and\ examination\ marks$

- Research question 2: What are the perceptions of the students regarding the technology-based interventions?

  $H_0: The\ students' perception\ of\ the\ intervention\ is\ neutral\ or\ negative$

  $H_a: The\ students' perception\ of\ the\ intervention\ is\ predominantly\ positive$

1.5 CHAPTER DIVISION

The chapters are divided as follows:

Chapter 1 consists of the overview of the study. Chapter 2 comprises the literature review. Chapter 3 contains the rationale for the empirical research, research design and methodology. Chapter 4 covers the quantitative data analysis and interpretation. Chapter 5 consists of the course evaluation by the students and Chapter 6 encloses the overview, conclusion, and recommendations.

1.6 ETHICAL CLEARANCE

Ethical clearance was obtained on 3 October 2014 with document number EC 140721-067 and it conformed to the requirements of the NAS Ethics Committee, (see Annexure 1).

CHAPTER 2
LITERATURE REVIEW

Chapter 2 consists of an exposition of the literature of the related theoretical framework – constructivism and three technology-based interventions, namely online homework, flipped learning and QT-clicker use.

The chapter starts with the foundation of the interventions, namely cognitive constructivism, and social constructivism. The leading theorists are identified and their contribution to this study is discussed. The strengths and weaknesses of online homework as an alternative to pen-and-paper homework are explained. Flipped learning is explored and the individual components are discussed. Lastly, the use of QT-clickers to enable active and cooperative learning for face-to-face engagement inside the classroom is explored in the literature.

2.1 CONSTRUCTIVISM – THE THEORETICAL FRAMEWORK

Constructivism is the theory that posits that students can create knowledge rather than receive knowledge. Cognitive constructivism is a comprehensive term for the theory of knowledge assimilated by a student's mind or mental processes, rather than received from external sources. Social constructivism is the theory of a student's learning that occurs because of a student's interactions in a group.

Constructivism is the theoretical underpinning of the three technology-based interventions, i.e. online homework, flipped learning, and the use of QT-clickers in class. The leading theorist among cognitive constructivists is Jean Piaget (1896 – 1980). Lev Vygotsky (1896 – 1934) can be considered the main social constructivist, while Jerome Bruner (1915 – 2016) combines the cognitive and social constructivism theory (Lutz & Huitt, 2004). Piaget perceives a child as an independent being who should acquire knowledge through self-discovery. Vygotsky observes a child as a "social being and cognitive development is led by social interactions". Cognitive development can be accelerated if the right scaffolding provided by an educated person is used within the zone of proximal development. The "zone of proximal development is the distance between actual development level and the potential development level" (Vygotsky, 1978, p. 84), e.g. in tertiary education, it denotes the difference between what a student can achieve with or without the help of an expert lecturer or peer. Bruner was strongly influenced by Vygotsky's theory of social constructivism (Zhou, 2020). He also believes that learning new concepts is influenced by active learning with the assistance of an educator, but the new concepts should be discovered by the learner, not told by the educator (Zhou, 2020). Bruner differs from Piaget by arguing that cognitive development should be seen as a continuous process instead of being categorised into separate stages (Tomic & Johannes, 1996).

The three constructivists mentioned above provide a significant basis to this study. Piaget's cognitive development theory influenced the design of this study's out-of-class activities of the students. Students were expected to use self-directed learning, i.e. pre-reading of the textbook and class notes and engaging in online pre-class assignments to explore and construct statistical concepts, in order to build their own prior knowledge. Relationships have been found between the prior knowledge and performance in statistics courses (Schutz et al., 1998). Our in-class activities are inspired by Vygotski and Bruner's social constructivism (Zhou, 2020). Students are engaged in various active learning activities in class, with QT-clickers playing a fundamental role. At face value, the large lecture groups seem contradictory to social constructivism perspectives, but students form small informal groups for peer discussion before they submit their answers to QT-clicker questions in class. The immediate feedback helps students to clarify their misconceptions (Dunn et al., 2012) and integrate it with their prior knowledge (Bransford et al., 2000, pp. 10-11, pp. 54-55). It is a continual scaffolding process building on prior knowledge in a student-centred environment, where a higher level of potential learning can be attained (Hannafin & Land, 2000). The lecturer helps students to engage in active learning through the use of scaffolding, i.e. providing comprehensive support with all the different tasks in class (Wood et al., 1976). As students take more ownership for their learning, the assistance can slowly be reduced and eventually removed when students perform better.

To improve learning, it is important to understand Bloom's taxonomy. The revised Bloom's taxonomy in 2001 consists of two divisions, namely the knowledge division and the cognitive process division (Krathwohl, 2002). The knowledge division is divided into four categories, i.e. factual, conceptual, procedural and metacognitive. The metacognitive knowledge category plays an important role in the students' learning strategy (Krathwohl, 2002; Pintrich, 2002). The vital component of metacognition is self-knowledge, amongst others. "Self-knowledge gives students an opportunity to assess their own strengths and weaknesses" (Pintrich, 2002, p.

222). Students should be able to reflect on their own learning, identify what they do not know and get help from their instructor. The cognitive process division is divided into six categories; the lower-level thinking skills represented by the bottom three and the higher level thinking skills represented by the upper three categories in Figure 1. Several cognitive research studies have agreed that learning can be improved if students can get involved in the higher level thinking skills of Bloom's revised taxonomy in Figure 1, i.e. analyse, evaluate and create (Krathwohl, 2002; Mayer, 2002; Pintrich, 2002).

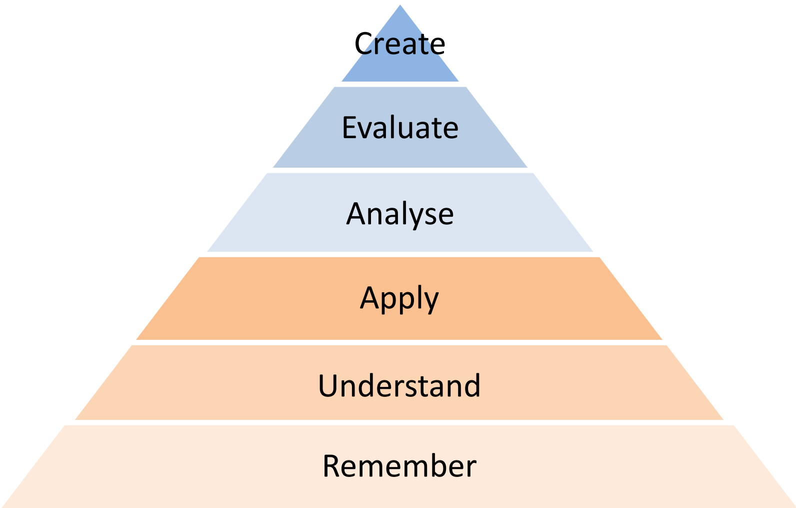


Figure 1: Bloom's Revised Taxonomy
Source: Based on an adaptation of Wilson (2016)

In the ideal constructivist classroom, the focus tends to shift from the lecturer to the students, called a student-centred classroom. Students are no longer perceived as empty vessels that should be filled with knowledge by the lecturer, instead: "Students learn by constructing knowledge rather than just receiving knowledge" (Garfield & Ben-Zvi, 2008, p. 46).

If students are given excess new information at any one time, it can lead to cognitive overload (Farrington, 2011). Students who attend a lecture without prior knowledge regarding the course content, can be strained by cognitive overload. In a study measuring the influence of prior knowledge (pre-stats knowledge and pre-math skills) amongst other factors, relationships were found between the prior knowledge and performance in statistics courses (Schutz et al., 1998).

Students in class cannot be seen as clean sheets; their existing knowledge influences their understanding of text (Lovett & Greenhouse, 2000). According to Bransford et al. (2000, pp.

10-11, p. 233, p. 267), learning is dependent upon prior knowledge and as we engage with new knowledge, the prior knowledge does not disappear; it is integrated with the new knowledge, i.e. the new knowledge is constructed on the prior knowledge. Instructors should be aware of incomplete prior knowledge, misunderstandings, fallacies and immature interpretation of concepts, which become the lecturer's challenge to build on inexpert ideas to make them more logical for students (Bransford et al., 2000, pp. 71-72, pp. 238-241; Lovett & Greenhouse, 2000).

Clark et al. (2006, p. 15, pp. 28-40) explain how the working memory and long-term memory interact. The working memory has limited capacity; therefore students should construct their own prior knowledge (Cowan, 2016). If prior knowledge is constructed, it will be stored in the long-term memory, which has a huge storage capacity. The working memory will then be released to receive new information and connect it to the prior knowledge in the long-term memory, which can deepen the learning process (Cowan, 2016).

Miller (1956) publicised limits to capture the amount of information that can be held in the working memory. Saaty and Ozdemir (2003) argue that the limits can even be stricter. More recent studies with varying views on the limited capacity of the working memory were recorded (Abeysekera & Dawson, 2015; Clark et al., 2006, pp. 29-33; Cowan, 1998, 2016; Farrington, 2011). The main conclusion is that working memory is restricted (Saaty & Ozdemir, 2003) and students should be taught how to construct their own knowledge, in order to optimise the working memory for problem solving.

Social constructivist theories of learning support the combination of the flipped classroom approach and active learning in the classroom (Milligan et al., 2013). The flipped classroom approach will be discussed in Section 2.3. Social constructivism implies that the learner is much more actively involved in a combined effort with the lecturer, trying to create or construct novel understanding. Vygotsky's (1978, p. 84) idea of the "zone of proximal development" is appropriate, in that a student can learn some things independently, but can also be challenged to a further level of understanding through scaffolding, i.e. supported by constructed tasks with a lecturer or peers in an active learning, inquiry-based lecture hall, using a problem-based approach.

In a constructivist classroom, constructivism is seen as a learning and not a teaching theory (Baviskar et al., 2009). Instructors often think if group work or active learning is used in class it constitutes a constructivist classroom, but that is not necessarily the case. A constructivist classroom should give students the best chance to optimise learning, no matter the technique used (Baviskar et al., 2009).

According to Baviskar et al. (2009, p. 541), there are four required characteristics of constructivism, namely "eliciting prior knowledge, creating cognitive dissonance, application of new knowledge with feedback, and reflection on learning". The four characteristics listed can be described to optimise student learning in a constructivist classroom. Students can be guided by the lecturer to think and say what they know about a subject. The lecturer should not give all the answers. This will help a student to identify their own prior knowledge to integrate the new knowledge correctly. A student should be made aware of the difference that can exist between their prior knowledge and the new knowledge. A student should apply the new knowledge, for example in group discussions or puzzles, where students could compare their understanding of new concepts with peers' or the lecturer's. Lastly, a student should reflect on the new knowledge that has been tested. A way of doing it is to explain the new concepts to peers.

2.2 ONLINE HOMEWORK

Student success in an undergraduate introductory statistics course depends on practise and feedback, typically in the form of a homework assignment, but whether it is pen-and-paper or online does not necessarily matter (Palocsay & Stevens, 2008).

Firstly, the various available options of online homework are discussed. Most statistics textbook publishers provide online homework platforms compatible with the institution's learning management system, e.g. Blackboard Learn. There are also open access online homework banks available, e.g. Webwork. In addition, there are still some lecturers who prefer to create their own online homework systems. Santoro and Bilisoly (2014) created an online homework system for mathematics and statistics by using the Mathematica program. The statistics lecturer in Santoro and Bilisoly (2014) followed a more data-driven, textbook-free approach, i.e. no access to any publisher's online homework platform is needed.

Research findings on the advantages of online homework versus pen-and-paper homework are varied and debated by several authors (Chow, 2015; Williams, 2012). Some students enjoy doing their homework online, but other students prefer pen-and-paper homework (Chow, 2015). Palocsay and Stevens (2008), for example, found that in a business statistics course the type of homework system used did not make a significant difference in students' performance. They constructed a regression model to predict the examination score and found that the most important predictor was the cumulative Grade Point Average (GPA) of students and not the homework medium. It is not conclusive that the advantages of an online homework system outweigh those of pen-and-paper homework (Chow, 2015; Williams, 2012). Increasing class sizes at universities, however, complicate graded pen-and-paper homework. Some authors found that although there is no significant difference in students' performance between online and traditional homework, online homework is beneficial for large classes (Doorn et al., 2010; Jonsdottir et al., 2017). Online homework is more effective than assigning ungraded pen-and-paper homework or not assigning homework at all (Santoro & Bilisoly, 2014).

The prominent advantages of online homework are automatic grading, consistency of graded assignments across the board, and instant and effective feedback (Burch & Kuo, 2010; Chow, 2015; Dillard-Eggers et al., 2008; Lunsford & Pendergrass, 2016). The immediate and effective feedback is valuable for both students and instructors. Students can learn from their mistakes and instructors can observe students' progress and change the lecture plan accordingly, i.e. revise previous concepts before new concepts are introduced (Chua-Chow et al., 2011; Lunsford & Pendergrass, 2016). Feedback is most influential when it addresses misconceptions rather than merely dealing with a lack of knowledge (Hattie & Timperley, 2007). Smolira (2008) argues that timely feedback may be a factor in the effectiveness of homework. He reasons that an online homework system with immediate feedback may be a solution to the problem of delayed feedback that is often associated with pen-and-paper homework. Students tend to forget what the previous homework involved when they move on to a new topic in the next class.

Secondary benefits of online homework are that students enjoy online homework; it is less likely that students will copy answers from peers as they get different versions of the same questions (Doorn et al., 2010), or a complete new set of questions with the Tutor-Web (Jonsdottir et al., 2017), and they can get sufficient practice opportunity and multiple attempts for an assignment (Palocsay & Stevens, 2008), where they could learn from mistakes and get alternate versions of the same assignment and topic, because of the pool of available questions (Doorn et al., 2010). In most online question pools, questions are assigned with equal probability to each assignment. Even if students work together in informal groups, it will increase student understanding because they could learn from their peers (Doorn et al., 2010).

Patron and Smith (2009) describe the benefit of multiple attempts of online homework for business statistics students who learn by repetition. After the first attempt a student can get immediate instructional feedback and will be able to increase quiz scores in a second attempt and learn the material in the process. Results show better student performance for students who made use of multiple attempts for online quizzes (Patron & Smith, 2009). Burch and Kuo (2010) found that the online homework students in an algebra course outperformed the traditional homework students in the examination, due to revisiting and reattempting online problems.

Lunsford and Pendergrass (2016) explain that online homework can increase student engagement and achievement, particularly in lower level statistics classes. Lunsford and Pendergrass (2016) motivate students to keep a notebook for online homework problems. They integrate online homework problems in the face-to-face classes for students to discuss, which would probably lead to deeper learning.

The disadvantages of online homework include technical issues, e.g. student access, server crashes and students' insufficient technical abilities (Doorn et al., 2010), although server and access problems are constantly decreasing and students' abilities are better than in the past. Other minor disadvantages perceived by business statistics students in an anonymous questionnaire regarding online assessment are that the lecturer is not available, online assessment can encourage dishonesty, and online assessment allowing multiple attempts discourages studying for the assessment beforehand (Patron & Smith, 2009). These are all valid perceptions, but not insuperable. The lecturer can be e-mailed or video-called, dishonesty can be limited by different versions of questions or time limits on assessment tasks, and a different grading system can be used, e.g. the final mark is the average of all the attempts instead of the best of all the attempts (Patron & Smith, 2009).

For some of the studies an overall significance in students' performance and choice of homework system (or computer-assisted instruction or not), could be found.

Chua-Chow et al. (2011) compared the Introductory Business Statistics class of Term 1 of 2009 (without e-homework) with Term 1 of 2010 (with e-homework). The 2010 class with e-homework outperformed the 2009 class without e-homework regarding final marks, increased class average and pass rate. A student survey completed by the 2010 students revealed a positive attitude towards e-homework. It helped them to be prepared for tests, it encouraged them to discuss statistics with peers and the immediate feedback increased their understanding of concepts.

Basturk (2005) used a quasi-experimental design to compare a computer-assisted group with a lecture-only group regarding mid-term and final examination scores in an introductory statistics course. The computer-assisted group attended an extra 40-minute class in the laboratory, where they engaged in exercises and tutorials using the data analysis package SPSS (Statistical Package for Social Sciences). Students also had to know how to analyse and interpret the SPSS computer output. The average scores of the computer-assisted group were higher than the lecture-only group. A noteworthy result was that moving from descriptive statistics (mid-term) to inferential statistics (final examinations), the learning gap between the computer-assisted and lecture-only groups enlarged, i.e. the average score of the computer-assisted group increased and the average score of the lecture-only group decreased. In a meta-analysis of 45 studies done by Sosa et al. (2011), the effectiveness of computer-assisted instruction compared to lecture-only instruction in statistics education is measured. The average effect size[2] of 0.33 implies that computer-assisted instruction is on average more successful than lecture-only instruction.

Several authors (Basturk, 2005; González & Birch, 2000; Larreamendy-Joerns et al., 2005; Tchantchane & Fortes, 2011) are of the opinion that computer-assisted instruction based on exercises that are realistic and relate to students' everyday life can reinforce their understanding of the course material and of difficult concepts.

Although some authors could not find an overall significant difference in students' performance and choice of homework system, others found a noteworthy link between performance and repeat or low-skilled students.

[2] Effect size guidelines for Cohen's $d = 0.2$ is small; $d = 0.5$ is medium and $d = 0.8$ is large

Smolira (2008) did not find a significant difference in introductory finance examination scores of the online homework and traditional homework group. Brewer and Becker (2010) also did not find an overall significant difference in various sections of college algebra of the online homework and traditional homework group. The only significant difference between the web-based and traditional homework groups were for the low-skilled students who obtained a higher mathematics grade. Another result worth mentioning, while not significant, is that more repeat or low-skilled students who used online homework passed the algebra course compared to their non-repeat or higher-skilled peers. An online homework system can similarly be used to motivate students and is valuable for low achievers (Brewer & Becker, 2010). Wooten and Dillard-Eggers (2013) state that high-performing or naturally motivated accounting students will perform no matter what type of homework system or learning method is used.

Jonsdottir et al. (2017) used a unique webpage called Tutor-Web (https://tutor-web.net) for web-based homework versus pen-and-paper homework. The web-based homework, a drilling system, can be used free of charge and is set up for several courses, including statistics and mathematics. Initially homework exercises were randomly drawn from a pool of questions which implies that everybody gets questions of the same difficulty level. As from 2011, experimental changes have been made to the Tutor-Web, i.e. the allocation algorithm, the grading procedure, and the type of feedback. In 2012 to 2013, Jonsdottir et al. (2017) changed the allocation algorithm to increase the questions' difficulty level as the student's learning process progressed. A student could only complete a session with a high grade if the most difficult questions could be done. Students could also get questions from previous sections to refresh their memories. The type of feedback changed from limited feedback restricted to the correct answer to detailed feedback including calculations. The advantage of the Tutor-Web is that the latest version (mobile website) can run on tablets and smart phones and an internet connection is only needed when downloading questions (Lentin et al., 2014). Students can complete the questions off-line and the answers will be stored on the student's device until the student is connected to the internet where the answers will be uploaded to the Tutor-Web server. Jonsdottir et al. (2017) conducted a cross-over experiment on an annual basis from 2011 to 2014, using 100 introductory statistics students to test whether the changes implemented in Tutor-Web were significant between web-based and pen-and-paper homework. They observed a significant positive difference in learning using test scores between web-based and pen-and-paper homework from 2011 to 2014.

2.3 FLIPPED LEARNING

Flipped learning emerged during the late 1980s (Baker, 2000; Strayer, 2007), due to the universal call for change in pedagogy (Moore, 1997). It was simplified by the technological movement, e.g. the advance of computing power of computers from the late 1950s and the free software movement from the 2000s that have changed the face of education (Bishop & Verleger, 2013). The initial principle of the flipped learning approach was to interchange the normal lecture and the homework environments. This is where the term inverse learning originated and since then the concept of flipped learning has been coined, based on the same principle. Baker (2000) presented a paper more than two decades ago, clearly discussing the out-of-class and in-class activities of a flipped classroom in the finest detail. Flipped learning in this study is also based on the principles of Baker's (2000) model of out-of-class and in-class activities.

Flipped learning is implemented around the globe in various disciplines, e.g. mathematics (Liou et al., 2016; Naccarato & Karakok, 2015), science (González-Gómez et al., 2016; Liou et al., 2016), and statistics (Chen et al., 2015; Peterson, 2016).

The major attribute of flipped learning is constructivism as its foundation (Chang, 2016). Constructivism is therefore the underpinning of flipped learning. The non-profit online community for instructors called the Flipped Learning Network (FLN, 2014) distinguishes flipped learning from a flipped classroom. A flipped classroom can consist of class work at home and homework in class, but without certain methodologies in place it is not flipped learning. Therefore, a flipped classroom can lead to flipped learning, but should not be used interchangeably. The four pillars of flipped learning (FLN, 2014) are: Flexible Environment, Learning Culture, Intentional Content and Professional Educator. The definition of flipped learning, according to FLN (2014, p. 1), is as follows: "Flipped learning is a pedagogical approach in which direct instruction moves from the group learning space to the individual learning space, and the resulting group space is transformed into a dynamic, interactive learning environment where the educator guides students as they apply concepts and engage creatively in the subject matter". Flipped learning allows students to learn concepts and content and construct basic knowledge in their own learning space, before attending the group learning space to explore and experience concepts by practical activities and more in-class time to maximise collaboration between lecturers, students and peers (Chang, 2016; Mason et al., 2013; Wilson, 2013). The integration of out-of-class and in-class activities is of the utmost importance for flipped learning (Naccarato & Karakok, 2015; Peterson, 2016). The flipped learning pedagogy is established and has become a flourishing domain for education research and practice development (Bishop & Verleger, 2013; Naccarato & Karakok, 2015; Overmyer, 2014).

Although terminology like flipped classroom or flipped learning has been commonly used for decades, it can be misleading or incorrectly interpreted by some educators and authors. Bogost (2013) criticises a flipped classroom as a condensed classroom where course material is shortened into pre-built summarised lectures and students are disadvantaged of not having access to fundamental material. He also postulates that a flipped classroom is related to Massive Open Online Courses (MOOCs), because the lecturer wants the students to watch a video lecture at home to come prepared to class the next day for problem-solving.

Flipped learning is often used in the same frame of reference or context of fully online, augmented, hybrid or blended learning (Halili & Zainuddin, 2015; Suleiman, 2018). Flipped learning is neither placing a traditional classroom course online, nor blended learning per se, although some instructors classify flipped learning as a special blended learning model (Halili & Zainuddin, 2015). Hybrid or blended learning is combining online learning with face-to-face learning, but the rationale of flipped learning is how the out-of-class online component is integrated into the overall methodology (Tucker, 2012).

The success recipe of flipped learning that contributes to improvement in students' learning outcomes is the interaction of the prior knowledge using an online homework system and active learning in class (Gross et al., 2015). This is the reason why flipped learning is in a different class to fully online, hybrid or blended learning.

Several studies have been done in different study fields, comparing the learning outcomes of traditional learning (face-to-face) to fully online or hybrid learning. Ary and Brune (2011), Driscoll et al. (2012), and Wilson and Allen (2011) compared fully online learning, i.e. learning and assessment online to traditional formats, and found that delivery mode does not count, i.e. there is no significant difference between student performance and student satisfaction regarding delivery modes. Bowen et al. (2015) compared a hybrid to a traditional learning group for an introductory statistics course and found no significant differences in learning outcomes between the two groups. Many more studies with similar outcomes show the necessity for a change in pedagogy, i.e. learning in a "constructivist" manner (Moore, 1997), of which flipped learning forms part.

Schwartz et al. (2016) point out that there is no "one-size-fits-all" flipped classroom, therefore no two flipped classrooms are the same. Nor do all lecturers implement a 100% flipped classroom at first, for example Wilson (2013) refers to her undergraduate statistics classroom as a "half" or "three-quarter" flipped classroom, where some of the lecture content forms part of the in-class activities. Vidic et al. (2011) implemented a flipped classroom approach in two of the three sections of their introductory probability and statistics course for engineers and kept the traditional teaching model in the third section to measure the benefits of flipped learning. Although there were no significant differences in examination scores between the flipped sections and the traditional section, students could experience active learning and a variety of resources, e.g. videos.

Haughton and Kelly (2015) evaluated the performance of a so-called flipped-hybrid model for introductory business statistics versus a traditional face-to-face lecture model. In the flipped-hybrid course, students had to view lecture material online preceding a once-a-week face-to-face lecture, where the instructor could revisit concepts and give more detailed explanations to clarify uncertainties. The flipped-hybrid model students performed better than the traditional model students, controlling for all covariates, although there were no significant differences in final grades or student perceptions. Research has shown that if better performance or learning outcomes are the focus of changing the pedagogy of the traditional course, then flipped learning could be an option (Harris et al., 2016).

The flipped classroom is an attractive alternative to the traditional classroom (Munir et al., 2018). A traditional lecture is lecturer-centred and knowledge is transferred to passive spectators, who are all treated the same (Bligh, 2000; Munir et al., 2018). Students easily get into a comfort zone of just receiving information, without really constructing and assessing the new information to better understand it (Garfield & Ben-Zvi, 2008). Another criticism of lecture-based teaching is that it does not promote long-term retention of threshold concepts or the application of concepts in real world settings (Bacon & Stewart, 2006). Davis and Minifie (2013) observed that if content is repeatedly revisited in different ways, i.e. before class in student preparation, in-class with different active learning techniques and possibly consolidation after class, then there is more potential for long-term retention and better learning outcomes.

A reality that should be considered to create better learning opportunities is the students' different personality types and learning styles. Students use different senses and learn in different ways (Munir et al., 2018). The traditional lecture approach fails to incorporate the different personality types and learning styles, which are better integrated by the flipped classroom model that uses multimedia (Dove, 2013; Lage et al., 2000). Gross et al. (2015) found that the flipped classroom increased the performance of students with lower GPAs and of female students in a physical chemistry course, because of the diversity of learning tools that better utilise all different learning styles.

Prensky (2001) declares that students have changed over the years. The Y-generation (Millennials) and Z-generation students are the so-called "digital natives" who are born and bred into technology. Instructors think that "digital natives" have attention-deficit problems, but these students are more open-minded and have different needs that could be addressed by innovative pedagogy and updated content (Prensky, 2001, 2011). Roehl et al. (2013) and Davis and Minifie (2013) propose the flipped classroom as a workable solution to capture the attention of the 21st century students through the use of active learning activities in class.

Flipped learning changes the lecturer-centred environment to a student-centred environment, where student motivation might be enhanced and cognitive load may be controlled (Chang, 2016). If the traditional lecture is just presented online and nothing else is changed, there will be no significant impact on learning (Abeysekera & Dawson, 2015). The change in transmission of information by using computer technology and the internet out-of-class is an essential process in flipped learning, because prior knowledge is created that can be used in the construction of knowledge which contributes to the effectiveness of learning (Clark et al., 2006, pp. 248-257).

Self-paced preparation in the form of lecture videos will not overload the cognitive capacity and might control the working memory better than traditional lectures (Abeysekera & Dawson, 2015; Clark et al., 2006, pp. 34-36, pp. 293-311). Even if students just control the pace of the videos, there will be improvement in the learning, because of a reduction of cognitive load (Clark et al., 2006, pp. 293-311). Highly intelligent students can fast-forward the video and should not be bored by listening to concepts that they already understand, but the less proficient students could prepare at a slower pace, i.e. rewind the video-recorded lecture and listen to it numerous times (Dove, 2013). Research has shown that if students' study sessions could be spaced out and not crowded, summative learning could be improved (Carpenter, 2012). Instructors should revisit concepts and introduce cumulative tests and examinations. Long-term retention of knowledge could be supported if students repeatedly revise course material by using the advantages of spaced-out sessions (Cepeda et al., 2008).

Strayer (2012) compares the learning environments of the flipped classroom with the traditional teaching model for a course in introductory business statistics and discusses how each of the models contributed to the strength and connectedness of classroom learning communities. The main finding is that the flipped classroom students preferred an innovative environment and cooperation with peers in class activities, while significantly fewer traditional classroom students associated group activities with a successful classroom learning community.

Hamdan et al. (2013), who support flipped learning, suggest that students should not passively receive material in class; they have to come to class prepared to collaborate with peers and the lecturer so that misconceptions can be corrected, difficult concepts can be revisited and mastered, and deeper learning can take place in class. Ben-Zvi (2007) advises that prior preparation in the form of pre-reading, videos or quizzes can be reinforced by collaborative discussions in Wiki. Wiki is a collaborative website where all users can edit and change content and have fruitful discussions using the discussion page that supplements the Wiki page. Statistics students can post questions after they have read the required chapter and make necessary connections using lecturers' or peers' timely feedback (Ben-Zvi, 2007).

Active learning forms an integral part of flipped learning. "Active learning can be defined as instructional activities involving students in doing things and thinking about what they are doing" (Bonwell & Eisen, 1991, p. 82). Learning is not a spectator sport, but an dynamic activity for students to engage in (Bonwell & Eisen, 1991; Sutherland & Bonwell, 1996). The learning environments of flipped classrooms are interchanged. Since the transfer of content is scheduled out-of-class, active learning, i.e. cooperative, collaborative, peer and problem-based learning all form part of the in-class activities. Baepler et al. (2014) sacrifice the traditional amphitheatre where they taught a large chemistry class for much smaller active learning classes. Recorded content was posted online for students to prepare and class time reduced by 66% in turn to incorporate active learning. They found that students identified better with the new learning environment and their performance was at least as good as with the traditional model.

Since the 1990s there were calls to reform statistical education by integrating cooperative learning activities with active learning in-class (Garfield, 1993, 2013). Cooperative learning involves students working face-to-face in formal groups. A few recent studies on the combination of cooperative learning in conjunction with the flipped classroom in different courses have been positively received by students, discussed subsequently. The first study used Q-Methodology on a small group of Taiwanese undergraduate statistics students to identify student perception factors of the use of cooperative learning in combination with flipped learning. Apart from a few unique statements characterising student perceptions of cooperative learning, their overall perception is that cooperative learning is an effective learning method to use as part of the flipped classroom approach (Chen et al., 2015). In another study, Foldnes (2016) performed two randomised experiments in a mathematics course where they implemented a flipped classroom with or without cooperative learning. In the experiment where students were allocated to the flipped classroom combined with cooperative learning, they outperformed the traditional classroom students. In the third study, cooperative learning formed an integral part of the engineering course's in-class activities of flipped learning, and apart from better grades, more than 90% of the students were pleased with the positive influence on their learning, e.g. development of thinking and problem-solving skills (Munir et al., 2018).

Research indicates that there are several studies where the implementation of a flipped classroom in an undergraduate statistics course had a significant influence on performance. In the study of Wilson (2013), a flipped statistics classroom was introduced for a small group of students with diverse majors taught by lecturers from the psychology department. Graded reading quizzes to motivate students to read the textbook were implemented successfully. An improvement in examination scores could be related to the amount of time spent out-of-class reading the textbook and constructing prior knowledge.

There is evidence that the implementation of a flipped classroom in an undergraduate business statistics course had a significant influence on performance (Shinaberger, 2017; Winquist & Carlson, 2014). Although the lecturer found the introductory business statistics flipped class enjoyable to teach, the majority of the students had negative attitudes, which is contradictory to their significant student grades and learning outcomes (Cilli-Turner, 2015). Peterson (2016) used a quasi-experimental study in a statistics course and showed that the flipped classroom students outperformed the traditional classroom students and were positive about the flipped classroom. The flipped mathematics classroom introduced in conjunction with cooperative learning resulted in highly significant increases in test and examination scores compared to the traditional classroom (Foldnes, 2016).

The key finding concerning flipped learning is the significance of the right combination of the out-of-class preparation activities with active learning activities in-class. Gross et al. (2015) believe that the combination of preparing online and then attending an active learning class forms the strength of the science of flipped classroom compared to the standard lecture classroom. The examination performance of the flipped classroom students increased significantly by 11.6% compared to the traditional classroom students. Another study by Movarec et al. (2010) that uses an approach called Learn before Lecture (LBL) in a large introductory biology class made a noteworthy contribution to flipped learning. Equally efficient LBL worksheets and video LBLs were developed for out-of-class preparation, which, if combined with in-class active learning, improved learning outcomes significantly, i.e. for the six LBL-related examination questions the average increase in performance was 21% compared to a 3% increase on all non-LBL examination questions.

Any new pedagogy has its challenges for both the lecturer and the students. The lecturer, also known as a facilitator in the flipped classroom, should be able to step down from the podium and be more in the background observing students who struggle and assist them with timely feedback (Bergmann & Aaron, 2012, p. 16). The facilitator should also deal with constructive criticism from students and allow for a disorganised learning space (Hamdan et al., 2013).

Another challenge from the instructor's point of view is that extensive time and resources are needed to develop online lecture videos and other flipped classroom materials (Akçay & Akçay, 2018; Roehl et al., 2013). The preparation and recoding of high-quality videos to maximise student engagement and contentment are time-consuming, but should be seen as a long-term investment (Akçay & Akçay, 2018; Balaban et al., 2016). Jensen et al. (2015) and Kay and Macdonald (2016) argue that increased learning in their studies could have been a result of active learning used for in-class activities of flipped learning. If it is true, then all the effort and time spent on creating out-of-class materials, e.g. videos, could have been focused to improve active learning activities to be integrated within the traditional learning classroom.

One of the most frequently reported challenges is the insufficient student preparation preceding lectures (Abeysekera & Dawson, 2015). Taylor (2011) found that the majority of students would not read the textbook prior to class if the lecturer was going to present similar material in class. Educators will have to motivate students to come prepared to class by using lecture videos or pre-reading, since prior knowledge is a vital facet of flipped learning (Balaban et al., 2016; Clark et al., 2006). For example, the first-year students in an introductory probability and statistics course for engineers did not prepare with the pre-class videos as proposed (Vidic et al., 2011). McCarthy (2016) found in an undergraduate 3D animation course that most of the older students (19+ years) preferred the flipped format, but more than half of the younger students (17 to 18 years) preferred the traditional format. The reality is that younger students are used to their high school learning practices and probably have limited knowledge of self-directed learning, while students in a traditional teaching model are not necessarily expected to take responsibility for their own learning, because the lecturer is the supplier of information.

Phillips and Phillips (2016) observe that student perceptions not only influence student learning experiences but provide good guidance for instructors of flipped learning. Bechter and Swierczek (2017) and Dove (2013) note that their students are positive and enthusiastic about flipped learning. They find the activities in-class constructive to learning that lead to better understanding of course material. One student in Dove's (2013, p. 5) statistics class said: “This has been the easiest class for me, and I attribute it to your video method and practise time in class”. The other students in class expressed a newly found confidence and interest in mathematics skills.

In conclusion, students in the majority of disciplines find lecture or tutorial videos supportive and they can watch it at their own pace and leisure, but the drawback is that they cannot ask the lecturer questions while watching the videos (Chen et al., 2015; McCarthy, 2016). Some students do not have the hardware, software or internet connection at home to watch videos, therefore it is the lecturer’s responsibility to make sure that students can download the videos and watch them at university or at home without an internet connection (Chen et al., 2015; McCarthy, 2016; Schwartz et al., 2016). Some students feel that they are overloaded by out-of-class activities (Akçay & Akçay, 2018; Chen et al., 2015). Bishop and Verleger (2013) observed that student perceptions from an inclusive survey are positive but contradictory, namely that students perceived improved learning in the flipped format compared to the traditional format, and although they preferred normal lectures to video lectures, they favoured interactive activities over lectures.

2.4 CLICKERS

The use of electronic voting systems in large lecture modules, especially in science, can be dated back to the 1960s (Judson & Sawada, 2002). Since then, lecturers using such systems have the same aim, namely active learning and instant feedback, particularly from students in large classes. Technology has evolved rapidly, therefore crude, hard-wired systems of the past paved the way for today's modern wireless, multi-function alternatives, although the underlying pedagogy has stayed the same in the majority of classes (Judson & Sawada, 2002).

A distinct feature of current classroom response systems (clickers) is the ability to display graphic representations of students' responses in the form of bar charts, which facilitates immediate feedback. The vast majority of formative assessment research on classroom response systems has been done in psychology (Lantz & Stawiski, 2014), biology (Smith et al., 2011), science (Baltaci-Goktalay, 2016), business subjects (Rana & Dwivedi, 2016), mathematics (Chen et al., 2010), and statistics (Kaplan, 2011). Only a few non-specific articles, none in statistics, could be found on summative assessment involving clickers (Han & Finkelstein, 2013; Hancock, 2010; Kay & LeSage, 2009; Premkumar, 2016; Wang et al., 2014). Studies on the use of clickers in flipped classrooms are even scarcer. Only two clicker studies could be traced, one in an English and the other in a statistics flipped classroom (Hung, 2017; McGee et al., 2016). Formative assessment is continuous assessment with or without grades. If graded, the formative assessment tasks contribute a small percentage to students' grades. Summative assessment is a formal, graded test or examination paper, written in a controlled environment without peer interaction.

Mazur (1997, pp. 9-32) implemented peer instruction supported by clickers in the 1990s. Mazur and his team later developed an interactive program, Learning Catalytics (Prensky, 2011), which needed a more sophisticated response tool in the form of smart devices, such as smart phones and tablets to accommodate advanced software. More recently, several papers on the use of smart phones and other internet devices as substitutes for traditional clickers have been published (Chou et al., 2017; Dunn et al., 2012; Hung, 2017).

Barnett (2006) divided the advantages of using a classroom response system for formative assessment into three categories, namely interactional, attitudinal, and pedagogical advantages, which are discussed accordingly.

Lasry (2008) argues that using peer instruction and low-technology flash cards or peer instruction or using high-technology clickers provide the same interactional benefit. However, while a lecturer can easily implement flash cards instead of clickers in small classes, with large classes there could be gain from clickers that provide automatic, accurate, anonymous and immediate feedback (Koppel & Berenson, 2009).

According to the *Guidelines for Assessment and Instruction in Statistics Education*, GAISE *project*, (American Statistical Association, 2016), lecturers should enhance active learning and use the appropriate technology to develop conceptual understanding. Clickers offer a good option, because they can promote active and fun learning instead of students passively listening to a lecturer (Caldwell, 2007; Dufresne et al., 1996; Hoekstra, 2008; Sharma et al., 2005; Wit, 2003).

Many lecturers initially used clickers to incorporate active learning as a one-way engagement in large classes (Crews et al., 2011; Littauer, 1972). It soon emerged that students prefer to use clickers with peers to learn from each other and the lecturer, discussing questions and exchanging ideas (Beatty, 2004; Bojinova & Oigara, 2013; Caldwell, 2007; Kay & LeSage, 2009; Smith et al., 2011; Yourstone et al., 2008).

Another benefit of the use of clickers is that it may help to improve student attitudes. A student's attitude towards a subject is of critical importance. Many students have a fear of subjects like mathematics, statistics and science (Fullerton & Umphrey, 2001). Positive attitudes when using clickers are observed in physics by Sharma et al. (2005). Barnett (2006) shared a similar experience with science students who demonstrated their clicker experience as fun, like playing a game, and convenient. Some students are often negative about the introductory statistics module under discussion, which is a service course with compulsory enrolment. Clickers could change students' attitudes towards the subject if they start to enjoy learning statistics and realise it is useful (Mateo, 2010; Mocko & Jacobbe, 2010; Titman & Lancaster, 2011).

Researchers found that the anonymity of clicker use results in students not having to embarrass themselves in front of their peers, thereby fostering positive attitudes and active participation in class (Draper & Brown, 2004; Freeman et al., 2006; Laxman, 2011; Sharma et al., 2005; Trees & Jackson, 2007).

The study by Amstelveen (2013) confirms that proper clicker usage can improve attitudes, which, in turn, has an influence on class attendance. There is a relationship between class attendance and academic achievement in an introductory statistics course (Wang & Englander, 2010). In a study by Credé et al. (2010) it was found that class attendance is the better predictor of grades than any other of the student characteristics, including hours studied and GPA scores.

Finally, we turn to the pedagogical advantage. It has been found that metacognition from clicker use has a positive impact on the learning process – learning occurs at a higher level and students can clarify their own misconceptions because of immediate feedback (Dunn et al., 2012; Forster, 2014; Lantz & Stawiski, 2014), and therefore gain better understanding of content (Barnett, 2006; Brady et al., 2013; Mayer et al., 2009). Mayer et al. (2009) suggest that a questioning-based teaching model, using clickers, can enhance academic performance and generative learning (constructing meaning) in large classes.

To develop higher order questions that will stimulate students' critical thinking takes time and effort (Beatty et al., 2006). Barnett (2006), Smith et al. (2011) and Büyükkurt et al. (2012) argue that the use of properly set questions and peer discussion combined with the lecturer's explanation can enhance student learning.

Wit (2003) implemented a 50:50 polling technique (like "who wants to be a millionaire?") in a statistics tutorial class. He posed a clicker question to the students with a certain concept and after the first voting round, removed all but two possible answers. He discussed why the eliminated answers were wrong and students voted again and then the second explanation would emphasise the correct answer to the question on the concept under discussion. For example, Wit (2003) posed a question in Figure 2 on the “assumption checks in a regression problem” to his students. After the first voting round he eliminated all the answers except number 2 and 4 (students frequently confuse linearity with normality of residuals), discussed the reason for the eliminated answers and let the students vote again, between normal and not very normal. He finally explained why the data are not very normal. Hence, clickers can effectively be used to explain difficult or confusing concepts.

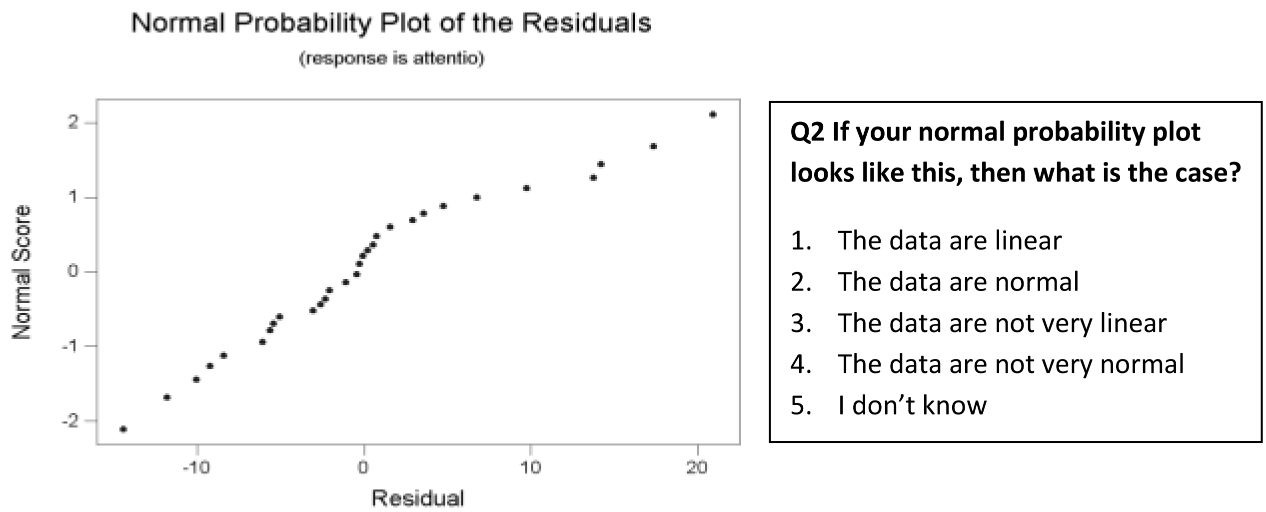


Figure 2: A clicker question on the normal probability plot of the residuals
Source: Wit (2003)

Kaplan (2008) used three types of clicker questions in a case study for a large statistics class, namely questions on common misunderstandings, review questions for revisiting topics, and class activity questions following from simulations to help students learn a concept. Students usually find the fundamental statistical concepts significant to inferential statistics challenging to understand. Kaplan (2011, p. 4) used simulations on a calculator and then students submitted the results via basic clickers that could only enter multiple choice options, called i-clickers, in a large lecture statistics class to improve the “conceptual understanding of statistical inference”. Twelve simulation activities are demonstrated by Kaplan (2011) in the study.

Studies on the use of clickers in flipped classrooms are scarce. Only two clicker studies could be traced, one in an English and the other in a statistics flipped classroom (Hung, 2017; McGee et al., 2016).

Hung (2017) used multiple choice questions in the form of *Kahoot!* quizzes in his gamified use of smart devices (clickers) called BYOD (bring your own device) in a flipped classroom. *Kahoot!* is a cloud-based automated response system where students can use any smart device instead of clicker devices. *Kahoot!* quizzes had a beneficial effect on student learning, i.e. the Kahoot group who used smart devices outperformed the non-Kahoot group who used PowerPoint slides and the raising of hands in the post-test. A perception survey was conducted to find out how both groups experienced the flipped learning. The results showed that the Kahoot group was consistently more positive than the non-Kahoot group.

McGee et al. (2016) used clickers for formative assessment and immediate feedback in a large flipped statistics classroom with JITT (Just-In-Time Teaching). Apart from watching videos, reading a chapter before class or active learning in class, students had to complete a short set of online questions before attending the next class. Based on the online answers, the instructor could immediately address and correct misunderstandings in the next class.

It can be argued that the use of clickers can have a positive influence on achievement. Several studies reported on the influence of clickers on performance and how the experimental group (with clickers) outperformed the control group (without clickers) on the final examination scores (Gauci et al., 2009; Kyei-Blankson, 2009; Majerich et al., 2011; Mayer et al., 2009).

Clickers do not always increase undergraduate statistics students' final grades, but can be used for large class engagement (Wilde, 2014). Forster (2014) found that compulsory assessments that count as little as 10% of the final mark, not only improved class attendance and in return engaged students in learning, but reduced the many student queries after class. Students started to enjoy statistics and had fun.

Instructors have varying conceptions of the difference between formative and summative assessment. According to Kay and LeSage (2009), formative assessment consists of questions and tests without grades, while summative assessment consists of formal tests with grades. Han and Finkelstein (2013) describe the use of Clicker Assessment and Feedback (CAF) for formative assessment as a measure of understanding course concepts without grades. Summative assessment is similar to formative assessment, but students' answers are linked to grades. Han and Finkelstein (2013) report on a large CAF supporting project that was introduced to improve the students' engagement in undergraduate modules. The 74 instructors who participated in the project were categorised as either formative or summative CAF users. Instructors' use of CAF for formative assessment is found to be more effective than summative assessment based on students' opinions of engagement and learning.

For this study the instructors' use of CAF for formative and summative assessment is equally important. Both types of assessments are graded, but formative assessment contributes a small percentage towards the semester mark. Formative assessment assists in active learning and revisiting of course concepts in class. In the past, summative assessment was done by multiple choice papers, but with QT-clickers, partial grade crediting became possible, to the advantage of the students.

Hancock (2010) discarded all paper tests and used PowerPoint polling where clicker questions are integrated into the PowerPoint lecture or tests for both formative and summative assessment. Tests consisted of a series of PowerPoint slides with Multiple Choice Questions (MCQs). Each slide is time-limited for all students, depending on the type of problem. The limitation is that students cannot go back to a specific question on a previous slide. Wang et al. (2014) also used a set of MCQs for summative assessment at the end of each unit. A compatible clicker software program, Exam View, was used to build questions of different difficulty levels for the summative assessment tests. Premkumar (2016) used self-paced polling of TurningPoint® technology for a complete MCQ paper, substituting the scannable answer sheet for clicker answers.

No articles could be found on using self-paced polling and written papers where students could have the benefit of revising and changing answers on a clicker, by typing in answers instead of simply selecting an MCQ option, and be advantaged by partial grade crediting. In this study the novel intervention of self-paced polling using QT-clickers and written papers is used, with the benefit of partial grade crediting.

An added benefit of the self-paced polling testing versus PowerPoint polling both using TurningPoint® technology is that tests and examinations can be taken at a student's own pace, i.e. answers can be reviewed and changed as the student works through the paper. Students can submit answers after each question or return to previous questions and submit only after completion of the paper. The only limitation is the three-hour cap on the paper as a whole.

Premkumar (2016) also used self-paced polling but did not change the MCQ paper to a written paper.

In summary, cognitive- and social constructivism form the underpinning of the technology interventions discussed. The value of online homework is not to replace pen-and-paper homework, but the positioning of online homework as preparation tool to interact with prior knowledge in a flipped learning environment. The advantages of the traditional clickers are discussed and its usage in an active learning class. More sophisticated response systems in the form of smart phones and tablets can be used as substitutes for clickers. The use of clickers in a flipped classroom setting and/or clickers for summative assessment are explained.

CHAPTER 3
RESEARCH DESIGN AND METHODOLOGY

3.1 INTRODUCTION

The teaching model of the Introductory Business Statistics course under discussion was restructured repeatedly because of low pass rates and was informed by recommendations of the GAISE report by the American Statistical Association (2005, revised 2016). The learning differentiation between the eight independent cohorts is shown in Figure 3. The researcher was involved in the actual teaching and implementation of specific teaching methods being compared, and was also involved in principled planning and shaping of the interventions from 2011 until 2018 and onwards. The traditional classroom model was still followed in 2011. In 2012, technology was used in the form of an online homework system (Aplia™, hereafter referred to as Aplia) with special focus on post-class online assignments. In 2013, the mode of delivery was changed by incorporating the online homework system into a flipped classroom model. The rationale behind the 2013 intervention was to change the mode of delivery to a constructivist classroom, while using the online homework system to assist with graded pre-class assignments of initial course content through pre-reading of the textbook and class notes. These strategies were informed by the success of the interventions as reported in the literature (e.g. Patron & Smith, 2009; Strayer, 2012). This study focused on the impact of the online homework system embedded within the flipped classroom. In this study, flipped classroom and flipped classroom embedded with online homework had been used interchangeably.

The status quo was maintained in 2014. QT-clickers have been introduced in 2015 and used since then to allow for active and cooperative learning in class. Students formed informal groups in problem-based lectures to collaborate with peers on certain clicker questions. A distinguishing feature of this study was that QT-clickers were used not only for formative assessment, but it was extended to summative assessment. In 2016 the pedagogy remained in place, but an extra tutorial session was implemented in addition to the three normal lectures, to spend more time on the application of difficult concepts. MindTap™ (hereafter MindTap), a cloud-based system, was embarked on in 2017, which incorporates the e-textbook, learning material, quizzes, videos and Excel assignments. The applications called CNow™ (hereafter CNow) and Aplia in MindTap are used for online pre- and post-class assignments. In 2018, apart from more attention to cooperative, peer learning in class, we decided to choose problem-based applications for class discussion that are mostly located within specific South African-based social, economic, and environmental contexts, e.g. the water crisis in the Western Cape[3] and Listeria[4]. The teaching model remained the same as in 2017, but we incorporated the switching of lecturers amongst the five groups.

This is a retrospective study. The cohorts consist of eight consecutive intakes of first-year statistics students from 2011 to 2018. The term cohort is used as proxy for the different technology interventions and the five intervention cohorts imply the initial introduction of the technology interventions in the five cohorts, i.e., 2011, 2012, 2013, 2015 and 2018. Finer detail regarding the eight cohorts is shown in Figure 3 and is further discussed in Section 3.2. The colours in Figure 3 denote initial changes to the pedagogy, classes, or homework system for a specific cohort.

[3] A period of severe water shortage experienced in 2017 in the Western Cape
[4] An outbreak of Listeria monocytogenes food poisoning in 2017 – 2018 in South Africa

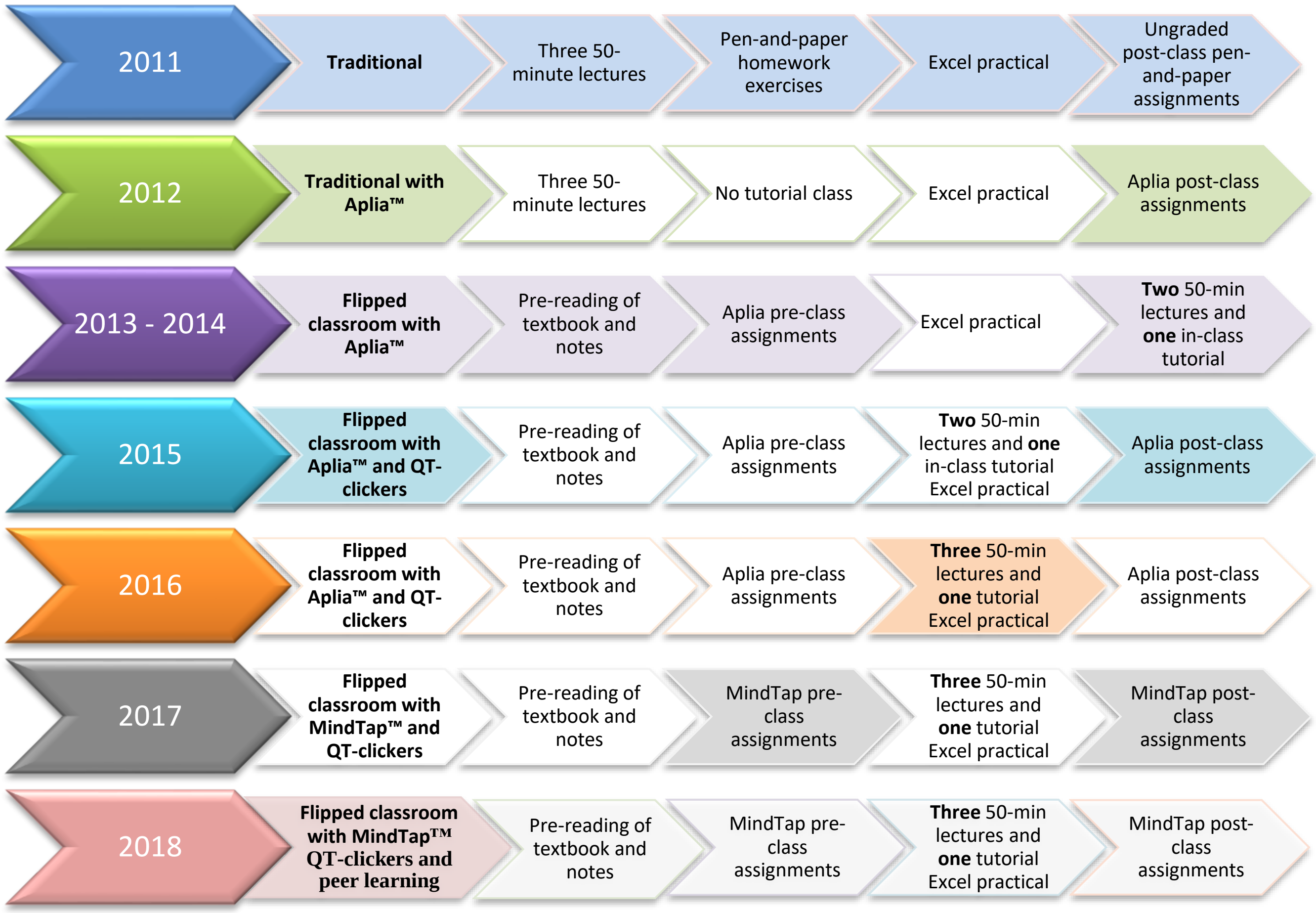


Figure 3: Learning differentiation between eight independent cohorts

3.2 RESEARCH INTERVENTIONS

2011 to 2012: Traditional lecture teaching strategies

A traditional teaching method was used where the lecturer provided the students with knowledge via PowerPoint slides and/or displayed transparencies on an overhead projector. Students were provided with incomplete versions of lecture slides on the Blackboard learn™ system, which they could complete during the lectures. An assignment book was available, containing post-class homework assignments including solutions, and it was the students' responsibility to complete the assignments after each chapter. A tutoring system was available for students who needed extra help, conducted by tutors who assisted students one-on-one on a first-come-first-served basis.

2012: Traditional model and online homework system (Aplia). In addition to the traditional lecture teaching strategies described above, the Aplia online homework system was introduced. Aplia is an online interactive homework system with immediate feedback. It is a platform of Cengage Learning, based on the current Introductory Business Statistics textbook (Aplia, 2012). Cengage Learning claims that "Aplia assignments connect statistics concepts to the real world". Aplia was implemented in 2012 as 14 post-class assignments, i.e. one per week, which counted towards the semester mark. Aplia was introduced to help students engage more with the many new and often difficult concepts of statistics. Each Aplia assignment is problem-based with realistic applications. The questions consist of a combination of MCQs, numerical calculations, missing words, drawing of graphs and Excel applications. Once a student has completed an Aplia assignment, it is graded, and immediate automated feedback is provided as step-by-step detailed scores with explanations of the correct formulae and answers. A similar assignment based on the same concepts with different questions, can be generated from the database. Up to three attempts are allowed. The average of the marks obtained for each attempt is entered as the final assignment mark.

2013 to 2014: Flipped classroom with embedded online homework system

The Aplia online homework system was incorporated into a flipped classroom model (Mazur, 2009) by using it for pre-class preparation. To encourage students to prepare for class, the assignment marks contributed to their semester marks. Our flipped classroom model consists of:

- Out-of-class preparation:
  - Prescribed pre-reading of relevant parts of the textbook and basic class notes.
  - Pre-class Aplia online homework assignments.
- In-class active learning:
  - Two 50-minute lectures.
  - A weekly tutorial in class.
- After-class consolidation:
  - A redesigned tutoring system.

The pre-reading of the textbook, basic class notes, and the pre-class Aplia online homework assignments go together. This means the students are expected to do the initial learning by themselves and then come to class prepared and ready to engage in activities that foster understanding. In class, the lecturer presents more extensive class notes and revisits complicated concepts using new examples. A weekly tutorial is implemented, where statistical concepts can be reinforced. Students have to attempt the tutorial exercises at home so that difficult concepts and misunderstandings regarding exercises can be addressed by the lecturer in class.

The one-on-one tutoring system used in the past for the traditional teaching method was substituted by a more student-accommodating system, managed by the tutors. The purpose of the redesigned tutoring system was to be an alternative option for students who, despite the flipped classroom, were still struggling. The tutoring system comprised three components, namely revision classes (groups), hot-spot sessions (one or a few students), and under-50 classes (students-at-risk), which could be attended on a daily or weekly basis, depending on the timetable.

The revision classes were in the form of group classes where lecturers explained difficult concepts and revised previous tests and tutorials and students could re-do similar exercises with the assistance of tutors. These were the most popular of the three components.

The hot-spot sessions were special one-on-one or one-on-a-few sessions, where specific problem questions were addressed and clarified. The hot-spot sessions were also well supported.

The under-50 classes, which targeted the most vulnerable group of students, were not well attended. A mid-term progress mark was calculated to identify students at risk, and they have been divided into the respective under-50 classes to help them to improve their marks. In the end, because of the poor attendance, the under-50 classes were offered to any student who needed help with the subject content.

2015 to 2018: QT-clickers and flipped classroom with embedded online homework system and peer learning activities in 2018

In 2015, the classroom response system most suitable to our needs, the QT-clicker of Turning Technologies, was implemented. A more standard alternative is the Response Card RF LCD of Turning Technologies that could also have been used, but its functionalities are very limited. The QT-clicker differs from the standard clicker in that it has a full-featured keypad that allows more complicated data to be entered, as opposed to the standard MCQ options. It is widely known that multiple choice options may lead to guessing or clever elimination of certain distractors for a question (McKenna, 2019). With QT-clickers, instead of providing students with a choice of answers, one (or more) of which is correct, students have to calculate answers which they submit via the keypad. It enables a wide variety of question types, from multiple choice (Question 2), true/false and numerical answers, which can be marked in a range (Question 8) to words (Question 1), sentences and essays, as can be seen in Figure 4. Students' answers can be tracked as each device has a unique serial number (Clicker ID in column B) allocated to an individual student. QT-clicker devices are registered on the Blackboard learning management system and the QT-clicker serial number is then linked to a unique student number, which prevent students from borrowing a peer's clicker.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Student	Clicker ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
2			RATIO; RATIO SCALE	D	10	7	25,50; 25, 50	23.81, [23.8, 23.82]	8	0.936, [0.935, 0.94]	4.5866, [4.586, 4.587]	QUANTITATIVE; CONTINUOUS; CONTINOUS
3	1	881C22	Ratio	D	10	7	25, 50	23.81	8	0.936	4.5866	Quantitative
4	2	8EBC02	Ratio	D	10	7	25,50	23.81	8	0.936	4.3866	Quantitative
5	3	8CE15E	Nominal	D	7	7	0.25 ,0.5	23.81	8	0.877	4.5667	Qualitative
6	4	8EBB9C	Nominal	C	10	7	25, 50	23.81	8	0.936	4.5866	Ordinal
7	5	87FF21	Ratio	C	10	7	25,50	23.42	8	-0.877	-0.3678	Quantitative
8	6	8835D4	Nominal	D	7	7	25,50	23.81	8	0.936	4.5866	Ratio
9	7	8815B3	Numerical	D	7	7	25 and 50	23.81	8	0.936	4.5866	Quantitative
10	8	8EC444	Interval	D	7	7	25 and 50	-	8	-	4.5866	Quantitive
11	9	8E52BA	Nominal	D	7	7	25,50	1	8	14	5.4713	ratio
12	10	8EBBBE	Ratio	C	10	7	25,50	23.81	8	0.468	4.5866	Ratio
13	11	8EC18F	Nominal	C	7	7	25	23.81	8	0.44	4.38	Catagorical
14	12	8552EB	Ratio	D	10	7	25, 50	23.81	8	0.936	4.5866	Quantitative
15	13	880D18	Ratio	D	7	7	25,50	33.33	15.5	0.003	0.0025	Qualitative
16	14	855834	Ratio	D	10	7	25,50	23.81	8	0.936	4.5866	Quantitative
17	15	8D3307	Ratio	D	7	7	33,57	14.29	8	1.753	4.5866	Nominal
18	16	C72EB1	Ratio	D	10	7	25, 50	35	8	0.877	4.5678	Quantitative
19	17	8CE0D6	Nominal	D	10	7	25, 50	1	8	0.876	3.8322	Cross-sectional
20	18	C69921	Nominal	D	10	7	25 and 50	23.81	8	0.936	4.5866	Quantitative

Figure 4: Extract of students' original examination answers on the Excel clicker report

Approximately 5% of the students in the 2015 - 2018 cohorts could not afford to purchase a clicker, but they were not penalised, since their class tests and semester tests were graded by hand and the marks manually captured in an Excel spreadsheet. They could actively participate in class with formative assessment questions and, when needed, handed their answers in on paper. It is not an ideal situation because of the loss of the anonymity advantage of QT-clickers.

In the 14-week semester of the period 2015 to 2018, the flipped classroom model consisted of three segments:

- Out-of-class preparation:
  - Prescribed pre-reading of relevant textbook sections and classnotes.
  - Pre-class Aplia or CNow online homework assignment with automated grading and immediate feedback.
- In-class active learning:
  - Problem-based real world exercises.
  - Clicker questions using TurningPoint® technology dispersed through the weekly three[5] 50 minute periods.
  - One 50-minute tutorial per week including clicker questions or a short clicker test using TurningPoint® technology.
  - One optional 50-minute Excel practical session in the laboratorium.
- After-class consolidation:
  - Post-class Aplia online assignment based on real world scenarios.
  - A revised tutoring system.

The flipped classroom pedagogy requires students to come to class prepared, where difficult concepts can be revisited and where clicker questions are used in class to test the students' understanding. TurningPoint® technology is used to integrate the clicker questions into the PowerPoint lecture. For some of the clicker questions in class, depending on the type, students have to submit an individual answer, indicated by a deleted group-symbol; if they are allowed to discuss with their peers, the group-symbol is used.

The immediate feedback after polling makes it possible for the lecturer to assess the overall level of understanding, while students know what they do not understand and they can compare themselves against their peers.

The tutorial classes are based on a worksheet posted on the Blackboard learning management system for students to prepare for the weekly tutorial. Students have to submit their worksheet answers via QT-clickers in the tutorial class. The lecturer can then give immediate feedback on misunderstood concepts or calculation errors. Some students simply copy worksheet answers from their peers and for this reason unannounced short clicker tests based on the worksheet are given in the tutorial class.

[5] In 2015 two 50 minute periods were used

The revised tutoring system is based on the redesigned tutoring system used from 2013 to 2014 without the under-50 classes and the hot-spot sessions. The revision group classes were most popular and well attended by the students, although some of the students prefer to see a tutor or lecturer one-on-one.

In 2018 it was decided to customise the in-class activities to the South African context. A few examples topics are SA's got talent, the Comrades marathon, social media, and the use of the Gautrain. In terms of rethinking and re-evaluating the ways in which we learn, teach and assess within the discipline of introductory statistics, data on a topic identified as of interest to our students, namely social media, have been collected from the current cohort themselves. This attempts to relate "home literacy" to the curriculum. Social media serves as a common theme for formal formative as well as summative assessment activities for the module. Student data as well as other resources are incorporated.

While using QT-clickers for individual as well as informal group work in class, we have decided in 2018 to give more attention to peer learning. Peer group tests using scratch cards were implemented during tutorial sessions. A scratch card consists of a number of rows denoting the different questions. Each row has four options or boxes covered in opaque coating and the box with a small star specifies the correct answer. The value of these cards are twofold; firstly the discussion of the group of students to get to an agreement before selecting the first box to scratch, and the instantaneous, educative feedback. The QT-clicker also made it possible to record students' attendance in normal lectures as well as tutor sessions. A final mark allocated to attendance formed part of the semester mark. We also implemented lecturer-switching amongst the five groups, to see if it has a positive effect on students' learning and performance.

3.3 RESEARCH DESIGN

The main goal of the study was to determine how effective the four technology-based interventions are in terms of student performance and according to the students' perception.

The initial purpose of the study, before the use of QT-clickers, was to measure the impact of the flipped classroom versus the traditional classroom model and the role that online homework plays in both models. QT-clickers have been used since 2015 and the aim was to investigate whether the use of QT-clickers for active learning extended to summative assessment could enhance student learning. In 2018, peer learning activities with localised content were introduced as part of active learning in class.

A mixed methods design was used to present a complete picture of the effectiveness of the four technology-based interventions by integrating the quantitative and qualitative analyses.

3.4 RESEARCH METHODOLOGY

3.4.1 PARTICIPANTS AND SAMPLING

Participants were mainly commerce (BCom) students, followed by science (BSc) students, while a small percentage of the remaining students studied consumer science, town and regional planning, social sciences and arts. The students were registered for the first semester module called Introductory Business Statistics (STK110) at UP in South Africa for the period 2011 to 2018. The eight cohorts used the same textbook, course outline and content. A group of four to five lecturers taught STK110 in the first semester, and typically at least three of the lecturers remained the same on an annual basis. The first four weeks of each academic year were used to revisit basic statistics and probability concepts taught in Grade 12 mathematics to ensure that the pre-intervention understanding of statistics concepts was similar for all cohorts.

This was an observational study and it was the aim to limit possible initial bias. The cohorts in the period 2011 to 2018 had certain prerequisites for prospective first-year students. Specific Advancement Placement Scores (AP Scores) based on the six best Grade 12 marks excluding life orientation are allocated to each study programme offered by the university. Although the Grade 12 mathemetics prerequisite of at least 60% was required from 2012 onwards, the system still allowed certain students with a Grade 12 mathematics mark below 60%, because of the specific programme requirements.

In order to make the consecutive cohorts comparable, it was decided to divide the students in each cohort into two samples. The sample of students satisfying all prerequisites was called the prerequisite sample, consisting of only new students with at least 60% for Grade 12 mathematics and an AP Score of at least 26. The sample with the remaining students who did not satisfy all the above-mentioned prerequisites was called the non-prerequisite sample. It must be noted that repeat students, who may have an advantage based on their prior exposure to the module, formed part of the non-prerequisite sample. The non-prerequisite sample was further divided into new and repeat students for use in certain statistical analyses. The repeat students were of great concern to us since the year 2000, and it would be reassuring if they have benefited from some of the interventions. All samples of students used for statistical analyses consisted of students who gained admission to the examination, that is a semester mark of at least 30%.

As mentioned in Section 3.1, the terminology “five intervention cohorts” denotes the five years when the technology interventions were initially introduced and will be referred to as such throughout the thesis.

3.4.2 DATA COLLECTION PROCEDURE

The eight cohorts under consideration can be separated into two periods, namely before QT-clickers (2011 to 2014) and after QT-clickers (2015 to 2018), which also had an influence on the form of assessment. Formative assessment in our study was continuous assessment which involved individual or group clicker questions in class, which formed part of PowerPoint polling slides. Formative assessment measured students’ understanding of concepts, and combined with the advantages of peer learning and immediate feedback, it renders an improved learning model. These formative assessment tasks contribute a small percentage to students’ grades. Formative assessment is also attained in the form of online homework assignments which constitute a larger part of the semester mark. Summative assessment is the formal, graded tests or examination papers, written in a controlled environment without peer interaction.

All tests were based on a subset of the syllabus while the examination paper was based on the entire syllabus. The semester mark included all formative assessment and summative assessment tasks (test marks), but excluded the examination mark. Students with a semester mark of at least 30% were admitted to the examination. The final mark was the maximum of a mark based on a ratio of 50:50 or a ratio of 40:60 between the semester mark and the examination mark. A final mark of at least 50% is required to pass the module and a supplementary examination is awarded for a final module mark of 40% to 49%.

The examination paper for every cohort was set by a team of examiners and tested the same learning outcomes, but the questions differed from year to year. Face validity was established by an internal panel that evaluated the final examination papers. In addition, an item analysis was performed post examination for each of the cohorts' examination papers to verify the compatibility and suitability of the individual questions.

The eight successive years from 2011 to 2018 were divided into two time periods regarding clicker use. Clicker use had an immense influence on the type of assessment.

Period: 2011 to 2014

Formative assessment

In 2011 no formative assessment was used. As from 2012, the Aplia online homework assignments contributed a small percentage (±16%) to formative assessment.

Summative assessment

Class tests, semester tests, practical tests (Excel) and the examination formed approximately 84% of summative assessment. Apart from written and hand-graded class tests, all other grading was done by the sole use of MCQs due to large classes of between 300 to 600 students. The written class tests resulted in delayed feedback because of human resource capacity constraints and it is hypothesised that this could have impacted negatively on student learning. The examination written in nine venues on campus consisted of approximately 64 MCQ questions with five distractors each. Each question counted one or two marks. The distractors focused on conceptual errors that students typically make. Students could do their calculations on the counter page, but it was not handed in for grading purposes. The three hours scheduled for the examination paper included the complesion of a scan sheet used to capture the answers for the MCQ papers. Optical mark recognition was used to read the scan sheets and an Excel file with the students' options was provided for further analysis.

Period: 2015 to 2018

Formative assessment

QT-clickers were used for all the cohorts during this period. Formative assessment was extended to clicker questions in class, clicker attendance, clicker questions or short tests in tutorials, online assignments using Aplia and CNow in MindTap™. Since 2015, the contribution of formative assessment to the semester mark increased to between 25% and 30%.

Summative assessment

The summative assessment section of the semester mark decreased to approximately 70%. Apart from practical tests, all class tests, semester tests and the examination were clicker-based assessment. The examination paper was more comprehensive and had approximately 55 free responses and a few MCQ questions, counting one or two marks each. Students had 2.5 hours to write the paper and 30 minutes to submit their answers on their QT-clickers in nine different venues on campus.

All students received a hard copy of the test or examination paper, which had to be turned in, showing their detailed calculations. The students' hard copies were only graded by hand if a power failure or any other technical malfunction occurred. The responses were submitted on their QT-clickers using self-paced polling, which were sent to a wireless receiver (dongle), collected and then sent to the lecturer's device. An Excel file was exported for the calculation of the final assessment marks. The few non-clicker papers of students who did not have QT-clickers were graded by hand and captured in an Excel file. The process of partial grade crediting could commenced by viewing the Excel sheet with all the responses. The answers to each question, located in a column in the spreadsheet, could then be sorted individually to identify common mistakes. For example, if the normal probability of a variable should be calculated and a student managed to get the correct z-value, but read the wrong probability from the normal table, then intermediate marks (partial credit) could be awarded.

Figure 5 shows an extract of twelve students' answers for Questions 8 to 14 of the June 2018 examination paper. The answers for Questions 1 to 7 were hidden, due to limited space. It should be noted that students were not penalised for incorrect spelling which could be amended afterwards, and some answers were marked in an interval due to decimal places. The answers for each question, located in a column in the spreadsheet, could then be sorted and filtered individually for partial grade crediting. A red cell denotes a wrong answer and a green cell a correct answer. Partial grades (half marks), denoted by orange cells, were given for Question 14 if the wrong sign was chosen or the correct sample proportion, but incorrect normal table values were used, i.e., $z_{0.01}$ one-sided value in stead of $z_{0.005}$ two-sided value. An example of a mistake that was carried and full marks awarded was Question 13 in Figure 5. For Question 13, the students had to calculate the value of the sample proportion and for Question 14 they had to use the calculated proportion to find the value of the margin of error for a 99% confidence interval, i.e. a wrong answer could be carried from Question 13 to Question 14. If the sample proportion was incorrect, they lost one mark, but if students used the incorrect value and did correct calculations for Question 14, they were awarded full marks, denoted by blue cells in an extra column created to the right of Question 14 (Q14C) in Figure 5. Partial grade crediting is an advantage of written papers with self-paced polling and a benefit to summative assessment, which was not possible for solely MCQ-based papers.

	A	B	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	Student	Clicker ID	Q8	Q8C	Q9	Q9C	Q10	Q10C	Q11	Q11C	Q12	Q12C	Q13	Q14	Q14C
2	1	856369	0.025		4.568	1	Quantitative		5.5		51	1	0.26	0.1597	
3	2	8E3943	0.606		4.4867		Qualitative		2.5		5.1		0.26	0.1598	
4	3	8E9437	0.936		4.5866		Quantitative		2.5		5.1		0.26	0.165	
5	4	8E52D0	0.936		4.5866		Quantitative		2.9		5.1		0.26	-0.1597	1
6	5	856DFE	1		4.6	2	Interval		25	0.5	35		0.12	0.1184	2
7	6	88423C	4		8.8		Quantitive	1	2.5		5.5		0.28	0.1116	
8	7	8CDFBA	1,248		4,7712		Quantitative		2,5		5,1		0,26	0,1598	
9	8	8E5204	0.877	0.5	4.5678	1	ratio		2.5		5.5		0.26	0.1597	
10	9	C71132	0.877	0.5	4.5678	1	Cross-sectional		25	0.5	45		0.71	9.9803	
11	10	854F67	0.025		4.7747		Quantitative		25	0.5	5.7		0.069	9.9609	
12	11	C7119E	0,228		4,3163		Ratio		2,5		5,1		6	0,7746	
13	12	8D0AE3	0,936		4,5866		Ratio		25	0.5	51	1	0,57	0,1803	2

Figure 5: An extract of partial graded answers of Questions 8 to 14 for twelve students

After partial grading was done for the whole paper, a final examination mark was calculated. The final module marks were calculated, linking the examination marks to the semester marks (Figure 6). In column BN of Figure 6, the final mark consists of the maximum of the ratios 50:50 and 40:60, used for the semester and examination marks respectively.

	A	BJ	BK	BL	BM	BN
1	Number	Semester Mark	Examination Mark	Ratio 50:50	Ratio 40:60	Final Mark
2	1	72	69.38	70.69	70.43	71
3	2	76	84.38	80.19	81.03	81
4	3	44	65	54.50	56.60	57
5	4	33	40.63	36.82	37.58	38
6	5	70	73.13	71.57	71.88	72
7	6	65	60.63	62.82	62.38	63
8	7	66	55.63	60.82	59.78	61
9	8	41	57.5	49.25	50.90	51
10	9	55	41.25	48.13	46.75	48
11	10	40	53.75	46.88	48.25	48
12	11	71	75.63	73.32	73.78	74
13	12	39	33.75	36.38	35.85	36
14	13	67	83.75	75.38	77.05	77
15	14	67	44.38	55.69	53.43	56
16	15	83	88.75	85.88	86.45	86

Figure 6: Contribution of the semester mark and examination mark to the final mark

The final module marks were sorted in ascending order and the borderline cases were colour-coded (Figure 7). The borderline cases were isolated and linked to the latest class list of the module for the students' personal details and to be able to search for the students' examination scripts. All borderline cases' scripts were remarked (hand-graded), i.e., 36% to 39% re-marked for possible entry to the supplementary examination, 46% to 49% re-marked for a pass, and 72% to 74% re-evaluated for distinctions.

	A	BJ	BK	BL	BM	BN
1	Number	Semester Mark	Examination Mark	Ratio 50:50	Ratio 40:60	Final Mark
2	943	46	25	35.50	33.40	36
3	24	46	28.13	37.07	35.28	37
4	86	30	43.75	36.88	38.25	38
5	166	52	25.63	38.82	36.18	39
6	587	60	32.5	46.25	43.50	46
7	387	58	36.88	47.44	45.33	47
8	10	40	53.75	46.88	48.25	48
9	887	39	56.25	47.63	49.35	49
10	896	66	75.63	70.82	71.78	72
11	274	80	65	72.50	71.00	73
12	25	62	81.25	71.63	73.55	74

Figure 7: Extract of borderline cases colour coded for re-marking

3.4.3 DATA ANALYSIS

A mixed methods approach was used to obtain a comprehensive picture of the effectiveness of the technology interventions. Both quantitative and qualitative analyses were done which will be discussed in the next sections. All statistical analyses were conducted by the researcher herself.

3.4.3.1 Quantitative analysis

The quantitative analysis comprises descriptive and inferential statistics. Descriptive statistics consists of summaries, tables, and graphs to give an overview of the important features of the eight cohorts. The analyses were performed using IBM SPSS Statistics 26 and R-studio. The conventional 5% level of significance was used, specified for all the statistical tests. Since the samples were very large, resulting in tests with too much power, the effect sizes are also reported.

The inferential statistics for Section 4.2.1 (Impact of the different technology interventions) are done for the following samples, namely prerequisite, non-prerequisite, and non-prerequisite divided into new and repeat students. It includes an ANOVA (Analysis of Variance) test with a single factor and chi-square tests for the five intervention cohorts. ANOVA is used to test for the equality of the population means (i.e. final mark means of the five intervention cohorts). Welch's $F-$test (a robust alternative of the $F-$test) was used because the homogeneity of the variance assumption was violated (Delacre et al., 2020). The Games-Howell post-hoc procedure was utilised for pairwise comparisons between selected cohorts. Games-Howell is accurate with unequal sample sizes (Field, 2009, p. 374).

The first four chi-square tests evaluated the equality of the population proportions (pass rates) for the five intervention cohorts. The Marascuilo multiple comparison procedure was used to test certain pairwise comparisons of all pairs of the population proportions (Anderson et al., 2020, pp. 537-540). To avoid an inflated Type I error, Bonferroni corrections were applied to the pairwise comparisons to obtain an overall level of significance of 0.05 (McDonald, 2014).

The next two chi-square tests assessed the relationship of the five selected cohorts with respect to their marks' distribution. For this chi-square analysis, the standardised residuals are also reported. The standardised residual can be roughly compared to a z-score for large samples. To facilitate interpretation of the interactions, the expected frequencies under the null hypothesis of no interaction are also included and compared to the observed counts.

In Section 4.2.2 (Impact of the QT-clickers), two chi-squared tests and McNemar's tests for paired data were conducted using only the 2014 and 2017 cohorts. The chi-squared tests measured the association between the examination marks for the 2014 cohorts compared to the 2017 cohort (with and without partial grading). The McNemar's test tested which grade intervals had different proportions between the examination marks for the 2017 cohort with and without partial grading.

Regression models using general linear models (GLM) were built to investigate the relationship between clicker use and examination performance. QT-clickers were already introduced in 2015, but it was decided to use the 2017 cohort as the intervention group, because of practical problems encountered with the implementation of QT-clickers in 2015[6] and 2016. The lecturers were unaware that the power saver option had to be disabled on the computers used for summative assessment in 2015. This led to uncaptured QT-clicker answers, and apart from concerned students and the majority of scripts graded by hand, an optional, compensational test opportunity was offered which disqualified the comparison of the 2015 to the 2014 cohort. In 2016 the political unrest on campus regarding Afrikaans as a language of tuition resulted in the 2016 cohort also not being credibly comparable to the 2014 cohort.

The intervention of using QT-clickers was evaluated along two lines: the pedagogical influence of the QT-clicker and the effect of partial grade crediting. It could be argued that a difference between the two cohorts may be attributable to the partial crediting implemented in the 2017 examination versus no partial crediting in the 2014 examination, and not to the use of the QT-clickers for formative and summative assessment per se. Therefore, the 2017 examination and semester test papers were re-graded in a multiple-choice fashion (i.e. without partial credit for intermediate steps) to allow for a comparison of the 2014 and 2017 summative papers. This allowed us to evaluate the impact of partial grade crediting on summative assessment by building two GLMs, namely the 2017 examination graded with and without partial credits, each against the 2014 MCQ examination. Several other GLM models were built to explore the

[6] A separate study was initiated to investigate the problem. Unfortunately the postgraduate student assigned to the study left the university without returning the questionnaires that she was tasked to capture.

role of the semester mark as the strongest predictor of the examination mark. We initially built simple models with only one predictor and finally a model with interaction terms to estimate the examination mark.

A final GLM was constructed to measure the effect of the technology interventions in conjunction with other predictors and interaction terms on the examination mark. A machine learning algorithm XGBoost (XGB) was used to construct a model from a different perspective, i.e., using an artificial neural network, to validate the results of the GLM model.

XGBoost is a supervised machine learning algorithm to build a regression tree that can be used as an alternative to or in combination with a statistical model such as GLM. In machine learning terminology, variables are referred to as features and will be used interchangeably. Gradient boosting is an essential machine learning method used in regression and classification problems (Niu et al., 2019). The model evolves using an iterative step-by-step approach, where in each step it introduces a new regression tree (new weak learner) to correct the residual errors in predictions made by the previous regression trees (Niu et al., 2019). Regression trees are added successively until no more improvements can be made. XGB (EXtreme Gradient Boosting) is an application of the gradient boosting method (Niu et al., 2019). XGB applies gradient descent for optimisation by improving the prediction accuracy at each optimisation step. This is done by moving iteratively towards the steepest direction to minimise a convex function. XGB minimises this standardised objective (Niu et al., 2019, p. 3):

$$L(\emptyset) = \sum_i l(y_i, \hat{y}_i) + \Omega(\emptyset) \tag{1}$$

"where $\emptyset$ is the learned parameter set, l is a differentiable convex loss function that measures the difference between the predictions $\hat{y}_i$ and the target y_i, and Ω is the regularisation term". XGBoost is used in Section 4.4.2 to build regression trees for the prerequisite sample and a comparison to GLM is done in Section 4.4.4, in the Comparison between GLM and XGB (prerequisite sample) section.

$$RMSE = \sqrt{mean \sum_i (y_i - \hat{y}_i)^2} \tag{2}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3}$$

For the XGBoost model, the root mean square error (RMSE) and R^2-value defined in Equations (2) and (3) respectively are selected to evaluate the XGB model performance. The evaluation results are reported in Section 4.4.2, the module results section. The objective is to create a model with a R^2 close to 1 and an RMSE that is as small as possible.

A recent method that helps to interpret the output of a complex model and that stood out amongst other machine learning techniques is Shapley Additive exPlanations (SHAP), which is developed by Lundberg and Lee (2017). SHAP evolved from game theory (Štrumbelj & Kononenko, 2014) and local explanations (Ribeiro et al., 2016), which provide a way to approximate the influence of each variable. With SHAP, the contribution of every variable (ϕ_i is the contribution of variable i) on the model's output $v(N)$ is assigned by focusing on their marginal contribution (Lundberg & Lee, 2017). The Shapley values are calculated by using Equation 4:

$$\phi_i = \sum_{S \subseteq N \{i\}} \frac{|S|!\,(n - |S| - 1!)}{n!} [V(S \cup \{i\}) - v(S)] \tag{4}$$

The SHAP(ϕ_i) value evaluates the importance of a variable in the model with and without the variable, in all possible orders that the variable can enter the model and taking into account the interaction with other variables (Lundberg & Lee, 2017). The SHAP interpretations can be seen in Section 4.4.3, in the Feature Analysis and Interpretation section.

3.4.3.2 Qualitative analysis

A qualitative component was added via surveys, including open-ended questions (Garfield & Ben-Zvi, 2008) and the use of focus groups to augment the quantitative analysis of the data. The qualitative analysis consists of a summary of the students' evaluation of the course. The perception of the value of the four interventions in the students' own words (students' voice) was polled.

The 2012 and 2013 cohorts were surveyed, and the results summarised in Table 25 to convey the students' voice regarding the Aplia online homework system and tutorials. The response rates were 86.1% for the 2012 cohort and 72.8% for the 2013 cohort.

It was also decided to compare the outcome of the 2017 and 2018 questionnaires with respect to the Aplia online homework system that was still being used in 2017 and the MindTap learning tool implemented in 2018.

The qualitative analysis component for QT-clickers consisted of ten survey questions to measure how students in the 2017 and 2018 cohorts perceived the use of QT-clickers for formative and summative assessment. The survey was administered during class in the last few weeks of the first semester of 2017 and 2018. The response rate was 79.5% for 2017 and 86.5% for the 2018 cohort. The tutoring system used in 2017 versus 2018 and the teaching approach were also surveyed. A copy of the questionnaires for the cohorts of 2012, 2013 to 2014, 2015 to 2016, 2017 and 2018 are available in Annexure 6.

CHAPTER 4
QUANTITATIVE DATA ANALYSIS AND INTERPRETATION

4.1 DESCRIPTIVE STATISTICS

This descriptive statistics section is divided into two; the first section provides a general discussion on the different cohorts and the second section discusses the selected variables.

4.1.1 COHORT DESCRIPTION

Tables and graphs are used to present and compare basic statistics of the eight independent cohorts from 2011 to 2018. Table 1 represents the information received for the years 2011 to 2018 from the Bureau for Institutional Research and Planning (BIRAP).

Table 1: Student information for STK110 from 2011 to 2018 created from BIRAP data

Year	Registered (A)	Dis-continued (% of A)	Registered for exam (B)	Without exam entrance (% of B)	Absent from exam (% of B)	Admitted to Supple-mentary exam (C) (% of B)	Supple-mentary exam passed (% of C)	Total passed (% of B)
2011	3318	275 (8.3)	3042 S1:2609 S2:433	256 (8.4)	57 (1.9)	677 (22.3)	244 (36.0)	2023 (**66.5**)
2012	2354	101 (4.3)	2253 S1:2039 S2:214	129 (5.7)	33 (1.5)	483 (21.4)	150 (31.1)	1589 (**70.5**)
2013	2244	77 (3.4)	2167 S1:2031 S2:136	72 (3.3)	31 (1.4)	275 (12.7)	49 (17.8)	1767 (**81.5**)
2014	1880	47 (2.5)	1833 S1:1780 WS:53	51 (2.8)	24 (1.3)	287 (15.7)	111 (38.7)	1485 (**81.0**)
2015	1945	126 (6.5)	1819 S1:1817 SE:2	111 (6.1)	36 (2.0)	189 (10.4)	132 (69.8)	1520 (**83.6**)
2016	1847	144 (7.8)	1703 S1:1649 WS:54	95 (5.6)	18 (1.1)	266 (15.6)	176 (66.2)	1354 (**79.5**)
2017	1703	57 (3.3)	1646	68 (4.1)	18 (1.1)	217 (13.2)	124 (57.1)	1344 (**81.7**)
2018	1851	96 (5.2)	1755	49 (2.8)	24 (1.4)	215 (12.3)	120 (55.8)	1496 (**85.2**)

S1 = Semester 1; S2 = Semester 2; WS = winter school; SE = Special exam

The student information in Table 1 includes both semesters for certain years (see column 4 in Table 1). Until 2013, STK110 was presented in the first semester and for unsuccessful students repeated in the second semester (called the anti-semester). Therefore, the students in the anti-semester formed a subset of the students in the first semester. As from 2014, a winter school substituted the anti-semester system, on the condition that at least 40 students registered for the winter school. From 2017, the winter school was phased out and STK110 was only presented in the first semester. It is notable from Table 1 that the number of registered students decreased significantly from 2011 to 2012. The reason for this could be the new mathematics prerequisite that was introduced in 2012. The number of students who discontinued in a specific year could have an influence on the number of repeat students for the next year. In 2016, the political unrest on campus regarding Afrikaans as a language of tuition, as well as the 2016 students being confronted with "fees must fall" strikes and riots on campus, influenced the education medium, quality and pass rate and resulted in a decline of approximately 8% in the student intake of 2017.

The STK110 examination information in Table 1 was reworked for the first semester only and is presented in Table 2 to differentiate between new and repeat students. According to the overall pass rates in Table 2, it is noticeable that the pass rate for STK110 in the first semester increased for the years 2011 to 2013, when calculated without the anti-semester marks. On the other hand, the pass rates decreased marginally in 2014 and 2016 without the contribution of the winter school marks, although the student numbers were below 60 in both winter schools.

Table 2: Reworked STK110 student information from 2011 to 2018 for the first semester

Year	**Registered for exam in first semester (B)**	**Total number of students (% of B)**		**No exam entrance* (% of B)**		**Absent for exam (% of B)**		**Failed exam** (% of B)**		**Passed exam (% of B)**		**Pass rate (% of B)**
		New	**Repeat**	**New**	**Repeat**	**New**	**Repeat**	**New**	**Repeat**	**New**	**Repeat**	
2011	2609	2123 (81.4)	486 (18.6)	155 (5.9)	62 (2.4)	57 (2.2)	18 (0.7)	440 (16.9)	121 (4.6)	1471 (56.4)	285 (10.9)	67.3
2012	2039	1704 (83.6)	335 (16.4)	75 (3.7)	44 (2.2)	15 (0.7)	9 (0.4)	306 (15.0)	101 (5.0)	1308 (64.1)	181 (8.9)	73.0
2013	2031	1741 (85.7)	290 (14.3)	40 (2.0)	25 (1.2)	16 (0.8)	14 (0.7)	189 (9.3)	66 (3.2)	1496 (73.7)	185 (9.1)	82.8
2014	1780	1643 (81.4)	137 (7.7)	34 (1.9)	17 (1.0)	14 (0.8)	10 (0.6)	228 (12.8)	43 (2.4)	1367 (76.8)	67 (3.8)	80.6
2015	1817	1658	159	87	24	25	11	124	28	1422	96	83.5

		(92.3)	(8.8)	(4.8)	(1.3)	(1.4)	(0.6)	(6.8)	(1.5)	(78.3)	(5.3)	
2016	1649	1466 (88.9)	183 (11.1)	57 (3.5)	38 (2.3)	12 (0.7)	4 (0.2)	182 (11.0)	51 (3.1)	1215 (73.7)	90 (5.5)	79.1
2017	1646	1466 (89.1)	180 (10.9)	39 (2.4)	29 (1.8)	11 (0.7)	8 (0.5)	182 (11.1)	34 (2.1)	1234 (75.0)	109 (6.6)	81.6
2018	1755	1590 (90.6)	165 (9.4)	37 (2.1)	12 (0.7)	15 (0.9)	3 (0.2)	171 (9.7)	21 (1.2)	1367 (77.9)	129 (7.4)	85.2

*Semester mark < 30% **Final mark < 50%

It is clear from Table 2 that there was a meaningful decrease in the percentage of new and repeat students who failed the examination in 2013, 2015 and 2018, which are the implementation of the flipped classroom, clickers, and peer learning activities interventions.

The descriptive statistics for the eight cohorts are based on students with exam entrance. The GLM models in Section 4.3 (Inferential Statistics: Statistical modelling (GLM)) are only based on the prerequisite sample of STK 110 students as shown in Table 3. A success is considered a final mark of attaining 50%, which is a pass mark at UP, South Africa.

Table 3: Descriptive statistics based on the prerequisite and non-prerequisite sample of students and with a final mark

Cohort	2011	2012	2013	2014	2015	2016	2017	2018
Prerequisite Total	1472	1346	1448	1455	1427	1274	1326	1475
Non-prerequisite Total	718	419	355	160	173	197	163	169
Prerequisite Final mark: $\bar{x}$ (s)	58.6 (14.1)	59.3 (13.8)	65.2 (14.6)	62.2 (14.9)	65.2 (15.2)	60.4 (14.7)	61.9 (15.0)	64.0 (15.3)
Non-prerequisite Final mark: $\bar{x}$ (s)	49.7 (10.3)	49.9 (9.3)	54.4 (10.5)	49.3 (9.7)	53.4 (12.0)	48.9 (12.4)	53.3 (12.6)	53.3 (9.7)
Prerequisite Pass rate	82.1	82.3	89.8	86.4	92.8	88.2	88.0	90.0
Non-prerequisite Pass rate	62.3	61.6	73.2	62.5	78.0	66.5	76.1	82.2

The clustered box plot in Figure 8 represents the final marks of the eight cohorts that distinguished between students satisfying all the prerequisites or not. Students without Grade 12 mathematics marks captured by the university were discarded. It is noticeable that for the

~~prerequisites~~ sample, the box plots of cohorts 2013, 2015 and 2018 look similar and without any outliers[7]. A link could be made to the introduction of the flipped classroom in 2013, clickers in 2015 and peer learning activities in 2018. The box plots of the non-prerequisite sample of 2015 and 2018 are comparable. For both the 2015 and 2018 cohorts, the middle 50% of the students lie approximately between 50% and 59%, but more inspiring is that at least 75% of the students had a final mark of above 50%. As mentioned with the prerequisite sample, it could be connected to clicker use in 2015 and peer learning activities in 2018.

[7] Outliers are denoted by circles and extreme outliers by an asterisk (*), i.e. a final mark exceeds the length of the box (interquartile range) more than three times

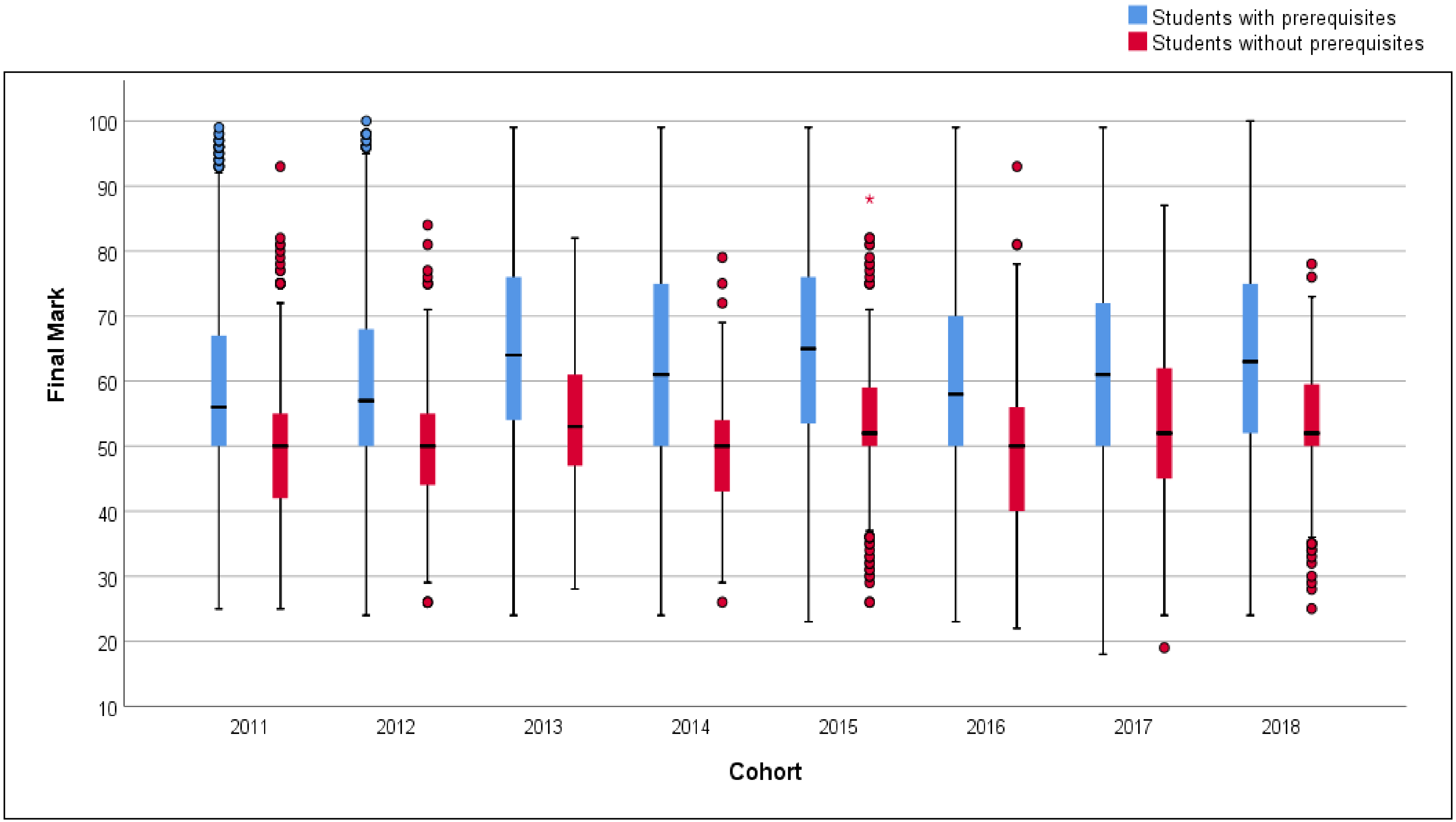


Figure 8: Clustered Box plot of final mark differentiated by prerequisite and non-prerequisite samples for the different cohorts

4.1.2 INTERVENTION AND VARIABLE DESCRIPTION

The interventions and selected variables based on their importance in statistical modelling (GLM) in Section 4.3 were extracted for each student and are summarised in Table 4 and discussed in this section.

Table 4: Descriptions of each intervention and variable in the different models

Variable in models	Intervention/ Variable	Description
Cohort	Online homework	First technology-based intervention introduced in 2012. Traditional teaching model used in combination with online homework. Online homework was a substitute for pen-and-paper homework. Students had three attempts and immediate valuable feedback available after each attempt.
Cohort	Flipped classroom	Second technology-based intervention implemented in 2013 with embedded online homework. Change in pedagogy which is still used for STK110 to this day.
Cohort or Clicker	QT-clickers	Third technology-based intervention introduced in 2015 and used as part of a flipped classroom. QT-clickers were initially meant for active learning in class but used more comprehensively for formative as well as summative assessment. (Clicker/No clicker)
Cohort	Peer learning activities	Fourth technology-based intervention in 2018 is an extension of the third intervention with problem-based, peer learning activities.
Mtongue	Mother tongue education	Mother tongue education is schooling in the language that children speak at home. Not everybody is so privileged to be taught at university in their mother tongue. It can have an impact on students' performance. (Mtongue/Non-Mtongue)
STK exposure	STK contact time	The STK110 contact time is divided into four categories according to the degree programme, i.e. number of months or years exposed to statistics tuition. It starts with six months (one semester) up to a maximum of three years for a basic bachelor's degree. (Six months/Nine to twelve months/Two years/Three years)
Residence	Stay in residence	Student accommodation at university. Approximately 32% of the STK110 students stay in residence. (In res/Not in res)
Repeat	Repeat students	Students who failed STK110 at least once.
NumTimes	Number of times in STK110	The number of times that a student repeated STK110 (1 to 6 times). One indicates the first time STK110 students, i.e. new student.
YrsReg	Years registered	The number of years students are registered at UP

AP Score	Gr 12 AP Score	A score based on the six best Grade 12 subjects, excluding life orientation.
Gender	Gender	Gender can be classified in more than two categories, but the data were obtained through BIRAP, which provided only binary gender categories, i.e. male or female, provided by the students themselves.
Race	Ethnicity	In the South African context Whites are often perceived as previously advantaged, therefore ethnicity is included in the model as a proxy for privilege. Students are divided into three ethnicity groups (African, White, other (Indian, mixed ethnicity)).

The interventions and variables that appear to have an influence on the final marks were considered with accompanying stacked bar graphs in Figures 9 to 16. For Figures 9 to 13, separate stacked bar graphs of the prerequisite and non-prerequisite samples are shown; for Figure 14 all students are combined in one graph; for Figure 15 only the non-prerequisite sample is used; and for Figure 16 only the repeat students are represented.

- **Flipped classroom**

  Flipped classroom model versus traditional model and/or online homework. Flipped classroom = 0 is traditional and/or online homework and Flipped classroom = 1 is only flipped classroom or a combination of flipped classroom with QT-clickers and/or peer learning activities. Figure 9 shows that the flipped classroom pedagogy improved student performance in terms of pass rate for both the prerequisite and non-prerequisite samples. For the prerequisite sample, the percentage students who failed decreased radically and several students moved into higher percentage categories. The non-prerequisite sample of students in the flipped classroom's marks increased and even moved into the 70% to 79% bracket.

- **QT-clicker**s

  QT-clicker use in a flipped classroom model with/without peer learning activities versus traditional model with/without online homework or flipped classroom only. Clicker = 0 is the traditional model with/without online homework or flipped classroom only and Clicker = 1 is QT-clicker used in a flipped classroom model with/without peer learning activities. It is evident from Figure 9 that using a QT-clicker has a positive influence on learning statistics. Fewer students failed using clickers compared to students not using them, moving more students into the higher percentage brackets, that is from 70% to 100% for the prerequisite sample and 60 to 80% for the non-prerequisite samples.

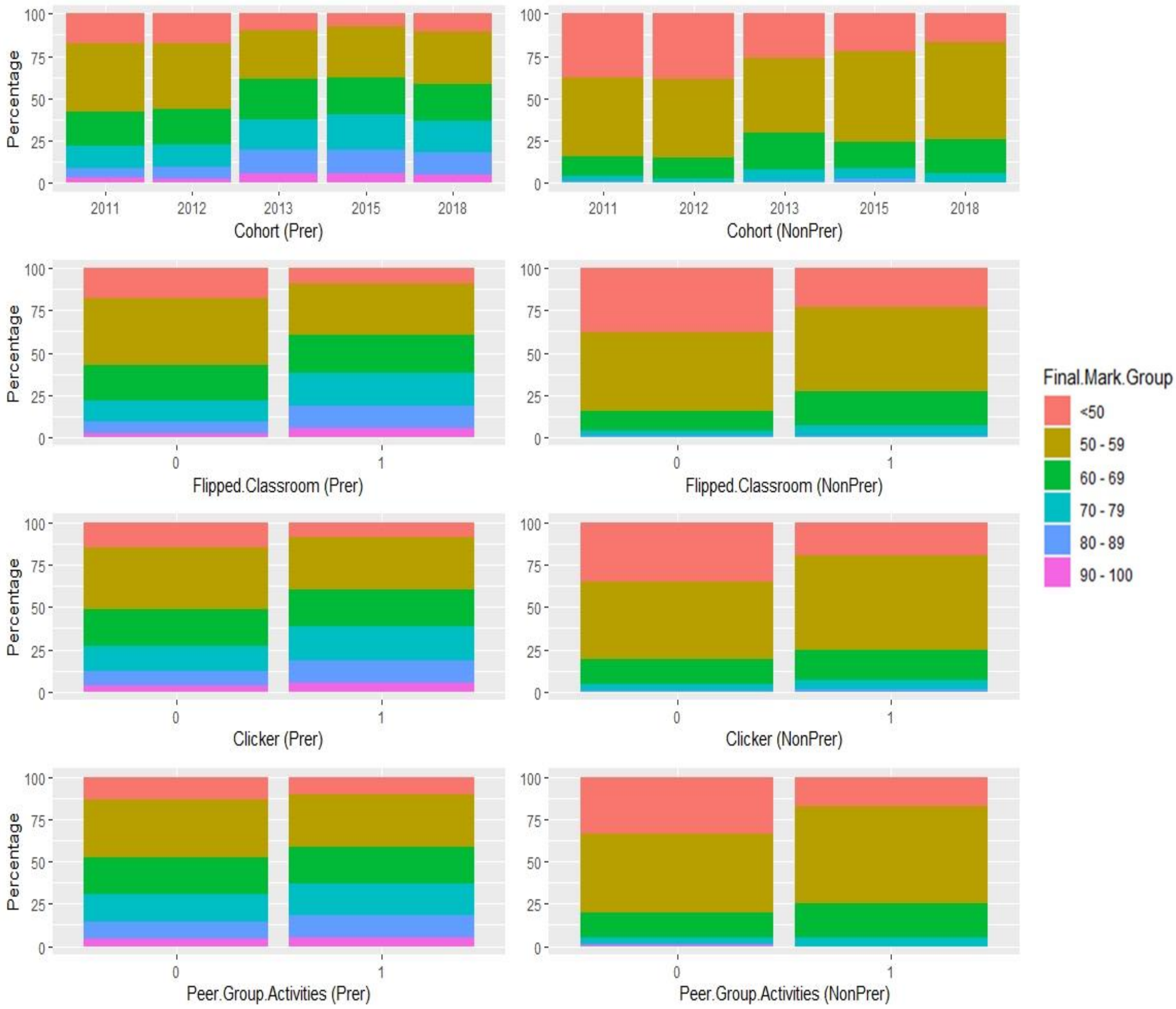


Figure 9: The effect of the technology interventions on performance per cohort and individually for the prerequisite and non-prerequisite samples

- **Peer learning activities**

  Peer learning activities that were introduced in 2018 (Peer learning activities = 1) had a beneficial effect on the higher percentage categories (cf. Figure 9), i.e. high achievers did even better and a minor effect was visible on the lower percentage categories compared to the other four years, when peer learning activities were not used (Peer learning activities = 0). The pass rate for all students in 2018 increased with approximately 2% compared to 2015. It is noticeable that peer learning activities had a positive impact on the non-prerequisite sample, which means that many more students passed compared to the students who were not exposed to peer learning activities.

Figure 9 **(**top two graphs) shows the effect of the interventions per cohort. On face value the 2015 cohort, the first-time QT-clicker users in the prerequisite sample, outperformed the other cohorts. For the prerequisite sample, the cohorts 2011 and 2012 look similar, and 2013, 2015 and 2018 alike. For the non-prerequisite sample, the 2011 and 2012 cohorts look almost alike, opposed to the 2013, 2015 and 2018 cohorts. The peer learning activities in 2018 had a major effect on these students' performance. The 2018 cohort had an increase in the pass rate for the non-prerequisite 2013.

- **Mother tongue education**

  In Figure 10 the variable Mtongue = Yes indicates mother tongue and Mtongue = No indicates non-monther tongue tuition. The proportion of students being educated in their mothing tongue is approximately 50%. For the prerequisite sample, students who were educated in their mother tongue performed better in the higher percentage categories 80% to 100% and were divided approximately equal in the lower categories, as can be seen in Figure 10. The opposite pattern is visible for the non-prerequisite sample for percentage categories 60% to 89%, except for the 90% to 100% category.

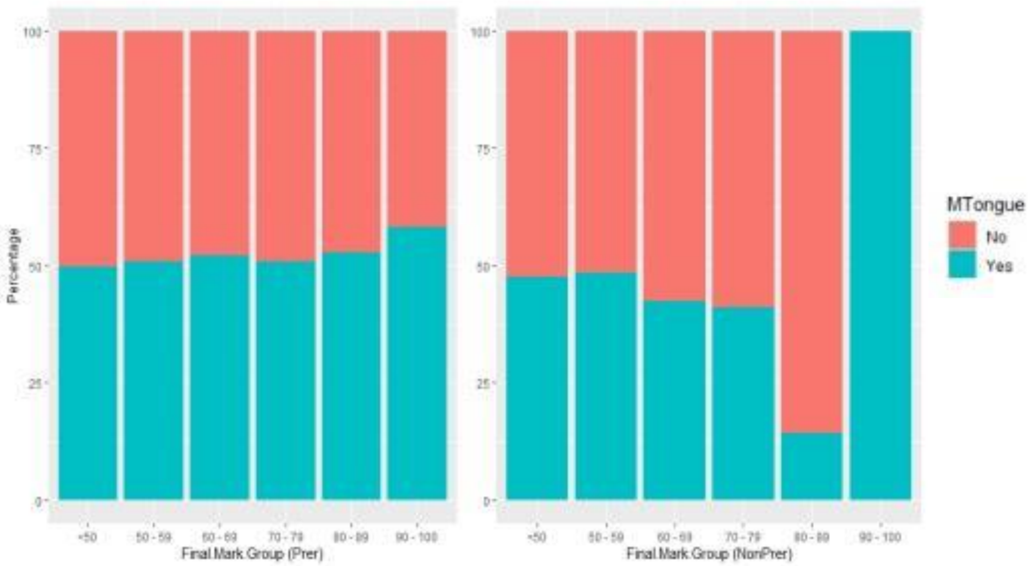


Figure 10: Performance of mother tongue students versus non-mother tongue students for the prerequisite and non-prerequisite samples

- **STK contact time**

  The key finding in Figure 11 (upper five bar charts) is that the prerequisite sample of students who only have statistics for six months (STK exposure = 1), outperformed the students who needed to take statistics for 9 to 12 months (STK exposure = 2). The reason for this finding could be the inclusion of the BCom (Accounting Science) students, who formed up to 30% of the six-month group. The admission requirement of the BCom (Accounting Science) programme for Grade 12 mathematics is between 10% to 20% higher than that of other BCom programmes.

  The remaining two groups, two-year statistics students (STK exposure = 3) and three-year statistics students (STK exposure = 4), approximately had the same performance pattern. The influence of the different technology interventions is visible in Figure 11. For example, in 2015 fewer 9 to 12 months and more two-year statistics students fell in the failure category.

  In Figure 11 (lower five bar charts) the performance of the non-prerequisite sample of students who needed to take statistics for 9 to12 months is more prominent. On face value, fewer students in this group failed in 2015 and 2018. This implies that the clicker and peer interventions had a positive effect on the performance of these students. The students who needed to take statistics for two or three years performed well in 2013 when the flipped classroom model was introduced.

Figure 11: The influence of STK exposure on performance for the prerequisite and non-prerequisite samples

- **Stay in residence**

  The proportion of students staying in residence versus those not staying in residence is approximately 1:3. For both prerequisite and non-prerequisite samples, the majority of students in the <50% category were not staying in residence; in all the remaining categories, the students not staying in residence outperformed the students who were staying in residence. For the prerequisite sample, the percentage of students staying in residence slightly increased as the percentage categories get higher (cf. Figure 12).

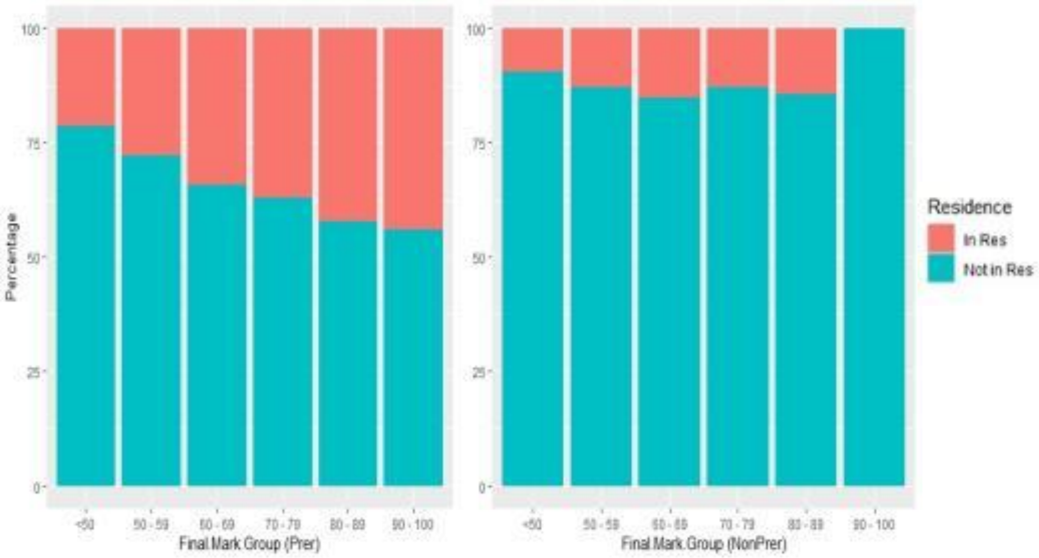


Figure 12: Performance of students in residence versus students not in residence for the prerequisite and non-prerequisite samples

- **Gender**

  Gender code = 0 denotes male students and gender code = 1 denotes female students. For the prerequisite sample, the female students performed better over all the percentage categories and cohorts compared to male students (cf. Figure 13), but for the non-prerequisite sample there is no noticeable difference in performance.

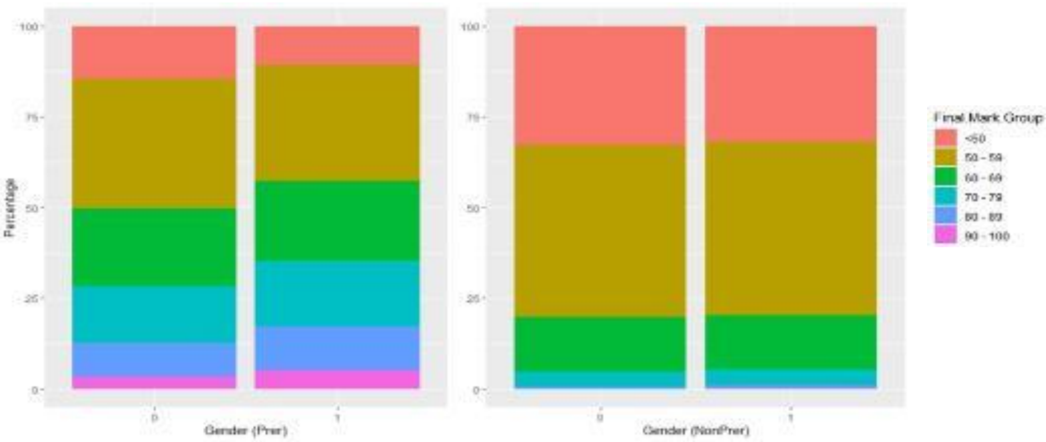


Figure 13: Performance of females versus male students over the percentage categories for the prerequisite and non-prerequisite samples

- **Gr 12 AP Score**

  Clustered bar graphs of each of the AP Score values for the entire sample of students are presented in Figure 14.

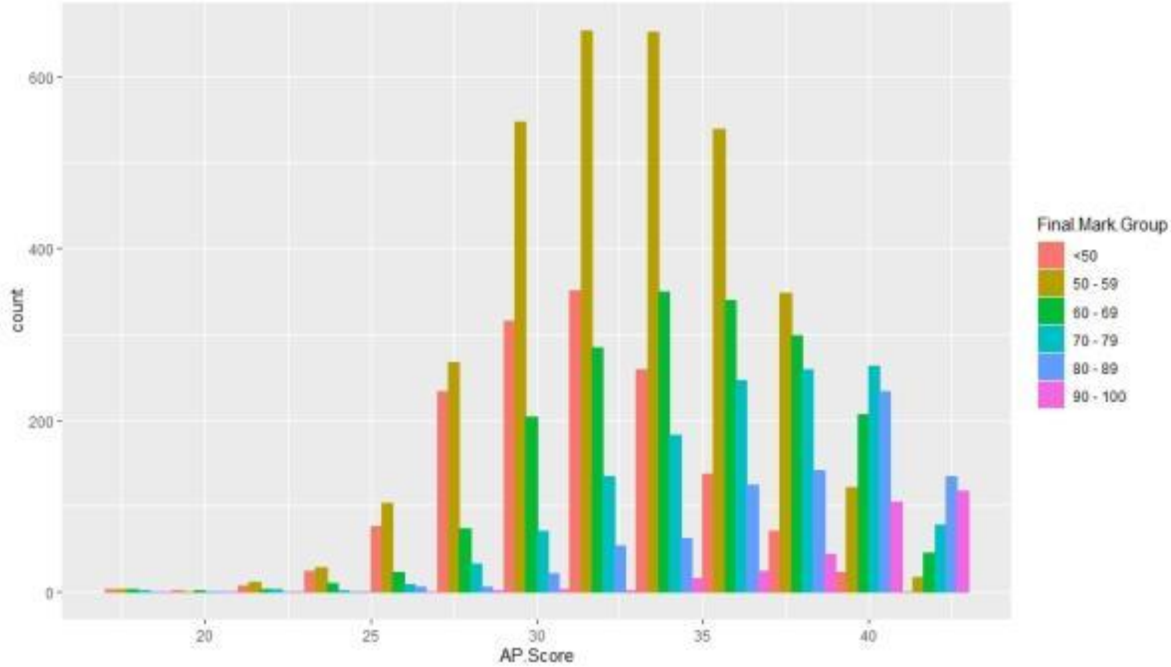


Figure 14: Clustered bar graphs of the marks distribution at each of the AP Score values for all the students

It is noticeable in Figure 14 that the final mark distribution of students with low AP Scores was skew to the right, i.e. at each low AP Score most students failed or fell in the lower percentage categories.

As the AP Score tended to 39, the distribution became approximately normally distributed and for AP Sores from 40 to 42, the final mark distribution tended to be skew to the left, i.e. more students at each high AP Score fell into the higher percentage categories.

- **Repeat students/Number of times in STK110**

  Repeat = No is new students and Repeat = Yes represents students who repeated the module once or more. One of the visions initiated in the previous decade was to reduce the number of repeat students.

Figure 15 shows an increase in repeat students from 2011 to 2018. In 2018, the percentage of new students was only about 20%. More repeat students compared to new students in the non-prerequisite group fell in the lower percentage categories, with no repeat students in the 90% to 100% bracket. In the bottom two graphs in Figure 15, it is visible that the new and first-time failing students are prominent. If a student failed STK110 more than once, they unfortunately tended to fall into a failing cycle, because they were relying on the fact that they have completed the module before.

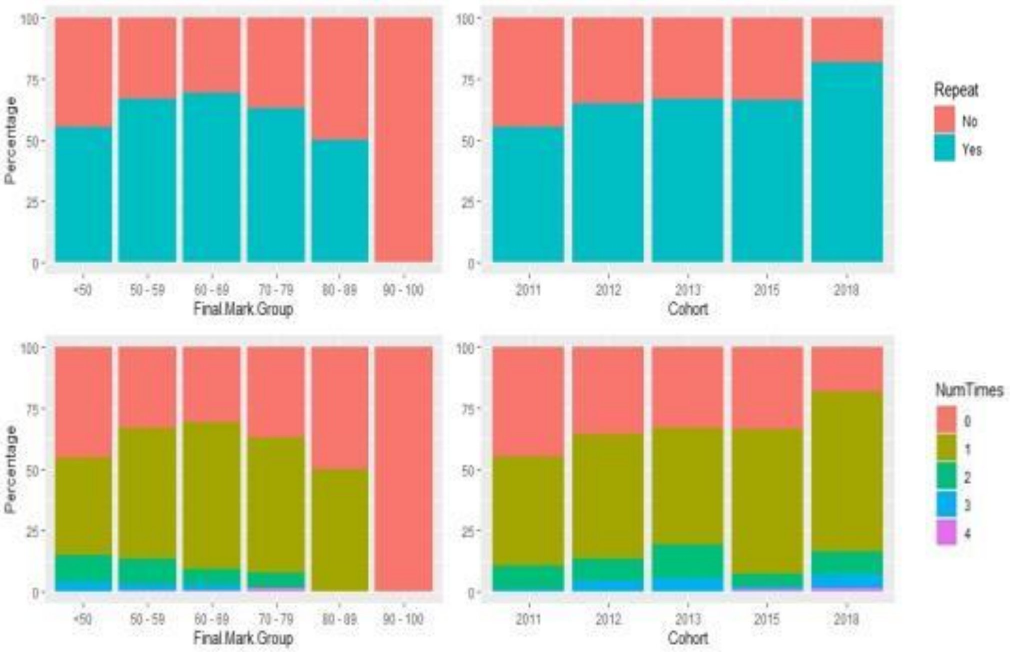


Figure 15: The effect of repeat students and the number of times in STK110 on performance for the non-prerequisite sample

The difference between a repeat student's maximum failing mark earned in previous years and the final mark when the repeat student passed for the eight cohorts from 2011 to 2018 is captured in Figure 16. It is important to note of the category which includes all students who did not have a previous final mark for STK110. This category includes students who were, for example, absent from the exam or did not get exam entrance for STK110. Larger differences were noticeable in 2011, 2013, 2015, 2017 and 2018, which implies that the technology interventions had a positive effect on the repeat students (cf. Figure 16). The flipped classroom, QT-clickers and peer learning activities made essential changes in the repeat students' pass rates in the decade from 2011 with technology-based interventions.

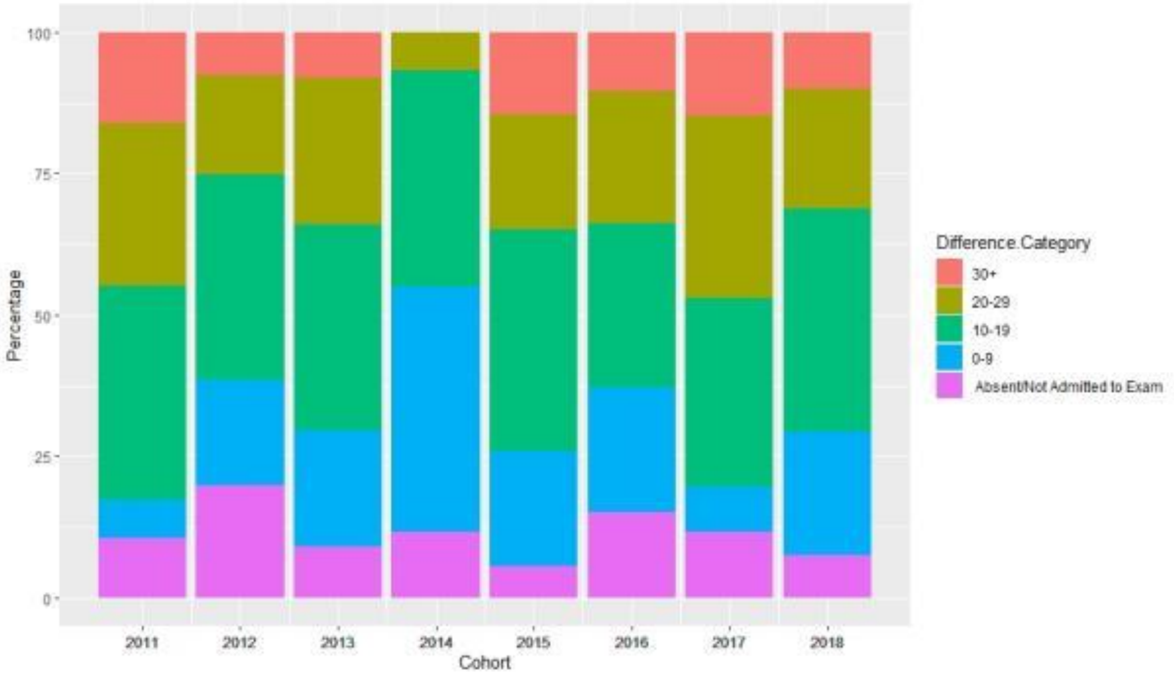


Figure 16: The difference between a repeat student's maximum failing mark and the pass mark from 2011 to 2018

4.2 INFERENTIAL STATISTICS: EFFECT OF INTEVENTIONS

ANOVA, chi-square and McNemar's tests were done to determine the impact of the different technology interventions on student performance and to explicitly determine the impact of the QT-clickers on student performance. The outcome of these tests will be discussed in the following two sections.

4.2.1 IMPACT OF THE DIFFERENT TECHNOLOGY INTERVENTIONS

The five cohorts used represent the different technology interventions implemented from 2011 to 2018. The cohorts under discussion are 2011 (traditional model), 2012 (traditional model with online homework), 2013 (flipped classroom model), 2015 (flipped classroom model with QT-clickers), and 2018 (flipped classroom model with QT-clickers and peer learning activities). Both the prerequisite and non-prerequisite samples were used for analyses. The descriptive statistics are given in Table 3. The following research question is analysed in this section:

Research question 1: How do the technology-based interventions impact on first year students' performance, as measured by their final marks and the pass rate?

Define the following population parameters:

μ_i = population mean of final marks for the i-th cohort

π_i = population proportion of students likely to pass for the i-th cohort

Research question 1a: $H_0: \mu_{2011} = \mu_{2012} = \mu_{2013} = \mu_{2015} = \mu_{2018}$

H_a: Not all population means are equal

ANOVA 1 and 2: Testing for the equality of the five population means of the technology-based interventions for the ***prerequisite and non-prerequisite samples with a final mark****.*

The assumption of equal variances of the ANOVA test was violated for both the prerequisite and non-prerequisite samples, with Levene statistic as $F(4; 7163) = 8.190, p < 0.0001, \eta^2 = 0.0382$[8] and $F(4; 1829) = 4.405, p = 0.002, \eta^2 = 0.0368$ respectively, both who denote a small to medium effect. The Welch robust test statistic for the prerequisite and non-prerequisite samples is $F(4; 3577.223) = 73.455, p < 0.0001$ and $F(4; 589.592) = 17.052, p < 0.0001$, respectively. Therefore, we concluded that the five population means were not all equal for both samples. The mean final mark difference between the cohorts is displayed in Table 5. The mean final mark difference between 2011 and 2013, 2015 and 2018, and between 2012 and 2013, 2015 and 2018 were significant at the 0.05 level of significance by applying the Games Howell post-hoc procedure, for both samples.

Table 5: Multiple comparisons by applying the Games Howell post-hoc procedure for the prerequisite sample and non-prerequisite sample (in brackets in blue) of the five intervention cohorts

Pairwise comparison		Mean Difference (I-J)	Std. Error	Significance (If $p < 0.01$)	95% Confidence Interval
Cohort I	**Cohort J**				
2011	2012	-0.640 (-0.230)	0.526 (0.598)	0.742 (0.995)	(-2.07; 0.80) (-1.86; 1.40)

[8] Effect size guidelines for Eta squared, η^2: 0.01 small effect, 0.06 medium effect and 0.14 large effect

2011	2013	-6.602 (-4.663)	0.531 (0.680)	<0.0001	(-8.05;-5.15) (-6.52;-2.80)
2011	2015	-6.567 (-3.690)	0.544 (0.991)	<0.0001	(-8.05;-5.08) (-6.41;-0.97)
2011	2018	-5.393 (-3.605)	0.541 (0.843)	<0.0001	(-6.87;-3.92) (-5.92;-1.29)
2012	2013	-5.962 (-4.433)	0.537 (0.722)	<0.0001	(-7.43;-4.50) (-6.41;-2.46)
2012	2015	-5.927 (-3.461)	0.550 (1.020)	<0.0001	(-7.43;-4.43) (-6.26;-0.66)
2012	2018	-4.754 (-3.376)	0.547 (0.878)	<0.0001	(-6.25;-3.26) (-5.78;-0.97)
2013	2015	0.035 (0.973)	0.555 (1.070)	1.000 (0.893)	(-1.48; 1.55) (-1.96; 3.91)
2013	2018	1.208 (1.058)	0.552 (0.935)	0.184 (0.790)	(-0.30; 2.72) (-1.51; 3.62)
2015	2018	1.173 (0.085)	0.565 (1.181)	0.230 (1.000)	(-0.37; 2.71) (-3.15; 3.32)

In conclusion, there was no difference between the mean final mark of the traditional model in 2011 and the traditional model with online homework in 2012. The flipped classroom, QT-clickers and peer learning activities used in 2013, 2015 and 2018 made a difference to the mean final mark compared to the traditional model with or without online homework, but did not differ among themselves.

The repeat students have been of concern to us for many years; therefore, the non-prerequisite sample was divided into new and repeat students and two separate ANOVA tests were done.

ANOVA 3 and 4: Testing for the equality of the five population means of the technology-based interventions for the ***new and repeat students in the non-prerequisite sample with a final mark****.*

The assumption of equal variances of the ANOVA test was also violated for both the new and repeat students in the non-prerequisite sample, with Levene statistic as $F(4;671) = 2.917, p = 0.021, \eta^2 = 0.0573$ and $F(4;1153) = 3.438, p = 0.008, \eta^2 = 0.0245$ respectively, both who denote a small to medium effect. The Welch robust test statistics for the new and repeat students were $F(4;140.005) = 9.003, p$ <0.0001 and $F(4;423.454) = 7.572, p$ <0.0001, respectively. Therefore, we conclude that the five population means were not all equal for new and repeat students. The mean final mark differences between the cohorts are displayed in Table 6. The mean final mark difference for the new students was between the 2011 and 2013, the 2011 and 2015, and the 2012 and 2013 cohorts. The mean final mark difference for the repeat students was between 2011 and 2013, 2012 and 2013, as well as 2012 and 2018, and were significant at the 0.05 level of significance by applying the Games Howell post-hoc procedure.

Table 6: Multiple comparisons by applying the Games Howell post-hoc procedure for the new and repeat students (in brackets in blue) of the five intervention cohorts

Pairwise comparison		Mean Difference (I-J)	Std. Error	Significance	95% Confidence Interval
Cohort I	**Cohort J**				
2011	2012	-1.251 (-0.784)	0.961 (0.759)	0.690 (0.840)	(-3.89; 1.38) (-1.29; 2.86)
2011	2013	-6.398 (-3.261)	1.235 (0.815)	<0.0001(0.001)	(-9.80;-3.00) (-5.49;-1.03)
2011	2015	-5.832 (-2.083)	1.598 (1.260)	0.004 (0.466)	(-10.27;-1.39) (-5.56; 1.39)
2011	2018	-4.439 (-2.413)	2.430 (0.905)	0.375 (0.062)	(-11.44;2.56) (-4.90; 0.07)
2012	2013	-5.147 (-4.045)	1.324 (0.861)	0.001(<0.0001)	(-8.79;-1.50) (-6.40;-1.69)
2012	2015	-4.581 (-2.867)	1.660 (1.290)	0.053 (0.177)	(-9.20; 0.04) (-6.42; 0.69)
2012	2018	-3.188 (-3.197)	2.476 (0.946)	0.700 (0.007)	(-10.29; 3.92) (-5.79;-0.60)
2013	2015	0.565 (1.178)	1.832 (1.324)	0.998 (0.900)	(-4.51; 5.64) (-2.47; 4.83)
2013	2018	1.959 (0.848)	2.595 (0.991)	0.942 (0.913)	(-5.42; 9.34) (-1.87; 3.57)
2015	2018	1.393 (-0.330)	2.781 (1.381)	0.897 (0.999)	(-6.46; 9.24) (-4.13; 3.47)

It is evident that the flipped classroom made a significant difference to the mean final mark if compared to the traditional model with and without online homework, for both the repeat and new students in the non-prerequisite sample. The implementation of QT-clickers made a significant difference to the mean final mark if compared to the traditional model, only for the new students in the non-prerequisite sample. The peer learning activities made a significant difference to the mean final mark if compared to the traditional model with online homework, only for the repeat students.

Research question 1b: $H_0: \pi_{2011} = \pi_{2012} = \pi_{2013} = \pi_{2015} = \pi_{2018}$

H_a: Not all population proportions are equal

Table **7** represents the success and failure rates for both the prerequisite and non-prerequisite sample of students with a final mark, for the intervention cohorts.

Table 7: Success and failure rates of the prerequisite and non-prerequisite sample of students

Cohorts	2011	2012	2013	2015	2018	Total
Prerequisite sample of students with a final mark						
Passed (%)	1208 (82.1)	1108 (82.3)	1300 (89.8)	1324 (92.8)	1322 (89.6)	6262
Failed (%)	264 (17.9)	238 (17.7)	148 (10.2)	103 (7.2)	153 (10.4)	906
Total	1472	1346	1448	1427	1475	7168

Non-prerequisite sample of students with a final mark						
Passed (%)	447 (62.3)	258 (61.6)	261 (73.5)	135 (78.0)	140 (82.8)	1241
Failed (%)	271 (37.7)	161 (38.4)	94 (26.5)	38 (22.0)	29 (17.2)	593
Total	718	419	355	173	169	1834

Chi-square 1 and 2: Testing for the equality of the five population proportions of the five intervention cohorts for both the ***prerequisite and non-prerequisite sample with a final mark***

The chi-square test concluded that the pass rates were different for the five cohorts in the prerequisite sample ($\chi^2(4) = 120.8949, p < 0.0001, \text{Cramer's } V = 0.1299$[9]) as well as the non-prerequisite sample ($\chi^2(4) = 48.5577, p < 0.0001, \text{Cramer's } V = 0.1627$), both of which denote a small effect. Marascuilo's pairwise comparisons (Anderson et al., 2020, pp. 537-540) were performed to compare the different combinations of the cohorts for both samples. The results are summarised in Table 8, using $\alpha = 0.05$ and ten pairwise comparisons.

Table 8: Post-hoc comparisons of the proportion successful students in the prerequisite sample and non-prerequisite sample (in brackets in blue) of the five intervention cohorts

Pairwise comparison	$\lvert\bar{p}_i - \bar{p}_j\rvert$	V_{ij} **the critical value**	**Significant if:** $\lvert\bar{p}_i - \bar{p}_j\rvert > V_{ij}$
2011 vs. 2012	0.0025 (0.0068)	0.0444 (0.0920)	Not significant
2011 vs. 2013	0.0771 (0.1126)	0.0394 (0.0911)	Significant
2011 vs. 2015	0.1072 (0.1578)	0.0373 (0.1118)	Significant
2011 vs. 2018	0.0756 (0.2058)	0.0393 (0.1053)	Significant
2012 vs. 2013	0.0746 (0.1195)	0.0403 (0.1028)	Significant
2012 vs. 2015	0.1046 (0.1646)	0.0384 (0.1215)	Significant
2012 vs. 2018	0.0731 (0.2127)	0.0403 (0.1155)	Significant
2013 vs. 2015	0.0300 (0.0451)	0.0324 (0.1208)	Not significant
2013 vs. 2018	0.0015 (0.0932)	0.0346 (0.1148)	Not significant
2015 vs. 2018	0.0315 (0.0481)	0.0323 (0.1318)	Not Significant

Although the effect size (Cramer's *V*) for both samples was small, the conclusion was that the pass rates of the 2013, 2015 and 2018 cohorts were significantly different from the 2011 and 2012 cohorts, but not from one another. This implies that the flipped classroom, QT-clickers and peer learning activities made a positive contribution to the increasing pass rates.

The non-prerequisite sample consisted of new students not meeting all the prerequisites as well as repeat students meeting or not meeting all the prerequisites. Table 9 distinguishes between

[9] Effect size guidelines for Cramer's V: V∈[0.1;0.3] small effect, V∈[0.4;0.5] medium effect and V>0.5 large effect

the new and repeat students in the non-prerequisite sample regarding success and failure rates in the five intervention cohorts.

Table 9: Success and failure rates of the new and repeat students in the non-prerequisite sample

Cohorts	2011	2012	2013	2015	2018	Total
New students in the Non-prerequisite sample with a final mark						
Passed (%)	171 (53.1)	87 (58.8)	85 (72.6)	46 (79.3)	21 (67.7)	410
Failed (%)	151 (46.9)	61 (41.2)	32 (27.4)	12 (20.7)	10 (32.3)	266
Total	322	148	117	58	31	676
Repeat students in the Non-prerequisite sample of students with a final mark						
Passed (%)	276 (69.7)	171 (63.1)	176 (73.9)	89 (77.4)	119 (86.2)	831
Failed (%)	120 (30.3)	100 (36.9)	62 (26.1)	26 (22.6)	19 (13.8)	327
Total	396	271	238	115	138	1158

Apart from the repeat students in 2018, the total number of new and repeat students in the non-prerequisite sample decreases for the cohorts in Table 9. The percentages for the new versus repeat students in the non-prerequisite sample not meeting a specific prerequisite were calculated across the five intervention cohorts as follows: 97.6% versus 30.5% did not meet the Grade 12 mathematic requirement; 10.1% versus 2.4% did not meet the AP Score requirement; 7.7% versus 1.6% did not meet both. It is evident that meeting the Grade 12 mathematic prerequisite had a more substantial influence on the pass rates than meeting the AP Score requirement. It is important to know which technology-based interventions, if any, had an impact on the pass rates of the new and repeat students in the non-prerequisite sample and were investigated by the following chi-square tests and multiple comparisons.

Chi-square 3 and 4: Testing for the equality of the five population proportions of the five intervention cohorts for both the ***new and repeat students in the non-prerequisite sample with a final mark.***

The chi-square test concluded that the pass rates were different for the five cohorts for the new students in the non-prerequisite sample ($\chi^2(4) = 24.0703, \mathrm{p} < 0.0001, \mathrm{Cramer's\ V} = 0.1887$), as well as the repeat students ($\chi^2(4) = 27.4872, \mathrm{p} < 0.0001, \mathrm{Cramer's\ V} = 0.1541$), both of which denote a small effect. Marascuilo's pairwise comparisons were performed to compare the different combinations of the cohorts for the new and repeat students in the non-prerequisite sample. The results are summarised in Table 10, using $\alpha = 0.05$ and ten pairwise comparisons.

Table 10: Post-hoc comparisons of the proportion successful new and repeat (in brackets in blue) students in the non-prerequisite sample of the five intervention cohorts

Pairwise comparison	$\lvert\bar{p}_i - \bar{p}_j\rvert$	V_{ij} **the critical value**	**Significant if:** $\lvert\bar{p}_i - \bar{p}_j\rvert > V_{ij}$
2011 vs. 2012	0.0568 (0.0660)	0.1512 (0.1149)	Not significant
2011 vs. 2013	0.1954 (0.0425)	0.1531 (0.1129)	Significant (Not significant)
2011 vs. 2015	0.2620 (0.0769)	0.1849 (0.1396)	Significant (Not significant)
2011 vs. 2018	0.1464 (0.1653)	0.2724 (0.1150)	Not significant (Significant)
2012 vs. 2013	0.1387 (0.1085)	0.1779 (0.1258)	Not significant
2012 vs. 2015	0.2053 (0.1429)	0.2058 (0.1503)	Not significant
2012 vs. 2018	0.0896 (0.2313)	0.2871 (0.1277)	Not significant (Significant)
2013 vs. 2015	0.0666 (0.0344)	0.2073 (0.1487)	Not significant
2013 vs. 2018	0.0491 (0.1228)	0.2881 (0.1259)	Not significant
2015 vs. 2018	0.1157 (0.0884)	0.3061 (0.1503)	Not Significant

From Table 9 it is noticeable that the pass rate for new students in the non-prerequisite sample increased in 2013 and 2015 with approximately 20% compared to 2011. The multiple comparisons in Table 10 confirm the results, namely that the new students' pass rates increased significantly with the change of pedagogy embedded with online homework in 2013, as well as the implementation of QT-clickers in 2015, compared to the traditional teaching model on its own in 2011.

The repeat students had a rather different scenario. As already mentioned, the majority of the repeat students met the prerequisites. It can be seen from Table 9 that the pass rate for the 2011 repeat students (traditional teaching model) started at 69.7%, decreased to 63.1% in 2012 despite the implementation of the online homework system, and increased again to 73.9% and 77.4% in 2013 and 2015, respectively. According to Table 10, there was no significant increase in the pass rates of the repeat students in 2013 and 2015, i.e. the flipped classroom and QT-clickers, compared to 2011 and 2012 cohorts. The only diverse pass rate belonged to the 2018 cohort which was significantly different from the 2011 and 2012 cohorts. This indicates that the implementation of peer learning activities, modification of in-class activities to the South-

African context and switching of lecturers presented in 2018, could have made a significant contribution to the increasing pass rate of the repeat students.

Research question 1c: H_0: There is an association between the final marks and the cohorts
H_a: There is no association between the final marks and the cohorts

Chi-square 5: Testing for an association between intervention cohorts and the marks distribution for both the ***prerequisite and non-prerequisite sample with a final mark.***

A cross-tabulation of the final marks distribution (6 groups) across the five intervention cohorts is summarised in Table 11. The prerequisite sample is shaded in light grey across the six marks distribution groups. The chi-square test results for the prerequisite sample revealed that there was a significant relationship between the final marks distribution and the different cohorts ($\chi^2(20) = 313.003, p < 0.0001, Cramer's\ V = 0.104$). It is important to note that the chi-square test for the non-prerequisite sample was also significant ($\chi^2(20) = 87.311\ p < 0.0001, Cramer's\ V = 0.109$), both of which denote a small effect. The chi-square test for the non-prerequisite sample was not reliable because 30% (9 out of 30) of the cells in the cross-tabulation had expected frequencies below five. The 80% to 89% and 905 to 100% brackets were combined, and the analysis rerun for the non-prerequisite sample. The result was a significant relationship and small effect size between the final marks distribution and different cohorts ($\chi^2(16) = 84.5278, p < 0.0001, Cramer's\ V = 0.107$).

Table 11: Chi-square results of the analysis of the final marks distribution across the five cohorts for both prerequisite and non-prerequisite samples

Final marks distribution	Statistics	Cohort					
		2011	**2012**	**2013**	**2015**	**2018**	**Total**
<50	Count	264	238	148	103	153	**906**
	Exp. Count	186.1	170.1	183.0	180.4	186.4	
	Std. Residual	**5.7**	**5.2**	**-2.6**	**-5.8**	**-2.4**	
	Count	271	161	94	38	29	**593**
	Exp. Count	232.2	135.5	114.8	55.9	54.6	
	Std. Residual	**2.5**	**2.2**	-1.9	**-2.4**	**-3.5**	

50-59	Count	587	523	415	437	459	**2421**
	Exp. Count	497.2	454.6	489.1	482.0	498.2	
	Std. Residual	**4.0**	**3.2**	**-3.3**	**-2.0**	-1.8	
	Count	332	195	155	93	97	**872**
	Exp. Count	341.4	199.2	168.8	82.3	80.4	
	Std. Residual	-0.5	-0.3	-1.1	1.2	1.9	
60-69	Count	304	283	344	313	319	**1563**
	Exp. Count	321.0	293.5	315.7	311.2	321.6	
	Std. Residual	-0.9	-0.6	1.6	0.1	-0.1	
	Count	84	52	79	27	34	**276**
	Exp. Count	108.1	63.1	53.4	26.0	25.4	
	Std. Residual	**-2.3**	-1.4	**3.5**	0.2	1.7	
70-79	Count	187	179	261	298	277	**1202**
	Exp. Count	246.8	225.7	242.8	239.3	247.3	
	Std. Residual	**-3.8**	**-3.1**	1.2	**3.8**	1.9	
	Count	26	9	24	10	9	**78**
	Exp. Count	30.5	17.8	15.1	7.4	7.2	
	Std. Residual	-0.8	**-2.1**	**2.3**	1.0	0.7	
80-89	Count	86	91	200	197	193	**767**
	Exp. Count	157.5	144.0	154.9	152.7	157.8	
	Std. Residual	**-5.7**	**-4.4**	**3.6**	**3.6**	**2.8**	
	Count	4	2	3	5	0	**14**
	Exp. Count	5.5	3.2	2.7	1.3	1.3	
	Std. Residual	-0.6	-0.7	0.2	**3.2**	-1.1	
90-100	Count	44	32	80	79	74	**309**
	Exp. Count	63.5	58.0	62.4	61.5	63.6	
	Std. Residual	**-2.4**	**-3.4**	**2.2**	**2.2**	1.3	
	Count	1	0	0	0	0	**1**
	Exp. Count	0.4	0.2	0.2	0.1	0.1	
	Std. Residual	1.0	-0.5	-0.4	-0.3	-0.3	
n_1 (prerequisite sample)		**1472**	**1346**	**1448**	**1427**	**1475**	**7168**
n_2 (non-prerequisite sample)		**718**	**419**	**355**	**173**	**169**	**1834**
TOTAL		**2190**	**1765**	**1803**	**1600**	**1644**	**9002**

Apart from 2011 (base year – traditional model), the other four cohorts represent the introduction of a technology-based intervention. Inspection of the results in Table 11 for the prerequisite sample (shaded in light grey), confirmed that the flipped classroom and the use of QT-clickers were effective in increasing the pass rate. The discrepancy between the observed and expected number of students in each cell in the cross tabulation, as quantified by the standardised residuals, is of particular interest (Agresti, 2013, p. 81), since standardised residuals larger than |2| identify the cells where significantly more (or fewer) cases were observed than expected under the null hypothesis of independence.

For example, the positive standardised residuals of **5.7** and **5.2** in the first two cells of Table 11 indicate that many more students failed the module in 2011 and 2012 than expected under the null hypothesis of no association. It is encouraging to notice that the opposite happened from 2013 to 2018 after the flipped classroom was introduced (negative adjusted residual of **-2.6** in 2013, **-5.8** in 2015, and **-2.4** in 2018), i.e. far fewer students than expected failed the module in the above-mentioned years. The standardised residuals followed the same pattern for the 2011 and 2012 cohorts and the opposite pattern for 2013, 2015 and 2018 in the other mark brackets. Although the Aplia online homework system implemented in 2012 could have made a difference to the 50% to 59% bracket (more students than expected passed the module), it seems not to have contributed positively to the higher mark intervals. Again, the negative significant standardised residuals of the 2013, 2015 and 2018 cohorts in the 50% to 59% bracket emphasise the impact of the flipped classroom and the use of QT-clickers on the higher results. Fewer students than expected fell into this category, but that was because more students than expected moved into the higher brackets as quantified by the positive residuals in these brackets. This is in contrast with the 2011 and 2012 cohorts, where fewer than expected students fell into the higher brackets, illustrated by the negative standardised residuals associated with these two cohorts. QT-clickers were introduced in 2015 for the first time and the 2015 cohort performed well in all the categories and outperformed the other cohorts in three of the six categories.

It was decided to subdivide the non-prerequisite sample into new and repeat students to make sure that we did not miss unexploited findings. Too many of the cells in the cross-tabulation had expected frequencies below five, therefore the three percentage brackets between 70% to 100% were combined, as can be seen in Table 12.

*Chi-square 6: Testing for an association between intervention cohorts and the marks distribution for both the **new and repeat students in the non-prerequisite sample with a final mark.***

Table 12: Chi-square results of the analysis of the final marks distribution across the five cohorts for the non-prerequisite samples being divided into new (n_1) and repeat (n_2) students

Final marks distribution	Statistics	Cohort					
		2011	**2012**	**2013**	**2015**	**2018**	**TOTAL**
<50	Count *New*	151	61	32	12	10	**266**
	Exp. Count	126.7	58.2	46.0	22.8	12.2	
	Std. Residual	**2.2**	0.4	**-2.1**	**-2.3**	-0.6	
	Count *Repeat*	120	100	62	26	19	**327**
	Exp. Count	111.8	76.5	67.2	32.5	39.0	
	Std. Residual	0.8	**2.7**	-0.6	-1.1	**-3.2**	
50-59	Count *New*	133	65	47	32	11	**288**
	Exp. Count	137.2	63.1	49.8	24.7	13.2	
	Std. Residual	-0.4	0.2	-0.4	1.5	-0.6	
	Count *Repeat*	199	130	108	61	86	**584**
	Exp. Count	199.7	136.7	120.0	58.0	69.6	
	Std. Residual	-0.1	-0.6	-1.1	0.4	**2.0**	
60-69	Count *New*	24	19	26	10	6	**85**
	Exp. Count	40.5	18.6	14.7	7.3	3.9	
	Std. Residual	**-2.6**	0.1	**2.9**	1.0	1.1	
	Count *Repeat*	60	33	53	17	28	**191**
	Exp. Count	65.3	44.7	39.3	19.0	22.8	
	Std. Residual	-0.7	-1.7	**2.2**	-0.5	1.1	
70-100	Count *New*	14	3	12	4	4	**37**
	Exp. Count	17.6	8.1	6.4	3.2	1.7	
	Std. Residual	-0.9	-1.8	**2.2**	0.5	1.8	
	Count *Repeat*	17	8	15	11	5	**56**
	Exp. Count	19.2	13.1	11.5	6.5	6.7	
	Std. Residual	-0.5	-1.4	1.0	**2.3**	-0.6	
n_1 (new students)		**322**	**148**	**117**	**58**	**31**	**676**
n_2 (repeat students)		**396**	**271**	**238**	**115**	**138**	**1158**
TOTAL		**718**	**419**	**355**	**173**	**169**	**1834**

The new student sample is shaded in light grey across the four marks distribution groups. For the non-prerequisite sample (cf. Table 11), more students failed in 2011 and 2012 and far fewer did not fail in 2013, 2015 and 2018 than expected under the null hypothesis of no association. With closer inspection when the non-prerequisite sample was divided into new and repeat students, more new students failed in 2011 compared to more repeat students who failed in

2012. Fewer new students failed in 2013 and 2015 and far fewer students failed in 2018, which meant more repeat students than expected in the 50% to 59% bracket. The flipped classroom with or without clicker use had a greater impact on the new students in the less than 50% bracket opposed to peer learning activities and lecturer-switching on repeat students in the less than 50% and 50% to 59% brackets.

For the non-prerequisite sample, fewer students in 2011 and many more students in 2013 fell in the 60% to 69% bracket than expected under the null hypothesis of no association. From Table 12, for the 60% to 100% bracket, both new and repeat students benefited from the flipped classroom with or without clicker use.

The impact of the different interventions on the prerequisite and non-prerequisite samples with the non-prerequisite sample being divided into new and repeat students was clearer when visualised by the three sets of line graphs in Figure 17.

The interpretation for the first set of line graphs in Figure 17 for the prerequisite sample is as follows: Although the final marks of 2012 were slightly better than those of 2011 (traditional model with and without Aplia respectively), the line graphs look similar. The graphs of 2013, 2015 and 2018, however, showed a dramatic improvement over the other two years. The percentage of unsuccessful students in 2013, 2015 and 2018 decreased by between 7.9% to 11.1% compared to the 2011 group. The proportion of unsuccessful students had more than halved in 2015 and almost halved in 2013 and 2018 compared to the base year (2011), while the percentage of successful students improved substantially over all the other categories. It is evident that when the flipped classroom was introduced from 2013, the students' performance changed drastically. In 2015, QT-clickers were introduced, while students still had a flipped classroom teaching model, with all the advantages of the imbedded online homework. This could explain why the student performance in 2015 outsmarted the 2013 cohort in certain categories: there was a scaffolding process of building on the previous successful intervention. This is in line with the conclusion from the ANOVA test 1 combined with the multiple comparisons in Table 5 and chi-square test 1 with post-hoc comparisons in Table 8.

.

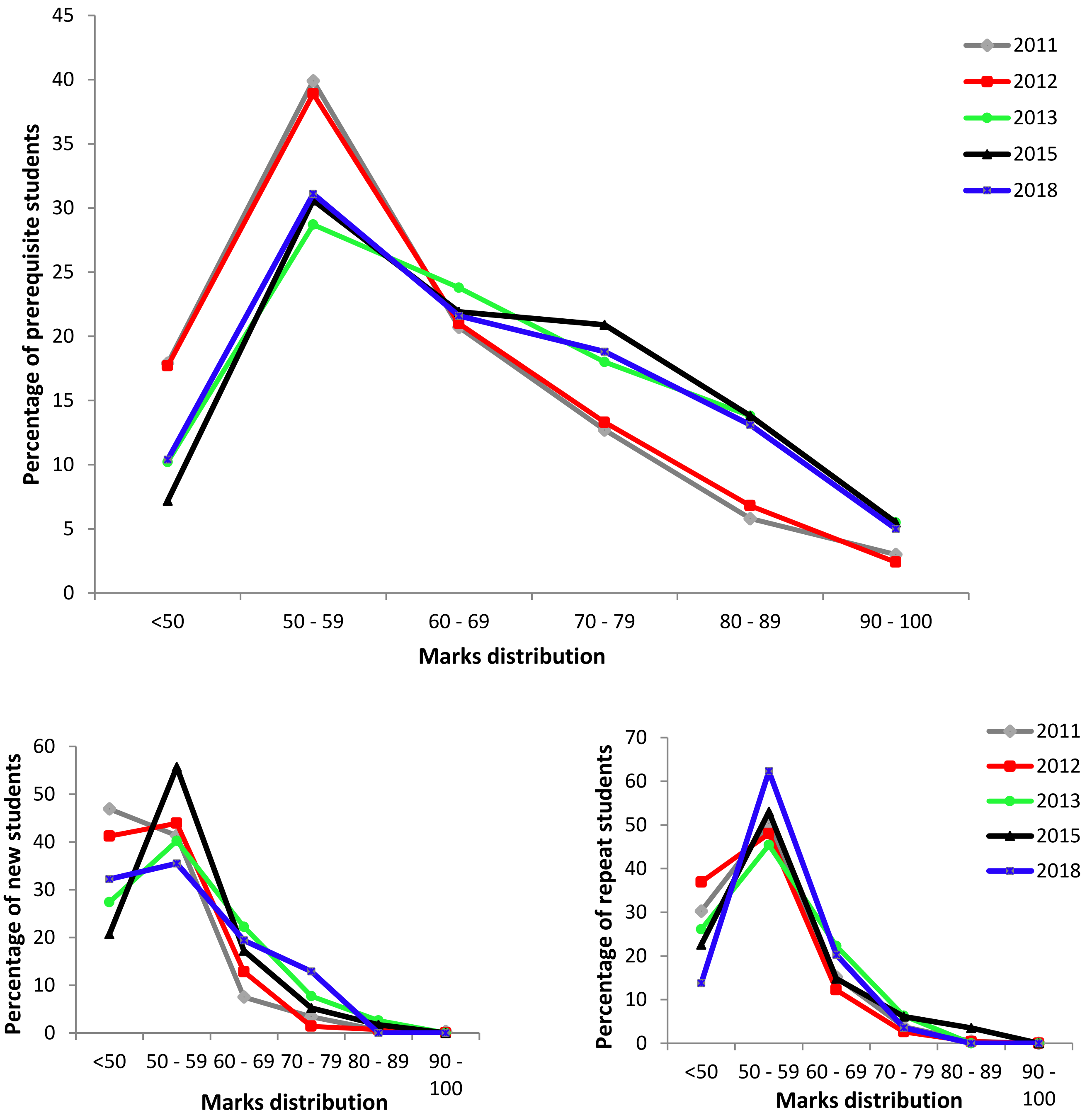


Figure 17: Marks distribution of the five cohorts for the prerequisite sample and non-prerequisite sample divided for new students and repeat students

The interpretation of the second set of line graphs for the new students in the non-prerequisite sample agreed with the outcome and interpretation of the corresponding ANOVA and chi-square tests. The most important interventions for the new students in the non-prerequisite sample were the flipped classroom and the use of QT-clickers. It is notable from the graph that far fewer students failed in 2013 (27.4%) and in 2015 (20.7%), and this implies a difference of between 19.5 to 26.2% compared to 2011. The 2015 cohort showed the highest peak of 55.7% of students in the 50% to 59% bracket, while with the 2013 and 2018 cohorts fewer students fell into the 50% to 59% bracket but moved into the 60% to 79% bracket. This implies that the flipped classroom and peer learning activities improved their performance.

The performance of repeat students is represented in the third set of line graphs, which is in accordance with the ANOVA and chi-square tests. Although the use of QT-clickers improved the performance of repeat students, the intervention that outsmarted all the others was peer learning activities in 2018 combined with South African customised context and switching of lecturers. In 2018 the failure rate of the repeat students for the five years decreased to 13.8% compared to the highest failure rate of 36.9% in 2012.

4.2.2 IMPACT OF THE QT-CLICKERS

A summary of the composition of the two cohorts 2014 (most recent cohort with no QT-clickers) and 2017 (third cohort with QT-clickers – see Section 3.4.3.1) is given in Table 13.

Table 13: Summary of the 2014 and 2017 cohort composition

Cohort	2014 (no QT-clickers)	2017 (QT-clickers – with partial grading)	2017 (QT-clickers – without partial grading)
Sample size	$n = 1448$	$n = 1176$	$n = 1176$
Proportion female %	52.5	48	48
Semester mark mean% (standard deviation %)	63.47 (15.07)	61.79 (14.53)	60.83 (14.51)
Examination mark mean% (standard deviation %)	59.32 (16.88)	61.32 (17.71)	59.52 (17.91)
Pass rate(%)[a]	86.5	88.4	85.6 – 88.4[b]

a. Pass rate is based on the final grade ≥ 50% which is a composite of the examination and semester mark, or on the supplementary examination mark.

b. 33 students who passed with partial grading failed without partial grading but would have qualified for a supplementary exam which they may or may not have passed.

Polygons based on deciles have been used in Figure 18 to give a better visual comparison between the distributions of the two cohorts. It shows that the 2017 cohort with partial grading outperformed the 2014 cohort from the $[60; 70)$ interval upwards. The 2017 cohort without partial grading also outperformed the 2014 cohort in the $[60; 80)$ and $[90; 100)$ intervals.

Two chi-squared tests were conducted to evaluate the association between the examination marks for the 2014 cohort compared to the 2017 cohorts with and without partial grading, as well as McNemar's tests for the paired data (the 2017 cohorts with and without partial grading). There was a significant relationship when first comparing the distribution of scores for the 2014 cohort to the 2017 cohort with partial grading ($\chi^2(8) = 37.343, p < 0.0001, Cramer's\ V = 0.119$). The grade intervals that contributed most to the significant difference between the two distributions for the 2014 cohort are $[10; 50)$, $[70; 80)$ and $[90; 100]$. The 2017 cohort, for example, had significantly fewer students with an examination mark in the interval $[40; 50)$ and several more students than expected under the null hypothesis of no association with an examination mark in the interval $[70; 80)$ and vice versa.

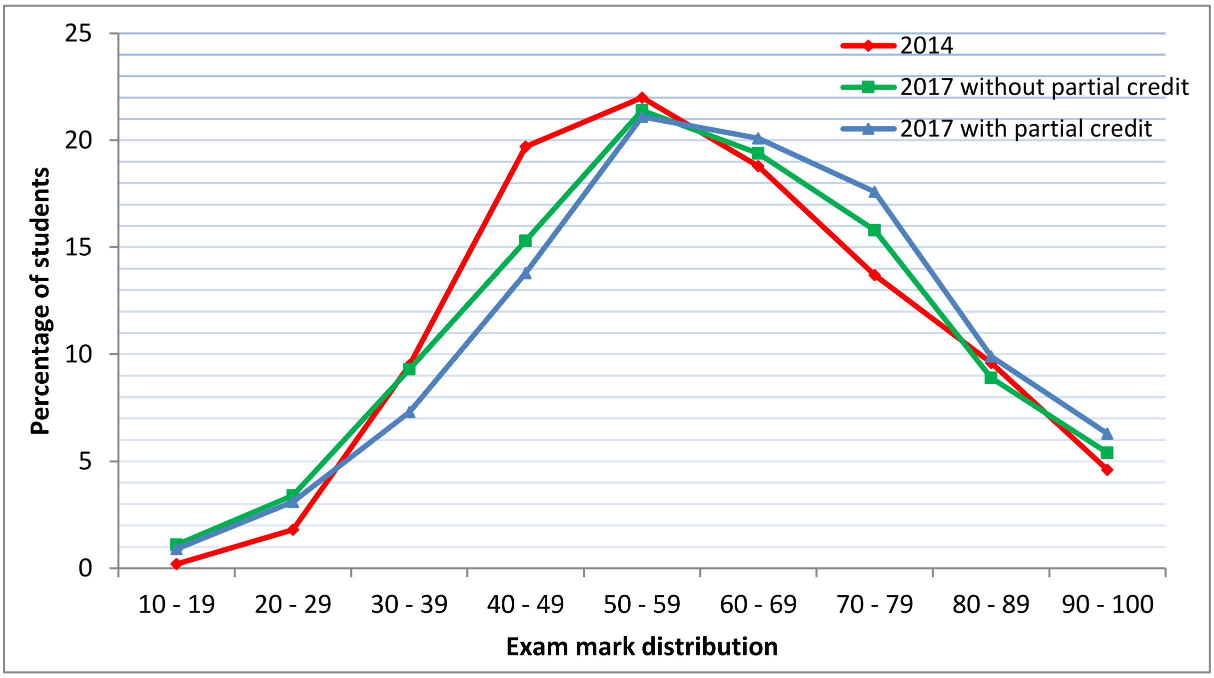


Figure 18: Frequency polygons of the distribution of examination marks for the 2014 and the 2017 cohorts with and without partial grading

Secondly, when comparing the distribution of scores for the 2014 cohort to the 2017 cohort without partial grading, a significant result was also found ($\chi^2(8) = 25.446, p < 0.001, Cramer's\ V = 0.098$). The grade intervals that contributed most to the significant difference between the two distributions are $[10; 30)$ and $[40; 50)$. We see that far fewer 2017 students than expected under the null hypothesis of no association achieved an examination mark in the interval $[40; 50)$, while many more 2014 students than expected fell in this interval.

Finally, McNemar's tests with a continuity correction were used to test which grade intervals, if any, had different proportions between the examination marks for the 2017 cohort with and without partial grading. The gains made by awarding partial credit were substantial, with a significantly higher proporion of students with examination scores in the intervals $[60; 70)$, $(p = 0.0133)$, $[70; 80)$, $(p < 0.0001)$, $[80; 90)$, $(p = 0.0015)$ and $[90; 100]$, $(p = 0.0026)$. There was also a significantly lower proportion of students with examination scores in the intervals $[30; 40)$, $(p < 0.0001)$ and $[40; 50)$, $(p < 0.0001)$.

4.3 INFERENTIAL STATISTICS: STATISTICAL MODELLING (GLM)

4.3.1 IMPACT OF THE QT-CLICKERS

The GLM has as its main objective the modeling of the effect of the intervention (clicker cohort in 2017 versus the control cohort in 2014) on the performance of students as measured by their examination marks with and without partial grading (Reyneke et al., 2021). The students of both cohorts were therefore included in the model and clicker use included as a dummy variable. The model allowed us to control for covariates that may also influence the outcome (i.e. examination mark). In regression analysis the covariates are included as independent (predictor) variables related to the dependent variable.

In the first part, two GLM models GLM I and GLM II with only main effects were constructed to evaluate the impact of partial grading (GLM I) versus no partial grading (GLM II) on examination performance. In the second part, a number of GLM models were built to investigate the dominance of Sem (semester mark) in the model. Finally, a model with interaction terms in addition to the main effects as in GLM I was constructed to model the effect of clicker use on examination performance for partial grading in 2017.

We identified 11 plausible covariates that were likely to have an effect on the examination mark, in addition to clicker use. We could not have the instructor role as a covariate in the models, because STK110 is a service course in many faculties, and it becomes impossible to have a timetable that suit all the students' needs, especially if they repeat STK110. Therefore, students sometimes attend more than one and up to four lecturers' classes.

The variables Gr 12 Math, Gr 12 Eng, AP Score and Sem (explained in Table 14) have been centred around the mean, because they do not contain a meaningful value of zero and with interaction terms being included in the model, centring can avoid multicollinearity issues. The assumptions for all the GLM analyses were tested and met (See Annexure 7.)

Table 14: Description of variables identified for GLM models

Variable	Type	Possible values
Dependent: Exam (examination mark)	Quantitative	0 – 100%
Independent: Clicker (clicker use)	Categorical, nominal	0 (No Clicker 2014) 1 (Clicker 2017)
Covariates:		
Sem (semester mark)	Quantitative	30 – 100%
Gr 12 Math (Grade 12 mathematics mark)	Quantitative	60 – 100%
Gr 12 Eng (Grade 12 English mark)	Quantitative	45 – 100%
Mtongue (mother tongue – home language versus tuition language)	Categorical, nominal	0 (Non-Mtongue - home language different from tuition language) 1 (Mtongue - home language same as tuition language)
AP Score (based on the six best Grade 12 marks excluding life orientation)	Quantitative	26 – 42
YrsReg (Years registered – the number of years a STK110 student has been registered at the university)	Quantitative	1 – 6
STK contact time (programme-dependent exposure to statistics)	Categorical, ordinal	1 (Six months) 2 (Nine to twelve months) 3 (Two years) 4 (Three years)
Class section (tuition language preference)	Categorical, nominal	0 (Eng class) 1 (Afr class)
Gender	Categorical, nominal	0 (Male) 1 (Female)
Ethnicity	Categorical, nominal	1 (African) 2 (White) 3 (other - Indian and mixed ethnicity)
Residence (stay in residence or not)	Categorical, nominal	0 (Not in res - Stay elsewhere) 1 (In res - Stay in residence)

Part I: Effect of partial grading

Two models were constructed: the GLM I (with partial grading) and GLM II (without partial grading) models explained 67.9% ($R^2 = 0.679$) and 67.6% ($R^2 = 0.676$) of the variation in the dependent variable and examination mark respectively. The coefficient of determination, R^2, can be interpreted as the percentage variability in the examination marks that can be explained by the predictors in the regression model. A stepwise procedure was followed to identify the significant covariates in the model. Five of the 11 covariates remained in both the GLM models, together with the independent variable of interest, i.e. clicker use. The results of GLM I are displayed in Table 15 and the results of GLM II in Table 16.

Table 15: Results of GLM I (with partial grading) in 2017

	Coefficients^a				
	Unstandardised Coefficients				
Model	$\boldsymbol{\beta}$	**Std. Error**	**Std.** $\boldsymbol{\beta}$	**t**	**Sig.**
Constant	57.622	0.686		82.105	< 0.0001
$\boldsymbol{X_{1,1}}$**: Clicker**	$\mathbf{3.337}$	0.388	0.096	8.601	< 0.0001
$\boldsymbol{X_{1,2}}$**: No Clicker**	0				
$\boldsymbol{X_2}$**: Sem**	0.866	0.016	0.744	52.984	< 0.0001
$\boldsymbol{X_3}$**: Gr12 Math**	0.175	0.030	0.090	5.844	< 0.0001
$\boldsymbol{X_{4,1}}$**: Mtongue**	-1.329	0.384	0.038	3.462	0.001
$\boldsymbol{X_{4,2}}$**: Non-Mtongue**	0				
$\boldsymbol{X_5}$**: AP Score**	0.239	0.077	0.049	3.097	0.002
$\boldsymbol{X_6}$**: YrsReg**	1.663	0.582	0.032	2.859	0.004
a. Dependent Variable: Examination mark ($R^2 = 0.679$)					

Table 16: Results of GLM II (without partial grading) in 2017

	Coefficients^a				
	Unstandardised Coefficients				
Model	$\boldsymbol{\beta}$	**Std. Error**	**Std.** $\boldsymbol{\beta}$	**t**	**Sig.**
Constant	57.330	0.693		80.858	< 0.0001
$\boldsymbol{X_{1,1}}$**: Clicker**	$\mathbf{2.361}$	0.396	0.068	6.002	< 0.0001
$\boldsymbol{X_{1,2}}$**: No Clicker**	0				
$\boldsymbol{X_2}$**: Sem**	0.869	0.016	0.746	52.687	< 0.0001
$\boldsymbol{X_3}$**: Gr12 Math**	0.183	0.030	0.094	6.070	< 0.0001
$\boldsymbol{X_{4,1}}$**: Mtongue**	-1.286	0.388	0.037	3.319	0.001

	Coefficients^{a}				
Model	**Unstandardised Coefficients**				
	$\boldsymbol{\beta}$	**Std. Error**	**Std.** $\boldsymbol{\beta}$	**t**	**Sig.**
$\boldsymbol{X_{4,2}}$**: Non-Mtongue**	0				
$\boldsymbol{X_5}$**: AP Score**	0.219	0.078	0.045	2.817	0.005
$\boldsymbol{X_6}$**: YrsReg**	1.560	0.588	0.030	2.654	0.008
a. Dependent Variable: Examination mark ($R^2 = 0.676$)					

The general linear model without partial grading is given by the following equation:

$$\hat{Y} = 57.330 + 2.361X_{1,1} + 0.869X_2 + 0.183X_3 - 1.286X_{4,1} + 0.219X_5 + 1.56X_6 \quad (5)$$

The most important finding from GLM II was that the mean examination mark for the 2017 students who used QT-clickers was 2.36% higher compared to the 2014 students who did not use QT-clickers, ceteris paribus. The unstandardised β-coefficients for the covariates of the GLM I model were comparable to the unstandardised β-coefficients for the GLM II model except for the clicker coefficient. The GLM I clicker regression coefficient for partial grading in 2017 was 3.337, which is almost 1% higher than without partial grading. It was found that students gained on average between 1% to 10% extra for a semester test or an examination paper when partially graded. The observed difference between the 2014 and non-partially graded 2017 scores could be due to the pedagogical advantage of QT-clickers, apart from partial grading.

Not surprisingly, the semester mark (term mark) was the best predictor of the examination mark in terms of its relative contribution to explaining the variance in the dependent variable (largest standardised regression coefficient in both models). For each 1% increase in the semester mark, the examination mark increased by 0.87%, controlling for all other predictors in the model, i.e. keeping the other terms constant. Following the semester mark, the use of QT-clickers was the next most important predictor for GLM I, followed by Gr 12 Math and vice versa for GLM II. This is an indicator that the use of QT-clickers is a key predictor of student performance as measured by the examination mark.

Part II: Dominance of the predictor Sem (*semester mark*) *for 2017 with partial grading*

It is obvious that Sem (semester mark) will be the stongest predictor of examination performance, well before exploring the data to assess the influence of the clickers and identify significant covariates that should be controlled for.

We constructed four additional GLM models to gain a better understanding of the role of Sem. Specifically, Sem was dropped from the model and only Clicker used as a predictor; thereafter, only Sem was included as a predictor in the model; finally, a model was constructed including all the identified covariates but excluding Sem. The results of GLM III, IV and V are captured in Table 17; the final model is GLM VI with interaction terms (cf. Table 18).

Table 17: Summary of three of the four additional GLM models

Model	**Dep variable**	**Covariates**	**Unstd $\boldsymbol{\beta}$**	**Sig.**	R^2	**Comment**
GLM III	Exam	Clicker	2.030	0.003	0.003	R^2 is poor
GLM IV	Exam	Sem	0.942	< 0.0001	0.655	R^2 is good for only Sem in the model
GLM V	Exam	All GLM I covariates except Sem			0.336	R^2 is lower than in GLM I

It is evident from GLM III that the mean examination mark for the 2017 students who used QT-clickers was 2.03% higher compared to the 2014 students who did not use QT-clickers.

We believe that the effect of clicker use was more evident towards the end of the semester and hence only truly quantifiable in the examination marks as indicated by GLM III.

For GLM IV we see that $R^2 = 0.655$, confirming the strong association between students' performance during the semester and their examination. This lower coefficient of determination $R^2 = 0.336$ for GLM V captures the importance of Sem in the model. The results of GLM V justified the investigation of a model with interaction terms, as it was evident that Sem plays a role, not only in predicting the examination mark, but it also influenced the other covariates' role in the model.

A model with all second and third order interactions was therefore built. Four second order interaction terms, namely Sem*Clicker, Sem*APS, Clicker*Mtongue and YrsReg*Math and none of the third order interaction tems were found to be significant. The results of GLM VI model with all the significant interaction terms are displayed in Table 18. The assumptions for GLM VI were tested and met (see Annexure 7 for complete output).

Table 18: Results of GLM VI (interaction model) with partial grading in 2017

	$\mathbf{Coefficients}^a$				
	Unstandardised Coefficients				
Model	$\boldsymbol{\beta}$	**Std. Error**	**Std.** $\boldsymbol{\beta}$	**t**	**Sig.**
Constant	56.726	0.728		77.877	< 0.0001
$\boldsymbol{X_{1,1}}$: **Clicker**	4.005	0.543	0.115	7.368	< 0.0001
$\boldsymbol{X_{1,2}}$: **No Clicker**	0				
$\boldsymbol{X_2}$: **Sem**	0.820	0.020	0.705	41.690	< 0.0001
$\boldsymbol{X_3}$: **Gr12 Math**	0.396	0.080	0.204	4.928	< 0.0001
$\boldsymbol{X_{4,1}}$: **Mtongue**	-0.673	0.515	-0.019	-1.306	0.192
$\boldsymbol{X_{4,2}}$: **Non-Mtongue**	0				
$\boldsymbol{X_5}$: **AP Score**	0.272	0.077	0.056	3.552	< 0.0001
$\boldsymbol{X_6}$: **YrsReg**	2.009	0.584	0.039	3.441	0.001
$\boldsymbol{X_{1,1}X_2}$: **Clicker*Sem**	0.106	0.026	0.060	4.026	< 0.0001
$\boldsymbol{X_{1,2}X_2}$: **No Clicker*Sem**	0				
$\boldsymbol{X_{1,1}X_{4,1}}$: **Clicker*Mtongue**	-1.520	0.768	-0.036	-1.979	0.048
All other combinations	0				
$\boldsymbol{X_3X_6}$: **Gr12 Math*YrsReg**	-0.225	0.069	-0.129	-3.247	0.001
$\boldsymbol{X_2X_5}$: **Sem*AP Score**	0.008	0.004	0.024	2.080	0.038
a. Dependent Variable: Examination mark ($R^2 = 0.683$)					

The general linear model is given by the following equation:

$$\hat{Y} = 56.726 + 4.005X_{1,1} + 0.82X_2 + 0.396X_3 - 0.673X_{4,1} + 0.272X_5 + 2.009X_6 + 0.106X_{1,1}X_2 - 1.52X_{1,1}X_{4,1} - 0.225X_3X_6 + 0.008X_2X_5 \quad (6)$$

The main objective was to interpret the effect of clickers on examination performance for both cohorts. Consider the following four options that the general linear equation reduces to:

1. No Clicker and Non-Mtongue:
   $\hat{Y} = 56.726 + (0.82 + 0.008X_5)X_2 + 0.396X_3 + 0.272X_5 + 2.009X_6 - 0.225X_3X_6$
2. No Clicker and Mtongue:

$$\hat{Y} = 56.053 + (0.82 + 0.008X_5)X_2 + 0.396X_3 + 0.272X_5 + 2.009X_6 - 0.225X_3X_6$$

3. Clicker and Non-Mtongue:

$$\hat{Y} = 60.731 + (0.926 + 0.008X_5)X_2 + 0.396X_3 + 0.272X_5 + 2.009X_6 - 0.225X_3X_6$$

4. Clicker and Mtongue:

$$\hat{Y} = 58.538 + (0.926 + 0.008X_5)X_2 + 0.396X_3 + 0.272X_5 + 2.009X_6 - 0.225X_3X_6$$

For the 2017 students who used clickers and were not educated in their mother tongue, there was a shift of approximately 4% in the intercept, with an accompanying increase of 0.106 in Sem (X_2), compared to the 2014 students who did not use clickers and were not educated in their mother tongue. Interaction terms Clicker*Sem and Sem*APS are of importance for interpreting the effect of the semester mark on the examination mark in combination with other covariates; however, both these two terms had a small impact on the examination mark. When comparing the two partial grading models, the distinct effect of Clicker on the examination mark was noticeable, i.e. for GLM I the coefficient was 3.337 and for GLM VI it was 4.005.

The estimated β-coefficient for Clicker*Mtongue of -1.52 captured the large effect of this interaction on the examination mark. It was evident that the 2017 students with clickers and non-mother tongue tuition, on average had a higher examination mark than those with mother tongue tuition. It could be that many students were already used to non-mother tongue tuition at school level, or students tended to work harder because of the language barrier at university level.

4.3.2 IMPACT OF THE FIVE TECHNOLOGY INTERVENTIONS PER COHORT

The five cohorts 2011 (traditional model), 2012 (traditional model with online homework), 2013 (flipped classroom model), 2015 (flipped classroom model with QT-clickers) and 2018 (flipped classroom model with QT-clickers and peer learning activities) based only on the prerequisite sample were used to construct a GLM to predict the examination mark.

In Table 19 below it is noticeable that the five cohorts were similar regarding their mean AP Score, Gr 12 Mathematics and Gr 12 English marks. Their STK110 mean semester marks were not all similar. The variables Gr 12 Math, Gr 12 Eng, APS score and Sem (cf. Table 19) have been group centered. The assumptions for GLM VII were tested and met. See Annexure 8 for complete output.

Table 19: The mean Grade 12 marks and STK110 semester marks for the five cohorts

Cohort	AP Score	Gr 12 Mathematics	Gr 12 English	STK110 Semester
2011	35.40	75.27	72.57	**54.81**
2012	35.47	73.05	73.32	**59.59**
2013	35.78	72.54	74.09	**66.08**
2015	35.82	74.21	73.44	**60.30**
2018	35.89	73.56	74.12	**63.91**

A model with all second and third order interactions was built. The few significant third order interactions had very small effect sizes, hence, for the sake of parsimony, it was decided to omit them from the model.

A summary of the categorical variables in the GLM VII model for the 7147 (7168 excluding 21 students without Gr 12 English marks) students who met the prerequisites in five cohorts is given in Table 20.

Table 20: Summary of categorical variables in the GLM VII model

Variable	Frequency (%)
Cohort 2011	1468 (20.5)
Cohort 2012	1342 (18.8)
Cohort 2013	1445 (20.2)
Cohort 2015	1422 (19.9)
Cohort 2018	1470 (20.6)
African	2475 (34.6)
White	4030 (56.4)
Other (Indian & Coloured)	642 (9.0)
Eng class	5432 (76.0)
Afr class	1715 (24.0)
Mtongue	3685 (51.6)
Non-Mtongue	3462 (48.4)
Not in res	4847(67.8)
In res	2300 (32.2)

The results of GLM VII model with all the significant interaction terms are displayed in Table 21.

Table 21: Results of GLM VII for the five cohorts

	Coefficients^a				
Model	**Unstd. Coefficients**				
	$\boldsymbol{\beta}$	**Std. Error**	**t**	**Sig.**	**95% Confidence Interval**
Constant	52.303	1.795	29.142	< 0.0001	(48.785 ; 55.822)
$\boldsymbol{X_{1,1}}$**: Cohort 2018**	3.230	2.231	1.448	0.148	(−1.143 ; 7.604)
$\boldsymbol{X_{1,2}}$**: Cohort 2015**	9.805	2.051	4.781	< 0.0001	(5.785 ; 13.825)
$\boldsymbol{X_{1,3}}$**: Cohort 2013**	4.076	2.054	1.987	0.047	(0.054 ; 8.098)
$\boldsymbol{X_{1,4}}$**: Cohort 2012**	−0.168	2.148	−0.078	0.938	(−4.378 ; 4.042)
$\boldsymbol{X_{1,5}}$**: Cohort 2011**	0				
$\boldsymbol{X_{2,1}}$**: African**	3.791	1.275	2.974	0.003	(1.292; 6.290)
$\boldsymbol{X_{2,2}}$**: White**	−0.200	1.129	−0.178	0.859	(−2.413 ; 2.012)
$\boldsymbol{X_{2,3}}$**: other**	0				
$\boldsymbol{X_{3,1}}$**: Eng class**	4.045	1.197	3.379	0.001	(1.699 ; 6.392)
$\boldsymbol{X_{3,2}}$**: Afr class**	0				
$\boldsymbol{X_{4,1}}$**: Mtongue**	2.899	1.060	2.735	0.006	(0.821 ; 4.977)
$\boldsymbol{X_{4,2}}$**: Non-Mtongue**	0				
$\boldsymbol{X_{5,1}}$**: Not in res**	0.224	0.613	0.365	0.715	(−0.978 ; 1.426)
$\boldsymbol{X_{5,2}}$**: In res**	0				
$\boldsymbol{X_6}$**: YrsReg**	4.544	0.903	5.031	< 0.0001	(2.773 ; 6.314)
$\boldsymbol{X_7}$**: Sem**	0.808	0.023	35.251	< 0.0001	(0.763 ; 0.853)
$\boldsymbol{X_8}$**: Gr 12 Eng**	−0.064	0.023	−2.753	0.006	(−0.110 ; −0.019)
$\boldsymbol{X_9}$**: AP Score**	0.481	0.059	8.195	< 0.0001	(0.366 ; 0.596)
$\boldsymbol{X_{10}}$**: Gr 12 Math**	0.081	0.036	2.267	0.023	(0.011 ; 0.152)
$\boldsymbol{X_7X_9}$**: Sem*AP Score**	0.005	0.002	2.279	0.023	(0.001 ; 0.010)
$\boldsymbol{X_{1,1}X_{3,1}}$**: 2018*Eng class**	−0.975	1.259	−0.774	0.439	(−3.444 ; 1.494)
$\boldsymbol{X_{1,2}X_{3,1}}$**: 2015*Eng class**	−0.483	1.040	−0.464	0.643	(−2.520 ; 1.555)
$\boldsymbol{X_{1,3}X_{3,1}}$**: 2013*Eng class**	2.112	1.019	2.072	0.038	(0.114 ; 4.110)
$\boldsymbol{X_{1,4}X_{3,1}}$**: 2012*Eng class**	−1.459	1.027	−1.421	0.155	(−3.472 ; 0.554)
$\boldsymbol{X_{1,5}X_{3,1}}$**: 2011*Eng class**	0				
All cohorts*Afr class	0				
$\boldsymbol{X_{3,1}X_6}$**: Eng class *YrsReg**	−2.800	0.885	−3.165	0.002	(−4.534 ; −1.066)

$X_{3,2}X_6$: Afr class *YrsReg	0				
$X_{1,1}X_{10}$: 2018*Math	-0.002	0.049	-0.038	0.970	$(-0.098\ ;\ 0.094)$
$X_{1,2}X_{10}$: 2015*Math	-0.062	0.051	-1.233	0.217	$(-0.161\ ;\ 0.037)$
$X_{1,3}X_{10}$: 2013*Math	0.119	0.050	2.368	0.018	$(0.021\ ;\ 0.218)$
$X_{1,4}X_{10}$: 2012*Math	0.026	0.051	0.515	0.607	$(-0.074\ ;\ 0.127)$
$X_{1,5}X_{10}$: 2011*Math	0				
$X_{1,1}X_{4,1}$: 2018*Mtongue	1.740	0.975	1.784	0.074	$(-0.172\ ;\ 3.652)$
$X_{1,2}X_{4,1}$: 2015*Mtongue	-1.073	0.989	-1.084	0.278	$(-3.012\ ;\ 0.867)$
$X_{1,3}X_{4,1}$: 2013*Mtongue	-0.974	1.018	-0.957	0.339	$(-2.969\ ;1.021)$
$X_{1,4}X_{4,1}$: 2012*Mtongue	0.870	1.040	0.836	0.403	$(-1.170\ ;2.909)$
$X_{1,5}X_{4,1}$: 2011*Mtongue	0				
All cohorts*Non-Mtongue	0				
$X_{1,1}X_{2,1}$: 2018*African	-0.158	1.705	-0.092	0.926	$(-3.499\ ;3.184)$
$X_{1,2}X_{2,1}$: 2015*African	-0.776	1.699	-0.457	0.648	$(-4.107\ ;2.555)$
$X_{1,3}X_{2,1}$: 2013*African	-4.776	1.741	-2.743	0.006	$(-8.189\ ;-1.363)$
$X_{1,4}X_{2,1}$: 2012*African	-3.862	1.838	-2.102	0.036	$(-7.465\ ;-0.260)$
$X_{1,5}X_{2,1}$: 2011*African	0				
$X_{1,1}X_{2,2}$: 2018*White	1.238	1.510	0.820	0.412	$(-1.722\ ;4.197)$
$X_{1,2}X_{2,2}$: 2015*White	0.661	1.495	0.442	0.658	$(-2.269\ ;3.591)$
$X_{1,3}X_{2,2}$: 2013*White	0.728	1.543	0.472	0.637	$(-2.298\ ;3.754)$
$X_{1,4}X_{2,2}$: 2012*White	1.298	1.630	0.796	0.426	$(-1.898\ ;4.493)$
$X_{1,5}X_{2,2}$: 2011*White	0				
All cohorts*other	0				
$X_{1,1}X_{5,1}$: 2018*Not in res	-1.742	0.874	-1.992	0.046	$(-3.455\ ;-0.028)$
$X_{1,2}X_{5,1}$: 2015*Not in res	-2.173	0.859	-2.529	0.011	$(-3.857\ ;-0.489)$
$X_{1,3}X_{5,1}$: 2013*Not in res	-0.231	0.861	-0.269	0.788	$(-1.920\ ;1.457)$
$X_{1,4}X_{5,1}$: 2012*Not in res	-1.798	0.885	-2.032	0.042	$(-3.533\ ;-0.064)$
$X_{1,5}X_{5,1}$: 2011*Not in res	0				
All cohorts*In res	0				
$X_{1,1}X_7$: 2018*Sem	0.089	0.031	2.897	0.004	$(0.029\ ;0.149)$
$X_{1,2}X_7$: 2015*Sem	0.023	0.031	0.757	0.449	$(-0.037\ ;0.084)$
$X_{1,3}X_7$: 2013*Sem	0.043	0.032	1.333	0.182	$(-0.020\ ;0.106)$
$X_{1,4}X_7$: 2012*Sem	-0.067	0.033	-2.055	0.040	$(-0.131\ ;-0.003)$
$X_{1,5}X_7$: 2011*Sem	0				
$X_{4,1}X_6$: Mtongue*YrsReg	-2.533	0.728	-3.478	0.001	$(-3.960\ ;-1.105)$
$X_{4,2}X_6$: Non-Mtongue*YrsReg	0				

a. Dependent Variable: Examination mark ($R^2 = 0.642$)

The data set was divided into a training (70%) and test (30%) data set using stratified random sampling, grouped according to the examination percentage intervals. The regression model in the training data set has a determination coefficient of $R^2 = 0.64$ and the regression model in the test data set a determination coefficient of $R^2 = 0.649$, although a few of the interaction terms in the regression model for the test set were insignificant.

The general linear model based on the complete data set is given by the following equation:

$$\hat{Y} = 52.303 + 3.23X_{1,1} + 9.805X_{1,2} + 4.076X_{1,3} - 0.168X_{1,4} + 3.791X_{2,1} - 0.2X_{2,2} + 4.045X_{3,1} + 2.899X_{4,1} + 0.224X_{5,1} + 4.544X_6 + 0.808X_7 - 0.064X_8 + 0.481X_9 + 0.081X_{10} + 0.005X_7X_9 - 0.975X_{1,1}X_{3,1} - 0.483X_{1,2}X_{3,1} + 2.112X_{1,3}X_{3,1} - 1.459X_{1,4}X_{3,1} - 2.8X_{3,1}X_6 - 0.002X_{1,1}X_{10} - 0.062X_{1,2}X_{10} + 0.119X_{1,3}X_{10} + 0.026X_{1,4}X_{10} + 1.74X_{1,1}X_{4,1} - 1.073X_{1,2}X_{4,1} - 0.974X_{1,3}X_{4,1} + 0.87X_{1,4}X_{4,1} - 0.158X_{1,1}X_{2,1} - 0.776X_{1,2}X_{2,1} - 4.776X_{1,3}X_{2,1} - 3.862X_{1,4}X_{2,1} + 1.238X_{1,1}X_{2,2} + 0.661X_{1,2}X_{2,2} + 0.728X_{1,3}X_{2,2} + 1.298X_{1,4}X_{2,2} - 1.742X_{1,1}X_{5,1} - 2.173X_{1,2}X_{5,1} - 0.231X_{1,3}X_{5,1} - 1.798X_{1,4}X_{5,1} + 0.089X_{1,1}X_7 + 0.023X_{1,2}X_7 + 0.043X_{1,3}X_7 - 0.067X_{1,4}X_7 - 2.533X_{4,1}X_6 \quad (7)$$

The main objective was to interpret the effect of the interventions on examination performance. A linear equation for each cohort was reduced from the general linear model above and can be seen in the next five equations.

- 2011 cohort: Traditional model

$$\hat{Y} = 52.303 + 3.791X_{2,1} - 0.2X_{2,2} + 4.045X_{3,1} + 2.899X_{4,1} + 0.224X_{5,1} + (4.544 - 2.8X_{3,1} - 2.533X_{4,1})X_6 + (0.808 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.081X_{10} \quad (8)$$

- 2012 cohort: Traditional model with online homework

$$\hat{Y} = 52.135 - 0.071X_{2,1} + 1.098X_{2,2} + 2.586X_{3,1} + 3.769X_{4,1} - 1.574X_{5,1} + (4.544 - 2.8X_{3,1} - 2.533X_{4,1})X_6 + (0.741 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.107X_{10} \quad (9)$$

- 2013 cohort: Flipped classroom

$$\hat{Y} = 56.379 - 0.985X_{2,1} + 0.528X_{2,2} + 6.157X_{3,1} + 1.925X_{4,1} - 0.007X_{5,1} + (4.544 - 2.8X_{3,1} - 2.533X_{4,1})X_6 + (0.851 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.2X_{10} \quad (10)$$

- 2015 cohort: QT-clickers within flipped classroom

$$\hat{Y} = 62.108 + 3.015X_{2,1} + 0.461X_{2,2} + 3.562X_{3,1} + 1.826X_{4,1} - 1.949X_{5,1} + (4.544 - 2.8X_{3,1} - 2.533X_{4,1})X_6 + (0.831 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.019X_{10} \quad (11)$$

- 2018 cohort: Peer learning activities and QT-clickers within flipped classroom

$$\hat{Y} = 55.533 + 3.633X_{2,1} + 1.038X_{2,2} + 3.07X_{3,1} + 4.639X_{4,1} - 1.518X_{5,1} + (4.544 - 2.8X_{3,1} - 2.533X_{4,1})X_6 + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10} \quad (12)$$

It is evident that the technology-based interventions have an effect on students' performance. Seven of the ten interaction terms included the covariate cohort. This is an indication that the technology-based interventions had an influence on many of the predictors. The post-class online homework system implemented in 2012 substituted ungraded pen-and-paper homework and although there was a slight increase in the pass rate, it did not have a major influence on students' performance. The positioning of the online homework system embedded in the flipped classroom, made a noticeable impact on students' performance in 2013. Online homework became a preparation instead of consolidation tool. The implementation of QT-clickers in 2015, to enhance active learning in class, had a significant effect on student learning. In 2018 the flipped classroom was further refined with peer learning, South-African based in-class activities and switching of lecturers. It made a contribution to the learning process, as can be seen from the results.

The only interaction with two continuous variables was between AP Score and Sem (semester mark). An interaction plot in Figure 19 was set up for the 2015 cohort and **c**an be used to clarify the relationship. The AP Scores considered for the interaction plot were divided into three categories, namely low, medium and high. Students with a minimum score of 26, a medium a score of 34 and the maximum score of 42 were taken into account (note that the AP Scores are centred at the mean of 35.82). The relationship between semester mark and examination mark stayed positive and increased as the semester mark increased from 60% to 100% (note that semester mark is centred at the mean of 74.21% for Figure 19). It is evident from Figure 19 that students with a low or medium AP Score's examination mark lagged the examination mark of students with a high AP Score by 5% to 16% as the semester mark increased. The other four cohorts' interaction plots were similar to 2015 for the same interaction term.

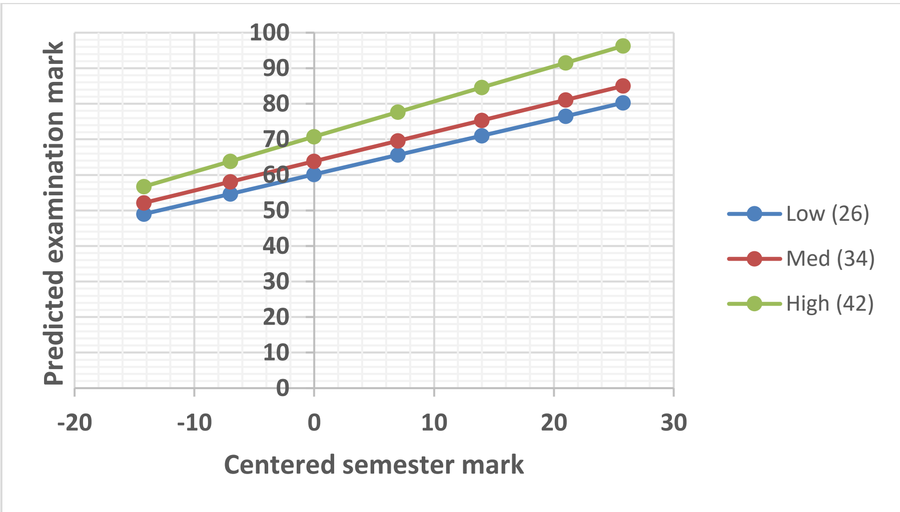


Figure 19: Interaction plot of the interaction term between AP Score and Sem for 2015

The core categorical predictors that played a significant role in predicting the examination mark were Mother tongue (MTongue), Residence (Not in res), Class section (Eng class) and Ethnicity (African,White).

All cohorts had a common coefficient for YrsReg(X_6), namely $(4.544 - 2.8X_{3,1} - 2.533X_{4,1})X_6$, which depended on the Class section and Mother tongue. For a mother tongue student in the English class the term reduced to $-0.789X_6$, which implies that for a one unit increase in the number of years registered, the average predicted examination mark would decrease by 0.789, keeping all other terms constant. If the other terms are kept constant, it is possible to explain the effect that changes in a specific predictor, i.e. number of years registered, have on the dependent variable, i.e. average predicted examination mark, without also having to take the effects of the other predictors into account. If a student was either non-mother tongue and/or in the Afrikaans class the coefficient of X_6 became positive, i.e. an increase in the average predicted examination mark, keeping all other terms constant.

The three key categorical predictors, i.e. Mother tongue (MTongue=1/Non-MTongue=0), Class section (Eng class=1/Afr class=0), and Residence (Not in res=1/In res=0). can be substituted into the 2018 cohort's equation (13) to demonstrate their role played in the change of the constant (in blue) and the coefficient of YrsReg(X_6) (in red). The eight possible combinations with three predictors each with two options are as follows:

1. MTongue = 1; Eng class = 1 and Not in res = 1

$$\hat{Y} = 61.724 - 0.789X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

2. MTongue = 1; Eng class = 1 and In res = 0

$$\hat{Y} = 63.242 - 0.789X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

3. MTongue = 1; Afr class = 0 and Not in res = 1

$$\hat{Y} = 58.654 + 2.011X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

4. MTongue = 1; Afr class = 0 and In res = 0

$$\hat{Y} = 60.172 + 2.011X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

5. Non-MTongue = 0; Eng class = 1 and Not in res = 1

$$\hat{Y} = 57.085 + 1.744X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

6. Non-MTongue = 0; Eng class = 1 and In res = 0

$$\hat{Y} = 58.603 + 1.744X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

7. Non-MTongue = 0; Afr class = 0 and Not in res = 1

$$\hat{Y} = 54.015 + 4.544X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

8. Non-MTongue = 0; Afr class = 0 and In res = 0

$$\hat{Y} = 55.533 + 4.544X_6 + 3.633X_{2,1} + 1.038X_{2,2} + (0.897 + 0.005X_9)X_7 - 0.064X_8 + 0.481X_9 + 0.079X_{10}$$

As an example, for the 2018 cohort, a comparison of mother tongue versus non-mother tongue students in the English class and in residence was as follows:

For the students educated in their mother tongue (Eq. 2), there was a shift of approximately 4.6% in the intercept, with an accompanying decrease of approximately 2.5% in YrsReg (X_6), compared to their counterparts not educated in their mother tongue (Eq. 6). In other words, the estimated examination mark for mother tongue students who were registered for one year was approximately 2.1% higher than their non-mother tongue counterparts. Approximately 91% of students are registered for one year. If students are registered for more than one year, the estimated examination mark of non-mother tongue students are higher than mother tongue students, e.g. for students registered for two years, the estimated examination mark for mother tongue students is approximately 0.4% lower than their non-mother tongue counterparts.

For students who were taught in their mother tongue the predicted examination marks were slightly higher than their counterparts in 2011 and 2012, and moderately higher than their counterparts in 2018, if they were only registered for one year. For the few students who were registered for more than one year, mother tongue education did not contribute to a higher predicted examination mark.

In the literature, mother tongue education is promoted, e.g. students educated in their mother tongue at university have an advantage over students who are not (Nyika, 2015). In reality, for two of the five intervention cohorts, namely 2011 and 2015, students not educated in their mother tongue on average marginally outperformed their mother tongue peers in the examination. The reason behind it could be that many of the students are educated from an early age in English, so that it becomes second nature.

The model predicted the English class students to perform better than the Afrikaans class students in 2011 and 2013, and marginally in 2015 and 2018, but in reality, the observed average examination marks showed that the English class students performed slightly better than the Afrikaans class students in 2011, 2015 and 2018, and vice versa in 2012 and 2013. It should also be considered that the size of the Afrikaans class decreased from approximately 46% to below 10% of the English class's size from 2011 to 2018. Many Afrikaans students preferred to take the module in the English medium, because they knew the business sector is English dominated. The predicted average examination marks for students in residence were higher compared to students who did not stay in residence for all students, but underestimated by the model compared to observed examination marks, i.e., students in residence outperformed students not in residence in the five intervention cohorts.

The scatter plot for the five intervention cohorts was overcluttered, therefore only the 2018 cohort is presented with a 45˚-line superimposed on the scatter plot in Figure 20. The scatter plot implies that the model tends to overestimate lower examination marks and slightly underestimates higher marks. A possible reason for overestimating lower examination scores could be that the semester mark was a key predictor in the regression model, and it was easier for a student to obtain a higher semester mark compared to the examination mark, because the semester mark consisted of both formative and summative assessment scores. If a student did not put in a sufficient effort to prepare for the examination, there would be a discrepancy between the actual examination mark and the predicted examination mark. Underestimation on high marks could also be due to high achievers who went the extra mile and put in an additional effort before the examination and being rewarded with higher marks than anticipated.

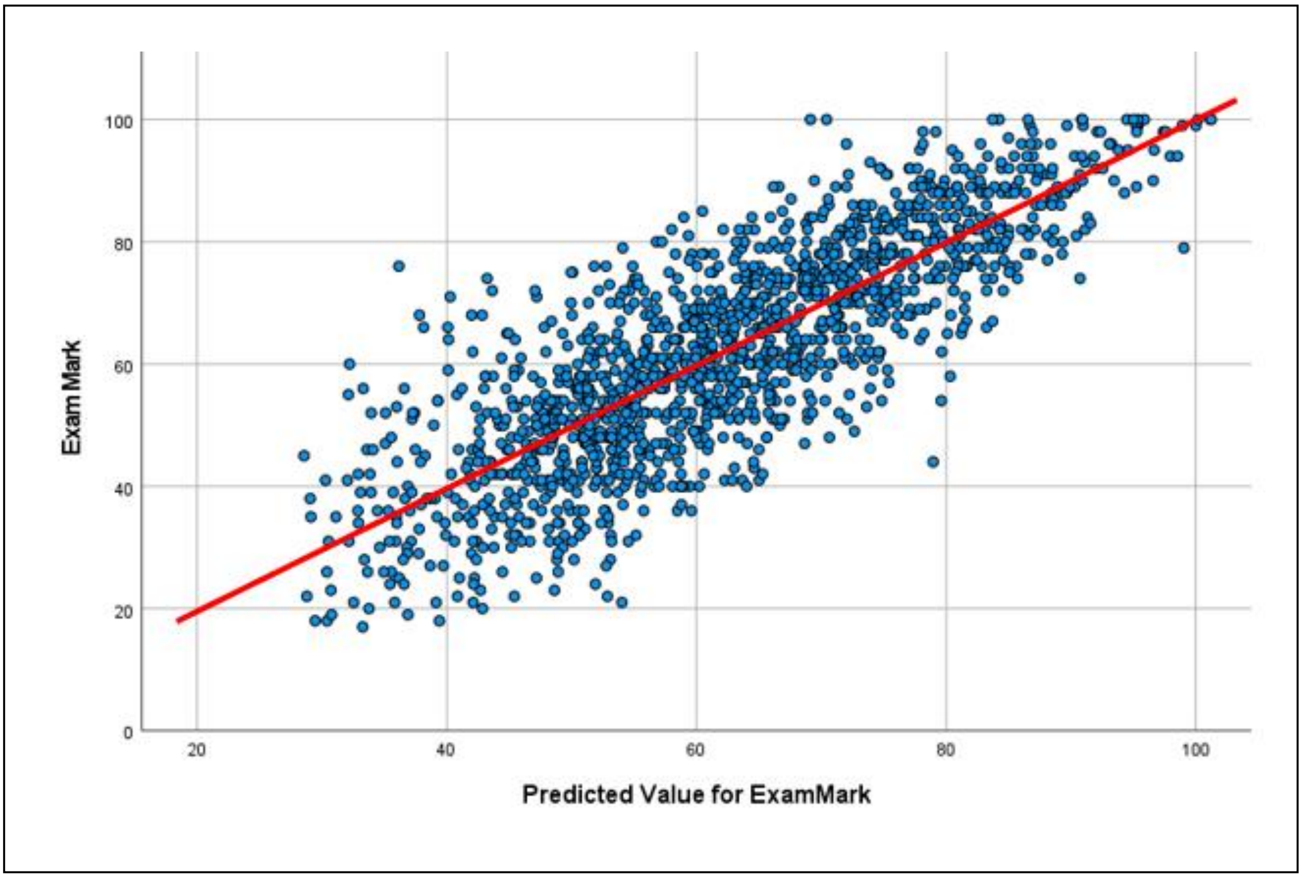


Figure 20: Scatter plot of examination marks by the predicted values of examination mark for the 2018 cohort

4.4 INFERENTIAL STATISTICS: MULTIVARIATE REGRESSION WITH XGBOOST

XGBoost (XGB) was used to build regression trees for the prerequisite sample of students.-A comparison of the results with XGB was be made with GLM VII in Section 4.4.4. A regression tree is a variant of a decision tree due to its continuous target or dependent variable, i.e., examination mark. The target variable's values were predicted by the various predictors (features) included in the model.

4.4.1 MODEL PARAMETERS

XGBoost has several parameters that need to be carefully chosen to maximise the model's performance. Parameter tuning is an essential part of XGBoost to avoid overfitting, but it can be a daunting task because it uses multiple parameters. The general parameter, booster, is used to specify the type of model to run at each iteration. The booster can either be tree or linear learners. A Monte Carlo study with 100 simulated samples showed that tree learners are superior to the linear learners for regression problems (Muller, 2018). The number of iterations specifies how many trees should be fitted into the model. The parameters that guided the booster at each step that was tuned in this study, also known as the booster parameters, were colsample_bytree, eta, gamma, max_depth, min_child_weight and subsample. Table 22 provides a list of hyper-parameter values that were evaluated during the XGBoost model building phase for the regression tree.

Table 22 : List of hyper-parameter values for XGBoost

Name	Description	Values Used
booster	Type of model	gbtree
colsample_bytree	Subsample ratio of features used to fit individual tree	(0.5,0.8,1)
eta (η)	Learning rate	(0.01,0.3,1)
gamma (γ)	Minimum loss reduction for further partition	(0.0,0.2,1)
max_depth	Maximum depth of a tree	seq(5,10)
min_child_weight (w_{mc})	Minimum weights of the instances required in a leaf	seq(1,10)
subsample	Subsample ratio of the training instances used to fit the individual tree	(0.8,1)
n_rounds	Number of iterations	500

The best possible XGBoost hyper-parameter values selected after cross-validation for this study were booster: gbtree, number of iterations: 500, colsample_bytree: 0.8, eta: 0.01, gamma: 0.2, max_depth: 5, min_child_weight: 1, subsample: 1.

4.4.2 MODEL RESULTS

The prerequisite dataset used in the regression consisted of 7168 observations. The split between the different cohorts can be observed in Table 23.

Table 23: Split of the five cohorts for the prerequisite sample

Cohorts	Frequency	Percentage
2011	1472	20.5
2012	1346	18.8
2013	1448	20.2
2015	1427	19.9
2018	1475	20.6
Total	7168	100

The values in the 80:20 samples were randomly selected proportional to each cohort. The model was trained on 80% of the data and the remaining 20% was used to test the model. The final performance metrics achieved for the test sample were:

RMSE = 10.445

R^2 = 0.643

An example of a regression tree (number 499) or the prerequisite sample is presented in Figure 21. Regression tree number 499 in Figure 21 has taken the residual errors of all the previous regression trees into consideration, therefore the 500 regression trees cannot be separated and should be interpreted as an ensemble. For the sake of brevity, race was used in Figure 21 – 27 as a proxy for ethnicity. The XGBoostExplainer package in R can be used to predict a single student's examination mark, as can be seen in Figure 27.

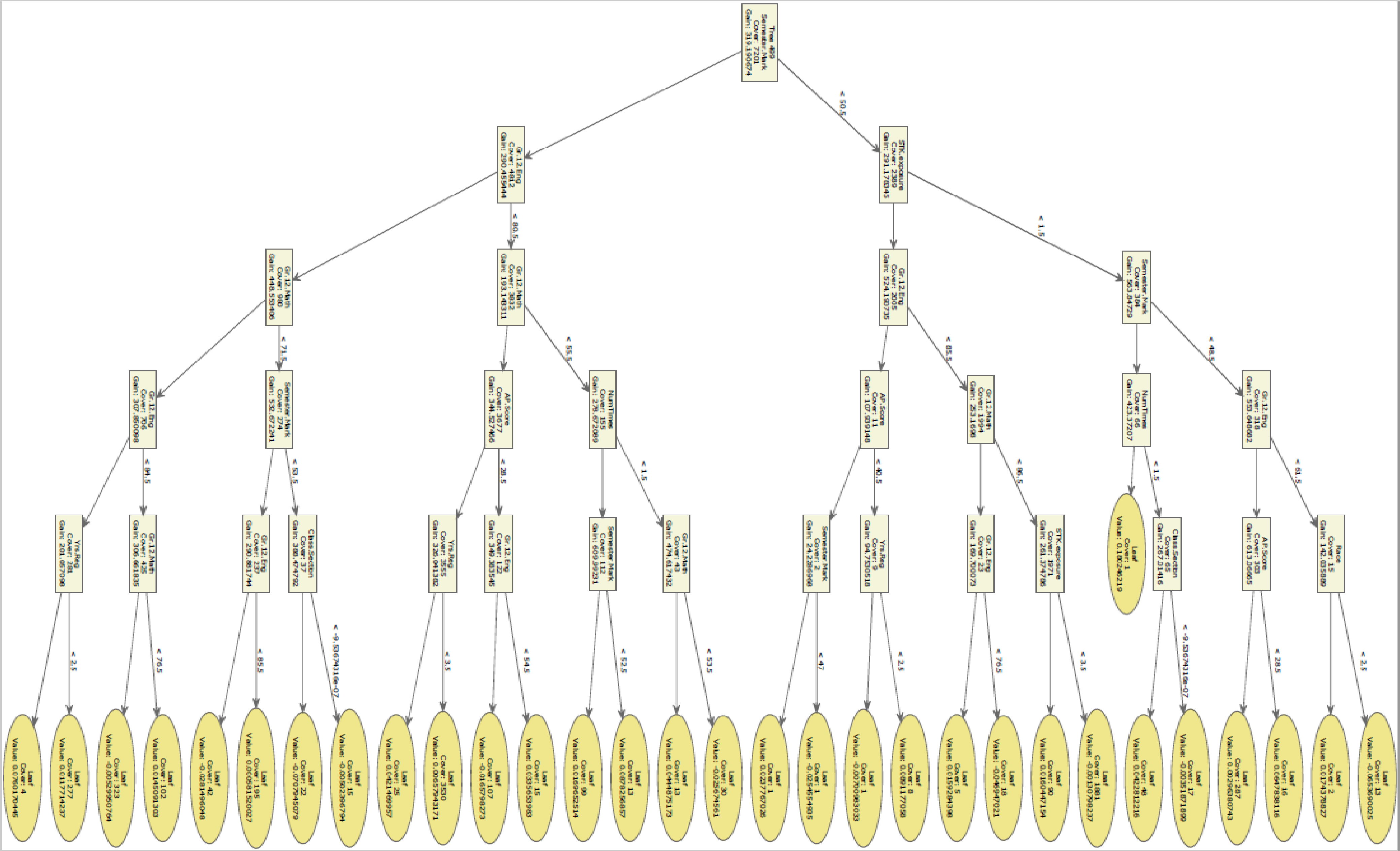


Figure 21: Example of a final regression tree for the prerequisite sample

4.4.3 FEATURE ANALYSIS AND INTERPRETATION

In Figure 22 a horizontal bar chart of the SHAP summary was plotted for predictors in descending order of importance to predict the examination mark for the regression tree.

It is notable that the semester mark was the main predictor with the greatest impact on the regression model, followed by the cohort, AP Score and Gr 12 Math. Apart from the semester mark, it implies the importance of the different technology interventions introduced in the five years, followed by Grade 12 prerequisites for STK110. Residence and Ethnicity were also of importance when the prerequisite sample was considered.

Figure 23 is an alternate representation of the SHAP summary plot that arranges variables (features) based on their importance to predict the examination mark. The variable value is colour coded; if the variable is binary, two colours are used, otherwise a colour gradient is used to denote all the values (cf. Figure 23). Clearly the lower (higher) the semester mark, the lower (higher) the examination mark will be. The variables AP Score and Gr 12 Math can be interpreted similarly to the semester mark. In general, the Grade 12 English marks were contradictory to the semester mark, Grade 12 mathematics mark and AP Score. On average, a higher Grade 12 English mark corresponded to a lower examination mark, but there were low (moderate) Gr 12 English marks that corresponded to low (moderate) examination marks.

It is difficult to interpret Figure 23 for the variable cohort. In Figure 24 the individual graph for the variable cohort is easier to understand. The SHAP value was positive for 2011 (Traditional model), then it dropped down to a negative value for 2012 (Traditional model with online homework), increased slightly, but still negative for 2013 (Flipped classroom), had a sharp increase to 4.5 for 2015 (QT-clickers with flipped classroom), and then dropped down to a negative value for 2018 (Peer learning activities, QT-clickers with flipped classroom). The pattern was rather unusual because it was expected that the examination marks of 2011 and 2012 should be similar with an increase in examination marks in 2013, 2015 and 2018.

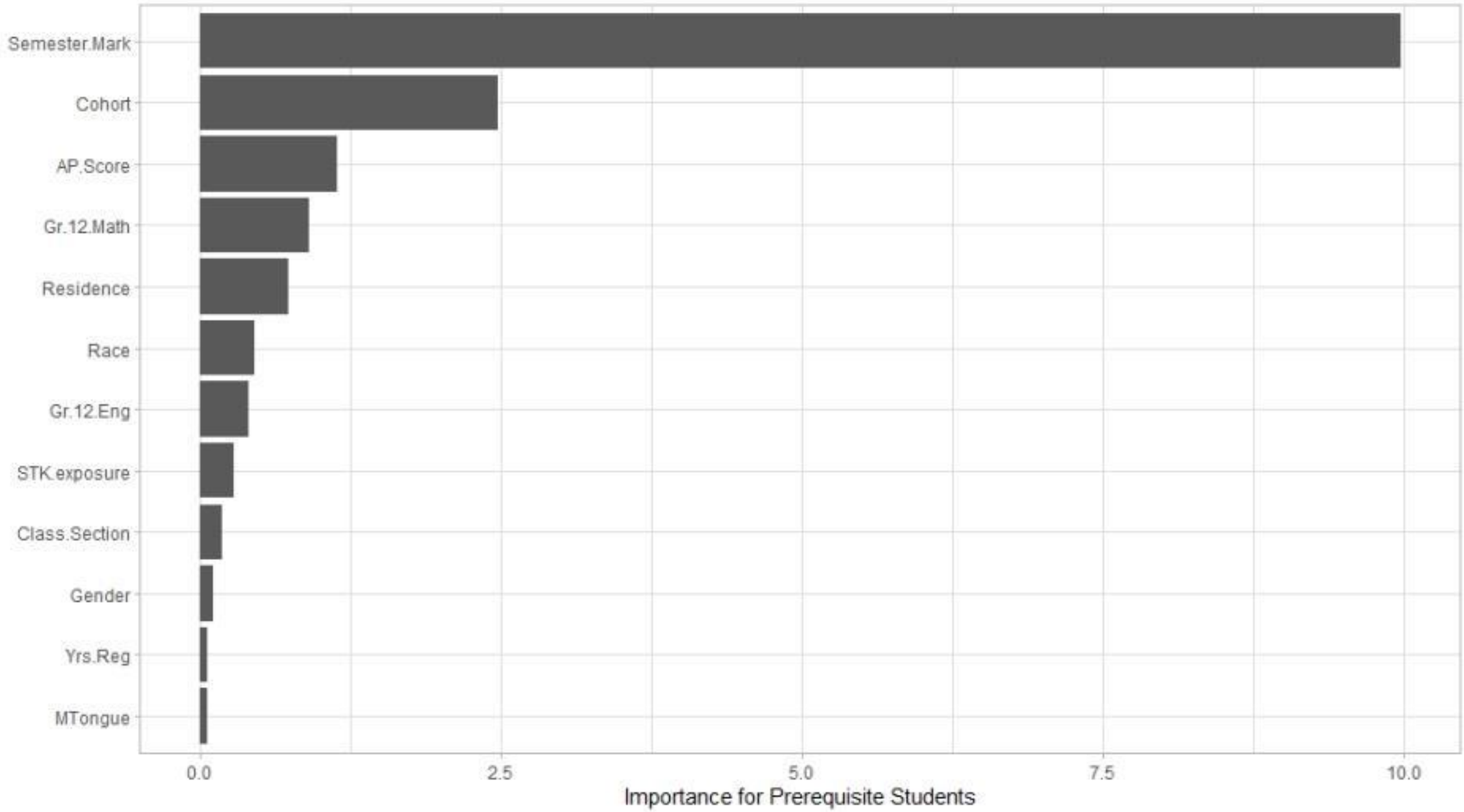


Figure 22: Bar chart of the twelve most important predictors of examination mark for the prerequisite sample

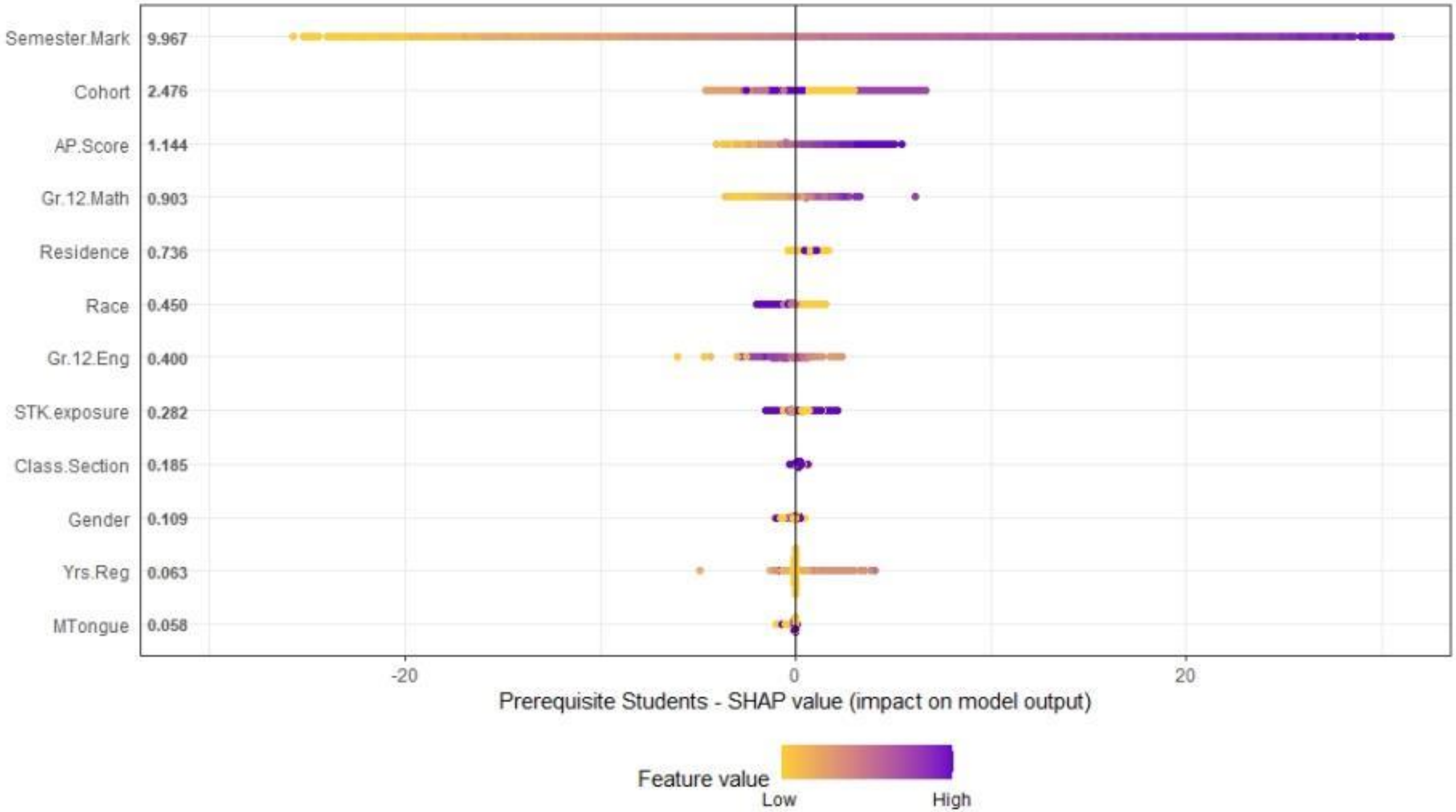


Figure 23: SHAP summary plot with most important predictors to predict examination mark for the prerequisite sample

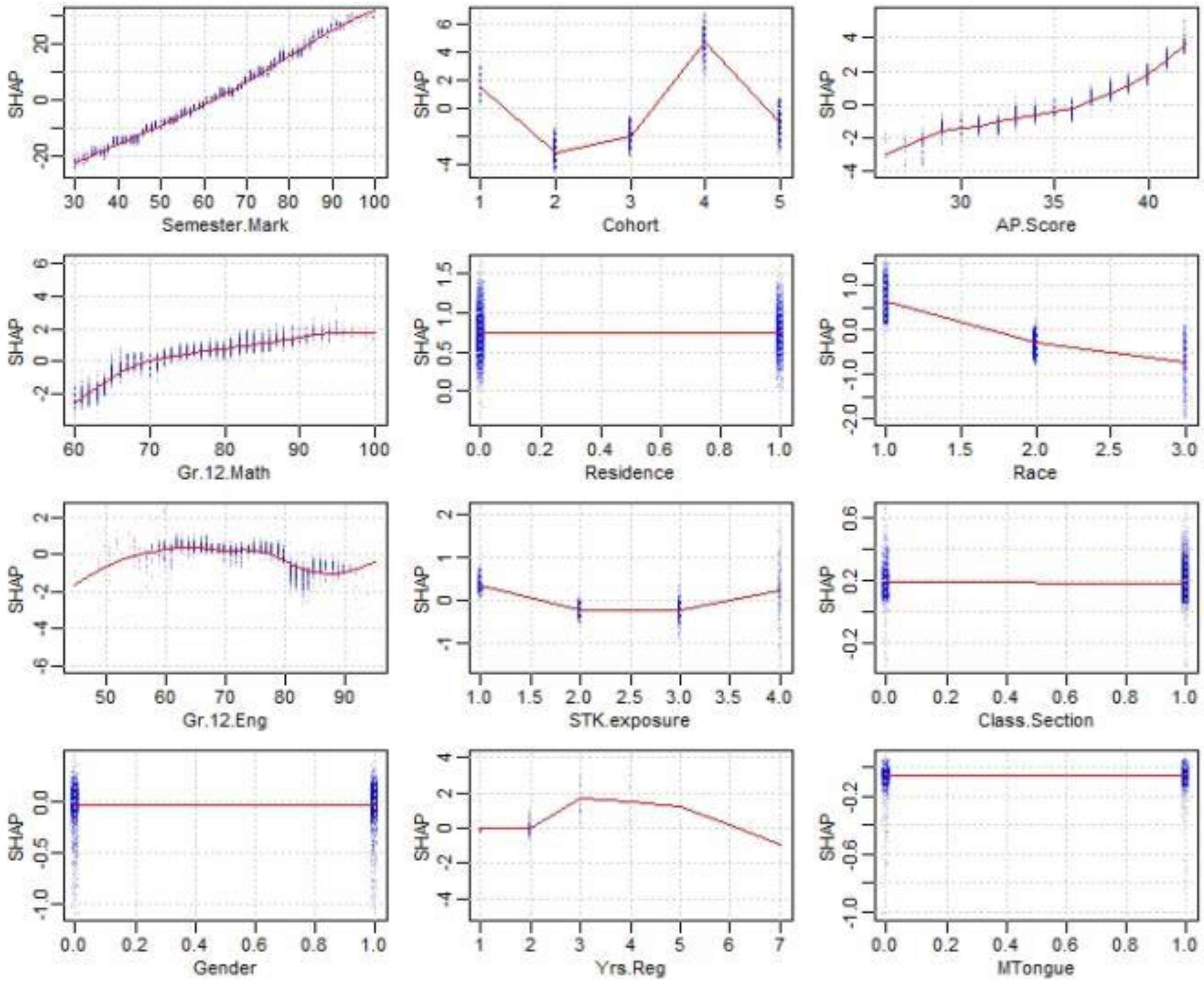


Figure 24: SHAP summary for individual variables in the model for the prerequisite sample

A violin plot in Figure 25 of the examination marks confirms the reason of the concern about the unusual pattern in the examination marks for the Cohort feature in Figure 24.

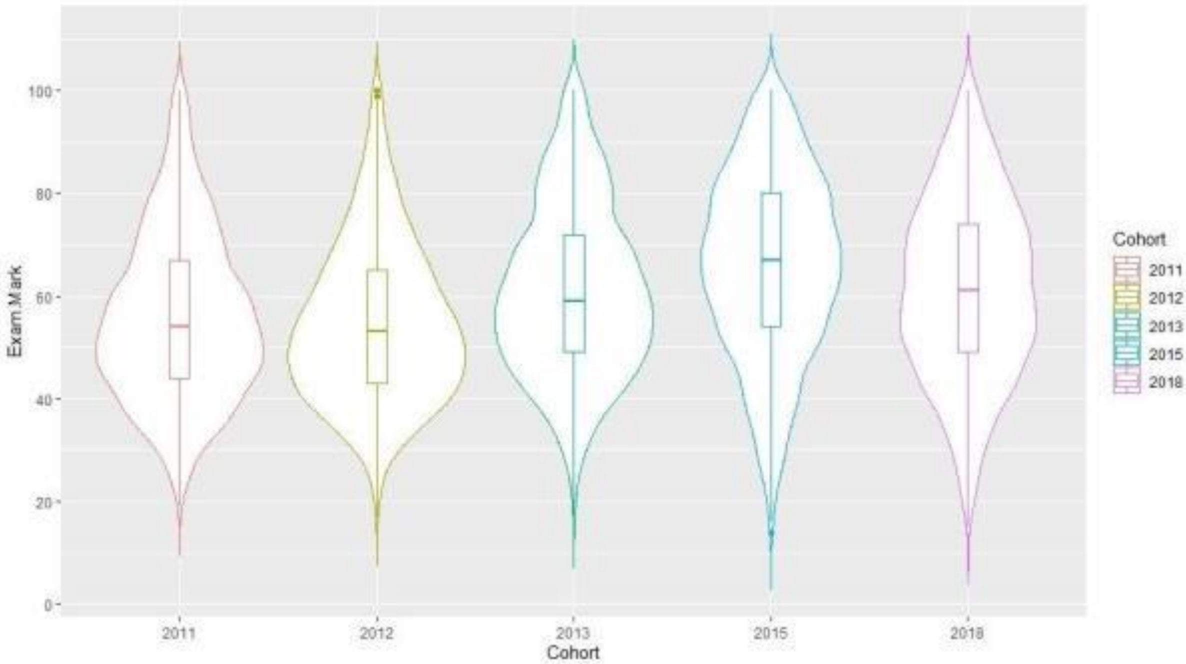


Figure 25: A simple violin plot of the distribution of examination marks for the five cohorts

At face value, the 2011 and 2012 violins look similar, although more lower examination marks are concentrated at the bottom half of 2012. The average examination mark for the 2011 cohort was 59% versus 57% for the 2012 cohort, and the middle 50% of the 2012 boxplot was slightly lower compared to the 2011 cohort's boxplot. Online homework did not improve the examination marks of the 2012 cohort compared to the 2011 cohort.

The examination marks increased considerably more from 2013 to 2018. The 2015 cohort, when clickers were introduced, showed the most improvement. The average examination mark was 67%, with the middle 50% of the boxplot between 56% and 80%. The violins for 2013 and 2018 look alike apart from small differences, like a slightly higher median in 2018.

For the variable Ethnicity in Figure 23, the low values of Ethnicity had a positive impact on examination mark and higher values had a slightly less positive impact on the examination mark. If the individual graphs of Ethnicity (cf. Figure 24)) are consulted, it seems that African students' examination marks were better than White and other students' examination marks.

Considering the observed examination marks, the African students did not outperform the White and other students.

In Figure 24, students who only took STK110 in the first semester and students who had to take statistics up to their second or third year had the greatest impact on the model. Students who had to take statistics in both semesters (STK110 and STK120) had the least impact on the model.

The Yrs reg (years registered) variable is complex to interpret. If the number of years registered is low to moderate (1 to 3 years), it will have a positive effect on the examination mark. In Figure 24, for the prerequisite sample, the SHAP value decreased suddenly after three years and then increased slightly after five years.

In Figure 24 it is notable that the variable Class section had a slightly higher SHAP value for 0 than for 1, which means the English group had a marginally higher impact on the model compared to the Afrikaans group. Mother tongue versus Non-mother tongue education had a similar effect on the examination marks.

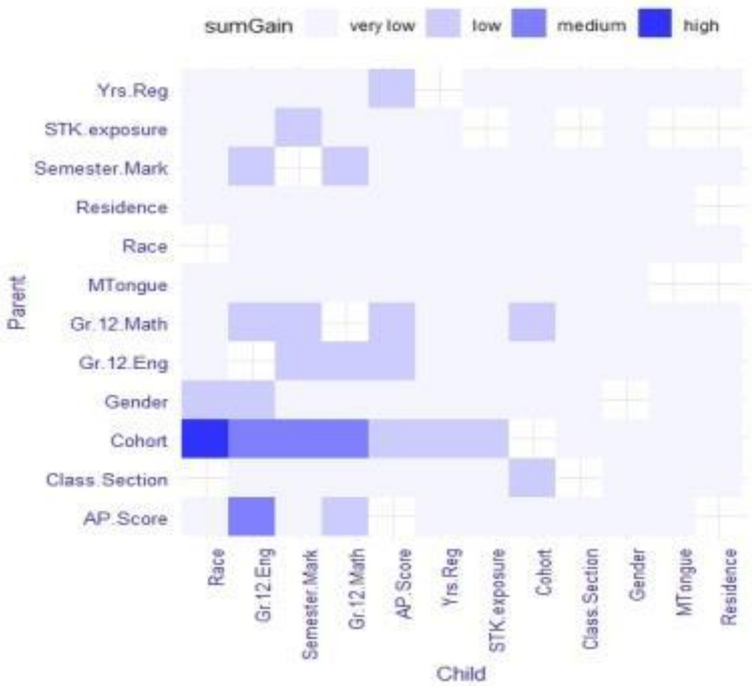


Figure 26: Graphical display of interactions between the different variables for the prerequisite sample

In Figure 26 the interactions between parent and child nodes are classified in shades of blue according to high, medium, low and very low.

The only high interaction that can be seen in Figure 26 is between Cohort and Ethnicity. The following medium interactions are visible between: Cohort and Gr 12 English, Cohort and Semester Mark, Cohort and Gr 12 Math and AP Score and Gr 12 English.

4.4.4 COMPARISON BETWEEN GLM AND XGB (PREREQUISITE SAMPLE)

The coefficient for determination for both models was comparable R^2_{GLM} = 0.642 and R^2_{XGB} = 0.643. For illustration purposes, a random sample of 30 students, six from each of the five cohorts, was chosen (cf. Table 24) and their actual examination marks as well as predicted examination marks by using GLM and XGB were provided (cf. Table 24). Figure 27 shows the XGB graphs created by using the XGBoostExplainer package for four of the 30 students (shaded in light grey in Table 24), where each student has its own set and order of predictors depending on the values assigned to the predictors per student. Although the models for GLM and XGB are very different, the predicted examination marks for 24 out of 30 (80%) of the students were comparable. GLM has a fixed equation with the same predictors and interaction terms for each student in the sample.

Similar to GLM, the semester mark was the most important predictor for XGB, which dominated the outcome of the predicted examination mark. The four students whose XGB graphs are presented in Figure 27 will be discussed. For student number 1(first graph), the most important predictors that added positive weights for this student were Semester mark 80%; AP Score 42; Cohort 2011 and Gender = 1 (female). There were also other less influential predictors that added negative weights, e.g. Gr 12 Eng, Residence and Class Section. Her predicted examination score for both models was close to 80%.

Student number 12 (second graph) had the following influential predictors, e.g. Semester mark 55%; Cohort 2012; AP Score 33 (all added negative weights to the final prediction), and Gr 12 Eng 69% and MTongue = 0 (non-mother tongue) added positive weights. The predicted examination score for both models was close to 50%, while the actual examination mark was

80%. This shows a student who underperformed during the semester but put in an immense effort towards the examination.

Student number 23 (third graph) had Semester mark 75% as the most important predictor, with Gr 12 Math, Gender =1 (female) and Ethnicity = 1 (African) as some of the variables that added positive weights to the predicted outcome. Predictors that added negative weights were, for example, AP Score 35, Cohort 2015, and Residence. Her actual examination mark is 70%. The XGB estimation is 72.38, which is a close approximation and GLM slightly overestimated her mark as 80.04%.

The fourth student is number 27 in Table 24, which corresponds with graph 4 in Figure 27. This was a student with good Grade 12 marks but was an under achiever at university. The important predictors providing both positive and negative weights for this student were, for example, Semester mark 57%; AP Score 40; Cohort 2018 and Residence. Both the models gave a predicted examination mark of approximately 60%, but the student achieved an actual mark of 40%. It could have been a student who relied on their good semester mark and did not put in a sufficient effort towards the examination.

Table 24: Comparison between XGB and GLM

Student #	Cohort	Actual	Predicted XGB	Predicted GLM
1	2011	80	80.13	82.00
2	2011	45	49.48	49.17
3	2011	65	66.42	67.31
4	2011	55	43.57	41.56
5	2011	70	61.34	62.39
6	2011	40	46.35	45.92
7	2012	45	44.85	43.38
8	2012	55	53.14	48.88
9	2012	70	56.94	59.28
10	2012	40	54.11	56.90
11	2012	65	75.47	75.89
12	2012	80	50.34	49.88
13	2013	45	55.79	57.61
14	2013	70	77.23	77.94
15	2013	65	63.07	62.19
16	2013	80	67.46	68.25
17	2013	40	43.79	38.87
18	2013	55	49.40	48.76
19	2015	56	65.22	65.57
20	2015	40	52.00	51.59
21	2015	46	52.61	60.61
22	2015	66	77.36	80.30
23	2015	70	72.38	80.04
24	2015	80	72.25	75.61
25	2018	70	53.80	49.84
26	2018	45	56.05	57.00
27	2018	40	58.40	60.20
28	2018	55	62.45	61.49
29	2018	65	71.76	69.33
30	2018	80	80.89	80.21

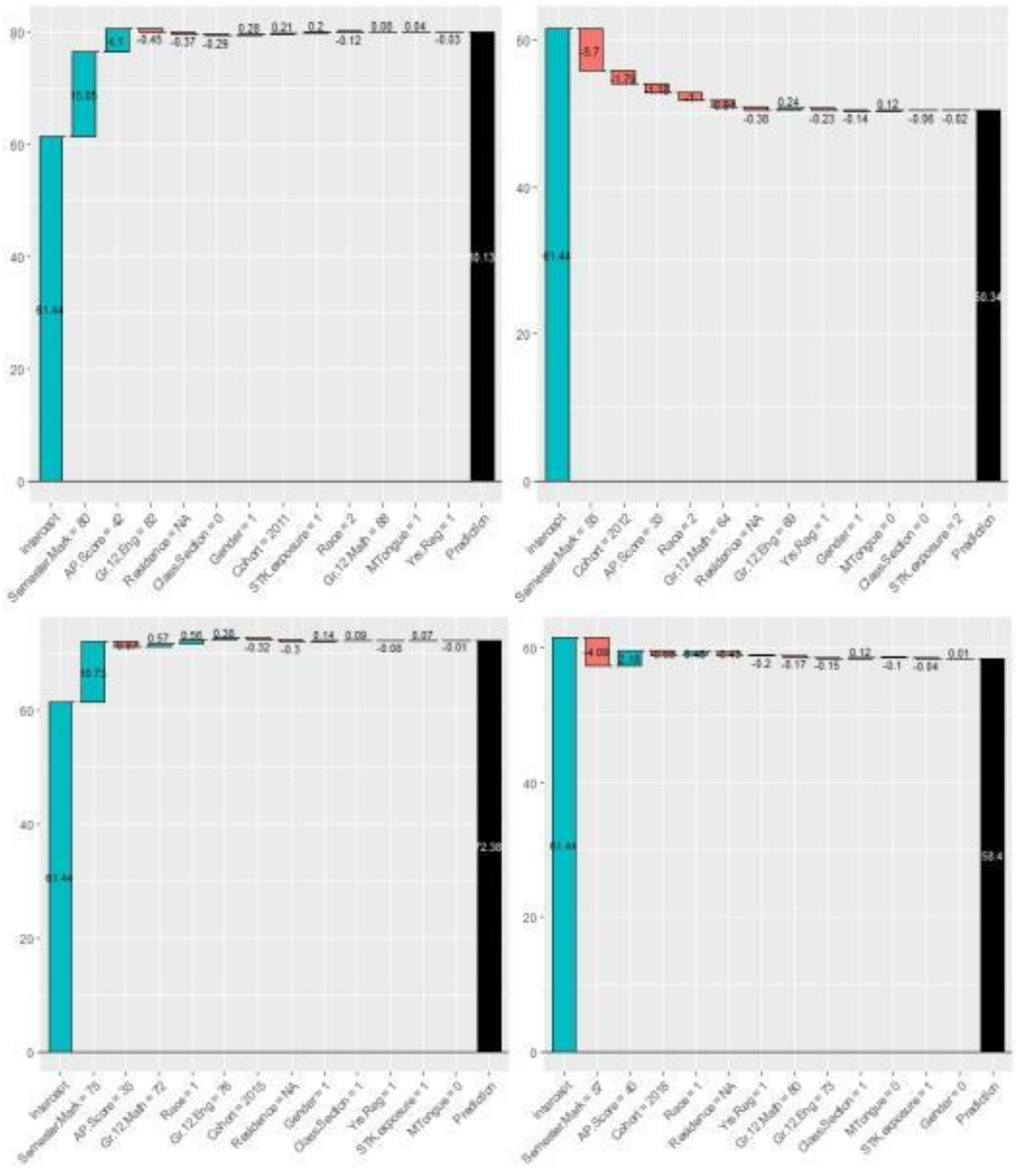


Figure 27: XGB predictions for four students

4.5 CONCLUSION

It is said that a picture is worth a thousand words. Without any statistical analyses, just by looking at the graphical representations in Chapter 4, it is evident that the two traditional model cohorts (2011 and 2012) had similar patterns and trends, and the flipped classroom model cohorts' (2013, 2015 and 2018) graphs and trends look alike. Even though from a practical perspective some of the effect sizes were low, the overall findings were favourable.

The mean final marks and pass rates of the traditional teaching model cohorts were lower than the mean final marks and pass rates of the flipped classroom model cohorts. There was no significant difference in mean final marks and pass rates within the different teaching cohorts. In essence, pedagogy matters! What is of interest is that repeat students benefited more from the flipped classroom and peer learning activities combined with lecturer-switching compared to new students who benefited from the flipped classroom and QT-clickers.

There is an association between the marks' distribution and the five intervention cohorts. Many more students in the traditional model cohorts failed STK110 compared to far fewer students in the flipped classroom model than expected under the null hypothesis of no association. The flipped classroom with/without QT-clicker use had a positive impact on the performance of the prerequisite sample, especially in the higher percentage brackets. In the non-prerequisite sample, in general, for the weaker students, more new students benefitted from the flipped classroom with/without QT-clicker use and repeat students from peer learning activities and lecturer-switching. For the stronger students, both new and repeat students benefitted from the flipped classroom with/without QT-clicker use.

Various GLMs were constructed to measure the effect of the technology interventions on the examination mark. The first two GLMs used the 2014 and 2017 cohorts to measure the effect of QT-clickers on examination mark regarding partial grading or not. The first two GLM models predicted a significantly higher examination mark for students with QT-clickers in 2017 with/without partial grading compared to the 2014 students who did not use QT-clickers, when keeping all the other covariates constant. In general, QT-clickers made a significant difference to student learning and pass rates.

The final GLM and XGBoost models were constructed to measure the effect of the five technology-based interventions on the examination mark for the prerequisite sample. Although the two algorithms for GLM and XGBoost are different, both have given similar predictions for the examination mark. The influence of various categorical predictors makes the interpretation of the effect of the interventions more complicated. In general, the flipped classroom with QT-clickers (2015 cohort) had the most significant effect on student learning, followed by the flipped classroom (2013 cohort) and peer learning activities combined with lecturer-switching (2018 cohort).

CHAPTER 5

QUALITATIVE EVALUATION

5.1 INTRODUCTION

Questionnaires and focus groups were used for certain cohorts to determine the perceptions of students regarding the interventions. The following research question is of importance:

Research question 2: What are the perceptions of the students regarding the interventions?

H_0: The students′ perception of the intervention is neutral or negative

H_a: The students′ perception of the intervention is predominantly positive

5.2 STUDENTS' VOICE ON ONLINE HOMEWORK AND THE FLIPPED CLASSROOM

We surveyed the 2012 and 2013 cohorts to get their input on the Aplia online homework system and tutorials. In the 2013 questionnaire, students had to rank their six conceptions of the pre-class Aplia assignments in order of importance. The students' comments were classified into a single category, and therefore the calculation of percentages is based on only the primary response, i.e. the first ranking of the conceptions of the student. Table 25 summarises the most important conceptions of the online assignments and the tutorials. The response rates were 86.1% (1755 students) for the 2012 cohort and 72.8% (1478 students) for the 2013 cohort.

Table 25: The 2012 and 2013 students' most important experiences with online assignments and tutorials

Students' most important experience	2012 students: (n =1755)	2013 students: (n = 1478)	
	Post-class Aplia online assignments	Pre-class Aplia online assignments	Class tutorials
Good preparation for tests resulted in improved marks	281 (16%)	251 (17%)	695 (47%)
It helped my understanding statistics concepts	597 (34%)	517 (35%)	621 (42%)
It forced me to read the textbook	-	473 (32%)	-
Various negative experiences	877 (50%)	237 (16%)	162 (11%)

The 2012 students used Aplia for the first time. From Table 25, half (34% + 16%) of the students indicated that they benefited cognitively from the Aplia system. Any technological initiative has its own problems. There were various negative experiences, like technological difficulties, some students found the exercises time-consuming, and English is not the home language of all the students. General problems reported by students included late registrations, internet problems and a lack of internet access at home.

In addition, the 2013 cohort's experiences and attitudes to the flipped classroom were surveyed. Their questionnaire was divided into three parts, namely the assignments, the tutorials, and the redesigned tutoring system. The types of questions asked for each part were similar. The first question was whether they took part in the activity. For the second question they had to rank six conceptions of the assignments and the tutorials in order of importance. For the tutoring system they had to identify the best type of tutoring section, i.e. hot spot, revision classes or tutors. An open-ended question for each of the three parts allowed students to add comments.

The quantitative data showed no significant improvement of performance for the 2012 cohort compared to the 2011 cohort. Table 25 shows that where students in 2012 saw the advantages of the post-class Aplia online assignments (only 50%), few linked these to an improvement in marks, but saw these assignments mainly as a tool for better understanding. The repositioning of the Aplia online homework as pre-class assignments in 2013 did not change its perceived contribution to an improvement in marks, nor in its contribution to improved understanding. However, an additional cluster of 2013 students saw benefits of the pre-class Aplia online assignments in the improvement of study procedures, i.e. the close reading of the textbook before class. This is a key characteristic of the flipped classroom. The data presented for the contribution of the post-class tutorials in 2013 suggests that this strategy was seen as contributing to good preparation for tests which resulted in improved marks. Thus, the tutorials, in conjunction with the pre-class online assignments, seem to be a key aspect in the significantly improved performance of the 2013 cohort compared to the 2012 cohort.

In 2014 the flipped classroom pedagogy was repeated, and similar feedback received compared to 2013. From 2015, with the implementation of QT-clickers, the format of the questionnaire changed. It was decided to compare the outcome of similar questions in the 2017 and 2018 questionnaires. It was important to evaluate reading of the textbook and class notes before students attempted the pre-class assignments. It was also essential to compare the 2017 and

2018 cohorts regarding the students' perception of the pre- and post-class assignments. In 2017, the Aplia online homework system was still used for pre- and post-class assignments, but in 2018 the MindTap leaning tool was introduced. MindTap incorporates the Aplia applications used in 2017 for post-class assignments and CNow application for pre-class assignments. The questionnaire was therefore adapted again.

Self-learning is a key component of the flipped classroom approach. The seven statements in the 2017 survey on Aplia and eight statements in the 2018 survey on MindTap where students could rate their agreement on a 0 – 10 scale, are summarised in Table 26 and Table 27. The scale from 0 to 10 was divided into three categories, 0 to 4 (low), 5 to 7 (moderate) and 8 to 10 (high).

Table 26: Survey statements on the students' perception of the Aplia online homework system (2017)

Rating	Percentage
S1: Textbook read before pre-class assignments attempted	
Low	23
Moderate	41
High	36
S2: Notes read before pre-class assignments attempted	
Low	18
Moderate	36
High	46
S3: Textbook is helpful and valuable	
Low	25
Moderate	33
High	42
S4: Pre-class assignments helped to go prepared to class	
Low	16
Moderate	36
High	48
S5: Pre-class assignments helped to explain difficult concepts	
Low	**30**
Moderate	**42**

Rating	Percentage
High	**28**
S6: Did better in Post-class assignments	
Low	**10**
Moderate	**33**
High	**57**
S7: Post-class assignments were interesting real life examples	
Low	20
Moderate	46
High	34

Table 27: Survey statements on the students' perception of the MindTap learning tool (2018)

Rating	Percentage
S1: Textbook read before pre-class assignments attempted	
Low	25
Moderate	39
High	36
S2: Notes read before pre-class assignments attempted	
Low	18
Moderate	34
High	48
S3: Access to a smart device	
Yes	97
No	3
S4: Textbook read on smart device	
Low	**37**
Moderate	**18**
High	**45**
S5: Pre-class assignments helped to go prepared to class	
Low	**15**
Moderate	**37**
High	**48**
S6: Excel understood better because of structured tutorials	
Low	19

Rating	Percentage
Moderate	29
High	52
S7: Problem-based post-class assignments reinforced difficult concepts	
Low	**9**
Moderate	**30**
High	**61**
S8: Watched videos on worked solutions for textbook exercises	
Never	27
Sometimes	51
Often	22

The statements common to 2017 with Aplia and 2018 with MindTap had very similar outcomes for the low, moderate, and high perceiving students. Approximately 23% in 2017 and 25% in 2018 of the students had low ratings for reading of the textbook sections before attempting the pre-class assignments. The average response for reading the textbook before attempting the pre-class assignments (Statement 1) was linked with the students' final marks in roughly four categories (cf. Table 28). Even for the students achieving distinctions in both years, reading of the textbook was moderately perceived. There was not much of a difference in perception of reading of the textbook over all categories. As the average response for the statement increased, so did the final marks.

Table 28: The average response of 2017 and 2018 students for reading the textbook before attempting the pre-class assignment

Final mark categories	<50%	50% – 69%	70% – 89%	90% – 100%
2017	6.1	6.0	6.8	7.0
2018	5.5	6.0	6.6	7.2

The low-perceiving students decreased by 7% when textbook reading changed to reading of the notes before attempting the pre-class assignments (Statement 2). The moderate- and high-perceiving students were almost equally spread for Statement 1 and Statement 2 in both 2017 and 2018.

With Statement 4 in the 2017 survey and Statement 5 in the 2018 survey, it is clear that students found the pre-class assignments valuable to help them to go prepared to class, but it did not help to explain difficult concepts to them (Statement 5 in 2017 survey). The pre-class assignments are textbook exercises at the end of each chapter to test elementary knowledge. Students have to come prepared to class with this basic knowledge so that the lecturer can revisit difficult concepts. Those 70% of the students who rated Statement 5 of 2017 as moderate or high would have had an advantage over the other students. The idea behind the post-class assignments was to reinforce difficult concepts with problem-based real-life examples. It is clear from the outcome of the 2017 and 2018 surveys that students' perception of the post-class assignments was particularly positive, especially in 2018.

With the implementation of MindTap in 2018, a novel feature was that the textbook could be downloaded onto a smart device, apart from the textbook that was accessible from the laptop (not downloadable as a pdf file). Therefore, we added two extra statements to the questionnaire regarding the textbook on a smart device. In 2018, 97% of students had access to smart devices, but it is noticeable that 37% of the students rated reading of the textbook on their smart devices as low, where 23% did not read the textbook on their smart devices at all.

Another useful quality of MindTap, apart from YouTube videos, is the custom-made videos based on difficult statistics concepts. There are also videos available on worked solutions of textbook exercises that could have been of much help if the students have watched them regularly. Only 22% of the students watched the videos often. Excel assignments in real time with accompanying structured tutorials were also embarked on in 2018, which 52% of the students perceived as highly valuable.

The open-ended Aplia/MindTap question was answered by only 43.6% of the 2017 students and 29.9% of the 2018 students. For the 524 students in 2018 who wrote a comment on MindTap™, 63.5% were positive. The 65.9% positive comments (473 out of 718) received for Aplia in 2017 were very similar. They felt the learning system was helpful and a useful application. The concept and Excel videos made a valuable contribution in 2018. Aplia and MindTap assisted with preparation, practise, revision and reinforcement of difficult concepts. Both were perceived as excellent learning platforms. Critical thinking made understanding better for certain students.

Approximately a third of the comments received for both Aplia and MindTap were negative. The general negative comments dealt with the teaching model. It is clear that students are not used to self-learning. Some students found the pre-class assignments stressful and difficult. They had to read the textbook and notes and then approach the pre-class assignments. Particular students also found the post-class assignments more difficult than the class exercises and time-consuming because of too many questions. Another problem for certain students was the e-book. Problems mentioned were the interface, navigation to the e-book, they could not see a full page on the e-book, and the issue of page numbers.

Although all students received a hard copy of the calendar with all the important information and due dates at the beginning of the semester, various students still wanted reminders and notifications closer to the online due dates. Specific students experienced technical problems with some of the browsers and drop-down menus used in Aplia and MindTap.

5.3 STUDENTS' VOICE ON THE USE OF A CLICKER

The 2015 to 2018 cohorts were all surveyed to obtain the students' perception of QT-clicker usage. The 2015 questionnaires could not be analysed, as explained in Section 3.4.3.1. The format of the questionnaire changed as we gained more experience over the years. The 2017 and 2018 questionnaires' results were analysed where students could compare their understanding of statistics course content with QT-clickers to their understanding of course content in other courses that do not use QT-clickers. Ten statements were posed on how students perceived QT-clickers anchored on a scale from 0 (don't agree) to 10 (fully agree). Where students did not fill in an answer, it was coded as a blank.

The ten QT-clicker statements in the survey that students could rate their agreement on a 0 to 10 scale, are as follows:

1. QT-clickers made class more fun and exciting.
2. I prefer to use a QT-clicker than to raise my hand during class activities.
3. The use of QT-clickers made me feel more inclined to engage with my peers in class.
4. The QT-clicker questions during a tutorial session inspired me to do my tutorial beforehand.
5. Response to QT-clicker questions and feedback improved my attention.

6. Response to QT-clicker questions and feedback in class helped me to be more involved and engaged in a large lecture hall.
7. QT-clickers contributed positively to my learning experience.
8. The use of QT-clicker feedback improved my understanding of course content.
9. The QT-clicker feedback enabled the lecturer to respond and explain difficult concepts that I might not have understood.
10. Answers to QT-clicker questions and feedback helped me to judge my own understanding of the course content.

The scale from 0 to 10 was divided into three categories, 0 to 4 (low), 5 to 7 (moderate), and 8 to 10 (high). For each of the ten statements, students' responses were categorised according to the above parsing of the scale. In Figure 28, a clustered bar graph represents the perceptions of the students in 2017 and 2018 (categorised as low, moderate and high) on the ten QT-clicker statements.

In general, the 2017 and 2018 averages for high perception of QT-clicker usage were 52.2% and 55.2% respectively. The overall increase of 3% could have been the effect of more group work with peers and South Afican-based problem scenarios used in class. There were exceptions like Statement 2, where 68% of students in 2017 and 71% of students in 2018 highly perceived QT-clickers a better medium of communicating their answers than hand raising. The moderate perceivers in 2017 and 2018 formed 32.9% and 29.9% of the students respectively. The 3% decrease in the moderate perceivers represents the same increase in the high perceivers. The average low perceivers in 2017 (14.8%) and 2018 (14.9%) remained almost the same.

A few data points in Figure 28 were flagged with asterisks to highlight the statements with noteworthy responses. The first statement identified for further discussion was "QT-clickers made class more fun and exciting". It is evident that not all students perceived QT-clicker usage as fun and exciting. Compared to the other nine statements, Statement 1 had the highest percentage of low perceivers, namely 22% in 2017 and 26% in 2018. The high perceivers decreased from 46% in 2017 to 41% in 2018. In the literature, some of the authors described clicker usage as fun and engaging, and it reduces boredom and repetition of lectures (Baltaci-Goktalay, 2016; Koenig, 2010; Mayer et al., 2009), but these students had differing opinions. A few students mentioned in the comments of the questionnaire of 2018 that we should rather use Kahoot! It could be that fun and exitement for students in class would rather be to make use of questions in the form of games and short quizzes with a surprise element (Hung, 2017). Every device and application have its advantages and disadvantages. It should be noted that Kahoot! is cellphone-based, more limited regarding the scope of features than QT-clickers, and the lack of wi-fi and data is an issue to consider.

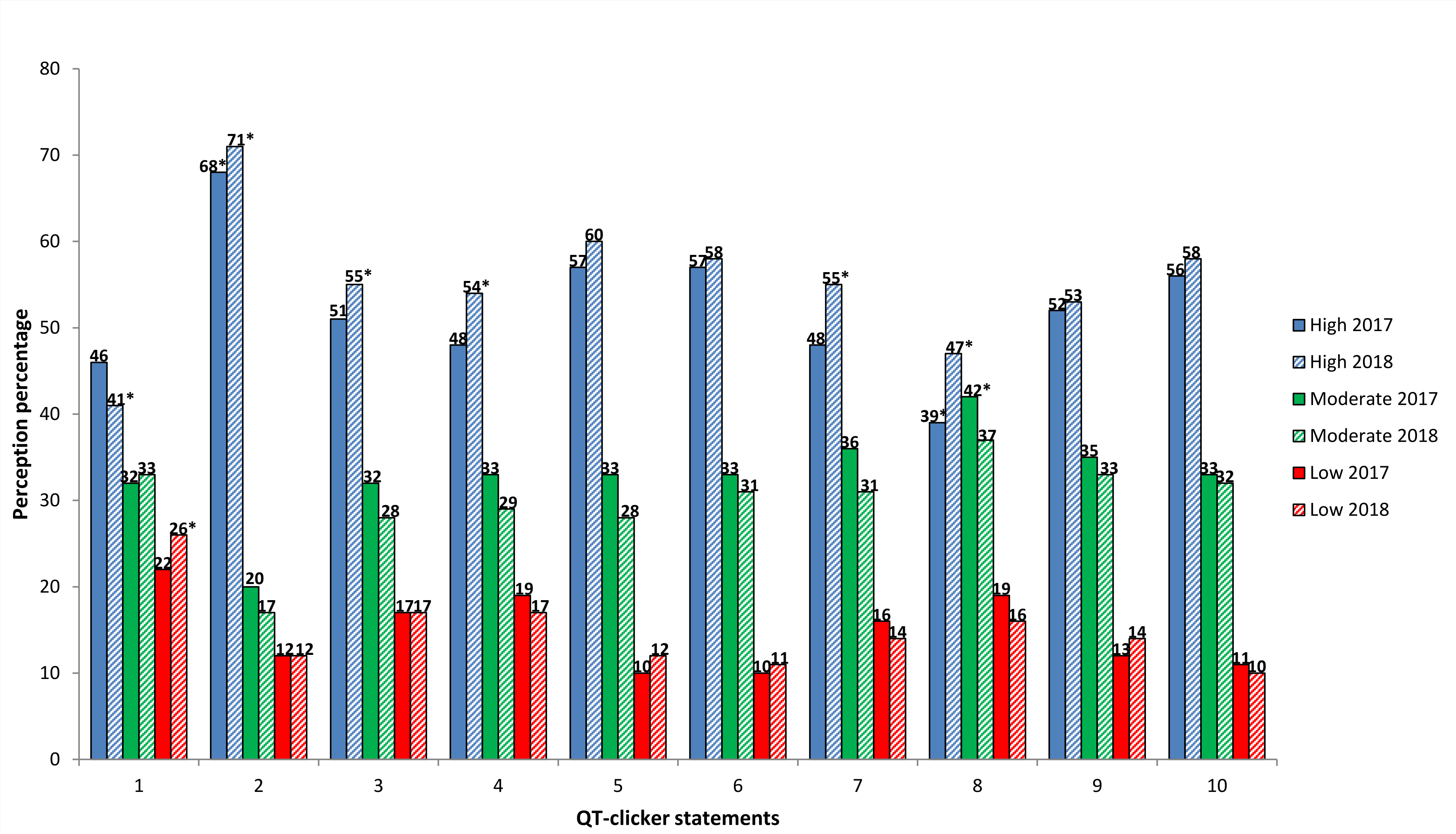


Figure 28: Clustered bar graph of students' perceptions on ten QT-clicker statements

The outcome of Statement 2 supported the other studies which have shown that students prefer to use a QT-clicker rather than raising their hands during class activities. Stowell and Nelson (2007) report that clickers are superior to hand-raising and flash cards, the reason being that clickers maximise participation, especially in large classes, because of anonymity (Hoekstra, 2008; Hwang & Wolfe, 2010), and students can relax and have confidence facing their peers or lecturer. In one of the focus groups a student said: “The students are shy and if you put up your hand to answer a question, the whole class turns and stares at you, which is embarrassing”.

The increase in the percentage of high perceivers from 2017 to 2018 is visible in four of the ten statements, namely Statement 3 (engagement with peers - from 51% to 55%), Statement 4 (inspired to do tutorial beforehand - from 48% to 54%), Statement 7 (positive contribution to learning experience – from 48% to 55%), and Statement 8 (improved understanding of course content – from 39% to 47%). As a result of the outcome of the 2017 questionnaire, we decided to implement more group work with peers, South African-based problems and made attendance of lectures and tutorials compulsory (attendance mark contributed 10% towards the semester mark). The increase in the higher perceiver percentages could have been a result in the refinement of the teaching model above.

The results of Statement 8 (whether QT-clicker feedback improves understanding of course content) were somewhat unforeseen in 2017, but improved by 8% for the high perceivers in 2018, which was encouraging. According to the literature, the purpose of feedback is to improve understanding of course content. The construction of more effective QT-clicker questions and simulations should be considered in future (Büyükkurt et al., 2012; Kaplan, 2011; Wit, 2003).

The student responses to the ten QT-clicker statements (average mark out of 10) were also compared and merged with their final marks. The statements were then sorted in ascending order of grade intervals. The averages for the different statements in the grade distribution were calculated and tabulated against the final marks of 2017 and 2018 in Table 29. If the difference between the 2017 and 2018 rating was more than $|0.4|$ per statement, it was highlighted in white and in bold face.

Table 29: Average statement response against final marks of the 2017 and 2018 cohorts

Grade intervals		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
0 – 39	**2017**	5.9	7.1	6.0	5.8	6.4	6.6	6.2	5.8	6.1	6.5
	2018	5.6	7.3	**6.4**	5.9	6.5	6.8	6.5	5.9	6.4	6.8
40 – 49	**2017**	6.0	7.5	6.6	6.6	7.0	7.2	6.6	6.2	6.7	6.8
	2018	**5.4**	7.5	6.3	**8.0**	6.8	7.0	6.4	6.5	**6.3**	6.8
50 – 59	**2017**	6.0	7.1	6.5	6.2	6.9	6.9	6.4	5.9	6.8	7.0
	2018	**5.6**	**7.5**	6.8	**6.7**	**7.4**	7.1	6.7	**6.4**	6.8	7.2
60 – 69	**2017**	6.9	8.2	7.3	7.1	7.7	7.7	7.4	6.9	7.6	7.7
	2018	**6.1**	7.9	7.1	6.9	**7.3**	7.4	7.3	6.8	**7.2**	7.4
70 – 79	**2017**	7.0	8.2	7.1	7.2	7.6	7.7	7.3	6.9	7.7	7.7
	2018	7.0	8.3	**7.7**	**8.0**	**8.0**	7.9	**8.0**	**7.6**	7.5	7.9
80 – 89	**2017**	7.5	8.9	7.8	7.8	8.2	8.2	7.9	7.3	7.9	8.3
	2018	**6.8**	**8.5**	7.8	7.9	**8.8**	8.4	8.1	**7.8**	8.1	8.5
90 – 100	**2017**	7.5	8.8	7.8	7.8	8.3	8.2	8.0	7.3	8.0	8.2
	2018	7.3	8.6	7.5	**8.2**	8.5	**8.8**	**8.4**	**8.0**	8.2	**8.8**

It is evident from Table 29 that the average student satisfaction of QT-clicker usage was positively correlated with the grade intervals of the final mark. It is worth noting that similar average statement responses could be bounded in broader bands of approximately 20% , i.e. 40% to 59%, 60% to 79% and 80% to 100%. It is also remarkable that the 2018 average rating for the 60% to 69% interval was lower than the 2017 average rating across all the statements. For certain statements, for example Statement 2, no matter what their grade distribution was, most students preferred not to raise their hands. Students from the weakest to the strongest achievers rated Statement 8 (how QT-clicker feedback improved mastery of course content) lower than expected in 2017, with an overall increase in average ratings for the 2018 cohort except for the 60% to 69% bracket. The lower rating of certain statements, i.e. Statement 8, will be investigated in follow-up studies.

The open-ended clicker question was answered by only 44.8% of the 2017 students and 25.8% of the 2018 students. A graphical representation of the primary response of the students can be seen in Figure 29.

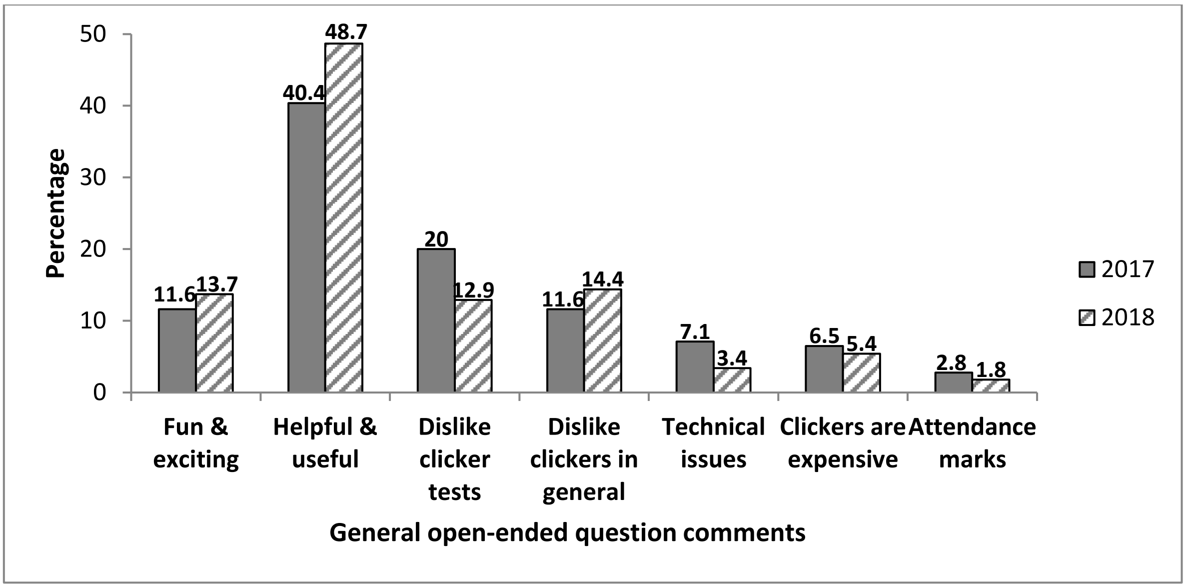


Figure 29: A clustered bar chart of students' general comments on clicker usage in 2017 versus 2018

The open-ended question produced several disparate comments. Fifty-two percent of students in 2017 and 62.1% of students in 2018 perceived the use of clickers as positive. Comments made were that clickers are useful, helpful, increased productivity and participation, learning and engagement in class, kept students focused, was a very good learning tool, among other. It made classes interactive, inspiring, fun, exciting, incredible and amazing learning experiences. A total of 11.6% of the students in 2017 and 14.4% of students in 2018 were negative in general regarding the use of clickers. They expressed their emotions in words such as that clickers were intimidating, stressful, frustrating, time-consuming, e.g. taking up teaching time, irrelevant, annoying, impractical and complicated. The perception was that clickers made them nervous and they did not like them. Twenty percent of the students in 2017 did not like summative assessment (writing formal tests) using clickers; the percentage decreased to 12.9% in 2018. Misconceptions regarding clicker usage in formal assessment were responsible for some of the comments, i.e. rounding of decimal places, marks not displayed after the test was written, poor marks in tests because of final answers on clickers, and no method marks, which could lead to poor results and a waste of time in tests. In reality students have never lost marks for incorrect spelling and after all the answers have been submitted, the lecturer could partially grade answers. All answers were marked in an interval, therefore they could not lose marks for incorrect decimal places. Students with clickers had at least an extra 10 minutes per hour to submit their answers on the clicker. Initially when we introduced clickers, students could see their final mark directly after the test, but we soon realised that it was better to partially grade the test first and then give them their final marks. Other categories of comments depicted in Figure 29 were students who experienced technical problems, students who felt that a clicker was very expensive and only used for one year in statistics, and a few students had problems with attendance marks. For example, if they forgot their clickers at home, they would not be able to do the questions in class and then lose marks on attendance. This was also a misunderstanding, because they could hand in a hard copy of their answers if they left their clickers at home.

In 2018, six focus groups of four to six students were randomly chosen and an independent education consultant had interviews with the students whose answers were recorded. We wanted to clarify some of the students' perceptions about the use of clickers. Some of the positive comments were as follows:

"I can get my marks immediately after submitting a test, hence reduced anxiety of having to wait for days before you see your results."

"See your marks on the spot."

"Have more time than non-clicker students to re-do and check my answers."

"They were very interesting, at first it looked like a phone, so it is easy to use and a fun interactive way to answer questions in class."

"It is a interactive way of teaching and simultaneously monitoring attendance."

"Helped a lot to improve interaction in class, especially because of the size of the class."

"Greater interaction between lecturer and students."

"I liked that the clicker provided the lecturer with clearer feedback on students' progress/ knowledge on a topic in class."

"Made it easier for the lecturer to know the topics the students struggle with."

"Helped to test my personal knowledge of the subject, as the response time of the system was fast."

"I found it to be very useful and well implemented."

"The whole sense of to be anonymous, so even if you got the answer wrong, no one will ever know, which also allowed me not to feel shy or embarrased in class."

The most common negative comments were:

"Clickers are very expensive, especially if it is only used for STK110."

"From time to time technical difficulties caused both lecturers and students to grow irritated and sometimes wasted a lot of time in class."

“Technology may be efficient, but it is not immune to malfunctions. At times clickers will not function as expected if they were second-hand. This is problematic, because it sets the precedent that one needs a brand new clicker to survive STK110 – which is unfair because it is already financially taxing.”

“Answers are either right or wrong – no method marks.”

“Clicker marking methods seemed very unfair to a number of students especially when it came to rounding off.”

“This one-way marking method of clickers caused numerous errors for students, which led to a number of students (including me) to loose marks.”

“I didn’t like how the clicker revealed my mark straight away, as it sometimes made me sad and dissapointed.”

“The fact that after writing, there still has to be a remark on the actual paper and in some cases your clicker mark doesn’t match the script.”

“Can incorrectly give you a wrong percentage after a test.”

“Can discourage you in a sense that if you keep on getting things wrong in class, you would assume that the module is difficult.”

5.4 STUDENTS’ VOICE ON THE TUTORING SYSTEM AND TEACHING APPROACH

The statements regarding students’ perception of the tutoring system were differently set for the 2017 and 2018 questionnaires. In Table 30 one statement combining all the different tutoring systems that were used in 2017 and a second statement to get feedback on the learning experience of the tutoring system are shown. In 2018 one statement with three sub-sections was used to separate the usage of the different tutoring systems, with a second statement to measure the value of the tutoring system.

Table 30: Students’ perceptions on the 2017 tutoring system

Rating	Percentage
S1: Tutoring system used, i.e. Excel, tutors, lecturers	
Yes	59
No	41
S2: Feedback on learning experience	
Positive	94
Negative	6

Rating	Percentage
S3: Revision classes - only students low progress marks	
Low	19
Moderate	23
High	58

The feedback on the tutoring system in 2017 was encouraging. Only 6% (36 out of 631) of the students provided negative comments like confusing, not helpful, struggle to find time, no interaction, too rushed and do not understand. The revision classes were aimed at students who did poorly (progress mark below 50%) and we wanted to revisit certain basic concepts. Some of the negative comments received were probably from students with good progress marks who should not have attended the classes. Some negative comments were classes were too full, too easy, should discuss all work, lecturer confusing, not helpful, more exam-related, more detailed slides and extra summary classes. In 2018 we decided to separate the use of the tutoring system (cf. Table 31).

Table 31: Students' perceptions on the 2018 tutoring system

Rating	Percentage
S1a: Made use of tutor consultation	
Yes	**25**
No	75
S1b: Made use of lecturer consultation	
Yes	**17**
No	83
S1c: Made use of revision classes	
Yes	**80**
No	20
S2: Tutoring system is valuable	
Low	24
Moderate	37
High	39

It is noticeable that only 25% of the students made use of the tutors and even fewer of the lecturers. Some students preferred not to be exposed on a one-on-one tuition system. The revision classes of 2018 were really popular, but different from 2017 in that they were aimed at all the students who needed extra help. Therefore, the reason the students rated the tutoring system in Statement 2 highly and moderately could possibly have been due to the preference of the revision classes.

We experienced achievements as well as weaknesses when we decided to extend the problem-based teaching approach in 2018. The textbook we had been using for the last few years was American-based. Apart from problem solving in class, it was decided to change all classroom problems to South African scenarios to help with understanding of context. More group work involving peers and switching of lecturers on a daily basis were implemented. The reason behind switching of lecturers was that we did not test for the professor-effect. We assumed with switching of lecturers that all students could then be exposed to good and/or different lecturing styles. Table 32 summarises the extension of the teaching approach.

Table 32: The flipped classroom: A problem-based teaching approach in 2018

Rating	Percentage
S1: SA scenarios helped understanding difficult concepts	
Low	18
Moderate	46
High	36
S2: Problem solving improved my understanding	
Low	**8**
Moderate	**35**
High	**57**
S3: Prefer working in group with peers	
Low	**18**
Moderate	**20**
High	**62**
S4: Switching lecturers worked for me	
Low	35
Moderate	28
High	37

The time and effort invested in creating South African-based problems seemed to have helped some students to better understand difficult concepts. Problem solving in class improved understanding (only 8% of the students gave a low rating). Specific students asked for lecturers to rather explain the work in class and leave the problem-solving part for the tutorial classes. Peer-based group work was also a good investment. Switching of lecturers gave several dissimilar outcomes. The majority of general comments was related to switching of lecturers. Some students rated it as helpful, because if one lecturer could not help, the next one should be able to do so, lecturers could be seen from different angles and students were exposed to different learning styles. Another group of students pleaded for one lecturer only, because they struggled to adapt to one learning style and as soon as they did, a new lecturer stepped in to offer the next class.

5.5 CONCLUSION

In conclusion, the second research question will be answered: What are the perceptions of the students regarding the interventions?

The first technology-based intervention implemented in 2012 was the Aplia online homework system. The students' perception was neutral (50:50), an even break between students who benefited cognitively from the online homework system opposed to students with negative perceptions. For this reason it was decided not only to change the pedagogy to a flipped learning model, but to use the online homework system as preparation tool. The students' perception changed to mostly positive. An overwhelmly 84% of the students were in favour of the new learning model compared to 16% with negative experiences. The class tutorials were also optimistically perceived by 89% of the students.

The third intervention embarked on in 2015, QT-clickers, was well received by the majority of students. According to the ten statements in the 2017 and 2018 questionnaires, between 74% to 90% of students rated the use of QT-clickers as moderate to high.

In 2018 a fouth intervention was embarked upon, namely a problem-based approach with peer learning activities was used in conjuction with the flipped learning and QT-clickers. The response of the students on Statement 2 of the questionnaire showed that problem solving definitely improved the understanding of statistics. Fifty-seven percent of the students rated the statement as high, 35% and medium and only 8% gave a low rating.

In summary, the answer to the second research question was that the overall students' voice regarding the technology-based interventions was certainly predominantly positive and optimistic.

CHAPTER 6
CONCLUSION AND RECOMMENDATIONS

6.1 OVERVIEW

In this reflective cohort study, five technology-based interventions grounded in constructivism as theoretical framework, were investigated with respect to student learning and performance.

In the first cohort a traditional teaching model was used. The second cohort consisted of a traditional teaching model and an online homework system. A flipped classroom, where the online homework system was used as preparation tool before attending class, formed the third cohort. The fourth cohort also used flipped learning with embedded online homework system but implemented QT-clickers for active learning in class. For the fifth cohort a problem-based approach, with peer learning activities, was added to flipped learning.

The first aim of the study was to measure the impact of the flipped classroom model versus the traditional classroom model and the role that online homework plays in both models. The key finding was that the online homework by itself did not make a significant contribution to students' performance, but how it was positioned within the flipped classroom model, that is before class interactions, followed by post-class tutorials.

Although there was a slight improvement in the 2012 cohort's performance who used an online homework system for the first time, we agree with Palocsay and Stevens (2008) that students' achievement is not related to the type of homework system used. Doorn et al. (2010) and Jonsdottir et al. (2017) also agree with Palocsay and Stevens (2008) but found that online homework was beneficial for large classes and we can fully endorse this view. In a student survey filled out by the 2012 cohort, half of our students indicated that they gained cognitively from the online homework system, which is in line with the conclusions drawn by Chua-Chow et al. (2011) for their Business Statistics class of 2010.

Why is the positioning of an online homework system key to making a significant contribution to students' performance? Gross et al. (2015) attribute the success of their science flipped classroom to the combination of preparing online before attending an active learning class, which is consistent with how we positioned the online homework system in the flipped classroom of 2013 (Reyneke et al., 2018).

If the online homework is used as preparation tool, lecturer-based learning is substituted for student-based learning, with self-learning and self-knowledge under the magnifying glass. The pre-class assignments helped students to reflect on their self-knowledge, which forms part of prior knowledge. Cognitive constructivism is in action when students discover, explore, and construct statistical concepts. This is in agreement with Vygotsky's (1978, p. 84) idea of the "zone of proximal development", in that a student acquires basic knowledge independently. Classes are attended by prepared students to rectify misconceptions and link them to the self-constructed prior knowledge in the long-term memory, releasing the working memory to ensure the minimum cognitive overload, commensurate with the findings of Cowan (2016). Relationships have been found between the prior knowledge and performance in statistics courses (Schutz et al., 1998).

In 2012, online homework substituted pen-and-paper homework and played the successor role in the traditional model. By changing the role of online homework to being a predecessor, the pedagogy was changed. According to Fullan and Langworthy (2014, p. 5), "the ultimate goal would be to alter whole education systems by getting the vast majority of members to use pedagogy and technology in new and integrated ways to achieve a new vision of deep learning". We agree with Fullan and Langworthy (2014), but it will be difficult for a whole education system to change; this study presents an example of positive change influencing the bigger educational systems gradually. The results showed that the flipped classroom was successful, but we soon realised that the problem-based class activities needed to be more active. Students wanted to take part in class activities without embarrassing themselves in front of their peers. It prompted us to introduce a classroom response system in the form of QT-clickers in 2015.

The second aim was to investigate the use of QT-clickers in a large first-year statistics module, following a flipped classroom teaching model. The intervention of using QT-clickers was evaluated with respect to the pedagogical influence of the QT-clicker and the effect of partial grade crediting. By using QT-clickers, students are forced to come up with answers in class and tests, instead of possible mutiple-choice guesses, helping students to achieve a better understanding of basic statistical concepts. Our study was in agreement with the research done by Dunn et al. (2012), Forster (2014), and Lantz and Stawiski, (2014). Our in-class small group activities using QT-clickers were inspired by Vygotski and Bruner's social constructivism (Zhou, 2020). Throughout the semester the immediate clicker feedback helped our students to clarify their misconceptions and if it was integrated with their prior knowledge, deeper learning was attained and impacted positively on their performance by the time they wrote the examination; illustrated by the results of the two basic GLM III and IV models. Similar to studies done by Kyei-Blankson (2009) and Mayer et al. (2009), the importance was the statistically significant increase in the students' examination marks that could be linked to the pedagogical influence of using QT-clickers.

Premkumar (2016) used clickers for summative assessment, but only for complete MCQ papers. We were the pioneers at the university using QT-clickers for summative assessment and written papers. No articles could be found using clickers for summative assessment and written papers, such that partial credit could be used. In this study, the novelty of awarding partial credit allowed instructors to better measure students' statistical understanding and get a more nuanced view of what students did and did not understand. Students also benefited since they could earn marks for the part of the question they could answer correctly and were not summarily penalised for the wrong answer.

In 2020, mobile clickers were introduced by Turningpoint technology and in March of that year students wrote their first semester test, where they had a choice to either use a QT-clicker or a mobile clicker. The limitations of mobile clickers on campus are discussed in Section 6.3. The COVID-19 pandemic hit the world early in 2020 and brought everybody to a standstill. In a short period of time we had to take our complete module online. We soon realised that because of the internet, mobile clickers would still be operational, no matter the distance of students. An initial internet limitation became an asset in online learning. Students registered their mobile clickers on the Turningpoint application on Blackboard and we used it for formative assessment online, like we used QT-clickers before COVID-19 in-class.

The use of clickers and the evidence of the benefits that it offers are not unique to this study, as shown by the existing rich body of literature on clicker use. The uniqueness of this study was not clicker use per se, but the benefits of QT-clickers from the perspective of a flipped classroom model, for both formative and summative assessment.

The explanation how clickers are used in our situation should be valuable for someone investigating either using QT-clickers in a flipped classroom setting and/or using QT- clickers for summative assessment. The results of this study should convince prospective users that using clickers in a flipped classroom for summative assessment will reinforce student learning.

The unforeseen lower ratings for certain statements in a clicker use survey filled out by the 2017 cohort, forced us to think from a new perspective. The fourth technology-based intervention was introduced in 2018; the learning activities in-class were modified. The third aim was to evaluate problem-based, peer learning activities with localised content and lecturer-switching within a flipped classroom environment with QT-clickers.

The result was that good students performed even better, and it had a surprisingly positive influence on the pass rate of the repeat students. As mentioned in Chapter 1, the repeat students have been of great concern to us since the year 2000, hence it was reassuring that the data showed that they had benefited from the alternative learning activities in 2018.

Regression models were built using GLM to assess the relationship between the technology-based interventions and examination mark. XGB was used as alternative machine learning method to predict the examination mark using a completely different angle. The results for both methods have shown that the QT-clicker intervention within a flipped classroom increased the average examination mark and outperformed the other intervention cohorts. The flipped classroom made a significant contribution to students' performance, but it lacked active learning in-class. Students had to come prepared to class with self-learned prior knowledge. A response system which protected a large class's anonymity was found in QT-clickers, implemented in 2015. The benefits of QT-clicker use through the lens of a flipped classroom environment, makes this study the first of its kind.

6.2 ANSWERING THE RESEARCH QUESTIONS

The main goal of the study was to evaluate the effect of the four technology-based interventions, through the lens of the research questions.

For research Questions 1a, b and c, the analyses were done for both the *prerequisite and non-prerequisite samples.* The non-prerequisite sample was further divided into *new* and *repeat students* and the analyses repeated for the two sub-groups.

Research Question 1a: Do the five intervention cohorts have different mean final marks?

It was found for both the *prerequisite and non-prerequisite samples* that there was a significant difference between the mean final marks of the five cohorts. The post-hoc test revealed that the mean final marks of the cohort with the traditional model and the traditional model with online homework did not differ, but their mean final marks were less than the other three intervention cohorts, i.e. flipped classroom, QT-clickers, and peer learning activities.

There was also a significant difference between the mean final marks of the five cohorts for the *new and repeat students* in the non-prerequisite sample. *New* students benefited more from the flipped classroom and QT-clickers, where *repeat* students benefited more from the flipped classroom and peer learning activities.

In summary, the flipped classroom teaching model made a difference to the mean final marks.

Research Question 1b: Do the five intervention cohorts have different pass rates?

For the *prerequisite and non-prerequisite samples* there was a significant difference in the pass rates of the five intervention cohorts. By using pairwise comparisons, the exact same deduction as for research Question 1a could be made, namely that the pass rates of cohort 2011 and 2012 were lower than the pass rates of cohorts 2013, 2015 and 2018, but they did not differ from one another. In short, the flipped classroom pedagogy had an influence on pass rates.

For the *new* and *repeat student*s there was also a significant difference in the pass rates of the five intervention cohorts. The pairwise comparisons exposed a similar pattern compared to Question 1a for *new* students, namely the positive influence on pass rates due to the flipped classroom and QT-clickers compared to the traditional model. The benefit of peer learning activities in a flipped classroom environment compared to traditional teaching methods contributed to better pass rates for the *repeat* students.

Research Question 1c: Is there an association between the five intervention cohorts and the marks distribution?

The chi-square test revealed a significant relationship between the final marks' distribution and the intervention cohorts for both samples, as well as for the new and repeat students of the non-prerequisite sample. Comparison of standardised residuals in the cross-tabulation obtained the same conclusion as reached by research Questions 1a and 1b. Many more students in cohorts 2011 and 2012 failed STK110 than expected under the null hypothesis of no association. Far fewer students in the cohorts 2013, 2015 and 2018 failed than expected.

For the *prerequisite sample*, the 2013, 2015 and 2018 cohorts' students in the 60% to 69% bracket moved to the higher percentage brackets. The effect of the QT-clicker was evident as the 2015 cohort outperformed the other cohorts in three of the six categories. In conclusion, the five cohorts can be divided into the two learning models, namely traditional versus flipped, where the flipped model is underpinned by constructivism and outperformed the traditional model.

For the *non-prerequisite sample*, which was further divided into new and repeat students, the flipped classroom implemented in 2013 and QT-clickers implemented in 2015 had a positive effect on the performance of new students in the failing bracket and both new and repeat students in the higher percentage brackets. The repeat students in the lower percentage brackets benefited from peer learning activities and lecturer-switching implemented in 2018.

Research Question 1d: Is there an association between QT-clicker use in 2017 versus 2014 cohorts and the examination marks?

Various GLMs were constructed to measure the effect of the intervention cohorts on the examination mark. GLM I and II measured the effect of QT-clickers on examination mark regarding partial grading. GLM I (with partial grading) predicted that students of the 2017 clicker cohort's examination marks were approximately 3.3% higher compared to the 2014 students who did not use QT-clickers, keeping all the other covariates constant. GLM II (without partial grading – like a MCQ paper) gave a similar result of a 2.4% higher examination mark, which highlights the pedagogical effect of clicker use. GLM III to VI measured the role of Sem (semester mark) in the regression model. Semester mark is the best predictor of the examination mark. Closer inspection in the form of basic GLMs (GLM III to GLM V) revealed that the clicker use effect could only be measured at the end of the semester in the examination marks (long-term effect). The semester mark played an essential role, not only in predicting the examination mark, but it also affected the other covariates' role in the model, as is noticeable in GLM VI with interaction terms. Apart from the influence of mother tongue versus non-mother tongue on clicker use, the significant effect of clicker use on the examination mark was evident.

Research Question 1e: Is there an association between the five intervention cohorts and the examination marks?

The final GLM and XGB models were constructed to measure the effect of the five intervention cohorts on the examination mark for the prerequisite sample. The R^2 value of the two models was comparable. Although the two algorithms for GLM and XGB were very different, both gave similar predictions for the examination mark. Cohort 2015 (QT-clickers) had the most significant effect on student learning, followed by cohort 2013 (flipped classroom), and 2018 (peer learning activities).

Research Question 2: What are the perceptions of the students regarding the interventions?

Students' perception of the online homework system as a post-class tool was neutral. The implementation of the flipped classroom with online homework as preparation tool was well perceived by the students. Most of the students could experience the value of the new pedagogy in their improved marks because of self-learning, i.e. preparation prior to class and reading of the textbook. The tutorials, in combination with the pre-class assignments, were the key combination for the improved performance and optimistically observed by the students.

The QT-clicker survey with the ten statements in the 2017 and 2018 cohorts shed light on how students perceived the use of clickers. Overall, according to the ten statements, the 2018 cohort was more optimistic about QT-clicker use than the 2017 cohort. Fewer students in the 2017 cohort disregarded QT-clickers to make classes more fun and exciting, compared to the 2018 cohort. Both cohorts valued the anonymity and immediate feedback to help them judge their own understanding of course content.

A large percentage of the 2017 cohort felt that QT-clicker feedback did not always improve understanding of course content. It was therefore decided to refine the teaching model in 2018 with problem-based, peer learning activities, as well as lecturer-switching, which were well perceived by most of the students. Many students rated the problem-based learning and working in groups with peers as moderate or high.

Negative perceptions of students decreased from 2017 to 2018, such as writing a semester test using a QT-clicker. In general, they perceived the use of QT-clickers mainly positive.

6.3 LIMITATIONS OF THE STUDY

Ideally, we would have liked to conduct a randomised experiment. However, from previous studies in the literature and as mentioned in Chapter 2, there is strong evidence that the four technology-based interventions, namely online homework, flipped learning, QT-clickers and peer learning activities, benefit students, and a randomised experimental design would knowingly disadvantage the group of students who were not allowed to use one or more of the interventions.

Mobile clickers were introduced in 2020 and the major difference between QT and mobile clickers is a radio signal connected to the lecturer's laptop needed for QT-clickers versus wi-fi or internet for mobile clickers. Not all venues had secure wi-fi, students do not always have enough data and with summative assessments, they can access documents from the Blackboard system and search for answers on the internet. Another problem was that students' cell phones were not always sufficiently charged for a semester test or three-hour examination.

The disadvantage of using mobile clickers for online assessment is that it could not count for marks, because it was not compulsory to attend no live lectures. Recordings of the live lectures were available afterwards. The students who attended the live lectures could participate in answering mobile clicker questions, assisting them to judge their own understanding of course content.

6.4 RECOMMENDATIONS

The research study proposed two final models, GLM VII and XGB to predict students' examination marks, given several predictors. Even though the results of the two models constructed were very similar, they both had moderate coefficients of determination ($R^2 = 0.642$ and $R^2 = 0.643$ respectively).

The following suggestions for future research that can complement the current study are:

- Nielsen et al. (2018) conducted a multiple regression model for their large undergraduate statistics class using predictors similar to this study, but added the following variables: Usefulness of statistics, Level of learner autonomy, Confidence in learning statistics, Math anxiety, Teacher effect and course rating. The GLM and XGB model used for this study overestimated low examination marks and underestimated high examination marks. As mentioned, some students had good semester marks, and as the semester mark was the main predictor of the examination mark, a student's predicted examination mark could be higher than the actual examination mark, for instance if they did not put in sufficient effort when preparing for the examination. Research on the measurement of certain problematic variables should be investigated, namely the measurement of the effort put in to prepare for the examination, examination anxiety and stress before and while writing the examination. A suggestion could be a six-point Likert scale. Students are hesitant to complete questionnaires before or after an examination. A few extra questions in the form of a short questionnaire at the end of the paper for bonus marks could motivate students to cooperate.
- In this thesis a retrospective cohort study was used. For ethical reasons we could not justify using a randomised experiment. The cohorts were imbalanced regarding certain covariates and to make the cohorts comparable, we divided the students into a prerequisite and non-prerequisite sample. In certain analyses only the prerequisite samples of the cohorts were compared. Propensity score matching techniques could be an alternative to adjust for confounding in an observational study. However, there are a few contradictory articles that would be stimulating to compare by using the data for this study (Elze et al., 2017; Glynn & Quinn, 2009; King & Nielsen, 2019).

6.5 FUTURE RESEARCH

Future research could:

- Investigate a breakdown of topics that have historically given students trouble (item analysis of threshold concepts could be used) and a comparison of two or more cohorts across those topics would be worth pursuing (Bulmer et al., 2007; Dunne et al., 2003; Khan, 2014).

- Explore if an online platform can be developed for summative assessment using a mobile clicker or device, in conjunction with a platform which secures the integrity of online assessments, e.g. Proctorio.com.
- Consider various modules world-wide that changed pedagogy from a traditional or flipped classroom model before COVID-19 to flipped learning during COVID-19. Many of the modules in the Statistics department of the UP adopted flipped learning. It would be valuable to investigate if lecturers realised the value of flipped learning and online teaching before or during COVID-19. Several articles have already been published in different subject areas (Campillo-Ferrer & Miralles-Martínez, 2021; Feijóo et al., 2021; Fogg & Maki, 2021; Latorre-Cosculluela et al., 2021).

6.6 VALUE AND CONTRIBUTION OF THIS STUDY

The novelty of this study (and thus the justification for this study) is that it considered the effect of the online homework system in combination with the flipped classroom. The repositioning of the online homework as pre-class assignments by the researcher could have made an improvement to study procedures, that is the close reading of the textbook and class notes before attempting the pre-class assignments. This process creates prepared students to perform well in problem-based, in-class activities, where they can construct their own knowledge. Students learn to spread learning activities through the semester, instead of cramming it all in before tests and examinations. Therefore, there will seldom be a cognitive overload. The timelier way to engage with course content could provide better performance in a flipped classroom.

The uniqueness of this study was that it expounded the benefits of QT-clicker use through the lens of a flipped classroom model and using these clickers for both formative and summative assessment. The exposition of how clickers are used within this environment should be valuable to someone considering either using QT-clickers in a flipped classroom setting and/or using QT-clickers for summative assessment. Based on the results of this study, prospective users can be fairly confident that using clickers in a flipped classroom for summative assessment is not detrimental to but, in fact, enhances student learning.

The unexpected in this study was the unforeseen increase in the pass rate in 2018 of the repeat students. Not all students are the same; they use different senses and styles, and in this study we could provide a group of students who were often left behind with something out of the ordinary.

REFERENCES

Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: definition, rationale and a call for research. *Higher Education Research and Development*, *34*(1), 1–14. https://doi.org/10.1080/07294360.2014.934336

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). Wiley-Interscience.

Akçay, G., & Akçay, M. (2018). The flipped classroom : A review of its advantages and challenges. *Computers and Education*, *126*, 334–345. https://doi.org/10.1016/j.compedu.2018.07.021

American Statistical Association. (2016). Guidelines for assessment and instruction in statistics education: College report. *Report*, *August 2005*, 1–61. https://doi.org/10.3928/01484834-20140325-01

Amstelveen, R. H. (2013). *The Effect of “Clickers” on Attendance in an Introductory Statistics Course : An Action Research Study*. Capella University.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2020). *Modern Business Statistics with Microsoft Excel, 7e*. Cengage

Aplia. (2012). Cengage. http://www.aplia.com/statistics/

Ary, E. J., & Brune, C. W. (2011). A comparison of student learning outcomes in traditional and online personal finance courses. *Journal of Online Learning and Teaching*, *7*(4), 465–474. http://sr.ithaka.org/?p=22464

Bacon, D. R., & Stewart, K. A. (2006). How fast do students forget what they learn in consumer behavior? A longitudinal study. *Journal of Marketing Education*, *28*(3), 181–192. https://doi.org/10.1177/0273475306291463

Baepler, P., Walker, J. D., & Driessen, M. (2014). It’s not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers and Education*, *78*, 227–236. https://doi.org/10.1016/j.compedu.2014.06.006

Baker, J. W. (2000). The “Classroom Flip”: using Web Course Management Tools to become the Guide by the Side. *11th International Conference on College Teaching and Learning*, 9–17.

Balaban, R. A., Gilleskie, D. B., & Tran, U. (2016). A quantitative evaluation of the flipped classroom in a large lecture principles of economics course. *Journal of Economic Education*, *47*(4), 269–287. https://doi.org/10.1080/00220485.2016.1213679

Baltaci-Goktalay, S. (2016). How personal response systems promote active learning in science education? *Journal of Learning and Teaching in Digital Age*, *1*(1), 47–54.

Barnett, J. (2006). Implementation of personal response units in very large lecture classes : Student perceptions. *Australasian Journal of Educational Technology*, *22*(4), 474–494.

Basturk, R. (2005). The effectiveness of computer-assisted instruction in teaching introductory statistics. *Educational Technology & Society*, *8*, 170–178. https://doi.org/1176-3647

Baviskar, S. N., Todd Hartle, R., & Whitney, T. (2009). Essential criteria to characterize constructivist teaching: Derived from a review of the literature and applied to five constructivist-teaching method articles. *International Journal of Science Education*, *31*(4), 541–550. https://doi.org/10.1080/09500690701731121

Beatty, I. D. (2004). Transforming student learning with classroom communication systems. *Educause Center for Applied Research*, *2004*(3), 1–13.

Beatty, I. D., Gerace, W. J., Leonard, W. J., & Dufresne, R. J. (2006). Designing effective questions for classroom response system teaching. *American Association of Physics Teachers*, *74*(1), 31–39. https://doi.org/10.1119/1.2121753

Bechter, C., & Swierczek, F. W. (2017). Digital storytelling in a flipped classroom for effective learning. *Education Sciences*, *7*(2), 61. https://doi.org/10.3390/educsci7020061

Ben-Zvi, D. (2007). Technology innovations in statistics education using wiki to promote collaborative learning in statistics education. *Technology Innovations in Statistics Education*, *1*(1), 1–19.

Bergmann, J., & Aaron, S. (2012). *Flip your Classroom. Reach Every Student in Every Class Every Day*. International Society for Technology in Education.

Bishop, J. L., & Verleger, M. A. (2013). The Flipped Classroom : A Survey of the Research The Flipped Classroom. *Proccedings of the Annual Conference of the American Society for Engineering Education*, 23.1200.1-23.1200.18. https://doi.org/10.1109/FIE.2013.6684807

Bligh, D. (2000). *What's the use of lectures?* Jossey-Bass.

Bogost, I. (2013). The condensed classroom. In *The Atlantic*. https://www.theatlantic.com/technology/archive/2013/08/the-condensed-classroom/279013/

Bojinova, E., & Oigara, J. (2013). Teaching and learning with clickers in higher education. *International Journal of Teaching and Learning in Higher Education*, *25*(2), 154–165.

Bonwell, C. C., & Eisen, J. A. (1991). *Active Learning: Creating Excitement in the Classroom.* ASHE-ERIC Higer Education Reports.

Bowen, W., Lack, K., Chingos, M., & Nygren, T. (2015). *Interactive Learning Online at Public Universities*. ITHAKA. https://doi.org/10.18665/sr.22464

Brady, M., Seli, H., & Rosenthal, J. (2013). "Clickers" and metacognition : A quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback devices. *Computers & Education*, *65*, 56–63. https://doi.org/10.1016/j.compedu.2013.02.001

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn: Brain, Mind, Exprience, and School: Expanded Edition*. National Academy Press. https://doi.org/10.17226/9853

Brewer, D. S., & Becker, K. (2010). Online homework effectiveness for underprepared and repeating college algebra students. *Journal of Computers in Mathematics and Science Teaching*, *29*(4), 353–371. www.learntechlib.org/p/33222

Bulmer, M., O'Brien, M., & Price, S. (2007). Troublesome concepts in statistics: a student perspective on what they are and how to learn them. *Assessment in Science Teaching and Learning Symposium, University of Sydney, September 28-29*, 139–144. http://openjournals.library.usyd.edu.au/index.php/IISME/article/view/6337

Burch, K. J., & Kuo, Y.-J. (2010). Traditional vs. online homework in college algebra. *Mathematics and Computer Education*, *44*(1), 53–63. http://library.capella.edu/login?url=http://search.proquest.com/docview/235939210?accountid=27965%5Cnhttp://wv9lq5ld3p.search.serialssolutions.com.library.capella.edu/?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ProQ&rft_val_fmt=info:of

Büyükkurt, M., Li, Y., & Cassidy, R. (2012). Using a classroom response system in an introductory business statistics course: Reflections and lessons learned. *Proceedings of Informing Science & IT Education Conference (ISI)*, 379–388.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE - Life Sciences Education*, *6*, 9–20. https://doi.org/10.1187/cbe.06

Campillo-Ferrer, J. M., & Miralles-Martínez, P. (2021). Effectiveness of the flipped classroom model on students' self-reported motivation and learning during the COVID-19 pandemic. *Humanities and Social Sciences Communications*, *8*(1). https://doi.org/10.1057/s41599-021-00860-4

Carpenter, S. K. (2012). Using spacing to enhance diverse forms of learning. *Educational Psychology Review*, 369–378.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning. A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, *1*(1), 1–26.

Chang, S. H. (2016). *The Marriage of Constructivism and Flipped Learning*. [Online submission]. ERIC.

Chen, J. C., Whittinghill, D. C., & Kadlowec, J.A. (2010). Classes that click: Fast, rich feedback to enhance student learning and satisfaction. *Journal of Engineering Education*, *99*(1), 159–168.

Chen, L., Chen, T.-L., & Chen, N.-S. (2015). Students' perspectives of using cooperative learning in a flipped statistics classroom. *Australasian Journal of Educational Technology*, *31*(6), 621–640. http://lib-ezproxy.tamu.edu:2048

Chou, P., Chang, C., & Lin, C. (2017). BYOD or not : A comparison of two assessment strategies for student learning. *Computers in Human Behavior*, *74*, 63–71. https://doi.org/10.1016/j.chb.2017.04.024

Chow, A. F. (2015). Online homework impact in undergraduate mathematics and business statistics courses. *Educational Studies*, *41*(3), 244–248. https://doi.org/10.1080/03055698.2014.961902

Chua-Chow, C., Chauncey, C. A., & McKessock, D. (2011). Enhancing the study of business statistics with an e-homework system. *Journal of Education and Vocational Research*, *1*(3), 96–105. https://doi.org/10.22610/jevr.v1i3.15

Cilli-Turner, E. (2015). Measuring learning outcomes and attitudes in a flipped introductory statistics course. *Primus*, *25*(9), 833–846. https://doi.org/10.1080/10511970.2015.1046004

Clark, R. C., Nguyen, F., & Sweller, J. (Eds.). (2006). *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load*. Pfeiffer.

Cobb, G. W. (1993). Reconsidering statistics education : A National Science Foundation Conference. *Journal of Statistics Education*, *1898*, 1–27. https://doi.org/10.1080/10691898.1993.11910454

Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum. *Technology Innovations in Statistics Education*, *1*(1), 1–16. http://www.escholarship.org/uc/item/6hb3k0nz%5Cnpapers3://publication/uuid/426B647B-BAD9-4278-8F02-990BD513D5FA

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in Cognitive Sciences*, *2*(3), 77–78. https://doi.org/10.1016/S1364-6613(98)01144-9

Cowan, N. (Ed.). (2016). *Working Memory Capacity* (Classic Ed). Routledge Taylor and Francis.

Cox, D. R. (1997). The current position of statistics: A personal view. *International Statistical Review/Revue Internationale de Statistique*, *65*(3), 261. https://doi.org/10.2307/1403368

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college : A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*(2), 272–295. https://doi.org/10.3102/0034654310362998

Crews, T. B., Ducate, L., Rathel, J. M., Heid, K., & Bishoff, S. T. (2011). Clickers in the classroom : Transforming students into active learners. *ECAR Research Bulletin*, *9*, 1–12.

Davis, K. S., & Minifie, J. R. (2013). Ensuring Gen Y students come prepared for class; then leveraging active learning techniques to most effectively engage them. *American Journal of Business and Management*, *2*(1), 13–19. https://doi.org/10.11634/216796061302228

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2020). Taking parametric assumptions seriously: Arguments for the use of welch's f-test instead of the classical f-test in one-way ANOVA. *International Review of Social Psychology*, *32*(1), 1–12. https://doi.org/10.5334/IRSP.198

Dillard-Eggers, J., Wooten, T., Childs, B., & Coker, J. (2008). Evidence on the effectiveness of on-line homework. *College Teaching Methods & Styles Journal (CTMS)*, *4*(5), 9–16. http://journals.cluteonline.com/index.php/CTMS/article/view/5548

Doorn, D. J., Janssen, S., & O'Brien, M. (2010). Student attitudes and approaches to online homework. *International Journal for the Scholarship of Teaching and Learning*, *4*(1). https://doi.org/10.20429/ijsotl.2010.040105

Dove, A. (2013). Students' perceptions of learning in a flipped statistics class. *Society for Information Technology & Teacher Education International Conference 2013*, 393–398.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, *20*(2), 81–94. https://doi.org/10.1111/j.1365-2729.2004.00074.x

Driscoll, A., Jicha, K., Hunt, A. N., Tichavsky, L., & Thompson, G. (2012). Can online courses deliver in-class results?: A comparison of student performance and satisfaction in an online versus a face-to-face introductory sociology course. *Teaching Sociolog*, *40*(4), 312–331. https://doi.org/10.1177/0092055X12446624

Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L. (1996). Classtalk: A classroom communication system for active learning. *Journal of Computing in Higher Education*, *7*(2), 3–47. https://doi.org/10.1007/BF02948592

Dunn, P. K., Richardson, A., McDonald, C., & Oprescu, F. (2012). Instructor perceptions of using a mobile-phone-based free classroom response system in first-year statistics undergraduate courses. *International Journal of Mathematical Education in Science and Technology*, *43*(8), 1041–1056. https://doi.org/10.1080/0020739X.2012.678896

Dunne, T., Low, T., & Ardington, C. (2003). Exploring threshold concepts in basic statistics, using the internet. *The International Association for Statistical Education, Statistics & the Internet (IASE/ISI Satellite, 2003)*, *about 1100*. http://iase-web.org/documents/papers/sat2003/Dunne.pdf

Elze, M. C., Gregson, J., Baber, U., Williamson, E., Sartori, S., Mehran, R., Nichols, M., Stone, G. W., & Pocock, S. J. (2017). Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 cardiovascular studies. *Journal of the American College of Cardiology*, *69*(3), 345–357. https://doi.org/10.1016/j.jacc.2016.10.060

Engelbrecht, J., & Harding, A. (2008). The impact of the transition to outcomes-based teaching on university preparedness in mathematics in South Africa. *Mathematics Education Research Journal*, *20*(2), 57–70. https://doi.org/10.1007/BF03217477

Engelbrecht, J., Harding, A., & Phiri, P. (2010). Are OBE-trained learners ready for university mathematics? *Pythagoras*, *0*(72), 3–13. https://doi.org/10.4102/pythagoras.v0i72.16

Farrington, J. (2011). Seven Plus or Minus Two. *Performance Improvement Quarterly*, *23*(4), 113–116. https://doi.org/10.1002/piq.20099

Fedderke, J. W., de Kat, R., & Luiz, J. M. (2000). Uneducating South Africa : The failure to address the 1910-1993 legacy. *Revue*, *46*(3), 257–281.

Feijóo, J. C. M., Suárez, F., Chiyón, I., & Alberti, M. G. (2021). Some web-based experiences from flipped classroom techniques in aec modules during the Covid-19 lockdown. *Education Sciences*, *11*(5). https://doi.org/10.3390/educsci11050211

Field, A. (2009). *Discovering Statistics using SPSS.* (3rd ed.). Sage.

Flipped Learning Network (FLN). (2014). *Flipped Learning Network (FLN) (2014) The Four Pillars of F-L-I-P ™*. www.flippedlearning.org/definition

Fogg, K. C., & Maki, S. J. (2021). A remote flipped classroom approach to teaching introductory biomedical engineering during COVID-19. *Biomedical Engineering Education*, *1*(1), 3–9. https://doi.org/10.1007/s43683-020-00001-4

Foldnes, N. (2016). The flipped classroom and cooperative learning: Evidence from a randomised experiment. *Active Learning in Higher Education*, *17*(1), 39–49.

Forster, M. (2014). Teaching Data Analysis in Large Classes Using Clicker Assessment. *Proceedings of the Ninth International Conference on Teaching Statistics (IASE)*.

Freeman, M., Blayney, P., & Ginns, P. (2006). Anonymity and in class learning: The case for electronic response systems. *Australasian Journal of Educational Technology*, *22*(4), 568–580. http://www.ascilite.org.au/ajet/ajet22/freeman.html

Fullan, M., & Langworthy, M. (2014). *A Rich Seam. How New Pedagogies Find Deep Learning*. Pearson.

Fullerton, J. A., & Umphrey, D. (2001). An analysis of attitudes toward statistics: Gender differences among advertising majors. *Association for Education in Journalism and Mass Communication*, 356–366.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, *1*(1), 1–9. https://doi.org/10.1080/10691898.1993.11910455

Garfield, J. (1995). How students learn statistics. *International Statistical Review/Revue Internationale de Statistique*, *63*(1), 25–34.

Garfield, J. (2013). Cooperative learning revisited: From an instructional method to a way of Life 1. Early experiences with cooperative learning. *Journal of Statistics Education*, *21*(2), 1–9.

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning. Connecting Research and Teaching Practice.* Springer.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372–396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education*, *10*(2). https://doi.org/10.1080/10691898.2002.11910665

Gattuso, L. (2006). Statistics and Mathematics: Is it possible to create fruitful links? *Proceedings of the Seventh International Conference on Teaching Statistics*, 1–6. www.stat.aukland.ac.nz/~iase/publications

Gattuso, L., & Pannone, M. A. (2002). Teacher Training in a Statistics Teaching Experiment. *Proceedings of the Sixth International Conference on Teaching Statistics*, 1–6. www.stat.aukland.ac.nz/~iase/publications

Gauci, S. A., Dantas, A. M., Williams, D. A., & Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, *33*, 60–71. https://doi.org/10.1152/advan.00109.2007.

Glynn, A. N., & Quinn, K. M. (2009). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis*, *18*(1), 36–56. https://doi.org/10.1093/pan/mpp036

González-Gómez, D., Jeong, J. S., Airado Rodríguez, D., & Cañada-Cañada, F. (2016). Performance and perception in the flipped learning model: An initial approach to evaluate the effectiveness of a new teaching methodology in a general science classroom. *Journal of Science Education and Technology*, *25*(3), 450–459. https://doi.org/10.1007/s10956-016-9605-9

González, G. M., & Birch, M. A. (2000). Evaluating the instructional efficacy of computer-mediated interactive multimedia: Comparing three elementary statistics tutorial modules. *Journal of Educational Computing Research*, *22*(4), 411–436. https://doi.org/10.2190/X8PQ-K0GQ-T2DR-XY1A

Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE Life Sciences Education*, *14*(4), 1–8. https://doi.org/10.1187/cbe.15-02-0040

Halili, S. H., & Zainuddin, Z. (2015). Flipping the classroom: What we know and what we don't. *The Online Journal of Distance Education and E-Learning*, *3*(1), 28–35. https://doi.org/10.1016/j.ijporl.2009.03.018

Hamdan, N., McKnight, P., McKnight, K., & Arfstrom, K. M. (2013). *A Review of Flipped Learning.* Pearson. https://doi.org/10.1055/s-0029-1198237

Han, J. H., & Finkelstein, A. (2013). Understanding the effects of professors' pedagogical development with Clicker Assessment and Feedback technologies and the impact on students' engagement and learning in higher education. *Computers and Education*, *65*, 64–76. https://doi.org/10.1016/j.compedu.2013.02.002

Hancock, T. M. (2010). Use of audience response systems for summative assessment in large classes. *Australasian Journal of Educational Technology*, *26*(2), 226–237.

Hannafin, M. J., & Land, S. M. (2000). Technology and student-centered learning in higher education: Issues and practices. *Journal of Computing in Higher Education*, *12*(1), 3–30. https://doi.org/10.1007/BF03032712

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., Baumer, B., Hall-Holt, O., Murrell, P., Peng, R., Roback, P., Temple Lang, D., & Ward, M. D. (2015). Data science in statistics curricula: Preparing students to "Think with Data". *American Statistician*, *69*(4), 343–353. https://doi.org/10.1080/00031305.2015.1077729

Harris, B. F., Harris, J., Reed, L., & Zelihic, M. M. (2016). Flipped classroom: Another tool for your pedagogy tool box. *Developments in Business Simulation and Experiential Learning*, *43*(1), 325–333. https://journals.tdl.org/absel/index.php/absel/article/view/3061

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.1111/j.1365-2923.2009.03542.x

Haughton, J., & Kelly, A. (2015). Student performance in an introductory business statistics course: Does delivery mode matter? *Journal of Education for Business*, *90*(1), 31–43. https://doi.org/10.1080/08832323.2014.968518

Hoekstra, A. (2008). Vibrant student voices: Exploring effects of the use of clickers in large college courses. *Learning, Media and Technology*, *33*(4), 329–341. https://doi.org/10.1080/17439880802497081

Hourigan, M., & O'Donoghue, J. (2007). Mathematical under-preparedness: The influence of the pre-tertiary mathematics experience on students' ability to make a successful transition to tertiary level mathematics courses in Ireland. *International Journal of Mathematical Education in Science and Technology*, *38*(4), 461–476. https://doi.org/10.1080/00207390601129279

Hoyles, C., Newman, K., & Noss, R. (2001). Changing patterns of transition from school to university mathematics. *International Journal of Mathematical Education in Science and Technology*, *32*(6), 829–845. https://doi.org/10.1080/00207390110067635

Hung, H. T. (2017). Clickers in the flipped classroom: Bring your own device (BYOD) to promote student learning. *Interactive Learning Environments*, *25*(8), 983–995. https://doi.org/10.1080/10494820.2016.1240090

Huntley, B. (2009). *Wits first-year pass-rate for mathematics dropped by 37%! Inflated matric results created unjustified expectations.* http://www.mathsexcellence.co.za/papers/Wits_first_year_pass_rate.pdf

Hwang, J. H., & Wolfe, K. (2010). Implications of using the electronic response system in a large class. *Journal of Teaching in Travel & Tourism*, *10*(3), 265–279. https://doi.org/10.1080/15313220.2010.503536

Jansen, J. D. (1998) Curriculum reform in South Africa: A critical analysis of outcomes-based education. *Cambridge Journal of Education*, *28*(3), 321-331. https://doi.org/10.1080/0305764980280305

Jensen, J. L., Kummer, T. A., & Godoy, P. D. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE - Life Sciences Education*, *14*(1), 1–12. https://doi.org/10.1187/10.1187/cbe.14-08-0129

Jonsdottir, A. H., Bjornsdottir, A., & Stefansson, G. (2017). Difference in learning among students doing pen-and-paper homework compared to web-based homework in an introductory statistics course. *Journal of Statistics Education*, *25*(1), 12–20. https://doi.org/10.1080/10691898.2017.1291289

Judson, E., & Sawada, D. (2002). Learning from past and present: Electronic response systems in college lecture halls. *Journal of Computers in Mathematics and Science Teaching*, *21*(2), 167–181.

Kantanis, T. (2000). The role of social transition in students’ adjustment to the first-year of university. *Journal of Institutional Research*, *9*, 100–110.

Kaplan, J. J. (2011). Innovative activities: How Clickers can facilitate the use of simulations in large lecture classes. *Technology Innovations in Statistics Education*, *5*(1), 1–15.

Kaplan, J. J. (2008). i>clicker Pedagogy Case Study. *Invited Paper for the Iclicker.Com Website*, 1–7. http:/iclicker.com/dnn/Portals/0/Kaplan_Case_Study.pdf

Kay, R., & LeSage, A. (2009). A strategic assessment of audience response systems used in higher education. *Australasian Journal of Educational Technology*, *25*(2), 235–249. https://doi.org/10.1016/j.compedu.2009.05.001

Kay, R., & Macdonald, T. (2016). *Comparing Flipped, Active, and Lecture-Based Teaching Approaches in Higher Education*. Conference Ed Media 2016. https://doi.org/10.13140/RG.2.1.4727.4481

Khan, R. (2014). Identifying threshold concepts in first-year statistics. *Education Research and Perspectives*, *41*, 217.

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435–454. https://doi.org/10.1017/pan.2019.11

Koenig, K. (2010). Building acceptance for pedagogical reform through wide-scale implementation of Clickers. *Journal of College Science Teaching*, *39*(3), 46–50.

Koppel, N., & Berenson, M. (2009). Ask the audience - Using Clickers to enhance introductory business statistics courses. *Information Systems Education Journal*, *7*(92), 1–18.

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory Into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104

Kyei-Blankson, L. (2009). Enhancing Student Learning in a Graduate Research and Statistics Course with Clickers. In *Educause review online*. http://www.educause.edu/ero/article/enhancing-student-learning-graduate-research-and-statistics-course-clickers

Lage, M., Platt, G., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *Journal of Economic Education*, *31*(1), 30–43.

Lantz, M. E., & Stawiski, A. (2014). Computers in human behavior effectiveness of clickers: Effect of feedback and the timing of questions on learning. *Computers in Human Behavior*, *31*, 280–286. https://doi.org/10.1016/j.chb.2013.10.009

Larreamendy-Joerns, J., Leinhardt, G., & Corredor, J. (2005). Six online statistics courses: Examination and review. *American Statistician*, *59*(3), 340–351. https://doi.org/10.1198/000313005X54162

Lasry, N. (2008). Clickers or flashcards: Is there really a difference ? *The Physics Teacher*, *46*(1), 242–244. https://doi.org/10.1119/1.2895678

Latorre-Cosculluela, C., Suárez, C., Quiroga, S., Sobradiel-Sierra, N., Lozano-Blasco, R., & Rodríguez-Martínez, A. (2021). Flipped classroom model before and during COVID-19: using technology to develop 21st century skills. *Interactive Technology and Smart Education*, (ahead-of-print). https://doi.org/10.1108/ITSE-08-2020-0137

Laxman, K. (2011). A study on the adoption of clickers in higher education. *Australasian Journal of Educational Technology*, *27*(8), 1291–1303.

Lentin, J., Jonsdottir, A. H., Stern, D., Mokua, V., & Stefansson, G. (2014). A mobile web for enhancing Statistics and Mathematics education. *Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS). ArXiv: 1406.5004*.

Liou, W. K., Bhagat, K. K., & Chang, C. Y. (2016). Beyond the flipped classroom: A Highly Interactive Cloud-classroom (HIC) embedded into basic materials science courses. *Journal of Science Education and Technology*, *25*(3), 460–473. https://doi.org/10.1007/s10956-016-9606-8

Littauer, R. (1972). Instructional Imp a low-cost electronic student response system technology brings instruction home. *Educational Technology*, *12*(10), 69–71.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician*, *54*(3), 196–206.

Luk, H. S. (2005). The gap between secondary school and university mathematics. *International Journal of Mathematical Education in Science and Technology*, *36*(2–3), 161–174. https://doi.org/10.1080/00207390412331316988

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *ArXiv:1705.07874*, *Section 2*, 1–10.

Lunsford, M. L., & Pendergrass, M. (2016). Making online homework work. *Primus*, *26*(6), 531–544. https://doi.org/10.1080/10511970.2015.1110219

Lutz, S. T., & Huitt, W. G. (2004). Connecting cognitive development and constructivism: Implications from theory for instruction and assessment. *Constructivism in the Human Sciences*, *1*(9), 67–90. http://www.edpsycinteractive.org/papers/cogdev.pdf

Majerich, D. M., Stull, J. C., Varnum, S. J., & Ducette, J. P. (2011). Facilitation of formative assessments using clickers in a university physics course. *Interdisciplinary Journal of ELearning and Learning Objects*, *7*, 11–24.

Mason, G., Schuman, T. R., & Cook, K. E. (2013). Inverting (Flipping) Classrooms – Advantages and Challenges. *120th ASEE Annual Conference Exposition*, 1–21. http://www.asee.org/public/conferences/20/%0Apapers/7171/download

Mateo, Z. F. (2010). Creating Active Learning in a Large Introductory Statistics Class using Clicker Technology. *Proceedings of the Eighth International Conference on Teaching Statistics (IASE)*.

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice*, *41*(4), 226–233.

Mayer, R. E., Stull, A., Deleeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*(1), 51–57. https://doi.org/10.1016/j.cedpsych.2008.04.002

Mazur, E. (1997). *Peer Instruction : A User's Manual*. Prentice Hall. https://doi.org/https://doi.org/10.1119/1.19265

Mazur, E. (2009). “Farewell lecture?” *Science*, *323*, 50–51.

McCarthy, J. (2016). Reflections on a flipped classroom in first year higher education. *Issues in Educational Research*, *26*(2), 332–350.

McDonald, J. H. (2014). *Handbook of Biological Statistics.* (3rd ed.). Sparky House Publishing.

McGee, M., Stokes, L., & Nadolsky, P. (2016). Just-in-time teaching in statistics classrooms. *Journal of Statistics Education*, *24*(1), 16–26. https://doi.org/10.1080/10691898.2016.1158023

McKenna, P. (2019). Multiple choice questions: answering correctly and knowing the answer. *Interactive Technology and Smart Education*, *16*(1), 59–73. https://doi.org/10.1108/ITSE-09-2018-0071

Miller, G. A. (1956). The Magical Number Seven Plus or Minus Two: Some limits on our capacity for processing information. *The Psychological Review*, *63*(2), 81–97.

Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *Journal of Online Learning and Teaching*, *9*(2), 149–159.

Mocko, M., & Jacobbe, T. (2010). The Effects of an Audience Response System (ARS) on Achievement and Attitudes towards Statistics in an Introductory Statistics Class. *Eighth International Conference on Teaching Statistics (IASE)*.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, *65*(2), 123. https://doi.org/10.2307/1403333

Movarec, M., Williams, A., Aguilar-Roca, N., & O’Dowd, D. K. (2010). Learn before lecture: A strategy that improves learning outcomes in a large introductory biology class. *Life Sciences Education*, *9*, 473–481. https://doi.org/10.1187/cbe.10

Muller, F. (2018). *XGBoost Tree vs. Linear*. [Blog post]. www.statworx.com/blog/xgboost-tree-vs-linear/

Munir, M. T., Baroutian, S., Young, B. R., & Carter, S. (2018). Flipped classroom with cooperative learning as a cornerstone. *Education for Chemical Engineers, 23*(May), 25–33. https://doi.org/10.1016/j.ece.2018.05.001

Naccarato, E., & Karakok, G. (2015). Expectations and implementations of the flipped classroom model in undergraduate mathematics courses. *International Journal of Mathematical Education in Science and Technology*, *46*(7), 968–978. https://doi.org/10.1080/0020739X.2015.1071440

Nielsen, P. L., Bean, N. W., & Larsen, R. A. A. (2018). The impact of a flipped classroom model of learning on a large undergraduate statistics class. *Statistics Education Research Journal*, *17*(1), 121–140.

Niu, X., Wang, L., & Yang, X. (2019). A comparison study of credit card fraud detection: Supervised versus unsupervised. *ArXiv*.

Nolan, D., & Lang, D. T. (2007). Dynamic, interactive documents for teaching statistical practice. *International Statistical Review*, *75*(3), 295–321. https://doi.org/10.1111/j.1751-5823.2007.00025.x

Nolan, D., & Speed, T. (2000). *Stat Labs: Mathematical Statistics Through Applications*. Springer. https://doi.org/10.1198/tech.2001.s635

North, D., & Scheiber, J. (2008). Introducing Statistics at school level in South Africa. The crucial role played by the National Statistics Office in training in-service teachers. *Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*, 1–6.

North, D., & Zewotir, T. (2006). Introducing Statistics at School Level in South Africa. *Proceedings of the Seventh International Conference on Teaching Statistics*, 1–6. www.stat.aukland.ac.nz/~iase/publications

Nyika, A. (2015). Mother tongue as the medium of instruction at developing country universities in a global context. *South African Journal of Science*, *111*(1–2), 1–5. https://doi.org/10.17159/sajs.2015/20140005

Overmyer, G. R. (2014). The flipped classroom model for college algebra: Effects on student achievement. In *ProQuest Dissertations and Theses*. https://search.proquest.com/docview/1615100148?accountid=15272

Palocsay, S. W., & Stevens, S. P. (2008). A study of the effectiveness of web-based homework in teaching undergraduate business statistics. *Decision Sciences Journal of Innovative Education*, *6*(2), 213–232.

Pather, S., Norodien-Fataar, N., Cupido, X., & Mkonto, N. (2017). First year students' experience of access and engagement at a University of Technology. *Journal of Economic Education*, *69*, 161–184. http://joe.ukzn.ac.za

Patron, H., & Smith, W. J. (2009). Improving student performance in business statistics courses with online assessments. *Journal of Economics and Finance Education*, *8*(1), 39–49.

Peterson, D. J. (2016). The flipped classroom improves student achievement and course satisfaction in a statistics course: A quasi-experimental study. *Teaching of Psychology*, *43*(1), 10–15. https://doi.org/10.1177/0098628315620063

Phillips, L., & Phillips, M. (2016). Improved student outcomes in a flipped statistics course. *Administrative Issues Journal*, *6*(1), 88–98. https://doi.org/10.5929/2016.6.1.5

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice*, *41*(4), 219–225. https://doi.org/10.1207/s15430421tip4104_3

Premkumar, K. (2016). Use of student response systems for summative assessments. *Creative Education*, *7*, 1851–1860.

Prensky, M. (2001). Digital natives, digital immigrants. *MCB University Press*, *9*(5), 1–6.

Prensky, M. (2011). A huge leap for the classroom. True peer-to-peer learning, enhanced by technology. *Educational Technology*, *Nov-Dec*, 1–4. https://www.marcprensky.com/writing

Rana, N. P., & Dwivedi, Y. K. (2016). Using Clickers in a large business class: Examining use behavior and satisfaction. *Journal of Marketing Education*, *38*(1), 47–64. https://doi.org/10.1177/0273475315590660

Reyneke, F., Fletcher, L., & Harding, A. (2018). The effect of technology-based interventions on the performance of first year university statistics students. *African Journal of Research in Mathematics, Science and Technology Education*, *7295*, 231–242. https://doi.org/10.1080/18117295.2018.1477557

Reyneke, F., Fletcher, L., & Harding, A. (2021). Enhancing a flipped statistics first year course by using QT-Clickers. *Journal of Statistics and Data Science Education*, *29*(1), 71–83. https://doi.org/10.1080/10691898.2021.1895694

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17 Aug.*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Roehl, A. M. Y., Reddy, S. L., & Shannon, G. J. (2013). The flipped classroom: An opportunity to engage millennial students through active learning strategies. *Journal of Family Consumer Science*, *105*(2), 44–50.

Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two. *Mathematical and Computer Modelling*, *38*(3–4), 233–244. https://doi.org/10.1016/S0895-7177(03)90083-5

Santoro, K., & Bilisoly, R. (2014). *Creating, Automating, and Assessing Online Homework in Introductory Statistics and Mathematics Classes*. JSM Central Publishing.

Schutz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences, 10*(4), 291–308. https://doi.org/10.1016/S1041-6080(99)80124-1

Schwartz, T. A., Andridge, R. R., Sainani, K. L., Stangle, D. K., & Neely, M. L. (2016). Diverse perspectives on a flipped biostatistics classroom. *Journal of Statistics Education*, *24*(2), 74–84. https://doi.org/10.1080/10691898.2016.1192362

Sharma, M. D., Khachan, J., Chan, B., & O'Byrne, J. (2005). An investigation of the effectiveness of electronic classroom communication systems in large lecture classes. *Australasian Journal of Educational Technology*, *21*(2), 137–154.

Shinaberger, L. (2017). Components of a Flipped classroom influencing student success in an undergraduate business statistics course. *Journal of Statistics Education*, *25*(3), 122–130. https://doi.org/10.1080/10691898.2017.1381056

Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE Life Sciences Education*, *10*(1), 55–63. https://doi.org/10.1187/cbe.10-08-0101

Smolira, J. C. (2008). Student perceptions of online homework in introductory finance courses. *Journal of Education for Business*, *84*(2), 90–95. https://doi.org/10.3200/JOEB.84.2.90-95

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research*, *81*(1), 97–128. https://doi.org/10.3102/0034654310378174

Stowell, J. R., & Nelson, J. M. (2007). Benefits of electronic audience response systems on student participation, learning, and emotion. *Teaching of Psychology*, *34*(4), 253–258. https://doi.org/10.1080/00986280701700391

Strayer, J. F. (2007). The Effects of the Classroom Flip on the Learning Environment: A Comparison of Learning Activity in a Traditional Classroom and a Flip Classroom That Used an Intelligent Tutoring System. [Doctoral thesis, Ohio State University], *68*(8-A), 3320. https://doi.org/10.1109/MILCOM.2003.1290218

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning Environments Research*, *15*(2), 171–193. https://doi.org/10.1007/s10984-012-9108-4

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x

Suleiman, N. (2018). Implementing blended learning and flipped learning models in the university classroom: A case study. *Teaching English with Technology*, *18*(4), 34–47.

Sutherland, T. E., & Bonwell, C. C. (1996). *Using Active Learning in College Classes: A Range of Options for Faculty*. Jossey-Bass.

Taylor, M. (2011). Teaching Generation NeXt: Methods and techniques for today's learners. In A collection of techniques of self-study and institutional improvement. *Higher Learning and Commission*, 113–119. http//www.taylorprograms.com/images/Techniques_article_2011.pdf.

Tchantchane, A., & Fortes, P. C. (2011). A multivariate survey analysis: Evaluation of technology integration in teaching statistics. *Proceedings Ascilite 2011*, 1227–1238.

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21stcentury: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, *20*(2), 1–56. https://doi.org/10.1080/10691898.2012.11889641

Titman, A. C., & Lancaster, G. A. (2011). Personal response systems for teaching postgraduate statistics to small groups. *Journal of Statistics Education*, *19*(2), 1–20.

Tomic, W., & Johannes, K. (1996). Three theories of cognitive representation and their evaluation standards of training effects. *Information Analysis*, *70*.

Trees, A. R., & Jackson, M. H. (2007). The learning environment in clicker classrooms: Student processes of learning and involvement in large university-level courses using student response systems. *Learning, Media and Technology*, *32*(1), 21–40. https://doi.org/10.1080/17439880601141179

Tucker, B. (2012). The Flipped Classroom. *Education Next*, *12*(1), 82–83.

Vidic, N. S., Clark, R. M., & Claypool, E. G. (2011). Flipped Classroom approach : Probability and statistics course for engineers. *American Society for Engineering Education*.

Vygotsky, L. S. (1978). Interaction between learning and development. *Mind and Society. Harvard University Press.*, 79–91.

Wang, Y., Chung, C., & Yang, L. (2014). Using Clickers to enhance student learning in mathematics. *International Education Studies*, *7*(10), 1–13. https://doi.org/10.5539/ies.v7n10p1

Wang, Z., & Englander, F. (2010). A cross-disciplinary perspective on explaining student performance in introductory statistics - What is the relative impact of procrastination? *College Student Journal*, *44*(2), 458–471.

Wessels, H. (2008). Statistics in the South Africn curriculum: Content, Assessment and Teacher Training. *Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*, 1–6.

Wilde, K. (2014). Clickers for Engagement in the Large Undergraduate Statistics Classroom: Do Clickers Improve Final Grades? *Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS)*.

Williams, A. (2012). Online homework vs. traditional homework: Statistics anxiety and self-efficacy in an educational statistics course. *Technolgy Innovations in Statistics Education*, *6*(1). https://escholarship.org/uc/item/32j2998k

Wilson, D., & Allen, D. (2011). Success rates of online versus traditional college students. *Research in Higher Education Journal*, *14*, 1–9. http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=70547715&site=ehost-live

Wilson, L. O. (2016). Anderson and Krathwohl Bloom's Taxonomy revised. *The Second Principle*, 1–8. https://quincycollege.edu/content/uploads/Anderson-and-Krathwohl_Revised-Blooms-Taxonomy.pdf%0Ahttps://thesecondprinciple.com/teaching-essentials/beyond-bloom-cognitive-taxonomy-revised/%0Ahttp://thesecondprinciple.com/teaching-essentials/beyond-bloom-cog

Wilson, S. G. (2013). The Flipped Class: A method to address the challenges of an undergraduate statistics course. *Teaching of Psychology*, *40*(3), 193–199. https://doi.org/10.1177/0098628313487461

Winquist, J. R., & Carlson, K. A. (2014). Flipped statistics class results: Better performance than lecture over one year later. *Journal of Statistics Education*, *22*(3), 10. https://doi.org/10.1080/10691898.2014.11889717

Wit, E. (2003). Who wants to be … The use of a personal response system in statistics teaching. *MSOR Connections*, *3*(2), 14–20.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wooten, T., & Dillard-Eggers, J. (2013). An investigation of online homework: Required or not required? *Contemporary Issues In Education Research – Second Quarter*, *6*(2), 189–199. http://www.cluteinstitute.com/

Xu, W., Zhang, Y., Su, C., Cui, Z., & Qi, X. (2014). Roles of technology in student learning of university level biostatistics. *Statistics Education Research Journal*, *13*(1), 66–76.

Yourstone, S. A., Kraye, H. S., & Albaum, G. (2008). Classroom questioning with immediate electronic response: Do Clickers improve learning? *Decision Sciences Journal of Innovative Education*, *6*(1), 75–88. https://doi.org/10.1111/j.1540-4609.2007.00166.x

Zewotir, T., & North, D. (2011). Opportunities and challenges for statistics education in South Africa. *Pythagoras*, *32*(2), 1–5. https://doi.org/10.4102/pythagoras.v32i2.28

Zhou, J. (2020). A critical discussion of Vygotsky and Bruner's Theory and their contribution to understanding of the way students learn. *Review of Educational Theory*, *3*(4), 82. https://doi.org/10.30564/ret.v3i4.2444

ANNEXURES

ANNEXURE 1 – Ethical clearance

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ETHICS COMMITTEE

Faculty of Natural and Agricultural Sciences

3 October 2014

Dr L Fletcher
Department of Statistics
University of Pretoria
0002

Dear Dr Fletcher

EC140721-067 Investigating different initiatives to address the success rate of large first level statistics module.

Your application conforms to the requirements of the NAS Ethics Committee. The review took longer because of the nature of the research on student performance. Please not that it is UP policy that student performance at UP cannot be used in a publication that might suggest UP as a benchmark or that UP students are a representative sample of students in the subject in South Africa. If an article is prepared for publication, it must have the approval of the Ethics Committee. It is also necessary to note that differentiation among students in terms of ethnic, gender or any other identity or criteria must be dealt with sensitively. Please take the liberty to call me should you wish to obtain more clarity in this regard.

Kind regards

Prof NH Casey
Chairman: Ethics Committee

Agriculture Building 10-20
University of Pretoria

Tel: 012 420 4107
Fax: 012 420 3290

ethics.nas@up.ac.za

ANNEXURE 2 – Gr12 Mathematics prerequisite

Table A2-1 summarises the results of the 2003 Grade 12 learners' mathematics marks versus their 2004 STK110 final marks.

Table A2-1 Average 2004 STK110 final mark versus the 2003 Grade 12 mathematics mark

GR 12 mathematics	Total	STK110-Mean	Standard deviation
Level-HG*			
A (80 – 100)	160	76.87	13.30
B (70 – 79)	139	62.33	11.56
C (60 – 69)	211	55.46	13.26
D (50 – 59)	236	**48.75**	11.69
E (40 – 49)	152	41.14	11.53
Level-SG*			
A (80 – 100)	161	**43.73**	12.19
B (70 – 79)	113	35.89	10.49
C (60 – 69)	93	32.15	9.31
D (50 – 59)	90	30.54	8.53
E (40 – 49)	11	39.18	10.94
Total	**1366**		

*Two levels for Grade 12 mathematics in South Africa, namely higher Grade (HG) and standard Grade (SG)

The following conclusions were deduced from Table A2-1:

1. Performance in STK110 is correlated with the mathematics results. The average STK110-mark drops over the mathematics categories for the HG students as well as for the SG students.
2. The average STK110 mark for Grade D students on HG is 48.75, which is higher than the average mark for Grade A students on SG, namely 43.73.
3. These results verify presumptions that students who obtained a D (i.e. 50% - 59%) on HG performed better than most of the SG students.
4. The prerequisite for STK120 (second semester module) is 40% for STK110. The average STK110 mark of SG students with a mathematics symbol of B or lower, is less than 40%, therefore most of those students could not continue with STK120.

Contingency tables were consequently constructed, and chi-squared tests done to assess if there was a statistically significant association between STK110 students' final marks and their mathematics performance.

Table A2-2: STK110 and Grade 12 mathematics HG

2004 STK110	**2003 Grade 12 mathematics (HG)**					
	A	B	C	D	E	TOTAL
FAIL (<50) Column %	2 1.25	15 10.79	57 27.01	106 44.92	113 **74.34**	293
PASS (50+) Column %	158 98.75	124 89.21	154 72.99	130 55.08	39 **25.66**	605
TOTAL	160	139	211	236	152	898

Table A2-3: STK110 and Grade 12 mathematics SG

2004 STK110	**2003 Grade 12 mathematics (SG)**					
	A	B	C	D	E	TOTAL
FAIL (<50) Column %	105 65.22	96 84.96	89 95.70	86 95.56	9 81.82	385
PASS (50+) Column %	56 **34.78**	17 **15.04**	4 **4.30**	4 **4.44**	2 **18.18**	83
TOTAL	160	139	211	236	152	468

The chi-squared test statistics for Table A2-2 ($\chi^2 = 271.77, df = 4$) and Table A2-3 ($\chi^2 = 55.03, df = 4$) were both highly significant ($p < 0.0001$), which confirmed the close relationship between first semester statistics and Grade 12 mathematics. As a result, higher Grade 12 mathematics marks as prerequisite for STK110 was implemented in 2005: Mathematics HG at least 50% (D+), previously at least 40% (E+), and Mathematics SG at least 70% (B+), previously at least 50% (D+).

ANNEXURE 3 – Timeline of events up to 2011

Figure A3-1 displays the pass rates and interventions of first-year statistics students. The solid black line displays the pass rates since 2001 for the total number of students who registered for the module STK110. The total is divided into the pass rates for:

- students who registered for the **first time** (fine dotted blue line); and
- students who **repeated the module** (coarse dotted red line).

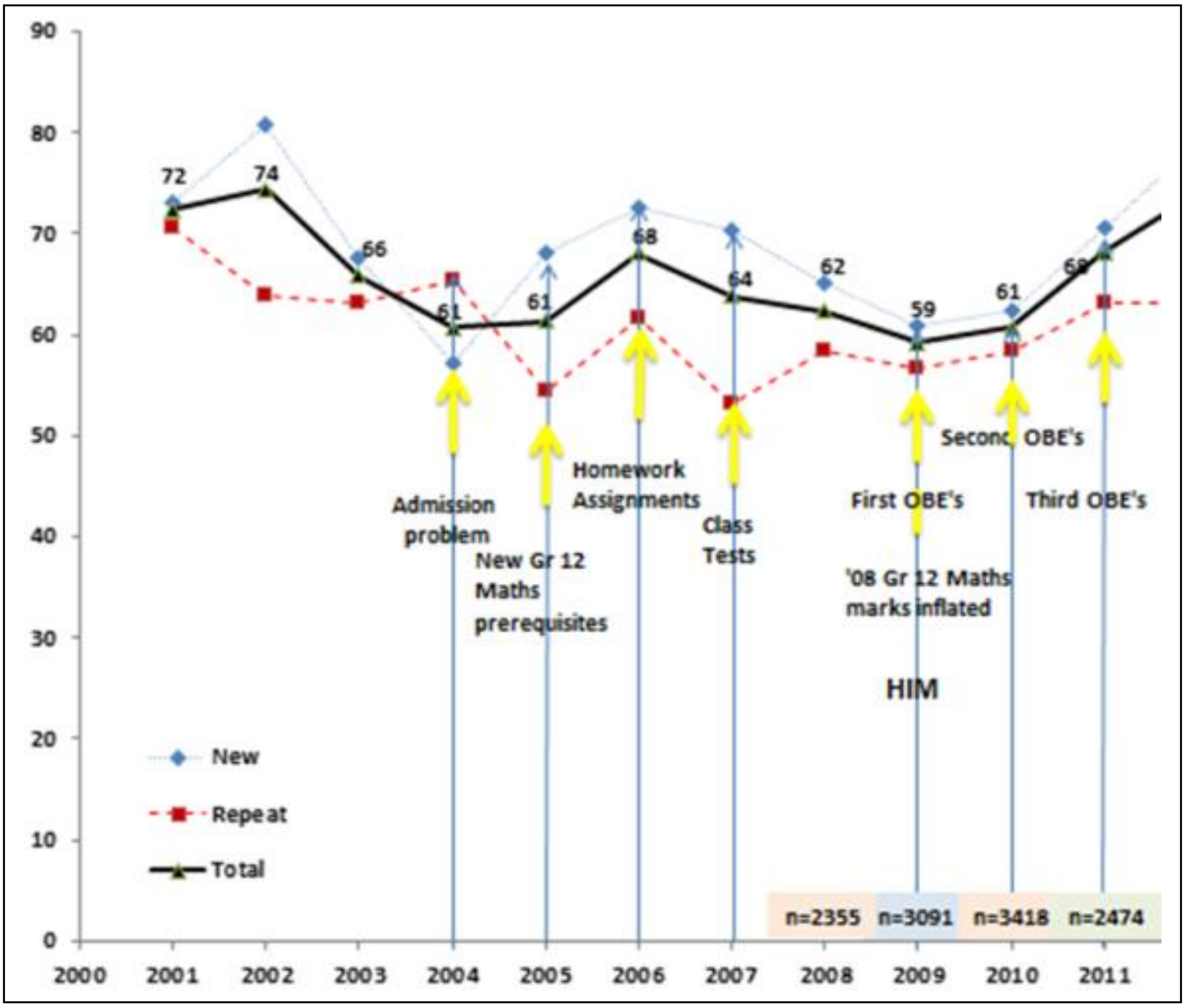


Figure A3-1: Pass rates and interventions of STK110

Various important events and interventions since 2004 are indicated on the graph by arrows. During the first part of the millennium there was a steady drop in the pass rates which was a cause for concern. The disturbingly low pass rate in 2004 (for the new students without the supplementary examination results) prompted a revision of the Grade 12 mathematics entrance criteria for the first year statistics students at UP. The higher Grade 12 mathematics prerequisites had a positive effect on the pass rate of the new students of 2005, but the pass rate of the repeat students dropped even further. Compulsory homework assignments were implemented in 2006. The pass rate of both new and repeat students improved, due to students who could copy their peers' assignments, which inflated their marks because of the substantial weight of the assignments towards the final mark. In 2007 the assignments were therefore replaced by class tests and the students' marks went down again.

ANNEXURE 4 – OBE (Outcome-Based Education)

Since 1995, which denoted the end of apartheid, after South Africa's first democratic election that resulted in a new government under a new dispensation, major changes took place in the South African school education system.

OBE was introduced into schools in South Africa in 1998, for all learners in Grades 1 to 6 and progressively phased in after that. The OBE system not only involved a change in approach, but also a change in curriculum. Learners who finished secondary school in 2008 formed the first group of students who had full exposure to OBE over their entire school career.

OBE introduced radical changes regarding the mathematics syllabus. Previously mathematics was offered at Higher Grade (HG) and Standard Grade (SG) at school. With the implementation of OBE, mathematics was only offered at one level, which led to a drop in mathematics standards of school leavers (Jansen, 1998). It is known that the mathematics marks of the 2008 Grade 12 learners were inflated (Huntley, 2009). This was done to partially compensate for ill-prepared teachers since many problems were encountered with teacher training (Gattuso, 2006; Gattuso & Pannone, 2002) in the new OBE system (North & Scheiber, 2008; Wessels, 2008). These students entered the university in 2009 and were our first intake of students who followed the new OBE curriculum (Engelbrecht et al., 2010).

ANNEXURE 5 – Logit model

A logit model was fitted to the data to explore the relationship between statistics achievement and Grade 12 mathematics marks in 2009 and 2010. STK110 achievement (dependent variable) is binary (Fail = 1 and Pass = 0). The independent variables are Grade 12 mathematics marks, categorised into four levels (50% to 59%, 60% to 69%, 70% to 79% and 80%+) and year matriculated with two levels (2009 and 2010) (Table).

Table A5-1: Three-way table of STK110 results by 2009 and 2010 Grade 12 mathematics results

STK110 achievement	Category of Gr12 mathematics mark in 2009/2010								Total
	50% - 59%		**60% - 69%**		**70% - 79%**		**80%+**		
	2009	2010	2009	2010	2009	2010	2009	2010	
Fail	233	107	220	99	120	62	25	12	878
Row%	26.5	12.2	25.1	11.3	13.7	7.1	2.8	1.3	100
Pass	136	88	320	220	401	285	395	351	2196
Row%	6.2	4.0	14.6	10.0	18.3	13.0	18.0	16.0	100
Total	369	195	540	319	521	347	420	363	3074

The odds of failing STK110 were modelled using a saturated logit model:

$$ln(odds) = \mu + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}, j = 1,2,3,4 \text{ and } k = 1,2 \tag{13}$$

Where

$\mu =$ the overall average effect of the levels

$\lambda_j^B =$ the effect of the $\mathrm{j^{th}}$ mathematics category, $j = 1,2,3,4$

$\lambda_k^C =$ the effect of the $\mathrm{k^{th}}$ year, $k = 1,2$

$\lambda_{jk}^{BC} =$ the interaction effect of the $\mathrm{j^{th}}$ mathematics category and the k^{th} year

Table A5-2: Maximum Likelihood Analysis of Varianc**e**

Source	DF	Chi-Square	*p* -value
Intercept	1	405.06	< 0.0001
Category	3	375.36	< 0.0001
Year	1	13.54	0.0002
Category*year	3	0.68	0.8779

Table A5-3: The logit model output

Effect	Level	Odds of Failing	Probability
Category	50% - 59%	1.4433	0.59
	60% - 69%	0.5562	0.36
	70% - 79%	0.2552	0.20
	80%+	0.0465	0.04
Year	2009	0.3865	0.28
	2010	0.2526	0.20
Category*Year	(50% - 59%)*2009	1.7133	0.63
	(50% - 59%)*2010	1.2158	0.55
	(60% - 69%)*2009	0.6875	0.41
	(60% - 69%)*2010	0.4500	0.31
	(70% - 79%)*2009	0.2993	0.23
	(70% - 79%)*2010	0.2175	0.15
	(80%+)*2009	0.0633	0.06
	(80%+)*2010	0.0342	0.03

The most important finding of this model is that the odds of students failing STK110, if their mathematics mark is in the category 50% to 59%, are 1.44 (Table A5-3). This translates to a probability of 0.59 to fail STK110, i.e. only a 41% chance to pass. As expected, the odds of failing decrease steadily as the Grade 12 mathematics result increases. From the maximum likelihood analysis of variance table (Table A5-2) all the effects were highly significant except for the interaction effect category*year ($p - value = 0.8779$). Although the interaction effect was not significant, it is notable to observe from

Table A5-3 that the probability to fail STK110 in 2009 was higher compared to 2010 across all the Grade 12 mathematics categories.

ANNEXURE 6 – Questionnaires

QUESTIONNAIRE - APLIA 2012

PLEASE ANSWER THE QUESTIONS BY CROSSING (X) THE RELEVANT BOX OR WRITING YOUR ANSWER IN THE GIVEN SPACE

The Statistics lecturers want to improve the pass rate of the first year students and need your honest input

1. What is your student number?

2. Did you take STK110 in the first semester of 2012?
If yes, go to Question 3, else proceed to Question 4

Yes	No

3. How did Aplia for STK110 influence the following outcomes:

a) Your semester mark		Decreased	Increased
b) Understanding of Statistics		Not better	Better
c) Perception of Statistics	Worse	Same	Better

d) General comment: ______________________________

____

4. How did Aplia for STK120/161 influence the following outcomes:

a) Understanding of Statistics		Not better	Better
b) Perception of Statistics	Worse	Same	Better

c) General comment: ______________________________

____

5. Aplia assignments were compulsory for students. Please give us a brief feedback on your experience of Aplia:

____

____

6. Did you experience Aplia differently in STK120/161 than in STK110? **If Yes, please explain in what way?**

Yes	No

____

____

____

QUESTIONNAIRE - Flipped classroom 2013 - 2014

We are in the process of implementing a flipped classroom approach and would value your feedback regarding a few concepts.

PLEASE ANSWER THE QUESTIONS BY CROSSING (X) THE RELEVANT BOX OR WRITING YOUR ANSWER IN THE GIVEN SPACE

What is your student number?

A. Pre-Class Aplia assignments:

1 Did you read the prescribed chapters in the textbook **before** you have attepted the assignments?

Yes	No

2 **Rank** your conceptions of the pre-Class Aplia assignments in order of importance.

A: Helped to explain difficult concepts
B: Forced me to read the textbook
C: Helped me to understand Statistics
D: It improved my marks
E: There was a language barrier
F: I have gone prepared to class

	Rank
1	
2	
3	
4	
5	
6	

3 General comment on pre-Class Aplia assignments: ____________________

____

____

B. TUTORIALS:

1 Did you complete the tutorial on each chapter **before** going to class?

Yes	No

2 **Rank** your conceptions of the tutorials (TUT's) in order of importance:

A: Helped with difficult concepts
B: Good preparation for tests
C: I didn't attend TUT classes
D: Helped me to understand Statistics
E: It improved my marks
F: Waste of time

	Rank
1	
2	
3	
4	
5	
6	

3 General comment on tutorials (TUT's): ____________________

____

____

C. TUTORING SYSTEM:

1 Did you make use of any part of our tutoring system ?

Yes	No

2 Which **one** of the following tutoring sections did you find the best?

A: Hot spot
B: Revision classes
C: Under 50 classes

A
B
C

3 General comment on the tutoring system: ____________________

QUESTIONNAIRE - Student progress 2015 - 2016

We want to evaluate a few learning tools and would value your feedback regarding certain concepts.

PLEASE ANSWER THE QUESTIONS BY CROSSING (X) THE RELEVANT BOX OR WRITE YOUR ANSWER IN THE GIVEN SPACE

A. APLIA

	Statement		Don't agree										Agree 100%
1	I read the prescribed sections in the textbook before I attepted the pre-class assignments.	1	0	1	2	3	4	5	6	7	8	9	10
2	I read the notes before I attempted the pre-class assignments.	2	0	1	2	3	4	5	6	7	8	9	10
3	I found the textbook helpful and valuable.	3	0	1	2	3	4	5	6	7	8	9	10
4	The pre-class assignments helped me to go prepared to class.	4	0	1	2	3	4	5	6	7	8	9	10
5	The pre-class assignments helped to explain difficult concepts.	5	0	1	2	3	4	5	6	7	8	9	10
6	I did better in the post-class assignments.	6	0	1	2	3	4	5	6	7	8	9	10
7	The post-class assignments were interesting real life examples.	7	0	1	2	3	4	5	6	7	8	9	10
8	I experienced technical problems with Aplia.	8	0	1	2	3	4	5	6	7	8	9	10

9 General comment on Aplia: ______________________________

____

____

B. CLICKERS

	Statement												
1	I prefer to use a clicker than raising my hand during class activities.	1	0	1	2	3	4	5	6	7	8	9	10
2	The clicker made class more fun and exciting.	2	0	1	2	3	4	5	6	7	8	9	10
3	The clicker questions in class helped me to learn.	3	0	1	2	3	4	5	6	7	8	9	10
4	The clicker questions in class help me to be more involved and engaged in a large lecture hall.	4	0	1	2	3	4	5	6	7	8	9	10
5	The clicker questions in class forced me to do my tutorial beforehand.	5	0	1	2	3	4	5	6	7	8	9	10
6	Clickers are a useful initiative to write tests and exams with.	6	0	1	2	3	4	5	6	7	8	9	10
7	I bought a second hand clicker and had problems with it.	7	0	1	2	3	4	5	6	7	8	9	10
8	My clicker's battery failed during a test.	8	0	1	2	3	4	5	6	7	8	9	10
9	I prefer to see my mark on my clicker after a test was written.	9	0	1	2	3	4	5	6	7	8	9	10

10 General comment on Clickers: ______________________________

____

____

C. GENERAL

1 Did you make use of any part of our tutoring system, i.e. Excel classes, revision classes or tutors? **1 | Yes | No |**

2 If Yes, please give us feedback on your learning experience:

____

____

3 Will you make use of an online tutoring system? **3 | Yes | No |**

4 A revision class was offered before the exam.
If you have attended the revision class, was it valueble? **4 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |**

5 General comments: ______________________________

____

____

QUESTIONNAIRE - Student progress 2017

We want to evaluate a few learning tools and would value your feedback regarding certain concepts.

PLEASE ANSWER THE QUESTIONS BY CROSSING (X) THE RELEVANT BOX OR WRITE YOUR ANSWER IN THE GIVEN SPACE

STUDENT NUMBER u _ _ _ _ _ _ _ _ _

Scale: Don't agree (0) – Agree 100% (10)

A. APLIA

1	I read the prescribed sections in the textbook before I attempted the pre-class assignments.	1	0	1	2	3	4	5	6	7	8	9	10
2	I read the notes before I attempted the pre-class assignments.	2	0	1	2	3	4	5	6	7	8	9	10
3	I found the textbook helpful and valuable.	3	0	1	2	3	4	5	6	7	8	9	10
4	The pre-class assignments helped me to go prepared to class.	4	0	1	2	3	4	5	6	7	8	9	10
5	The pre-class assignments helped to explain difficult concepts.	5	0	1	2	3	4	5	6	7	8	9	10
6	I did better in the post-class assignments.	6	0	1	2	3	4	5	6	7	8	9	10
7	The post-class assignments were interesting real life examples.	7	0	1	2	3	4	5	6	7	8	9	10

8 General comment on Aplia: ______________________________

____

B. CLICKERS

1	Clickers made class more fun and exciting.	1	0	1	2	3	4	5	6	7	8	9	10
2	I prefer to use a clicker than raising my hand during class activities.	2	0	1	2	3	4	5	6	7	8	9	10
3	The use of clickers made me feel more inclined to engage with my peers in class.	3	0	1	2	3	4	5	6	7	8	9	10
4	The clicker questions during a tut session inspired me to do my tutorial beforehand.	4	0	1	2	3	4	5	6	7	8	9	10
5	Response to clicker questions and feedback improved my attention.	5	0	1	2	3	4	5	6	7	8	9	10
6	The clicker questions and feedback in class helped me to be more involved and engaged in a large lecture hall.	6	0	1	2	3	4	5	6	7	8	9	10
7	Clickers contributed positively to my learning experience.	7	0	1	2	3	4	5	6	7	8	9	10
8	The use of clicker feedback improved my understanding of course content.	8	0	1	2	3	4	5	6	7	8	9	10
9	The clicker feedback enabled the lecturer to respond and explain difficult concepts that I might not have understood.	9	0	1	2	3	4	5	6	7	8	9	10
10	Answers to clicker questions and feedback helped me to judge my own understanding of the course content.	10	0	1	2	3	4	5	6	7	8	9	10

11 General comment on Clickers: ______________________________

____

C. GENERAL

1 Did you make use of any part of our tutoring system, i.e. Excel classes, tutors or lecturers? 1 Yes No

2 If Yes, please give us feedback on your learning experience:

____

3 Revision classes were offered for studens with low progress marks. If you have attended the revision class, was it valuable? 3 0 1 2 3 4 5 6 7 8 9 10

4 General comments: ______________________________

____

QUESTIONNAIRE - Student progress 2018

We want to evaluate a few learning tools and would value your feedback regarding certain concepts.

PLEASE ANSWER THE QUESTIONS BY CROSSING (X) THE RELEVANT BOX OR WRITE YOUR ANSWER IN THE GIVEN SPACE

STUDENT NUMBER | u | | | | | | | | |

Scale: 0 = Don't agree … 10 = Agree 100%

A. MINDTAP

1	I read the prescribed sections in the textbook before I attempted the pre-class assignments.	1	0	1	2	3	4	5	6	7	8	9	10
2	I read the notes before I attempted the pre-class assignments.	2	0	1	2	3	4	5	6	7	8	9	10
3	I have access to a smart device.	3	Yes	No									
4	If yes in 3, I read the textbook on the smart device.	4	0	1	2	3	4	5	6	7	8	9	10
5	The pre-class assignments helped me to go to class prepared.	5	0	1	2	3	4	5	6	7	8	9	10
6	I understand Excel better because of the structured tutorials.	6	0	1	2	3	4	5	6	7	8	9	10
7	The problem-based post-class assignments reinforced difficult concepts	7	0	1	2	3	4	5	6	7	8	9	10
8	I watched the videos on worked solutions for textbook exercises.	8	Never	Sometimes	Often								

9 General comments on MindTap: ____________________

B. CLICKERS

1	Clickers made class more fun and exciting.	1	0	1	2	3	4	5	6	7	8	9	10
2	I prefer to use a clicker than raising my hand during class activities.	2	0	1	2	3	4	5	6	7	8	9	10
3	The use of clickers made me feel more inclined to engage with my peers in class.	3	0	1	2	3	4	5	6	7	8	9	10
4	The clicker questions during a tut session inspired me to do my tutorial beforehand.	4	0	1	2	3	4	5	6	7	8	9	10
5	Response to clicker questions and feedback improved my attention.	5	0	1	2	3	4	5	6	7	8	9	10
6	The clicker questions and feedback in class helped me to be more involved and engaged in a large lecture hall.	6	0	1	2	3	4	5	6	7	8	9	10
7	Clickers contributed positively to my learning experience.	7	0	1	2	3	4	5	6	7	8	9	10
8	The use of clicker feedback improved my understanding of course content.	8	0	1	2	3	4	5	6	7	8	9	10
9	The clicker feedback enabled the lecturer to respond and explain difficult concepts that I might not have understood.	9	0	1	2	3	4	5	6	7	8	9	10
10	Answers to clicker questions and feedback helped me to judge my own understanding of the course content.	10	0	1	2	3	4	5	6	7	8	9	10

11 General comment on Clickers: ____________________

C. FLIPPED CLASSROOM - A problem-based teaching approach

1	SA based problem scenarios assist in understanding difficult concepts.	1	0	1	2	3	4	5	6	7	8	9	10
2	Being involved in solving problems in class improve my understanding.	2	0	1	2	3	4	5	6	7	8	9	10
3	I prefer working in a group with peers, e.g. the scratch-card test.	3	0	1	2	3	4	5	6	7	8	9	10
4	Switching lecturers for different lectures in the week worked for me.	4	0	1	2	3	4	5	6	7	8	9	10

5 General comments: ____________________

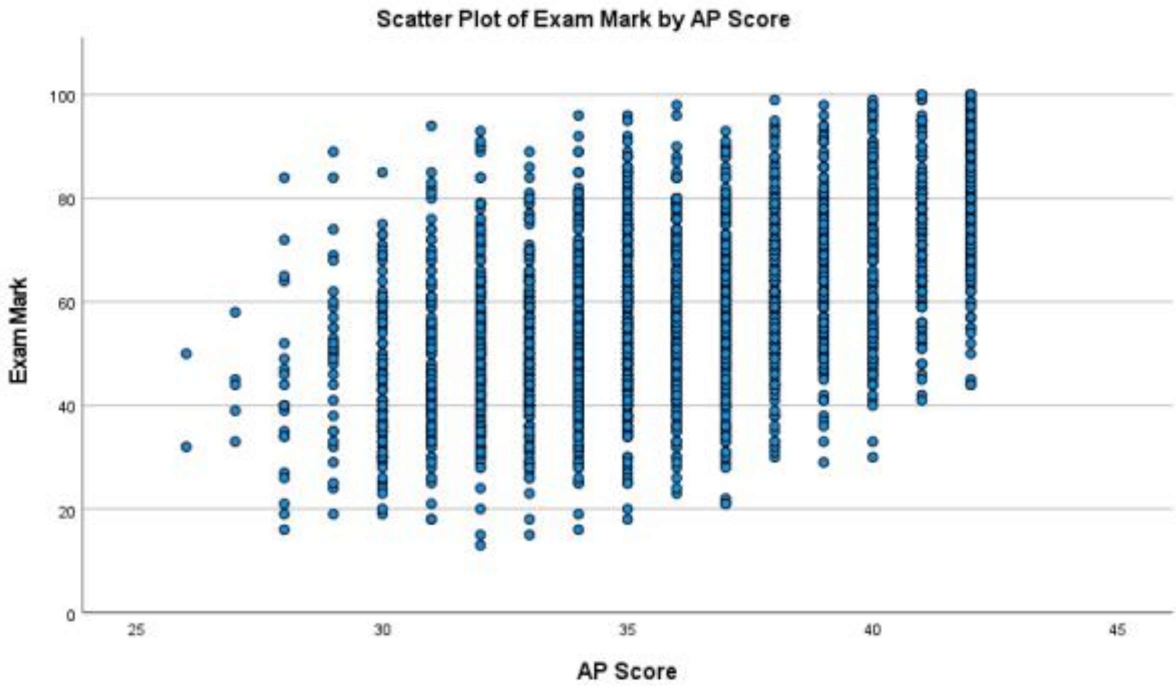
D. TUTORING SYSTEM

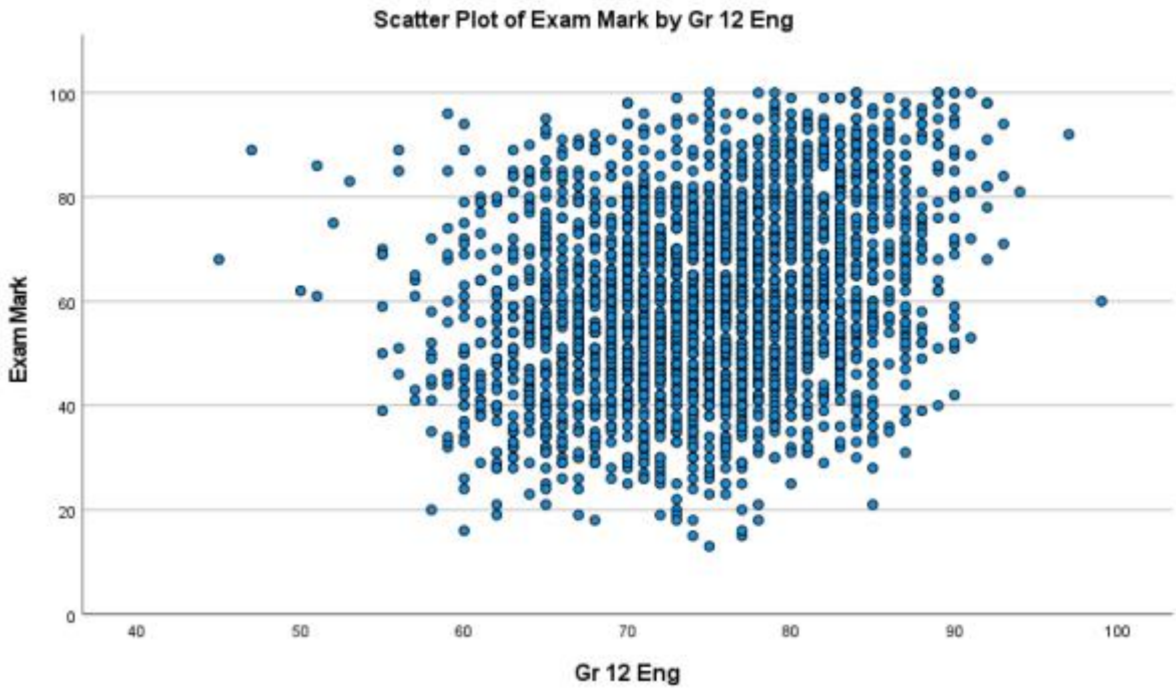
1 I made use of the tutoring system:	a. Tutor consultation	1a	Yes	No
	b. Lecturer consultation	b	Yes	No
	c. Revision classes	c	Yes	No

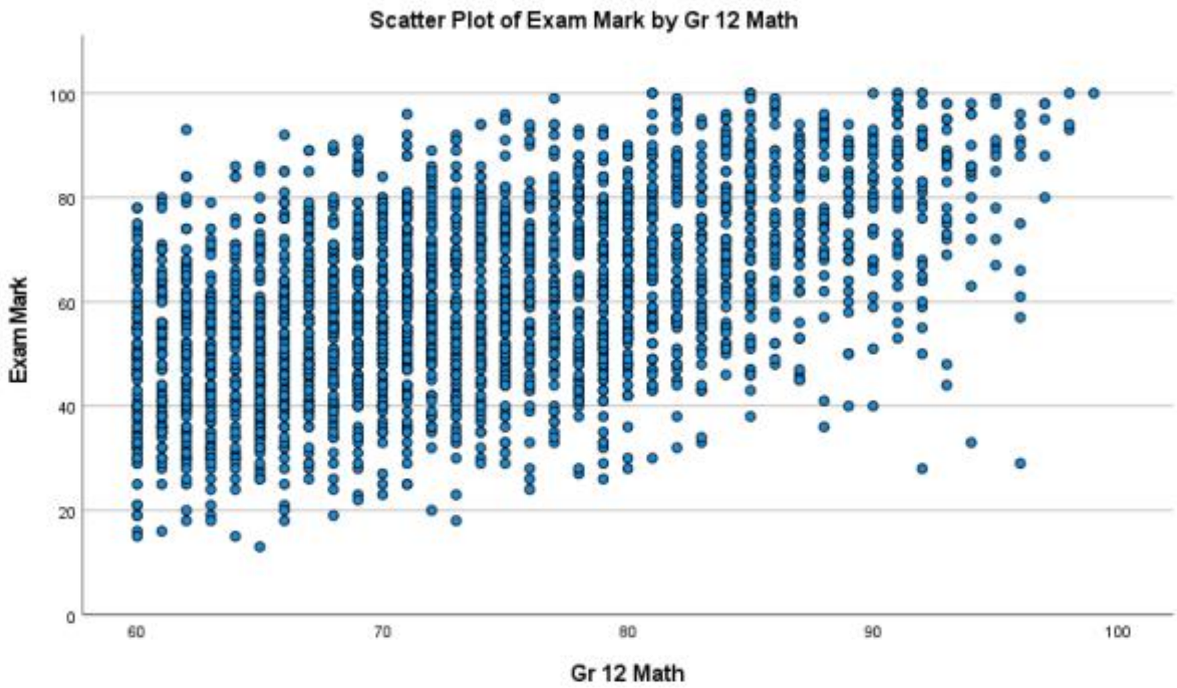
2	The tutoring system is valuable to me.	2	0	1	2	3	4	5	6	7	8	9	10

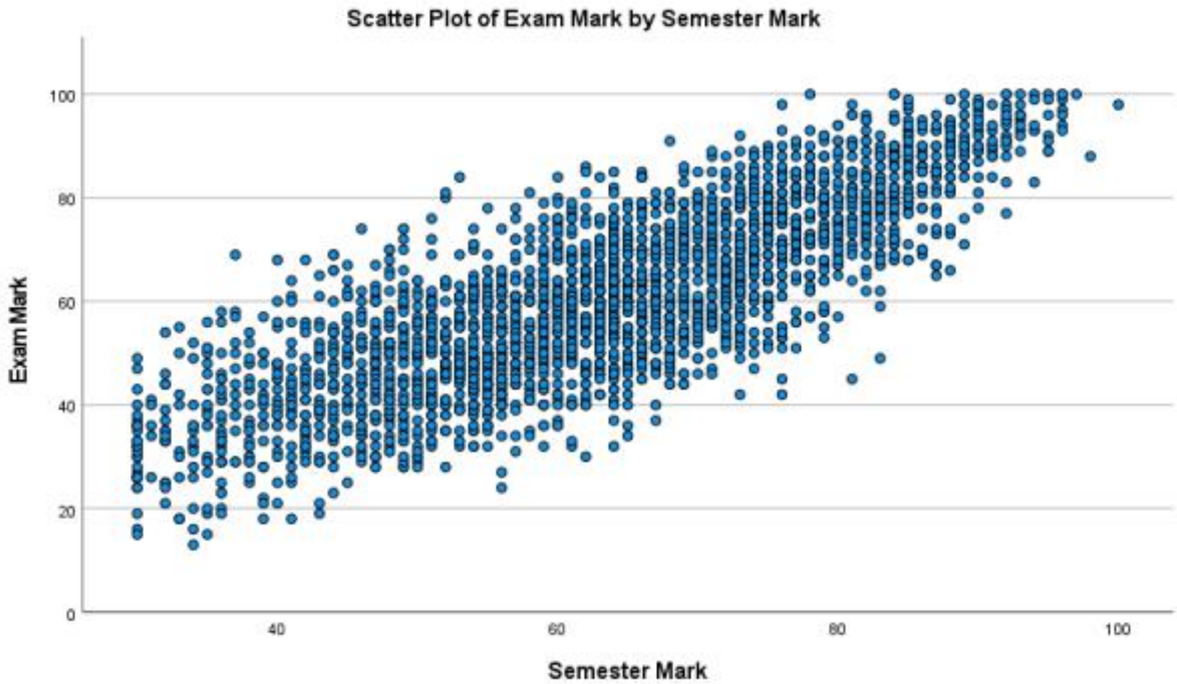
ANNEXURE 7 – Assumptions for GLM I to VII models based on the 2014 and 2017 cohorts in Section 4.3.1

Assumption 1: The relationship between the independent variables and the dependent variable is linear. The scatter plots for the quantitative variables are shown.









Assumption 2: There is no multicollinearity in the data. According to the collinearity statistics, assumption 2 has been met (VIF scores for quantitative variables are below 4 and tolerance scores above 0.2). According to the Pearson's correlations for the quantitative variables and Spearman's correlations for the categorical variables, none of the independent variables has a correlation coefficient above 0.7 (See tables below).

Correlations

Spearman's rho		Mtongue	Clicker	Gender code	Class Section	STK exposure
Mtongue	Correlation Coefficient	1.000	.055**	.072**	.270**	.009
	Sig. (2-tailed)	.	.005	.000	.000	.655
	N	2624	2624	2624	2624	2624
Clicker	Correlation Coefficient	.055**	1.000	-.044*	.172**	.008
	Sig. (2-tailed)	.005	.	.024	.000	.677
	N	2624	2624	2624	2624	2624
Gender code	Correlation Coefficient	.072**	-.044*	1.000	-.030	-.044*
	Sig. (2-tailed)	.000	.024	.	.130	.023
	N	2624	2624	2624	2624	2624
Class Section	Correlation Coefficient	.270**	.172**	-.030	1.000	.076**
	Sig. (2-tailed)	.000	.000	.130	.	.000
	N	2624	2624	2624	2624	2624
STK exposure	Correlation Coefficient	.009	.008	-.044*	.076**	1.000
	Sig. (2-tailed)	.655	.677	.023	.000	.
	N	2624	2624	2624	2624	2624

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations

		Yrs Reg	Sem_cen	Math_cen	AP Score_cen	Eng_cen
Yrs Reg	Pearson Correlation	1	-.099**	-.032	-.111**	-.077**
	Sig. (2-tailed)		.000	.101	.000	.000
	N	2624	2624	2624	2624	2622
Sem_cen	Pearson Correlation	-.099**	1	.537**	.574**	.316**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	2624	2624	2624	2624	2622
Math_cen	Pearson Correlation	-.032	.537**	1	.663**	.316**
	Sig. (2-tailed)	.101	.000		.000	.000
	N	2624	2624	2624	2624	2622
AP Score_cen	Pearson Correlation	-.111**	.574**	.663**	1	.664**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	2624	2624	2624	2624	2622
Eng_cen	Pearson Correlation	-.077**	.316**	.316**	.664**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	2622	2622	2622	2622	2622

**. Correlation is significant at the 0.01 level (2-tailed).

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Sem_cen	.622	1.607
	Math_cen	.498	2.007
	AP Score_cen	.297	3.368
	Eng_cen	.529	1.891
	Yrs Reg	.982	1.019

a. Dependent Variable: Exam Mark

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions					
				(Constant)	Sem_cen	Math_cen	AP Score_cen	Eng_cen	Yrs Reg
1	1	2.561	1.000	.00	.05	.05	.04	.04	.00
	2	1.954	1.145	.02	.00	.00	.00	.00	.02
	3	.770	1.823	.00	.17	.11	.01	.40	.00
	4	.472	2.329	.00	.72	.45	.01	.02	.00
	5	.199	3.589	.00	.05	.39	.93	.54	.00
	6	.044	7.657	.98	.00	.01	.01	.00	.98

a. Dependent Variable: Exam Mark

Assumption 3: The values of the residuals are independent. The Durbin-Watson statistic shows that assumption 3 has been met for GLM I (Durbin-Watson = 2.012) and is similar for the other GLM models.

Model Summary[g]

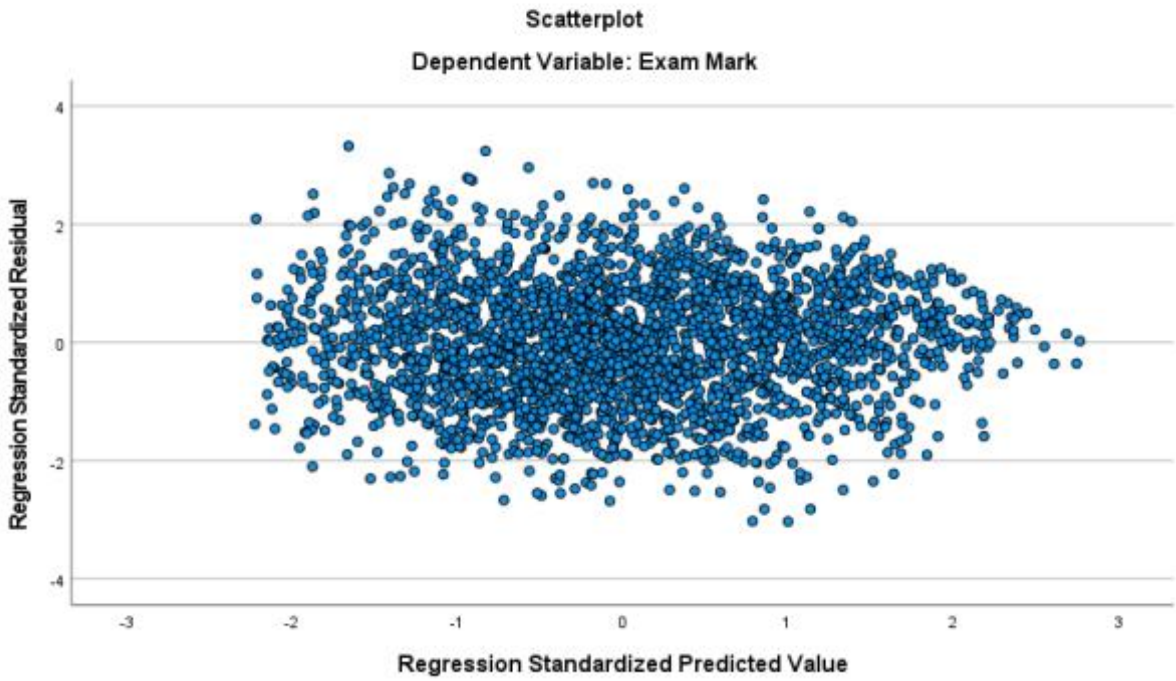
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
GLM I	.824[f]	.679	.679	9.793	2.012

a. Predictors: (Constant), Sem_cen

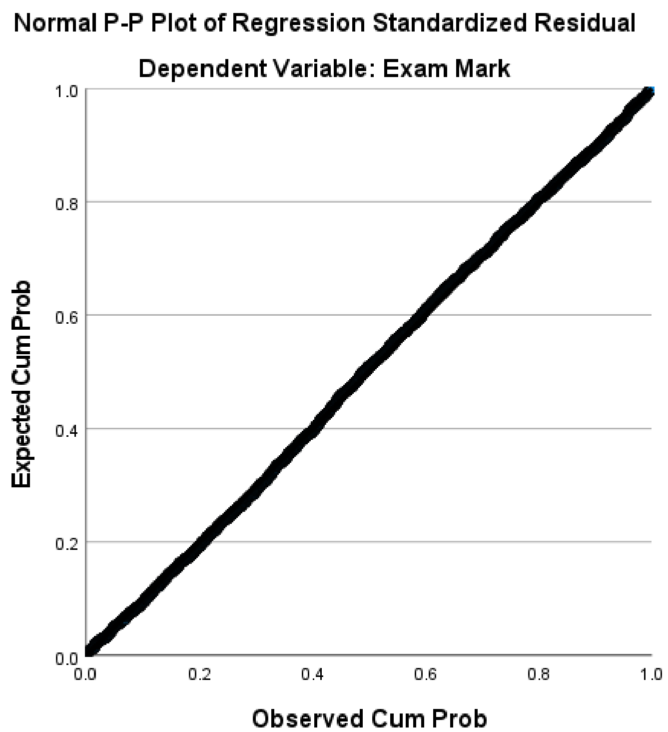
f. Predictors: (Constant), Sem_cen, Clicker, Math_cen, MTongue, AP Score_cen, Yrs Reg

g. Dependent Variable: Exam Mark

Assumption 4: The variance of the residuals is constant. The scatter plot of the standardised residuals against the standardised predicted values shows no discernible irregular patterns. The assumption of homoscedasticity is met.



Assumption 5: The values of the residuals are normally distributed. The P-P plot for the model shows that the assumption is met.



Assumption 6: There are no influential cases biasing the model. Cook's Distance values are all way below one. There are a few values with |standardised residuals| >3. These are students with low semester marks and high examination marks or vice versa. They could not be excluded from the data set.

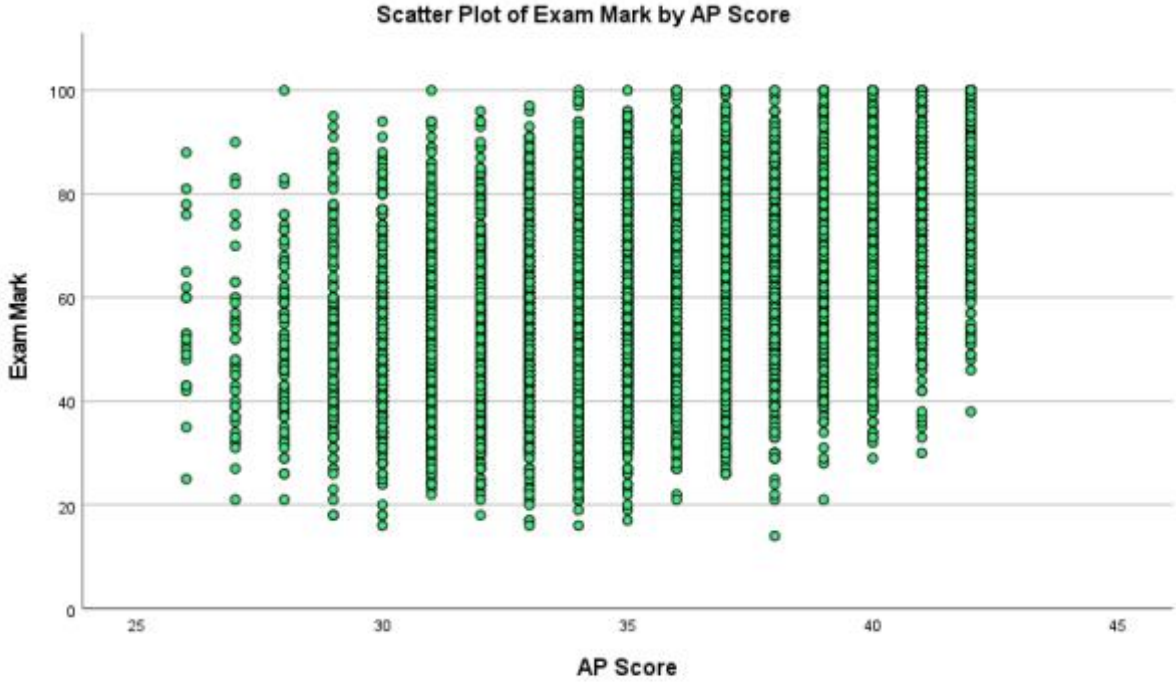
Residuals Statistics[a]

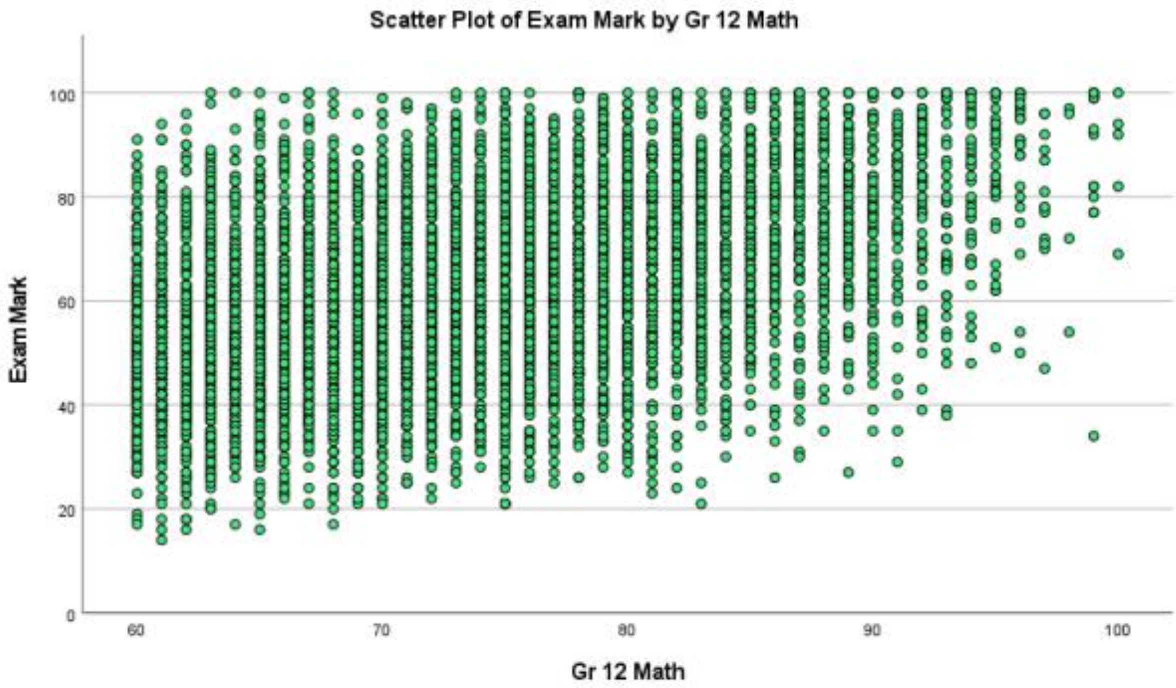
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	25.30	96.03	60.24	14.241	2624
Std. Predicted Value	-2.453	2.513	.000	1.000	2624
Standard Error of Predicted Value	.316	2.899	.486	.143	2624
Adjusted Predicted Value	25.25	96.02	60.25	14.243	2624
Residual	-40.492	33.241	-.012	9.788	2624
Std. Residual	-4.135	3.395	-.001	1.000	2624
Stud. Residual	-4.262	3.399	-.001	1.001	2624
Deleted Residual	-43.027	33.327	-.013	9.821	2624
Stud. Deleted Residual	-4.276	3.406	-.001	1.002	2624
Mahal. Distance	1.732	228.640	6.008	8.250	2624
Cook's Distance	.000	.162	.000	.004	2624
Centered Leverage Value	.001	.087	.002	.003	2624

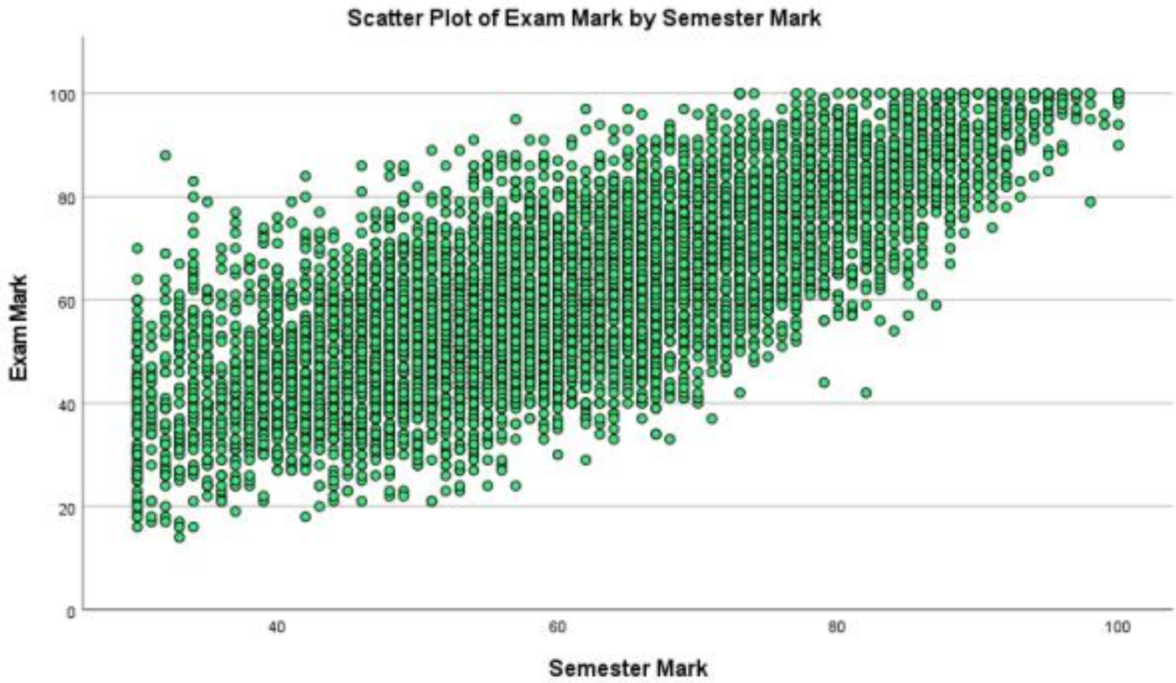
a. Dependent Variable: Exam Mark

ANNEXURE 8 – Assumptions for the GLM VII model based on the five intervention cohorts in Section 4.3.2

Assumption 1: The relationship between the independent variables and the dependent variable is linear. The scatter plots for the quantitative variables are shown.







Assumption 2: There is no multicollinearity in the data. According to the collinearity statistics, assumption 2 has been met (VIF scores for quantitative variables are below 3 and tolerance scores above 0.3). According to the Pearson's correlations for the quantitative variables and Spearman's correlations for the categorical variables, none of the independent variables has a correlation coefficient above 0.7 (See tables below).

Correlations

Spearman's rho		Race code	Class Section	Mtongue	Residence	Cohort order
Race code	Correlation Coefficient	1.000	-.280**	.603**	.189**	.018
	Sig. (2-tailed)	.	.000	.000	.000	.127
	N	7168	7168	7168	7168	7168
Class Section	Correlation Coefficient	-.280**	1.000	-.242**	-.047**	-.179**
	Sig. (2-tailed)	.000	.	.000	.000	.000
	N	7168	7168	7168	7168	7168
Mtongue	Correlation Coefficient	.603**	-.242**	1.000	.116**	.031**
	Sig. (2-tailed)	.000	.000	.	.000	.009
	N	7168	7168	7168	7168	7168
Residence	Correlation Coefficient	.189**	-.047**	.116**	1.000	-.005
	Sig. (2-tailed)	.000	.000	.000	.	.675
	N	7168	7168	7168	7168	7168
Cohort order	Correlation Coefficient	.018	-.179**	.031**	-.005	1.000
	Sig. (2-tailed)	.127	.000	.009	.675	.
	N	7168	7168	7168	7168	7168

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

		Yrs Reg	Sem_cen	Eng_cen	AP Score_cen	Math_cen
Yrs Reg	Pearson Correlation	1	-.083**	-.049**	-.097**	-.031**
	Sig. (2-tailed)		.000	.000	.000	.009
	N	7168	7168	7147	7168	7168
Sem_cen	Pearson Correlation	-.083**	1	.249**	.489**	.486**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	7168	7168	7147	7168	7168
Eng_cen	Pearson Correlation	-.049**	.249**	1	.658**	.285**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	7147	7147	7147	7147	7147
AP Score_cen	Pearson Correlation	-.097**	.489**	.658**	1	.624**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	7168	7168	7147	7168	7168
Math_cen	Pearson Correlation	-.031**	.486**	.285**	.624**	1
	Sig. (2-tailed)	.009	.000	.000	.000	
	N	7168	7168	7147	7168	7168

**. Correlation is significant at the 0.01 level (2-tailed).

Coefficients[a]

Model		Collinearity Statistics Tolerance	VIF
1	Sem_cen	.701	1.426
	Eng_cen	.540	1.852
	AP Score_cen	.337	2.971
	Math_cen	.546	1.831
	Yrs Reg	.986	1.014

a. Dependent Variable: Exam Mark

Assumption 3: The values of the residuals are independent. The Durbin-Watson statistic shows that assumption 3 has been met for GLM VII (Durbin-Watson = 2.006).

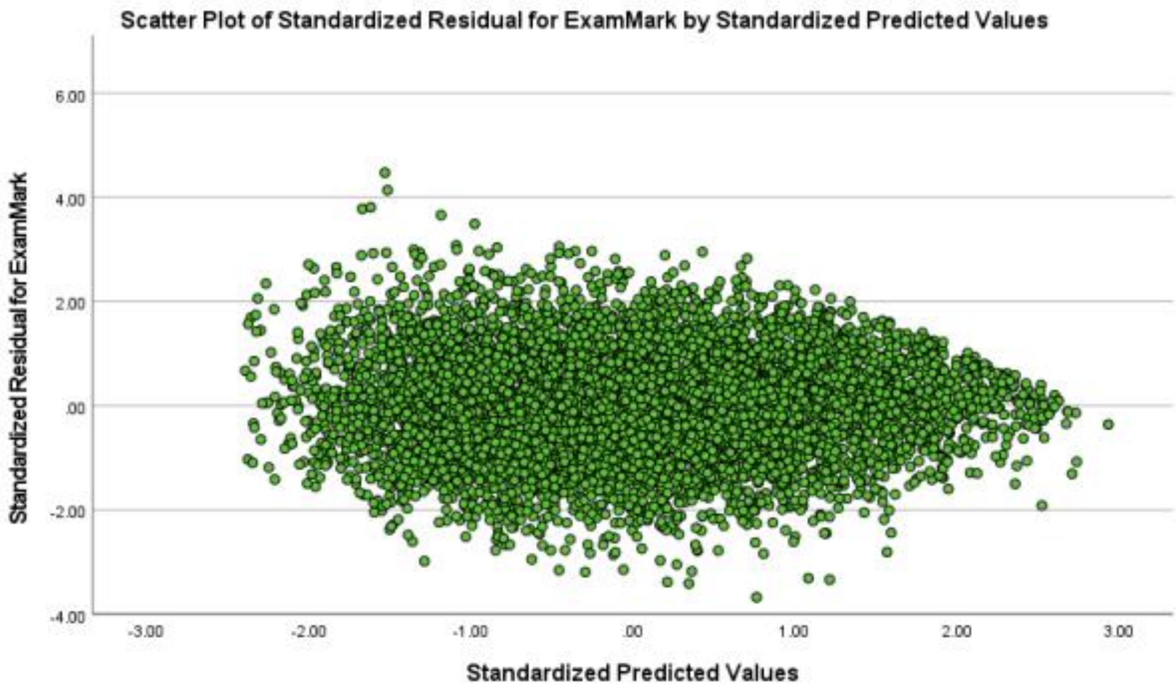
Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.801[a]	.642	.639	10.460	2.006

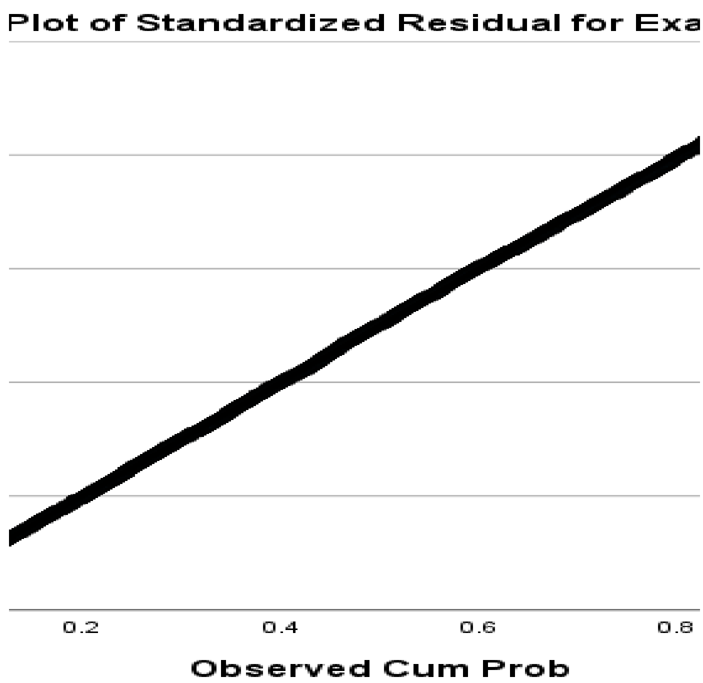
a. Predictors: (Constant), Cohort2018_ Residence, Cohort2015_Math, Cohort2012_Sem, Cohort2013_Sem, Mtongue_YrsReg, Cohort2018_Sem, Sem_AP Score, Cohort2012_ClassSection, YrsReg, Cohort2015_ Residence, Eng_cen, Cohort2013_White, Cohort2013_African, Cohort2018_African, Cohort2018_Math, Cohort2012_White, Cohort2015_African, Cohort2013_Math, Cohort2012_Math, Cohort2015_Sem, Residence, Class Section, Cohort2015_Mtongue, Cohort2018_Mtongue, Cohort2015_White, Cohort2018_White, AP Score_cen, Cohort2013_Mtongue, Cohort2012_African, Cohort2013_Residence, Cohort2012_Mtongue, Cohort2013_ClassSection, Cohort2012_Residence, Cohort2015_ClassSection, Math_cen, White, Sem_cen, Cohort2018_ClassSection, African, ClassSection_YrsReg, Mtongue, Cohort2015, Cohort2013, Cohort2012, Cohort2018

b. Dependent Variable: Exam Mark

Assumption 4: The variance of the residuals is constant. The scatter plot of the standardised residuals against the standardised predicted values shows no obvious irregular patterns. The assumption of homoscedasticity is met.



Assumption 5: The values of the residuals are normally distributed. The P-P plot for the model shows that the assumption is met.



Assumption 6: There are no influential cases biasing the model. Cook's Distance values are all below one. There are a few values with |standardised residuals| >3. These are students with

low semester marks and high examination marks or vice versa. They cannot be excluded from the data set.

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	27.78	103.82	61.83	13.951	7147
Std. Predicted Value	-2.441	3.010	.000	1.000	7147
Standard Error of Predicted Value	.521	3.685	.817	.191	7147
Adjusted Predicted Value	27.48	103.86	61.83	13.951	7147
Residual	-38.409	46.778	.000	10.427	7147
Std. Residual	-3.672	4.472	.000	.997	7147
Stud. Residual	-3.681	4.507	.000	1.000	7147
Deleted Residual	-38.599	47.504	.001	10.496	7147
Stud. Deleted Residual	-3.684	4.513	.000	1.000	7147
Mahal. Distance	16.699	885.889	44.994	28.294	7147
Cook's Distance	.000	.007	.000	.000	7147
Centered Leverage Value	.002	.124	.006	.004	7147

a. Dependent Variable: Exam Mark

ANNEXURE 9 – R-code for Multivariate regression with XGBoost and SHAP values in Section 4.4

R-code for Multivariate Regression

```
---
title: "Data Import"
output: html_document
---

```{r import packages}
#install.packages("dplyr")
#install.packages("ggplot2")
#install.packages("nnet")
#install.packages("caret")
#install.packages("rpart.plot")
#install.packages("rattle")
#install.packages("ROSE")
#install.packages("DMwR")
#install.packages("mice")
#install.packages("EIX")
#install.packages("UBL")
#install.packages("Metrics")
#install.packages("rlang")
#install.packages("purrr")
library(dplyr)
library(ggplot2)
library(caTools)
library(nnet)
library(caret)
library(rpart.plot)
library(rattle)
library(ROSE)
library(DMwR)
library(gdata)
library(mice)
library(xgboost)
library(gmodels)
library(EIX)
```
```

```
library(Metrics)
library(UBL)
library(rpart)
library(SHAPforxgboost)
library(tidyverse)
library(caret)
source("shap.R")
library(xgboostExplainer)
set.seed(111)

```

```
```{r Import the csv file}
final = read.csv("Final.csv", stringsAsFactors = FALSE)
final = subset(final,Prerequisites %in% c("1"))
```

```{r Remove unnecessary columns}
final = final[, !(colnames(final) %in%
c("Term","Student.Number","Gender.Desc","Supplementary.Mark","Academic.
Plan","Home.Language.Desc","Language.of.Preference.Desc","Academic.Plan
.Code","Final.Mark","Ethnic.Group.Desc","Prer","Online.Homework","Flipp
ed.Classroom","Clicker","Peer.Group.Activities","Repeat","Prer","Prereq
uisites","Repeat.comb"))]

#,"","Repeat","Number.of.times.in.STK.110","Repeat.comb","Prerequisites
","Prer","Offering.Language.Desc","Semester.Mark" ))]
```

```{r deal with missing values}
final = unknownToNA(x=final, unknown=c("Unknown"))
pMissing = function(x){sum(is.na(x))/length(x)*100}
apply(final,2,pMissing)


init = mice(final[c("Gr.12.Eng","Offering.Language.Desc")], maxit=0)
meth = init$method
predM = init$predictorMatrix
```
```

````
meth[c("Offering.Language.Desc")]="polyreg"
meth[c("Gr.12.Eng")]="pmm"

imputed = mice(final[c("Gr.12.Eng","Offering.Language.Desc")],
method=meth, predictorMatrix=predM, m=5)
imputed = complete(imputed)
final[c("Gr.12.Eng","Offering.Language.Desc")] = imputed[,1:2]

rm(imputed, init, predM,meth, pMissing)

final = final[, !(colnames(final) %in% c("Offering.Language.Desc"))]

```

```{r convert data to int & do correlation matrix}
final = final[-c(11)]
```

```{r make variables right type}


final$Cohort = factor(final$Cohort,
            levels = c('2011','2012','2013','2015','2018'),
            labels = c(1, 2,3,4,5))

final$Residence = factor(final$Residence,
            levels = c('Out of Res','In Res'  ),
            labels = c(0,1))

final$Gender = factor(final$Gender,
            levels = c('0',  '1'),
            labels = c(0, 1))

final$STK.exposure = factor(final$STK.exposure,
            levels = c('1',  '2','3','4'),
            labels = c(1, 2,3,4))

final$MTongue = factor(final$MTongue,
            levels = c('0',  '1'),
````

```
             labels = c(0, 1))

final$ Race = factor(final$ Race,
             levels = c('1',  '2','3'),
             labels = c(1, 2,3))

final$Class.Section = factor(final$Class.Section,
             levels = c('0',  '1'),
             labels = c(0, 1))

#str(final)
```

```{r Order dataset and split dataset in training and test set}
final = final[order(runif(7168)),]
split = sample.split(final$Exam.Mark, SplitRatio = 0.80)
training_set = subset(final, split == TRUE)
test_set = subset(final, split == FALSE)
rm(split)
#final2 = subset(final,Prerequisites %in% c("1"))

#final2 = final2[order(runif(7168)),]
#split = sample.split(final2$Exam.Mark, SplitRatio = 0.80)
#training_set = subset(final2, split == TRUE)
#test_set = subset(final2, split == FALSE)
#rm(split)

  #training_set_smoteBalan <- ?SmoteRegress(Exam.Mark ~., training_set,
thr.rel = 0.01, dist = "HEOM", C.perc ="balance")
```

```{r Do XGBoost on All Data & All Groups}
# Create numeric labels with one-hot encoding
cv.ctrl <- trainControl(method = "repeatedcv", repeats = 1,number = 3)
train<- as.matrix(training_set, rownames.force=NA)
test<- as.matrix(test_set, rownames.force=NA)
train <- as(train, "sparseMatrix")
test <- as(test, "sparseMatrix")
```

```
train_Data <- xgb.DMatrix(data = train[,1:13], label =
train[,"Exam.Mark"])

xgb.grid <- expand.grid(nrounds = 500,
            max_depth = seq(5,10),
            eta = c(0.01,0.3,1),
            gamma = c(0.0,0.2,1),
            colsample_bytree = c(0.5,0.8,1),
            min_child_weight=seq(1,2,3),
            subsample=1)

xgb_tune <-train(Exam.Mark ~.,
         data=training_set,
         method="xgbTree",#"xgbLinear",
         metric = "RMSE",
         trControl=cv.ctrl,
         tuneGrid=xgb.grid)
#?xgboost
print(xgb_tune)
params <- list(booster = "gbtree", max_depth = 5,eta =0.01,gamma =
0,colsample_bytree = 0.8,min_child_weight=1,subsample=1)
xgb_model = xgboost(params = params, data = train_Data, nrounds = 500)

test_data <- xgb.DMatrix(data = test[,1:13], label =
test[,"Exam.Mark"])

prediction <- predict(xgb_model, newdata = test_data)

RMSE = function(m, o){
 sqrt(mean((m - o)^2))
}
RMSE(prediction,test_set$Exam.Mark)
rss <- sum((prediction - test_set$Exam.Mark) ^ 2) ## residual sum of
squares
tss <- sum((test_set$Exam.Mark - mean(test_set$Exam.Mark)) ^ 2) ##
total sum of squares
1 - rss/tss

```
```
```

~~~~
```{r}
interactions<-interactions(xgb_model, test_data, option =
"interactions")
head(interactions)
plot(interactions)

xgb.plot.tree(feature_names = train@Dimnames[[2]], model =
xgb_model,trees=499)

```

```{r}
shap_values <- shap.values(xgb_model = xgb_model, X_train =
train[,1:14])
shap_values$mean_shap_score
xgb.importance(model = xgb_model)

shap_long = shap.prep(xgb_model = xgb_model, X_train =
as.matrix(train[,1:14]))
shap_long <- shap.prep(shap_contrib = shap_values$shap_score, X_train =
as.matrix(train[,1:14]))
shap.plot.summary(shap_long)

#ggplot(final, aes(x= Term, y=Exam.Mark))+ geom_violin()

## Calculate shap values
shap_result = shap.score.rank(xgb_model = xgb_model,
               X_train =as.matrix(train[,1:13]),
               shap_approx = F
               )

## Plot var importance based on SHAP
var_importance(shap_result, top_n=12)+labs(y= "Importance for
Prerequisite Students")



## Prepare data for top N variables
shap_long = shap.prep(shap = shap_result,
~~~~

```
              X_train = as.matrix(train[,1:13]),
              top_n = 12
              )
## Plot shap overall metrics
plot.shap.summary(data_long = shap_long)+labs(y= "Prerequisite Students
- SHAP value (impact on model output)")


xgb.plot.shap(data = as.matrix(train[,1:13]), # input data
       model = xgb_model, # xgboost model
       features = names(shap_result$mean_shap_score[1:12]), # only top
10 var
       n_col = 3, # layout option
       plot_loess = T # add red line to plot
)


install.packages("remotes")
remotes::install_github("davidADSP/xgboostExplainer")

explainer = buildExplainer(xgb_model,train_Data, type="regression",
base_score = 0.5, trees_idx = NULL)
pred.breakdown = explainPredictions(xgb_model, explainer, test_data)

cat('Breakdown Complete','\n')
weights = rowSums(pred.breakdown)
pred.xgb = 1/(1+exp(-weights))
cat(max(prediction-pred.xgb),'\n')

idx_to_get = as.integer(145)
test_set[idx_to_get,-14]
showWaterfall(xgb_model, explainer, test_data, data.matrix(test_set[,-
14]) ,idx_to_get, type = "regression")

write.csv(test_set,'test_set.csv')
```
```

R-code for SHAP values

```
library(gridExtra)
```
```

```
library(grid)
library(ggplot2)
library(cowplot)

indivusers = read.csv("indivusers.csv", stringsAsFactors = FALSE)
indivusers = indivusers[, !(colnames(indivusers) %in%
c("Term","Student.Number","Gender.Desc","Supplementary.Mark","Academic.
Plan","Home.Language.Desc","Language.of.Preference.Desc","Academic.Plan
.Code","Final.Mark","Ethnic.Group.Desc","Prer","Online.Homework","Flipp
ed.Classroom","Clicker","Peer.Group.Activities","Repeat","Prer","Prereq
uisites","Repeat.comb"))]
indivusers = indivusers[, !(colnames(indivusers) %in%
c("Offering.Language.Desc"))]
test_set2 = indivusers[-c(11)]
test2<- as.matrix(test_set2, rownames.force=NA)
test2 <- as(test2, "sparseMatrix")
test_data2 <- xgb.DMatrix(data = test2[,1:13], label =
test2[,"Exam.Mark"])


idx_to_get145 = as.integer(1) #11028832 - Student 1
idx_to_getw145 = showWaterfall(xgb_model, explainer, test_data2,
data.matrix(test_set2[,-14]) ,idx_to_get145, type = "regression")


idx_to_get1310 = as.integer(2) #12032302 - Student 12
idx_to_getw1310 = showWaterfall(xgb_model, explainer, test_data2,
data.matrix(test_set2[,-14]) ,idx_to_get1310, type = "regression")



idx_to_get774 = as.integer(3) #15227805 - Student 23
idx_to_getw774 = showWaterfall(xgb_model, explainer, test_data2,
data.matrix(test_set2[,-14]) ,idx_to_get774, type = "regression")



idx_to_get768= as.integer(4) #18065334- Student 27
```

```
idx_to_getw768 = showWaterfall(xgb_model, explainer, test_data2,
data.matrix(test_set2[,-14]) ,idx_to_get768, type = "regression")
```

```
grid.arrange(idx_to_getw145,idx_to_getw1310,ncol=2)
grid.arrange(idx_to_getw774,idx_to_getw768,ncol=2)
```