AOSIS

The effects of examination-driven teaching on mathematics achievement in Grade 10 school-based high-stakes examinations

CrossMark
click for updates

Authors:
Onyumbe Okitowamba[1]
Cyril Julie[1]
Monde Mbekwa[1]

Affiliations:
[1]School of Science and Mathematics Education, University of the Western Cape, South Africa

Corresponding author:
Cyril Julie,
cjulie@uwc.ac.za

Read online:
Scan this QR code with your smart phone or mobile device to read online.

Various efforts are underway to improve achievement in high-stakes examinations in school mathematics. This article reports on one such initiative which focuses on the development of quality teaching of school mathematics by embedding it within an examination-driven emphasis. A quantitative approach was used to analyse the performance of Grade 10 learners in three consecutive end-of-year school-based examinations set by the initiative. Results indicate a trend in a positive direction over the three-year period. Nevertheless, there was a discernible decrease between the first and second administration of the examinations. It is concluded that examination-driven teaching holds a promise for enhancing achievement in high-stakes school mathematics examinations if sensibly and sensitively implemented.

Introduction

Underachievement in school mathematics is a concern in most countries in the world. Watson and De Geest (2012) sketch the situation of underachievement in mathematics by drawing attention to 'identifiable groups of students, such as those with different language backgrounds and those from lower socioeconomic rankings [who] underachieve in national and international tests' (p. 213). Regarding the situation in South Africa, Reddy and Janse van Rensburg (2011) highlight two relevant characteristics in the South African education system. The first is that the national average mathematics achievement score for different grade levels across the schooling system is similar and stable, around 30% to 40% at different grades. The second is that there is a high differentiation of the educational performance of students from various socio-economic backgrounds.

By stating that 'methods of approaching this issue range from macro-changes in policy, curriculum and assessment to institutional change, provision of extra teaching and micro-advice about inclusive teaching in classrooms', Watson and De Geest (2012, p. 213) draw attention to the efforts embarked on to address the issue in the United Kingdom. The reference to 'provision of extra teaching' is akin to interventions, such as additional classes offered by universities and NGOs to selected learners in Grades 10 to 12 and extra tuition and vacation tutoring schools offered by the provincial education departments to improve achievement outcomes in the National Senior Certificate (NSC) Mathematics examinations (Reddy, Berkowitz, & Mji, 2005).

The seriousness with which countries take the improvement of achievement in mathematics of learners from low socio-economic and historically disadvantaged sectors of a country's demographic makeup is evident in current reform initiatives in school mathematics. For example, in Australia a desktop review was conducted to identify 'gaps in current pedagogical approaches and learning resources for the teaching of mathematics to inform the Mathematics by Inquiry initiative' (Australian Academy of Science, 2015).

One issue that had to be addressed in this initiative in Australia was linked to the teaching of socio-economic and historically disadvantaged groups. The commission given to the Australian Academy of Science by the Australian Government's Department of Education and Training was specifically stated as 'which pedagogical approaches have been shown to work with specific groups under-represented in advanced mathematics at senior secondary level (girls, Indigenous, disadvantaged students)?' (Australian Academy of Science, 2015, p. 17). It is also now commonplace in research reports that there is an explicit disaggregation of results along the lines of gender, socio-economic status and, where relevant, language diversity of participating cohorts in the research. The Australian situation is different from the South African one, since in South Africa much effort is invested in the improvement of achievement in mathematics of low socio-economic and historically disadvantaged groups.

The popularisation of a programme of teaching adopted to enhance achievement in marginalised groups in a high-stakes mathematics examinations is vividly portrayed in the 1988 film *Stand and Deliver*, depicting Jaime Escalante's work with disadvantaged Latino-American students in east Los Angeles. In the movie the producers obviously used their creative licence to render a fictionalised account of the real situation. However, it is widely (see, for example, Jesness, n.d.) reported that at the start of Escalante's programme only two of the five students who wrote the Advanced Placement Calculus examination passed. The pass rate steadily increased and in 1982 eighteen students passed. The film focuses primarily on the 1982 cohort of students. Escalante's methods of teaching and ways of working with students are described in Escalante and Dirmann (1990).

In Southern Africa there is a paucity of research related to efforts to enhance the achievement in mathematics in high-stakes examinations of students from low socio-economic environments. This does not imply that such projects and efforts do not exist. Many projects report on the impact of their initiatives to improve achievement in high-stakes school mathematics (see Reddy et al., 2005). What is not visible in the reports of these projects and efforts are issues such as the underlying pedagogical and theoretical underpinnings of these projects. In addition to project reports there are some research-based projects on learner achievement in high-stakes end-of-year mathematics and the professional development of teachers. Mogari, Kriek, Stols and Iheanachor (2009), for instance, report on such a study. Although the teachers in the study reported the professional development activities with which they were involved, there is no clear indication of the theoretical underpinnings.

This article reports on a classroom-based project to improve achievement in high-stakes examinations. The mentioned underpinnings of the project and results of learner achievement over three years are presented.

Brief description of the underlying project and the research question

The project, the Local Evidence-Driven Improvement of Mathematics Teaching and Learning Initiative, has as part of its aims the increase in the number of learners taking Mathematics as an examination subject for the NSC examination, an increase in the pass rates and an improvement in the quality of the passes in the participating schools. The project developed an intentional teaching model (Julie, 2013) for guiding instructional practices in classrooms. Mathematics teachers from secondary schools in low socio-economic areas – Bellville South, Bishop Lavis, Bonteheuwel, Elsies River, Gugulethu, Heideveld, Kleinvlei, Langa, Manenberg, Mfuleni and Strand – in the Cape Peninsula participate in the project. The project focuses on the development of high-quality mathematics teaching to improve achievement in mathematics. A belief underlying the project is that improvement of teaching can lead to an enhancement of achievement in high-stakes examinations.

Generally the project operates by offering workshops and institutes attended by participating teachers. Workshops are conducted after school and are usually of approximately two hours duration. Two to three workshops are held per term for the first three terms of the school year.

Institutes are extended and residential gatherings held normally from a Friday afternoon to Sunday lunchtime. Two institutes per year were held for the three years, 2012 to 2014. Overall the teachers were engaged in 64 hours of Continuing Professional Development activities for the three years for which results in the high-stakes school-based mathematics examination were tracked.

The content of the professional development activities focused on pedagogical issues such as analysis of lesson excerpts, discussions around dilemmas teachers face in their teaching, searching for ways to address these dilemmas and the design of lessons. Another feature of the content of the Continuing Professional Development is that in most of the meetings teachers worked on mathematical problems with the aim of developing their mathematicalness – flexible ways of dealing with mathematics. The mathematics of the tasks is explored and discussed. The ways teachers worked with the tasks and the facilitation are then discussed in relation to how teachers can engage learners in doing mathematics.

An example of a dilemma teachers face that was raised by teachers is that of learners not really doing homework. The purposes of homework were then discussed. One of the purposes offered was consolidation of completed work. This was connected in the discussions to the issue of forgetting. The outcome of the deliberations around the issue led to the development of a strategy for which the term 'spiral revision' was coined. Basically this consists of learners being presented with two to three exercises of previously covered work which they have to complete in class. This has to be done 3–4 periods per week in about 7–10 minutes before dealing with the lesson for the day. 'Spiral revision' is the project's version of 'distributed practice' (see, for example, Johnson & Smith, 1987; Seabrook, Brown & Solity, 2005; Smith & Rothkopf, 1984) which meta-analysis of meta-analytic studies found as one of the aspects that contributed towards enhancing achievement (Hattie, 2009). The other purposes of homework were not addressed and teachers generally used their own ways of dealing with these purposes.

Other pedagogical aspects engaged with during the workshops and institutes were clarity to both teachers and learners of the intentions or goals of a lesson, the use of feedback and provision of opportunities to work with different problem types. These are also aspects which Hattie's (2009) meta-analytic work showed had moderate to high effect sizes related to achievement.

The objective of the project, as stated above, is the improvement of achievement in high-stakes examinations. High-stakes examinations, which are discussed below, thus played a structuring role within which the above pedagogical

aspects were dealt with as shown by using the example of quadratic inequalities. This brought the issue of examination-driven teaching into the picture. The research question being reported on in this article is:

Does an examination-driven teaching strategy improve achievement in high-stakes school-based end-of-year summative mathematics examinations in Grade 10?

High-stakes examinations

As is evident from the research question, the notion of a high-stakes examination is one of the constructs of importance in this article. Various notions of high-stakes examinations exist. These are normally linked to the purposes of the examinations.

Howie (2012, p. 82) classifies three kinds of assessments around high-stakes examinations – 'Classroom assessment, System assessments [*and*] Public examinations' – with their purposes, frequencies, test cohorts and subject area coverage. She does not include school-based end-of-year summative assessments in her classification of assessments. In a recent survey on assessment in mathematics Suurtamm et al. (2016), include the last-mentioned assessments and view it as 'increasingly play[*ing*] a prominent role in the lives of students and teachers as graduation or grade promotion often depend on students' test results' (p. 4).

In this article, a high-stakes examination is one that has direct consequences, positive or negative, for the examinees. Particularly for Grade 10 learners, the school-based end-of-year mathematics examination has consequences such as promotion to Grade 11 or not and the right to continue taking Mathematics as an examination subject for the NSC examination. Non-continuation with Mathematics up to Grade 12 is a major issue. For the research reported here, of the 403 learners in the five participating schools who wrote the 2012 project-designed examination in Grade 10 only 280 proceeded to write the 2014 NSC Mathematics. This is an instance of the decrease in taking Mathematics from Grade 10 to 12 in a cohort of learners. Adler and Pillay (2017) indicate that in one of their project schools only 22% of learners who took Mathematics in Grade 10 proceeded with Mathematics as an examination subject 3 years later. Various reasons for learners' non-continuation with Mathematics up to Grade 12 are offered. Some of these are: failing Mathematics in Grade 10 but being promoted to Grade 11 due to fulfilling the promotion rules and dropping Mathematics for Mathematical Literacy and failing Grade 11 (Adler & Pillay, 2017). The importance of taking Mathematics, and being successful, as an NSC examination subject–for access to tertiary studies, job opportunities, etc.–attests to the high-stakes nature of the school-based end-of-year Mathematics examination in Grade 10.

For school-based end-of-year summative assessments the focus is on learners. As indicated above, success or not in these examinations has consequences for them. At a very basic level success (and the level of the success) or failure on these assessments decides whether or not learners will be able to proceed from Grade 10 onwards to be awarded a certificate of worth that they can use after their completion of schooling. The NSC in South Africa is such a certificate. The consequence of having at least this certificate is that it greatly enhances the chances of school-leavers to obtain employment and access to further studies. In this regard Statistics South Africa (2016, p. xiv) reports that 'those without matric constituted more than 58% of the unemployed among the black Africans and coloured population'. Matric refers to the exit level grade (or Grade 12) of the South African non-compulsory schooling system. Success in Grade 10 is the entry point for learners in their journey of pursuit to obtain this valued certificate.

In this article, high-stakes examinations are viewed as those that allow learners to progress from one grade to another and in particular to the exit level, Grade 12. As mentioned, the high-stakes examination is the school-based summative examination learners write at the end of the school year, Grade 10 for the purposes of this study. This examination is normally internally constructed and marked. To ensure quality, the head of the mathematics department of the school normally moderates both the construction and the marking of the examination. Further quality assurance and consistency across schools are ensured through a process of external moderation by the mathematics curriculum advisors of the Department of Education (Jacobs, Mhakure, Fray, Holtman, & Julie, 2014). Regarding the administration of the end-of-year school-based examination, similar security procedures such as the preparation, printing and release of the examination on the day it is written, as is the case for the NSC examinations, are followed. The staff of the school does invigilation during the writing, normally with the subject teachers not invigilating their own examination. School-based summative end-of-year examinations thus follow the processes and procedures that approximate those of the high-stakes NSC examinations. Thus, in terms of purpose and the entire set of processes and procedures, the end-of-year Grade 10 examination is a high-stakes examination.

Examination-driven teaching as underpinning of the project

According to Julie (2013), examination-driven teaching is normally viewed as 'teaching the content of previous examinations and anticipated questions that might crop up in an upcoming examination of the subject' (p. 1). Examination-driven teaching is a controversial issue. Debates about it abound. Opponents of examination-driven teaching argue that it leads to the fragmentation of knowledge, the restriction to low-level content, the fostering of the loss of disciplinary coherence, mitigation against flexible knowing, curriculum contraction, deskilling of teachers and the inhibition of making sound instructional decisions due to the predominant psychometric paradigm underlying high-stakes examinations (Davis & Martin, 2006; Shepard & Dougherty, 1991; Van den Heuvel-Panhuizen & Becker, 2003). Davis and Martin (2006) also draw attention to examination-driven teaching being

more often adopted as a preferred instructional approach to address low performance of learners from low socio-economic environments.

Proponents of examination-driven teaching, on the other hand, draw attention to its advantages for improving achievement outcomes. These include clarity of instructional goals, cost-effectiveness, motivation and examination assistance for learners by providing clarity on the kinds of problems they can expect to encounter in a high-stakes examinations and the feedback that examinations-driven teaching provides to teachers for instructional decision-making (Popham, 1987; Shepard & Dougherty, 1991).

Notwithstanding the debates about examination-driven teaching, there are considered positions about the structuring roles examinations exert on instructional practices. One such position is that examinations play a major role in the constitution of legitimate and valued school mathematics knowledge. Bishop, Hart, Lerman and Nunes (1993, p. 11) contend that 'examinations operationalise the significant components of the intended mathematics curriculum, so they tend to determine the implemented curriculum.' According to Julie (2013) 'the intended and interpreted curricula provide only boundaries of content to be dealt with but the implemented curriculum is heavily driven by the examined curriculum' (p. 6) and the examined curriculum eventually drives what is taught regardless of what the aims of the curriculum are.

The recognition of the structuring effects of the examined curriculum provides a strong argument that in order for teaching to comply with meaningful learning, examinations must be changed (Burkhardt & Pollak, 2006; Van den Heuvel-Panhuizen & Becker, 2003). Julie (2013) presents an argument that examination-driven teaching can contribute towards meaningful learning. He states that examination-like questions and mathematics problems that learners are exposed to during classroom teaching can be changed to questions that elicit process skills and develop critical and conceptual thinking skills. Swan and Burkhardt (2012, p. 5), concurring with Julie, state that if items that require that learners demonstrate their ability of critical and conceptual understanding are included in high-stakes examinations then 'teachers who teach to the test [*can*] deliver a rich and balanced curriculum'. A careful study of the NSC Mathematics examination indicates that such questions are part of the examination. Admittedly, the number and variety of such items need to be increased in assessments in all grades in the Further Education and Training band. To realise this increase, attention must be given to the percentages of items prescribed to be at certain levels of cognitive demand in the mathematics curriculum documents.

In South Africa there is currently an emergence of the use of large-scale systemic assessments to structure continuing professional development initiatives for mathematics teachers. Shalem, Sapire, and Huntley (2013) worked with teachers to do curriculum mapping of large-scale assessments which led to teachers reflecting on their instructional practices with respect to the content taught and the level of cognitive demand

that is focused on. The Data Informed Practice Improvement Project of Brodie (2013) focused on teachers engaging with misconceptions and errors resulting from learners' responses to examination items. This was followed by teachers designing lessons based on the analysis of the examination items. The lessons were implemented in their classrooms and teachers reflected on the efficacy of their implementation to address the identified errors and misconceptions.

In the project of interest in this article, examinations are used in a similar way to those described in the foregoing paragraph. Learners' responses in examinations are used to reflect on difficulties learners display in examinations, design of activities to address such difficulties and backward mapping from the high-stakes NSC Mathematics examination to provide focus for teaching in lower grades. For example, in the Curriculum and Assessment Policy Statement document (Department of Basic Education, 2011, p. 13) the content related to quadratic inequalities is given as solve 'quadratic inequalities in one variable and interpret the solution graphically'. This has to be dealt with in Grade 11 and manifests itself as 'Solve for x: $(x + 1)(4 - x) > 0$' as a level 1 (lowest level) question in the NSC Mathematics examination. In Grade 10 it is prescribed that quadratic graphs of the form $y = af(x) + q$, where $f(x) = x^2$, should be dealt with but the solution of quadratic inequalities is not. This is understandable given the restriction, as stated in the aforementioned sentence, for graphs of quadratic functions in Grade 10. In the project, learners are exposed to solving quadratic inequalities under the topics dealing with the real number system. The graphs of quadratic functions, without specifying the defining expression, are given and learners have to solve quadratic inequalities with a generically specified defining expression as given Figure 1.

Tasks of the nature given in Figure 1 are done during the first week of the first term when learners have not yet dealt with the quadratic functions and their graphs. This is followed up when quadratic functions of the form indicated for Grade 10 are taught and items related to quadratic inequalities are included in the year-end examination as indicated in Figure 2.

It is the contention of the project that if learners start engaging with questions that they will encounter in the ultimate NSC Mathematics examination as early as Grade 10 then they will have high levels of fluency to deal with the cognate problems in the NSC Mathematics examination.

In this section we presented an indication of how examination-driven teaching is conceived and implemented as underpinning in the project. The next section discusses the research design.

Research design

A quantitative design was adopted in this study because learners' scores are used to describe the phenomenon being investigated. The study is a trend study where results of the same phenomenon are tracked over a period. It is different
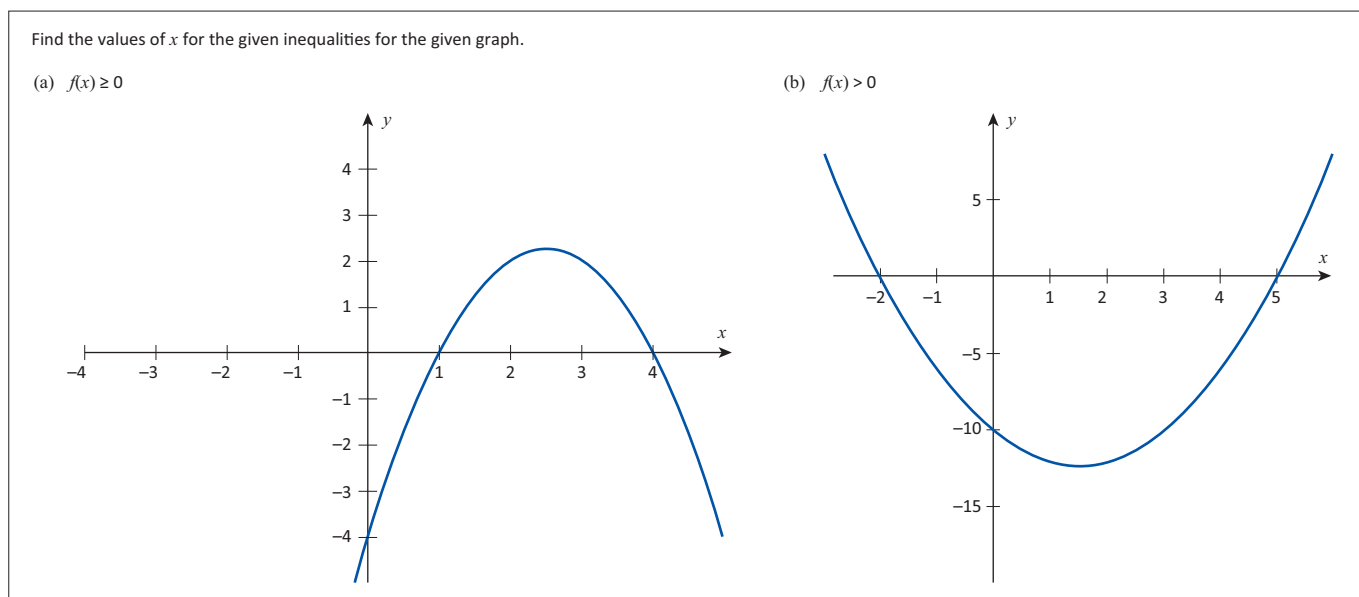
Find the values of x for the given inequalities for the given graph.

(a) $f(x) \geq 0$

(b) $f(x) > 0$

FIGURE 1: Task on quadratic inequalities when dealing with the real number system.

Question 6

6.1 Consider the function $f(x) = x^2 - 4$

6.1.1	Calculate the x-intercepts of f	(2)
6.1.2	Write down the coordinates of the turning points of f.	(2)
6.1.3	Draw a sketch graph of $f(x) = x^2 - 4$. Show all the intercepts with the axes.	(3)
6.1.4	For which values of x will $f(x) \geq 0$.	(2)
6.1.5	For which values of x will $f(x)$ be increasing as x increases?	(2)
6.1.6	Write down the range of f.	(2)

Source: Okitowamba, O. (2015). *Tracking learners' performances in high-stakes Grade 10 mathematics examinations.* Unpublished doctoral dissertation, University of the Western Cape, Bellville, South Africa (p. 154). Retrieved from http://hdl.handle.net/11394/5655

FIGURE 2: Examination item on the quadratic function (the quadratic inequality item is 6.1.4).

from a tracer study which follows the results of the same cohort over a period. The trend of the overall mathematics scores in the end-of-year summative school-based Grade 10 examination over three years – 2012 to 2014 – was thus investigated.

Trend studies are appropriate in situations of curriculum stability. The Trends in International Mathematics and Science Study (TIMSS) project does trend studies (Martin, Mullis, & Chrostowski, 2004). An important requirement of trend studies is that the items used in the assessment instrument should be similar in kind and degree. End-of-year summative school-based Grade 10 examinations are such and thus appropriate as instruments for use in the research reported here. Two methods, Rasch modelling and other statistical methods, were used. These methods are described in the section on analysis procedures.

Sample and sampling procedure

The sample was an opportunistic sample of five schools whose teachers were involved in the Continuing Professional Development initiative. Ten schools were initially involved in the project. After the first year of implementation, the

participation of one of the schools was terminated due to unsatisfactory participation in project activities. Not all the schools wrote the project-set common examinations for the reporting period. The reasons for this are: (1) the timing and availability of the common question papers in that some schools had their examination timetables ready before common agreed examination dates could be negotiated, (2) the standard of question papers was deemed too high in terms of their cognitive demand according the judgement of the teachers of their learners' cognitive levels and (3) the Grade 10 learners of one school were not available in 2013 and 2014 because they had gone to another school following a prior arrangement. It needs to be borne in mind that teachers' participation is voluntary and so the decision to write the common examination or not rests with the schools. Voluntary participation and the right to withdraw from research activities or part of it are important ethical principles in a research project involving human participants. This was made clear to teachers at the start of the Continuing Professional Development initiative. This resulted in five schools who wrote the project common examination for the three years.

A possible threat emanating from working with samples over different years is that the characteristics of the cohorts of participants might change and have a confounding effect. Major confounders are normally race, gender, age, class size and school type. Regarding gender, although the names of the learners appeared on the scripts, the difficulty of using names as a signifier for gender is highly problematic. 'Cyril', for example, can either be male or female. Furthermore, in a school-based examination learners do not indicate their gender. Gender dimensions were thus not included. Other confounding factors that might be linked to the contexts of the schools might have changed. However, the nature of schools in South Africa is such that the enrolments are reasonably stable with regard to socio-economic status and

demographic composition. Our own observations during classroom support visits revealed no observable change along these lines.

In line with common practice for school-based end-of-year summative Mathematics examinations, the examinations are governed by the assessment guidelines as described in the Curriculum and Assessment Policy Statement. The Curriculum and Assessment Policy Statement document describes modalities such as the topics and their weightings to be covered and percentage of marks to be allocated to the different levels of cognitive demand. The school-based end-of-year summative Mathematics examination comprises two papers of 2 hours duration each. The first paper deals with the topics (their weightings given between brackets): algebra and equations (and inequalities) (30 ± 3), patterns and sequences (15 ± 3), finance and growth (10 ± 3), functions and graphs (30 ± 3) and probability (15 ± 3). The topics dealt with in the second paper are: statistics (15 ± 3), analytical geometry (15 ± 3), trigonometry (40 ± 3) and Euclidean geometry and measurement (30 ± 3). The examinations adhered to these guidelines and were thus similar in kind and degree.

To protect anonymity the schools are named A, B, C, D and E. Table 1 gives the number of learners in the different schools for the period 2012 to 2014.

It can be observed from Table 1 that a small number of learners who wrote the first paper did not write the second one.

In this study, the scores for learners who missed a paper were treated as missing data. Thus the total number of learners was taken as 403 for 2012, 381 for 2013 and 406 for 2014.

The presented results are thus not representative of all the schools participating in the Local Evidence-Driven Improvement of Mathematics Teaching and Learning Initiative project or of Grade 10s in the Western Cape province or in South Africa. Therefore, to generalise about the outcomes for the entire province or for the country requires careful consideration if the results are to be more broadly applied. It also needs to be borne in mind that there are many interventions, for which information was not gathered for the participating schools, addressing the low performance of learners in school mathematics. Thus, the results may be confounded by influences of such interventions. However, it is well known that many interventions at Grade 10 level

TABLE 1: Number of learners per school, per year and per paper.

School	2012		2013		2014	
	Paper 1	Paper 2	Paper 1	Paper 2	Paper 1	Paper 2
A	64	64	124	124	91	91
B	94	93	63	63	90	90
C	68	68	36	35	22	22
D	78	78	42	39	109	108
E	101	100	116	113	95	95
Total	**405**	**403**	**381**	**374**	**407**	**406**

focus on selected learners with potential and short-term teacher initiatives focus on the enhancement of subject matter knowledge. The focus of the underlying project as referred to above was on the development of quality teaching.

Data and data collection

The data were the scores learners obtained in the end-of-year Grade 10 Mathematics examinations. These scores comprise 75% of the total mark of 200 that is awarded for Mathematics. The other 25% is compiled from tasks, tests and the mid-year examination.

Participating mathematics educators, mathematicians, mathematics teachers and mathematics curriculum advisors set the examinations. The mathematics educators and the mathematics curriculum advisors firstly designed draft items. These were discussed with teachers at workshops to ensure that there was fairness with regard to the topics that were covered in their teaching. Upon reaching consensus, the examination papers and the memoranda of marking were moderated by the participating mathematicians. Figure 2 is an example of an item of the examination.

The project staff designed the final versions of the examination papers and electronic versions were dispatched to schools for them to put in a format as required by the schools. For example, most schools follow the format where the cover page of their examination papers must have the school's emblem on it.

In order to prevent leakage of the examination papers, the school management teams were approached to timetable the examinations for the same date and time. The five schools agreed to this request. As is normal for school-based end-of-year examinations the responsible mathematics teachers of the schools marked the scripts. Except for two schools, the same teacher taught Grade 10 for the three years. For the one school where this was not the case, the school uses the strategy of a teacher 'taking the learners through' from Grade 10 to 12. This school had two teachers involved and they both attended all the project activities. The other school changed the teacher responsible for teaching Mathematics in Grade 10 in 2012 due to the responsible teacher being on maternity leave for the first half of 2013. The other teacher taught Grade 10 for 2013 and 2014 and also attended all project activities.

To ensure consistency of marking across the five schools, a $1\frac{1}{2}$ -day common marking session was held in 2012. This was not repeated for 2013 and 2014 since the same teachers who were involved in the 2012 administration were those for 2013 and 2014. It was assumed that the teachers would mark the scripts according to the procedures applied for the 2012 examination. Further, the project had access to the scripts to record the marks and no observable deviations from agreed-upon marking procedures developed in 2012 were found in 2013 and 2014.

The marked scripts were collected from the schools once all the administrative procedures that schools are required to do were completed. The score for each item – the sub-sections of a question – for each learner from a school was captured. Therefore, the only data recorded were the scores as reflected on the scripts of the learners.

After the collection of the common examination scripts, the data were checked, cleaned and coded as described in Okitowamba (2015).

Ethical considerations

The university's research ethics committee cleared the project of which this particular study is a part with the ethics registration number 11/9/33. The project was also approved by the Western Cape Education Department through a memorandum of understanding between the university and the Western Cape Education Department. In order to maintain anonymity, the names of learners were not used and not recorded in the data files. The scripts were assigned numbers for purposes of checking during the data cleaning phase. The names of schools were also anonymised as indicated in Table 1.

Data analysis

Rasch procedures

There is a vast body of literature in which Rasch measurement theory is broadly explained, from its origin to its applications (Andrich, 1978; Bond & Fox, 2010; Dunne, Long, Craig, & Venter, 2012; Griffin, 2007; Long, 2011; Rasch, 1960/1980; Wilson, 2005; Wright & Stone, 1979). An important use of Rasch analysis is the computation of '"measures" that can … be used with parametric statistical tests' (Boon, Staver & Yale, 2014, p. 3). Furthermore, application of Rasch procedures provides a solution 'of measuring changes across time in achievement … at the same grade level over several years' (Scantlebury, Boone, Kahle, & Fraser, 2001, pp. 649–650).

Similar to the study reported here, Rasch measurement was used to compare different cohorts of students in Australia and to detect improvement in students' mathematics achievement in lower secondary schools over time (Afrassa & Keeves, 1999). This method fits the purpose of the investigation as it can help to detect whether the trend in achievement for different cohorts of learners is in a positive or negative direction. A more in-depth discussion regarding the implementation of Rasch analysis for this research is given by Okitowamba (2015).

The software WINSTEPS 3.4.1 (Linacre, 2008) was used for Rasch measurement. Given that the scores were not dichotomous and for these examinations partial scores are awarded, the procedures for Rasch partial credit model for polytomous data were applied.

Reliability and validity

Concerning reliability, Rasch measurement provides person reliability index and item reliability indexes. These indices

TABLE 2: Reliability index by year.

Reliability index	2012	2013	2014
Item reliability	0.98	0.96	0.98
Person reliability	0.83	0.87	0.86

were specifically calculated from the raw data for this article and are presented in Table 2.

The acceptable range for reliability coefficients of an instrument is that it should be greater than or equal to 0.70 (McMillan & Schumacher, 2010). The indexes in Table 2 indicate that reliabilities were within the acceptable range for the three cohorts of learners in this study. The near equality of these coefficients is also indicative of the examinations functioning in a similar fashion for the three years.

Regarding validity, content validity was assured through the construction of the test by the set of 'experts' – the teachers, mathematics educators, mathematicians and the mathematics subject advisor. The examinations were administered under normal conditions for examinations during the end-of-year examination period and followed the same processes and procedures for such examinations. There were no deviations from the way end-of-year examinations are conducted and administered. This implies that the ecological validity – 'the degree to which an assessment of events, activities, participation, or environments reflects everyday life expectations' (Crist, 2015, p. 2) – of the study was high. Reasonably high ecological validity of studies is, according to Black and Wiliam (1998), an important determinant for acceptance by teachers of ideas and notions emanating from research.

In the Rasch analysis

> construct validity focuses on the idea that the recorded performances are reflections of a single underlying construct [*and*] fit indices help the investigator to ascertain whether the assumption of unidimensionality holds up empirically. (Bond & Fox, 2010)

For a test to satisfy unidimensionality only the items within the range 0.5 and 1.5 are deemed productive for measurement (Linacre, 2008, p. 227). The ranges for the infit and outfit statistics for 2013 were 0.85 to 1.14 and 0.62 to 1.28 respectively. Thus a condition for construct validity satisfied. This was also the case for the 2014 test where the range for the infit statistic was 0.23 to 1.44 and 0.29 to 1.4 for the outfit one. It was not the case for the 2012 test–infit statistic range 0.37 to 1.97 and outfit statistic range 0.26 to 1.75. For the 2012 examination, four items fell outside the range and were not included in the analysis. These items were spread over the two papers. In total they accounted for 2.5% of the total marks. They were at the 'easy' end of the person-item map. It was deemed that the contribution to the total score of the four items was negligible in the sense that it would not affect the analysis detrimentally. The items were excluded from the analysis rendering all three complying with a condition for construct validity. The 2012 examination had 74 items (405 persons), the 2013 one 82 items (82 persons) and there

were 62 items (407 persons) for the 2014 examination. Fit statistics provided by Rasch analysis is a quality control mechanism to determine whether data holds the assumption of unidimensionality (each item contributes to measure only one construct that represents one attribute or ability at a time). The data used thus measured what it was intended to measure (construct validity).

Other statistical analysis tests

Statistical significance between the scores of 2012 and 2013, 2013 and 2014 and 2012 and 2014 was determined using *t*-tests. The software SPSS version 23 was used.

'Effect sizes' were calculated to assess the effectiveness of the intervention. According to Coe (2002, p. 1):

'Effect size' is simply a way of quantifying the size of the difference between two groups [*and*] … particularly valuable for quantifying the effectiveness of a particular intervention. … It allows us to move beyond the simplistic, 'Does it work or not?' to the far more sophisticated, 'How well does it work in a range of contexts?'

Different formulae exist for determining effect size (see, for example, Cohen, 1988; Kerby, 2014; McGraw & Wong, 1992). In this study, the one that was used by Afrassa and Keeves (1999, p. 4) was utilised because of the similarity of this research and the study reported here. The formula is:

$$ES = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2 + s_2^2}{2}}}$$

where

$\overline{x_1}$ = estimated mean for group one (year X)
$\overline{x_2}$ = estimated mean for group two (year X + 1)
s_1 = standard deviation of the mean of group one
s_2 = standard deviation of the mean of group two

Since this study mirrors the one by Afrassa and Keeves (1999), the interpretation of the effect sizes given by them (p. 4) are adopted. Namely:

$ES < 0.20$:   the size of effect is trivial
$0.20 \leq ES < 0.5$:   the size of effect is small
$0.50 \leq ES < 0.80$:   the size of effect is medium
$ES \geq 0.80$:   the effect size is large

There are other interpretations for acceptability of effects sizes. Hattie (2009), for example, accepts an effect size as useful if it exceeds 0.4.

Results and discussion

The results of the analysis are presented in this section. They are given for three periods: 2012 to 2013, 2013 to 2014 and 2012 to 2014. Table 3, Table 4 and Table 5 present the outcomes of the various statistical analyses.

TABLE 3: Mean scores of the five schools for the three cohorts.

Year	*N*	Mean	*SD*	*SEM*
2012	405	26.14	13.483	0.670
2013	381	22.32	13.276	0.680
2014	407	35.42	16.478	0.817

Source: Okitowamba, O. (2015). *Tracking learners' performances in high-stakes Grade 10 mathematics examinations*. Unpublished doctoral dissertation, University of the Western Cape, Bellville, South Africa (pp. 93–94). Retrieved from http://hdl.handle.net/11394/5655

TABLE 4: *t*-test for mean difference for cohorts 2012 and 2013.

Period	*t*	*df*	Sig. (two-tailed)	*MD*	*SE* difference	95% confidence interval of the difference	
						Lower	Upper
2012–2013	4.000	784	0.00**	3.821	0.955	1.946	5.696
2013–2014	−12.243	786	0.00**	−13.105	1.070	−15.206	−11.004

Source: Okitowamba, O. (2015). *Tracking learners' performances in high-stakes Grade 10 mathematics examinations*. Unpublished doctoral dissertation, University of the Western Cape, Bellville, South Africa (pp. 93–94). Retrieved from http://hdl.handle.net/11394/5655
**, significant for *p* < 0.001

TABLE 5: The effect sizes for different periods.

Period	Effect size
2012–2013	−0.29
2013–2014	0.88
2012–2014	0.62

Source: Okitowamba, O. (2015). *Tracking learners' performances in high-stakes Grade 10 mathematics examinations*. Unpublished doctoral dissertation, University of the Western Cape, Bellville, South Africa (pp. 93–95). Retrieved from http://hdl.handle.net/11394/5655

Table 4 indicates that the differences were significant for the three periods. For the period 2012 to 2013 the mean score declined. It is conjectured that the introduction of the Curriculum and Assessment Policy Statement in Grade 10 in 2012 contributed to the low average score in 2012. Furthermore, regarding the contribution of the project, it required teachers to use teaching strategies with which they were not fully conversant. Given that teaching is a habit, it is widely accepted that the appropriation of other strategies takes time. For the period 2013 to 2014, the mathematics mean scores improved from 22.32% to 35.42% which was a significant increase. Contributors to this improvement are obviously the increased familiarity with the curriculum and an emerging implementation of some of the strategies linked to examination-driven teaching (Julie, 2016). The trend also reveals a significant positive increase for the period 2012 to 2014. Figure 3 presents the trends graphically.

The calculated effect sizes are given in Table 5.

Table 5 shows that the effect from 2012 and 2013 was trivial. As is argued by Julie (2016) this decrease follows the pattern of any improvement initiative where there is an initial deterioration before improvement. The effect size from 2013 to 2014 was large and between 2012 and 2014 it was medium. Taking the average effect of 0.40 as a crude indicator of an overall effect then it is small but within Hattie's (2009) cut-off for a reasonable effect.

Conclusion

In light of the study outcomes reported above, there are positive indicators that the trend of mathematics achievement in school-based end-of-year summative examinations moves
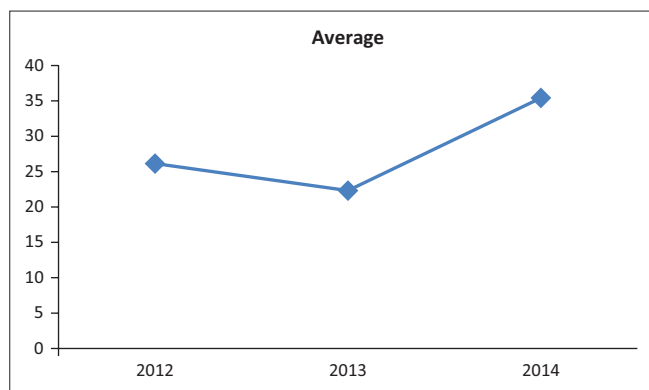
FIGURE 3: Trend of learners' mathematics performance over time.

Source: Okitowamba, O. (2015). *Tracking learners' performances in high-stakes Grade 10 mathematics examinations.* Unpublished doctoral dissertation, University of the Western Cape, Bellville, South Africa (p. 95). Retrieved from http://hdl.handle.net/11394/5655

in a positive direction over time when an examination-driven teaching strategy is employed. However, this improvement is not necessarily immediate and the nature of examination-driven teaching must be carefully considered and crafted to counter a minimalist view of teaching to the test. Examination-driven teaching is definitely not the only contributor since enhanced achievement in high-stakes examinations is closely related to socio-economic status and ultimately the desirable achievement results that are sought will only materialise if socio-economic inequalities are substantially reduced.

Acknowledgements

References

Adler, J., & Pillay, V. (2017). Setting the scene: School M, Mr T, the lesson and data. In J. Adler & A. Sfard (Eds.), *Research for educational change: Transforming researchers' insights into improvement in mathematics teaching and learning* (pp. 25–38). London: Routledge.

Afrassa, T.M., & Keeves, J.P. (1999). Changes in students' mathematics achievement in Australian lower secondary schools over time. *International Education Journal, 1*(1), 3–21. Retrieved from http://ehlt.flinders.edu.au/education/iej/articles/v1n1/afrassa/afrassa.pdf

Andrich, D. (1978). *Rasch models for measurement.* Beverly Hills, CA: Sage.

Australian Academy of Science. (2015). *Desktop review of mathematics school education pedagogical approaches and learning resources.* Canberra: AST. Retrieved from https://docs.education.gov.au/system/files/doc/other/trim_review_paper_2_-_aas_-_final.pdf

Bishop, A.J., Hart, K., Lerman, S., & Nunes, T. (1993). *Significant influences on children's learning of mathematics.* Science and Technology Education Document Series No. 47. Paris: UNESCO. Retrieved from http://www.unesco.org/education/pdf/323_47.pdf

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bond, T.G., & Fox, C.M. (2010). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Routledge.

Boon, W.J., Staver, J.R., & Yale, M.S. (2014). *Rasch analysis in the human sciences.* Dordrecht: Springer.

Brodie, K. (2013). The power of professional learning communities. *Education as Change, 17*(1), 5–18. https://doi.org/10.1080/16823206.2013.773929

Burkhardt, H., & Pollak, H. (2006). Modelling in mathematics classrooms: Reflections on past developments and the future. *ZDM—The International Journal on Mathematics Education, 38*(2), 178–195. https://doi.org/10.1007/BF02655888

Coe, R. (2002, September). *It's the effect size, stupid: What effect size is and why it is important.* Paper presented at the British Educational Research Association Annual Conference, Exeter. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.htm

Cohen. J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Crist, P.A. (2015). Framing ecological validity in occupational therapy practice. *The Open Journal of Occupational Therapy, 3*(3), Article 11. https://doi.org/10.15453/2168-6408.1181

Davis, J., & Martin, D.B. (2006). Racism, assessment, and institutional practices: Implications for mathematics teachers of African American students. *Journal of Urban Mathematics Education, 1*(1), 10–34. Retrieved from http://ed-osprey.gsu.edu/ojs/index.php/JUME/article/view/14

Department of Basic Education. (2011). *Curriculum and assessment policy statement.* Pretoria: DBE.

Dunne, T., Long, C., Craig, T., & Venter, E. (2012). Meeting the requirements of both classroom-based and systemic assessment of mathematics proficiency: The potential of Rasch measurement theory. *Pythagoras, 33*(3), Art. #19, 16 pages. https://doi.org/10.4102/pythagoras.v33i3.19

Escalante, J., & Dirmann, J. (1990). The Jaime Escalante math program. *The Journal of Negro Education, 59*(3), 407–423. Retrieved from http://www.jstor.org/stable/2295573

Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation, 33*(1), 87–99. https://doi.org/10.1016/j.stueduc.2007.01.007

Hattie, J.A.C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Howie, S. (2012). High-stakes testing in South Africa: Friend or foe? *Assessment in Education: Principles, Policy & Practice, 19*(1), 81–98. https://doi.org/10.1080/0969594X.2011.613369

Jacobs, M., Mhakure D., Fray, R.L., Holtman, L., & Julie, C. (2014). Item difficulty analysis of a high-stakes mathematics examination using Rasch analysis. *Pythagoras, 35*(1), Art. #220, 7 pages. https://doi.org/10.4102/pythagoras.v35i1.220

Jesness, J. (n.d.). Stand and deliver revisited: *The untold story behind the famous rise – and shameful fall – of Jaime Escalante, America's master math teacher.* Retrieved from https://endteacherabuse.org/Escalante.html

Johnson, D.M., & Smith, B. (1987). An evaluation of Saxon's algebra. *Journal of Educational Research, 81*(2), 97–102. https://doi.org/10.1080/00220671.1987.10885804

Julie, C. (2013). Can examination-driven teaching contribute towards meaningful teaching? In D. Mogari, A. Mji, & U.I. Ogbonnaya (Eds.), *Proceedings of the ISTE International Conference on Mathematics, Science and Technology Education* (pp. 1–14). Pretoria: UNISA Press.

Julie, C. (2016). *Does a CPD initiative focusing on the development of teaching to enhance achievement outcomes in high-stakes mathematics examinations work?* Bellville: University of the Western Cape. Retrieved from https://ledimtali.wixsite.com/ledimtali/reports

Kerby, D.S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Innovative Teaching, 3*(1), 1–9.

Linacre, J.M. (2008). *A user's guide to WINSTEPS®.* Retrieved from http://www.winsteps.com/winsteps.htm

Long, C. (2011). *Mathematical, cognitive and didactic elements of the multiplicative conceptual field investigated within a Rasch assessment and measurement framework.* Unpublished doctoral dissertation, University of Cape Town, Cape Town, South Africa. Retrieved from http://hdl.handle.net/11427/10892

Martin, M.O., Mullis, I.V.S., & Chrostowski, S.J. (Eds.). (2004). *TIMSS 2003 Technical Report.* Boston, MA: TIMSS & PIRLS International Study Center.

McGraw, K.O., & Wong, J.J. (1992). A common language effect size statistic. *Psychological Bulletin, 111*(2), 361–365.

McMillan, J.H., & Schumacher, S. (2010). *Research in education: Evidence-based inquiry.* New York, NY: Pearson.

Mogari, D., Kriek, J., Stols, G., & Iheanachor, O.U. (2009). Lesotho's students' Achievement in mathematics and their teachers' background and professional development. *Pythagoras, 70*(3), 3–15. https://doi.org/10.4102/pythagoras.v0i70.33

Okitowamba, O. (2015). *Tracking learners' performances in high-stakes Grade 10 mathematics examinations.* Unpublished doctoral dissertation, University of the Western Cape, Bellville, South Africa. Retrieved from http://hdl.handle.net/11394/5655

Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68*(9), 679–682. Retrieved from http://www.jstor.org/stable/20403467

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: The University of Chicago Press.

Reay, D., & Wiliam, D. (1999). 'I'll be a nothing': Structure, agency and the construction of identity through assessment. *Educational Research Journal, 25*(3), 343–354. https://doi.org/10.1080/0141192990250305

Reddy, V., & Janse van Rensburg, D. (2011). Improving mathematics performance at schools. *HSRC Review, 9*(2), 16–17. Pretoria: HSRC. Retrieved from http://www.hsrc.ac.za/en/review/June-2011/improving-mathematics

Reddy, V., Berkowitz, R., & Mji, A. (2005). *Supplementary tuition in mathematics and science: An evaluation of the usefulness of different types of supplementary tuition programmes.* Pretoria: HSRC. Retrieved from http://www.hsrc.ac.za/en/research-data/view/1983

Scantlebury, K., Boone, W., Kahle, J.B., & Fraser, B.J. (2001). Design, validation, and use of an evaluation instrument for monitoring systemic reform. *Journal of Research in Science Teaching, 38*(6), 646–662. https://doi.org/10.1002/tea.1024

Seabrook, R., Brown, G.D.A., & Solity, J.E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*(1), 107–122. https://doi.org/10.1002/acp.1066

Shalem, Y., Sapire, I., & Huntley, B. (2013). Mapping onto the mathematics curriculum – an opportunity for teachers to learn. *Pythagoras, 34*(1), Art. #195, 10 pages. https://doi.org/10.4102/pythagoras.v34i1.195

Shepard, L.A., & Dougherty, C.K. (1991). *Effects of high-stakes testing on instruction.* American Educational Research Association, 1–39. Retrieved from http://nepc.colorado.edu/files/HighStakesTesting.pdf

Smith, S.M., & Rothkopf, E.Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*(3), 341–358. https://doi.org/10.1207/s1532690xci0103_4

Statistics South Africa. (2016). *Quarterly labour force survey: Quarter 2: 2016.* Pretoria: Statistics South Africa. Retrieved from http://www.statssa.gov.za/publications/P0211/P02112ndQuarter2016.pdf

Suurtamm, C., Denisse, R., Thompson, D.R., Kim, R.Y., Moreno, L.D. Sayac, N., ... Vos, P. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment.* Springer Open. https://doi.org/10.1007/978-3-319-32394-7_1

Swan, M., & Burkhardt, H. (2012). A designer speaks: Designing assessment of performance in mathematics. *Educational Designer: Journal of the International Society for Design and Development in Education, 2*(5), 1–41. Retrieved from http://www.educationaldesigner.org/ed/volume2/issue5/article19/

Van den Heuvel-Panhuizen, M., & Becker, J. (2003). Towards a didactic model for assessment design in mathematics education. In A.J. Bishop, M.A. Clements, M.C. Keitel, J. Kilpatrick, & F.K.S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 689–716). Dordrecht: Kluwer Academic.

Watson, A., & De Geest, E. (2012). Learning coherent mathematics through sequences of microtasks: Making a difference for secondary learners. *International Journal of Science and Mathematics Education, 10*(1), 213–235. https://doi.org/10.1007/s10763-011-9290-3

Wilson, M. (2005). *Constructing measures: An item response modelling approach.* London: Lawrence Erlbaum.

Wright, B.D., & Stone, M.H. (1979). The measurement model. In B.D. Wright & M.H. Stone (Eds.), *Best test design* (pp. 1–17). Chicago, IL: MESA Press.