THE EFFECTS OF DISTRIBUTED PRACTICE ON TWO GRADE 10 MATHEMATICS CLASSES

By

PIERRE SUTHERLAND

(Under the direction of John Olive)

ABSTRACT

The purpose of this study was to investigate the effects of distributed practice on test performance and summer learning loss in two grade 10 mathematics classes in a South African public high school. Two teachers each taught a control and a treatment class. This study focused on the two treatment classes in which students took short, in-class tests at the start of class on 37 occasions throughout the third and four quarters of the academic year. In-class test items were similar to homework problems and arranged over an expanding time interval (1-2-4-8-16-32 days). Comparison between the control and treatment classes was inconclusive on test performance and insignificant in terms of summer learning loss ($p = 0.057$). However, enhanced strategic competence (Kilpatrick, 2001) was suggested by student responses on certain examination items. Finally, comments made by the teachers on overall effects of the study were compared to treatment and examination scores.

INDEX WORDS: Curriculum, Distributed Practice, Expanded Retrieval Practice, Fluency, Mathematics, Strategic Competence, Repeated Exposure, Secondary, Sequencing, Summer Learning Loss, Test Performance, Testing

THE EFFECTS OF DISTRIBUTED PRACTICE ON TWO GRADE 10 MATHEMATICS CLASSES

by

PIERRE SUTHERLAND

B.Sc., Randse Afrikaanse Universiteit, South Africa, 2004

B.Sc. (Hons.), University of Pretoria, South Africa, 2006

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2012

THE EFFECTS OF DISTRIBUTED PRACTICE ON TWO GRADE 10 MATHEMATICS CLASSES

by

PIERRE SUTHERLAND

Major Professor: John Olive

Committee: Martha Carr
James W. Wilson

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2012

DEDICATION

For my father: thanks for the letter.

ACKNOWLEDGEMENTS

I am grateful to all the students who participated in this study: not a single one of the 60 students opted out and many took on the challenge with vigor and enthusiasm. I am equally grateful for both of the teachers Mr. Reed and Mrs. Smith because they made the in-class testing a priority in spite of their workloads, and made it work. Also, as we were separated by an ocean and had to make do with lengthy email and phone conversations, I must thank my mother for being a valuable and supportive liaison.

I would like to thank my advisor, John Olive, and my committee members, Jim Wilson and Martha Carr, for showing me a level of scholarly professionalism, patience, and geniality worth aspiring to; I thank you for bearing with me. I would like to thank Bettie St. Pierre for playing an integral part in transitioning me to routined, scholarly writing. Many thanks all around.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The purpose of this study was to investigate the effects of distributed practice on test performance and summer learning loss in two grade 10 mathematics classes in a South African public high school. Specifically, did the testing improve test performance in comparison to classes without the testing? What were the effects of the testing on stronger or weaker candidates in the treatment groups respectively? What other effects, besides examination performance, did the testing have on students and teachers, if any?

Each of the two grade 10 mathematics classes, of about 30 students each, belonged to one of two teachers working together in planning and teaching the mathematics curriculum. These teachers both taught two grade 10 mathematics classes: one was the treatment group and the other was the control group. On a normal day, the students of a treatment group would start their mathematics class with a short test with an item from two days ago, another from four days ago, eight days ago, and so forth; this test was very similar to one they were given the day before in preparation for this test. The length of the test was determined by how long these items have been building up e.g., the 37th test had five items (2-4-8-16-32). Afterwards, the responses were handed in and a short discussion might have followed concerning some of the test items; these responses were later graded and captured to a database. This procedure took between four and eight minutes and instruction followed. These tests were born out of a technological variation of an old theme: distributed practice. It draws, as will be discussed later, on literature concerned with memory, summer learning loss, frequent testing, and mathematical proficiency.

In the literature review I looked into some of the psychology on memory – specifically distributed practice – highlighting the contention between two of its popular forms (Karpicke & Roediger, 2010) and reasons supporting expanding intervals between tests rather than equal intervals. The work of Rohrer & Pashler (2007) was also important in applying this mainly rote-learning concept (e.g. memorization of vocabulary or images) to the realm of mathematical knowledge. Also, the timing of this study allowed for investigating a possible connection with reducing summer learning loss (Bakle, 2010).

The design and method section of this paper looks into how a theme in the mathematics curriculum was broken down into daily topics and how these topics were linked to test items and sequenced according to the expanding distributed schedule. As the specific content of these items are important, I spend some time looking into and giving examples of several items. Some practical challenges of implementing the testing, along with capturing individual-level data, are addressed in this section.

This study may contribute to literature because research in the area of distributed practice in mathematics is not abundant; neither is the application of this practice to several themes or topics building up over time[1] and that over a non-trivial period of time (Cepeda, Coburn, Rorher, Wixted, Mozer, & Pashler, 2009).

Finally, I report on the findings by looking at data from student performance on in-class tests and exams, a pre- and post-test administered before and after the summer break, and interview data where claims made by the teachers are compared to the testing data. This last comparison also highlighted how changes appear to have occurred in the experimental classes but was evasive in terms of measurement or correlation with performance data.

[1] Contrasts Spiral Curriculum (Harden, 1999) as items are neither related nor increasing in difficulty.

CHAPTER 2

BACKGROUND AND LITERATURE

The researcher. I taught at the public high school in which this study took place for two and a half years before continuing graduate studies in the United States. I had built up rapport with my mathematics colleagues and especially with my mentor, Mr. Reed. I had their trust and they were willing to try something new. This study allowed me to experiment with an idea I touched on while teaching at this school and to remain connected with friends and colleagues. I was struck by how much information students had to contend with. It was not uncommon to see a student simultaneously texting, listening to music, and having a conversation: activity concerned mainly with short-term memory. I felt that a lesson also required little more than a short-term treatment since it was largely self-contained and unrelated to prior or future work; I wanted to give students more opportunities to connect the current lesson to prior work without adding to their workload but rather by restructuring class time reasonably unobtrusively. My thinking was also influenced by a program I developed using Microsoft Excel for the purpose of a warm-up activity. This Excel program generated arithmetic tests, which were copied onto a transparency and shown to the class, one item at a time, using a mask (a timer was also set on a negotiated time from 500 to 2000 milliseconds). The students seemed to enjoy the short break from routine and had small successes in these one-minute tests. Other effects seemed to include faster note taking, more conversation (from shy students), and seemingly better focus. In addition, emphasis was placed on a student's individual improvement rather than getting a perfect score. With this study, I wanted to incorporate course content into this warm-up activity without pilfering from instruction; the in-class tests (appendix A) were the result of this endeavor. This is my perspective that drove the design and implementation of the study.

Literature. The development of the in-class testing drew on four topics in the literature: memory, summer learning loss, the testing effect, and mathematical proficiency. Below I outline these themes and the connections between them, and, finally, how they influence design.

Memory and testing. Memory refers to *any relatively lasting storage of information in the brain, which is currently hypothesized to involve processes of encoding, storage, and retrieval of the information* (Matsumoto, 2009). Research on practices that enhance memory and learning has been evident for well over a century beginning around the time of Ebbinghaus (1913/1883), whose work is especially significant in this study, as he is accredited with the “spacing effect” (Pavlik, 2008) that relies on dispersed study sessions – as opposed to contiguous study sessions – to improve performance. I refer to this spaced practice as *distributed practice*, and I differentiate between equal interval practice and expanding interval practice. The interval refers to the time between study sessions, or exposures, to some information or concept. If students were to revisit, say, a list of vocabulary words every four days, this would be equal interval and noted as (4-4-4-4-4-4); in contrast, should the student revisit the vocabulary on the next day, then two days thereafter, then four, then eight etc. it would be expanding interval and noted as (1-2-4-8-16-32). Massed practice, or cramming, is essentially the opposite to this distributed practice; in this form of practice students are exposed to some content once and typically for longer than a distributed session.

It is recognized, in a variety of learning models, that distributed practice of information is better remembered than the massed practice thereof (Sisti, Glass, & Shors, 2007). In addition, distributed practice generally creates more learning opportunities without using more time, and these extra opportunities (when presented in a similar way to an assessment) can engender the self-fulfilling *testing effect:* students who are tested more often generally outperform students who are tested less often (Bangert-Drowns, Kulik, & Kulik, 1991). There is, however, contention – and lacking in empirical evidence – between which form of distributed practice, equal or expanding, is better (Karpicke & Roediger, 2010). It is because of this contention in the literature that expanding interval practice was chosen over equal interval practice: the prior saved time due to a slower item build up (this is discussed in the methodology).

Mathematical Proficiency. To claim that memorization improves performance in mathematics is ambitious; however, a case can be made for repeated exposure to mathematical items to enhance proficiency. To investigate the connection between mathematical proficiency and memory I will explain what is meant by proficiency and then identify areas that might be influenced. Kilpatrick (2001) defines mathematical proficiency as an encompassing term for the key components, or strands, necessary for anyone to learn mathematics successfully. These strands, along with brief descriptions, are:

- *Adaptive Reasoning*: capacity for logical thought, reflection, explanation, and justification;

- *Strategic Competence*: ability to formulate, represent, and solve mathematical problems;
- *Conceptual Understanding*: comprehension of mathematical concepts, operations, and relations;
- *Productive Disposition*: Habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy;
- *Procedural Fluency*: skill in carrying our procedures flexibly, accurately, efficiently, and appropriately.

I position the influence of this distributed practice mainly within the strand of strategic competence and draw on student responses to substantiate. I take strategic competence, furthermore, to mean:

> …students are often presented with clearly specified problems to solve, outside of school they encounter situations in which part of the difficulty is to figure out exactly what the problem is. Then they need to formulate the problem so that they can use mathematics to solve it. Consequently, they are likely to need experience and practice in problem formulating as well as problem solving. They should know a variety of solution strategies as well as which strategies might be useful for solving a specific problem. (Kilpatrick, 2001, p. 124)

Summer Learning Loss. Although popular opinion holds that summer learning loss is widely attributed to the need for agricultural labor needed during the summer, Gold (2002) offers an insightful argument for bureaucratic and political motivators and suggested that more affluent members of society flee cities during the warm summer months and want to take their children with them. It would seem that the long summer break still echoes some advantage that higher socio-economic status students have over their economic opposites through summer programs and activities. The literature suggests there is a more pronounced learning loss for “math facts” and spelling over the summer break than for other tested skill areas; furthermore, socio economic status has less of an impact on the rate of forgetting in mathematics than in reading. Combatants of summer learning loss included Cooper (2003) making a case for viable solutions in the form of increasing the number of days in the school year, or at least a reorganization of the school year. This study investigates whether distributed practice might aid in this impasse.

Memory, mathematics, and summer learning loss. As outlined in the previous sections, research on memory focuses primarily on the retention of information acquired through rote learning rather than skill acquisition or problem solving. Furthermore, Cepeda et al. (2009) pointed out that these studies generally unfold over time periods that are near trivial (e.g. hours, days, or weeks) when compared to the time between learning and testing in schools (e.g. weeks or months). Recent literature, however, shows that some benefits of distributed practice, rather than massed practice, *can* transfer to learning in subjects like mathematics. Pashler (2007) stated that *spacing principles applicable to declarative memory tasks*[2]*, seem to extend beyond declarative memory for facts and associations to at least some forms of mathematics skill learning*. Rohrer’s 2006 study suggested some evidence of this mathematical skill learning where students worked on ten combinatorial problems (combinations of letters in a sequence) either massed into one

[2] Declarative memory is defined by the Cambridge Dictionary of Psychology as *The process which underlies the ability to recall and verbalize information.*

study session or spaced over two study sessions. The findings were consistent with Pashler as *Spacers* outscored *Massers*; more specifically, no effects were found after one week, but substantial effects were found after four weeks. Rohrer & Pashler's data (2007) also showed that the benefits depend on both the *interval between study sessions and the interval between study and test.* These findings propose that it is possible to increase skill-related retention without increasing study time but rather by reorganizing the content of study sessions (Footnote 1).

In addition to skill related benefits, Litman & Divachi (2008) found – in a 24-hour study – that distributed practice affords a *savings in forgetting* over massed practice. More importantly, these savings in forgetting were specific to *relational memory* and not item memory, where relational memory is defined as *memory for the contextual or associative aspects of a prior experience* (Cohen & Eichenbaum, 1995).

Possible contribution to the literature. The design of this study drew on the outlined literature in an attempt to pilot a practical implementation of distributed practice that continually adapts to the implemented curriculum on a day-to-day basis, that is, the actual curriculum delivered in class. This study intends to contribute to literature by:

- further investigating the effects of distributed practice on performance in school level mathematics;
- investigating the effects of distributed practice of a non-trivial period of time called for by Cepeda, et al. (2009);
- applying this practice to several topics in the curricula concurrently, on site, as part of the regular curriculum; and finally,
- investigating the effects of distributed practice on summer learning loss.

In addition to these possible contributions, some interview data are shared on less tangible residues of the study concerning both students and teachers e.g. effects on attitude, confidence, and classroom discourse.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

This study met with some constraints, as does any study dealing with external influence on a real-world problem: the teachers did not have regular access to internet or personal email, we communicated roughly once a week with the aim of administering about four tests every week. In addition, the teachers, Mr. Reed and Mrs. Smith (pseudonyms), were busy and pressured to deliver the curriculum.

To be clear on the terminology used:

- Exam: a test taken by all students at the end of a term (or quarter). An examination might be split into two parts, or tests, as was the case in the third term where the two tests dealt with algebra and geometry, and trigonometry respectively. Every student in this grade and course took these exams.
- Quizzes and portfolio work: a teacher might administer a quiz once every two weeks of her own design and may or may not share this assessment with other teachers to do the same. Portfolio work can be any assignment – usually a worksheet or project – that would ideally be done in class or assigned as homework; students might have to work in a group or individually.
- Pre-test and post-test: These two tests refer specifically to the second test of the third term that was significantly shortened and re-administered in the new academic year.
- Treatment tests (Appendix A): I will refer to these tests exclusively as *practice* or *in-class* tests. These tests were specifically designed for the study and only administered to the students in the experimental classes, as often as possible. Both these types are very similar as the practice test prepares students for the in-class test, and was assigned the day before the in-class test was taken. It is not homework and students may or may not have worked through it in preparation for the in-class test the following day.

I would generate the test items, compile these items into tests according to the distributed schedule, and email them to the administrative office of the school. Here, the practice-test and in-class test were printed side-by-side on an A4 page and given to Mr. Reed. After the first round of four tests had been emailed, printed, copied, administered, graded, and captured – taking nearer to 10 minutes of class time rather than the budgeted four minutes – I gave Mr. Reed the option to abort the study, which he refused out of hand. Mrs. Smith and Mr. Reed endured, creatively adjusting and tweaking until a working system was in place. I credit any success of this study as a testament to their rigor, and less so to the meek ambitions of this scholar-in-training.

Site of study. This study was conducted at a public high school (grade 8 through 12) in South Africa with 1500 students and 80 faculty members. The languages of instruction are Afrikaans and English, with the latter students being the minority. At the time of the study, there were roughly 300 students in each grade and class sizes range from 25 to 40 students. Usually, there were eight Afrikaans classes and two English classes in a grade. Both experimental groups and one control group are Afrikaans, with one control group English. Students have the option of receiving instruction in either language. Some subjects, like Computer Science or Geography, are taught to mixed classes and upon revisiting I enjoyed watching teachers and many students switch between languages almost effortlessly during instruction. This study focused on four classes of grade 10 students and their teachers: The two teachers were selected because they both had two grade 10 mathematics classes: one for control and one for experiment. The two teachers in this study are Mr. Reed and Mrs. Smith. Mr. Reed has been teaching mathematics at this high school for fourteen years and is often in charge of curriculum and lesson planning for grade 10 mathematics. Mrs. Smith has been teaching for over ten years and mathematics for the last four of those years.

During this year Mrs. Smith shadowed Mr. Reed's lesson plans one day later: this means that Mr. Reed would teach a lesson and Mrs. Smith would teach the same content and assign the same homework to her class the next day, this allowed them to share lessons and homework solutions. This was fortunate for the study in terms of reasonably consistent pedagogy and assessment. All four classes were taught the same lessons and assigned the same tasks and the experimental groups also took the in-class tests on similar schedules.

Participants. The students in all four classes are those who performed well enough in the 9th grade to continue on to mathematics in grade 10, failure to do so would have them placed in the mathematics literacy class. Mathematics literacy is compulsory for all students, unless they are taking the regular mathematics course, and is less demanding. Throughout the year, these students may also be moved to mathematics literacy should their grades drop below 40%, which is the passing grade. Grades in South African schools are measured in terms of percentages throughout the year and a letter grade is assigned at the end of the year. Class means for mathematics in grade 10 may range from 45% to 65% with a given grade above 80% constituting a distinction.

One of the most significant changes to the life of a tenth grader in the regular mathematics course is the change in composition of her grade. In grade nine 20% of the final grade is from the examination score and 80% from class work and portfolio work (projects, quizzes and homework), whereas in grade 10, 80% of the final grade comes from the examination score and 20% from the portfolio work. This sudden shift to high stakes examinations is arguably the biggest change in expectation of the student than any other year in high school, and in no other subject is it more difficult for students to maintain their previous grades. I remember the day that the tenth graders receive their first examination scores during the first term: it is a particularly grim and sobering day for them!

A typical day. Students worked on a 6-day roster with 8 periods per day: Monday might be a “Day 1” timetable on the first day of school but would then be a “Day 6”-timetable the following week. In courses like mathematics, students had 8 periods in a 6-day cycle; this generally means one or two *double* periods and the rest made up out of *single* periods. There are two double periods in a six-day cycle and these occasions were usually the best opportunity to administer the in-class tests; the other days were more susceptible to circumstances that made in-class testing problematic. The school day runs from 7:30am until 2pm with two breaks of about 20 minutes and 30 minutes respectively. A teacher will typically teach around 32 - 40 periods of the total 48 periods in a 6-day cycle. This high school focuses on performing well in sport, academia, and culture in the “macro” category; this is the biggest school category and it is based on the number of students attending the school. During certain parts of the year, a teacher might experience a classroom less than half full due to drafting practices in rugby or cricket, choir performances or play rehearsals. Ideally, a school period is about 40 minutes but can be influenced significantly by the day of the week and school activities: For instance, on Fridays, there is a religious morning-service followed by some team announcements and perhaps an outtake of a play to market some cultural activity. In this case, period lengths might drop to 25 minutes and be further shortened by the minutes taken for students to move between classes and taking out books before the lesson. The background information here serves to show that the *daily* in-class tests could simply not be administered everyday due to a low number of students in a class or the length of a period (Figure 1 shows the agenda for a typical day).

VRYDAG	27	FRIDAY
DAG/DAY - 3		
Personeelkamer/Staff room: Me/Ms [redacted]		
Detensie/Detension: [redacted]		
MR Bekendmaking Dagbestuur/MC Announcement Management		
Koshuisraad: Nominasies sluit/Hostel Council: Nominations closes		
Redaksie: Bekendmaking/Editorial Committee: Announcement		
Rugby: Roodepoort Rugbyklub Hoërskool Rugbyreeks @ Roodepoort Rugby Klub		
Rugby: Roodepoort Rugbyklub High School Tournament @ Roodepoort Rugby Club		

Figure 1: A day from the school planner.

Timeline. The South African school year runs from mid- January until mid-December, there is a winter break of three weeks and two mid-semester breaks of about a week each. These mid-semester breaks partition each semester into two terms. The decision was made to implement the testing during the final two terms of the 2011 academic year. This way the number of items generated would be less, but a significant amount of testing should be able to take place; also, we thought that the new tenth graders would have enough to deal with in the first two terms! Results from the in-class tests did not count toward a student's grade and they were able to opt-out of the study at any time; fortunately, none did. All students were also entered into a lottery every week where the winner received merchandize (like a mug, T-shirt, baseball etc. ranging from five to ten dollars). IRB approval was received in July of 2011 and testing started shortly thereafter.

Student grades and assessment. Students received homework almost daily and it was graded the following day, for this, they would receive a *complete*, *partially complete*, or *incomplete* grade. In addition to homework, a class would ideally take a review test or quiz roughly every two weeks. Finally, an examination is administered at the end of every term that constitutes 80% of the student's grade.

Pre-test and post-test. The third term examination was used as the pre-test, and selected problems from this test were administered – identically – during the first term of the next year, constituting the post-test. Therefore, the data used in this study consisted of the 37 in-class tests, performance on individual items on the pre- and post-test, and examination scores. Treatment occurred during the third and fourth terms where the third term examination consisted of two tests: the first dealt with topics from the term one and term two and the second test dealt with topics exclusively from the third term. This second test of the third term had a strong match between in-class test items and examination items. The post-test, unfortunately, was assigned as homework to students now in the 11th grade, about half of whom were not part of the original study due to new class assignments. In the end, we were able to gather data on 15 students from each of three of the four classes (two experiment classes and one control class). Once the testing was completed, the two teachers, Mr. Reed and Mrs. Smith, were interviewed on their experience and opinion of the testing, their perceived effect on the class in general, and possibly on summer learning loss. Students were not interviewed. Figure 2a outlines the timeline of the study and figure 2b shows the distribution of the in-class tests.

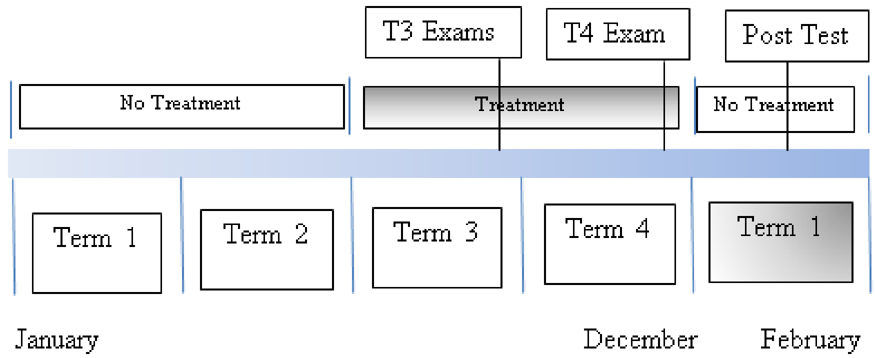


Figure 2*a: Study timeline*

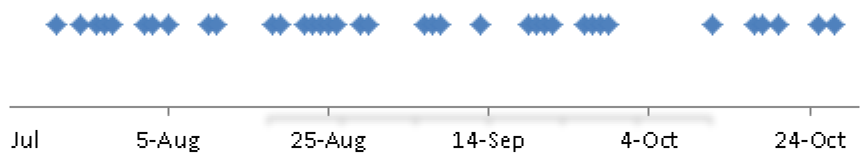


Figure 2b: Treatment timeline

Once interviews were conducted with Mr. Reed and Mrs. Smith, I also interviewed the two teachers who had the majority of the original treatment students in their grade 11 classes: for the sake of simplicity, I will refer to them as Mrs. Post-Reed and Mrs. Post-Smith. The purpose of these interviews was to get feedback from the teachers regarding the students in the experimental classes to gauge for possible differences in the behavior of the classes and to get a feel for the teachers' attitude toward the study and possible effects.

Some numbers. There were 187 students in grade 10 mathematics who were dealt into seven classes with student numbers ranging from 10 to 34 per class. The four groups used in the study were: Mr. Reed (30 experiment/30 control) and Mrs. Smith (26 experiment/33 control); for the post-test, data was available for Mr. Reed (15/18) and Mrs. Smith (23/0).

All in-class tests contained a sum of 137 items throughout 37 tests (these tests would contain 2, 3, 4, or 5 items each), each preceded by a similar practice test containing the same amount of items, ultimately yielding 274 “exposures” spread over 74 tests. At the end of the study, for each of the 60 experiment students, I had up to 137 item scores, 4 term grades and 24 scores for selected examination problems, totaling a little less than 10,000 data points.

Selection criteria. One of the main reasons that this grade and these classes were selected for the study instead of the 11^{th} and 12^{th} grade was that the tenth grade was the first time that a student’s grade is largely determined by test performance. Another reason was because the latter grade scores are used in university placement where the 10^{th} grade scores are not. Furthermore, I taught this curriculum for the full extent of my two and a half years at the school and felt most familiar with this content as well as student misconceptions. We chose these four classes specifically to be able to control for teacher differences: no other grade 10 mathematics teacher had two classes. This led to the first significant factor that hindered comparison – pointed out to me by Mr. Reed during the onset of the study – the class means of the treatment and control class vary greatly: the two treatment classes are two of the best performing classes in the grade while the control classes were two of the worst performing classes. A case was made for reversing treatment and control in both cases but the teachers were wary of this adjustment. By the end of the study, the in-class test took up nearly five hours of instruction, translating into more than a full cycle’s worth of instruction – the equivalent of more than a one-week holiday. Had the items been focused on preparing students for specific test items, the decision might have been reversed. It is a desperate situation and the novelty of frequent in-class testing, combined with the amount of time that it took away from normal instruction, was something that troubled the teachers because a performance dip might be the difference between a student staying in the mathematics course or changing over to mathematics literacy.

Item Design. Items were created using Microsoft Word along with screenshots taken of Geometer's SketchPad files, Cabri 3D documents, excel spreadsheets, data and charts from WolframAlpha.com, and any online document or image that was publicly available. Item design was driven by creating content that would challenge a student to understand the problem more than trying to memorize a specific pre-test item. For instance, the pre-test item and its corresponding in-class test item might be:

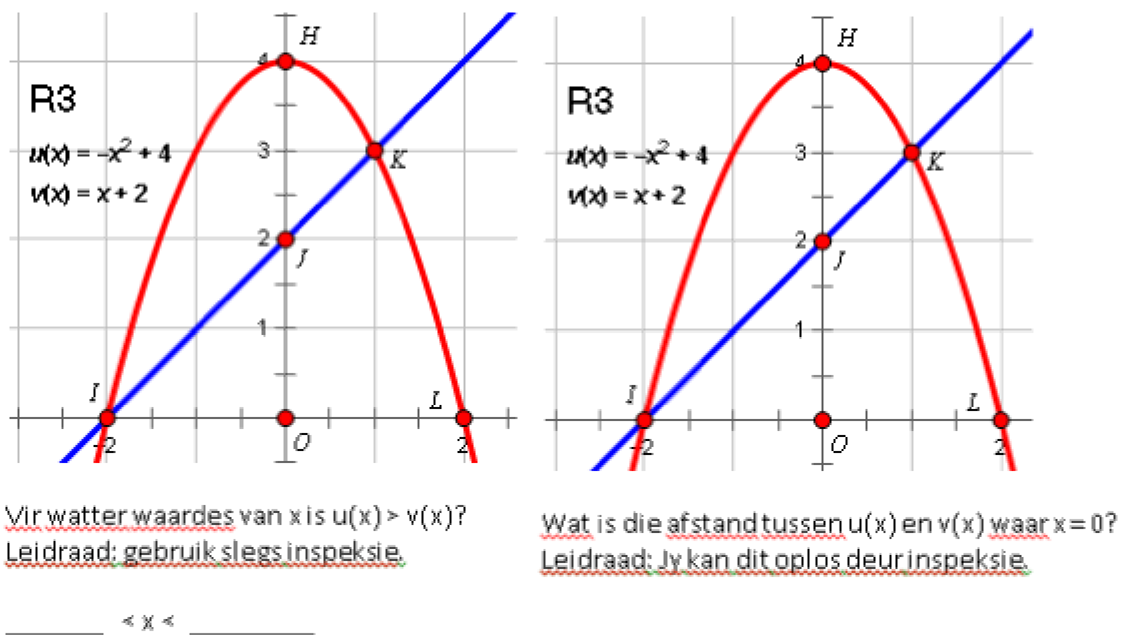


Figure 3: Practice item and its corresponding in-class item.

The item on the left in figure 3 prompts the student for the domain where u(x) > v(x); this item is related to *graphs and functions* and the item on the right asks for the distance between u(x) and v(x) where x = 0, an item still within the context of *graphs and functions* but possibly prompting a connection with the distance formula. This kind of items attempt to draw connections between topics in the curriculum that are seldom directly related. Items were used for various purposes: making connections between topics, gauging student understanding of a given topic, or even to prompt discussion by posing a problem in a novel way or context.

Algebra and geometry. Another barrier that I have found to be particularly crippling to students is algebraic manipulation of expressions and equations, but I could not think of a way to incorporate this into the existing items – without taking up even more time – other than posing problems like figure 4. This figure has four algebraic terms and asks the student which term represents area. The solution is contentious as a student might think of area units or linear units or, perhaps, propose some other solution. This specific solution is secondary to making such a connection or for stimulating discussion.

Gegee dat x, y, z rieele getalle is, watter uitdrukking(s) stel oppervlakte voor?

a.) z^2 b.) z

c.) x^2y d.) xy^2

Figure 4: Algebra and geometry

Compound interest and estimation. This was another topic addressed during the treatment and I decided to focus on estimation, notation, and place-value. Figure 5 shows examples of the items used: the student must match a word problem to the correct equation. This item does not require the student to memorize the formula but prompts for some understanding of its use.

R4: Susan bele R **a** teen **b** % vir **c** jaar, kwartaaliks saamgestel, watter vergelyking stel die som die beste voor?

a.) $a\left(1+\frac{\frac{c}{4}}{100}\right)^{4b}$ c.) $a\left(1+\frac{\frac{b}{4}}{100}\right)^{4c}$

b.) $b\left(1+\frac{\frac{c}{4}}{100}\right)^{4a}$ d.) $a\left(1+\frac{4b}{100}\right)^{\frac{c}{4}}$

R4: Sandra bele R **x** teen **p** % vir **y** jaar, kwartaaliks saamgestel, watter vergelyking stel die som die beste voor?

a.) $x\left(1+\frac{\frac{p}{4}}{100}\right)^{4y}$ c.) $x\left(1+\frac{\frac{y}{4}}{100}\right)^{4p}$

b.) $p\left(1+\frac{\frac{y}{4}}{100}\right)^{4x}$ d.) $x\left(1+\frac{4p}{100}\right)^{\frac{y}{4}}$

Figure 5: Compound interest.

Figure 5 is a good example of an estimation item, a skill that I felt was underrepresented in the curriculum at that time: the student must differentiate between two alternatives and determine the effect of interest rate and time. The equations might seem similar to a student, but option A is roughly $200 interest twice whereas option B is also roughly $200 but four times. Figures 6 and 7 also deal with the same topic.

Skat watter belegging is meer werd?

A) $20\,000\left(1+\frac{1}{100}\right)^{2}$

B) $10\,000\left(1+\frac{2}{100}\right)^{4}$

C) Hulle is dieselfde werd

Figure 6: Estimation and compound interest.

R4) Watter teken sal die stelling korrek maak?

$2500(1.05)^{20}$? $5000(1.05)^{10}$

a.) ≥
b.) ≤
c.) ≡
d.) Geen een van die bogenoemde

Figure 7: Estimate which inequality will make this statement true.

R2) Watter stellings is waar van die volgende belegging? (t beteken *toegepas*)

$$P = 2000\left(1+\frac{2}{100}\right)^{20}$$

a.) R2000 teen 2% jaarliks t. vir 20 jaar
b.) R2000 teen 4% halfjaarsliks t. vir 10 jaar
c.) R2000 teen 8% kwartaaliks t. vir 5 jaar
d.) R2000 teen 24% maandeliks t. vir 20 maande

Figure 8: Multiple-choice item where all the solutions are correct.

Figures 6 and 7 each had one solution, but figure 8 is different because every solution is correct; that is, the equation might represent $2000 invested at a rate of 2% compounded annually for 20 years or it might represent a rate of 8% compounded quarterly for 5 years. This item attempts to show that a concept may take on more than one, correct, form.

Statistics. From my experience with the statistics curriculum, I wanted to focus on items that helped communicate the practical uses of statistics and important differences between types of statistics. For instance, figure 9 is a collection of items that helped explain or highlight the applicability of certain statistics based on the context of the problem, and also presenting data to the student that would prompt interpretation.

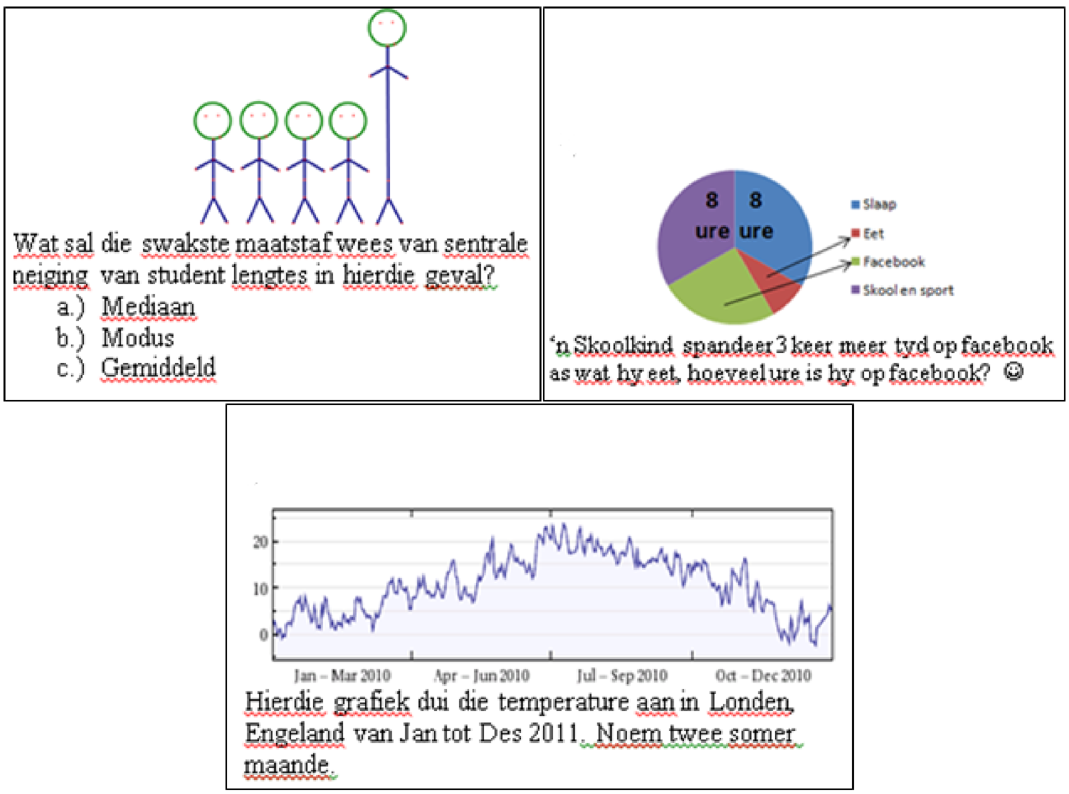


Figure 9: Statistical items: Median, mean, and mode (top-left); charts (top-right); and, graphs based on real-word data (bottom).

These figures outline some of the content addressed in the study and there are more examples of items and in-class test in appendix A. Table 1 displays the final number of items devoted to each theme in the order that the topics were presented to the students along with the amount of class periods devoted to each theme. *Graphs and functions* received more attention as requested by Mr. Reed about two months into the study.

Table 1

Total number of items per theme and periods devoted to a theme.

Theme	Items	Periods
Basic Trigonometry	10	9
Trigonometry word problems	21	3
Graphs and functions	38	6 + 4 (revision)
Interest	24	6
Area and volume	25	6
Statistics	20	4

Data Collection. There were four classes – of about 30 students each – involved in the study: two control, and two experiment. I will refer to them as Reed-Experiment, Reed-Control, Smith-Experiment, and Smith-Control. As outlined in the design, the primary data of this study (collected for each student in all four classes) are:

- per item scores for each of the 37 in-class tests;
- four aggregated term grades;
- per item scores for selected problems from examination 2, term 3 (*pre-test*);
- per item scores for said problems, re-taken after the summer (*post-test*).

Mrs. Smith had the students grade their own work, whereas Mr. Reed personally graded the students' work: over 4000 individual items in less than four months in addition to his workload. Items were generally scored as correct (1) or incorrect (0). Should an item have more than one question, the points scaled between 0 and 1; for instance, scoring 3 out of 4 questions, of the same item, right would result in a score of 0.75. These results were then captured to a spreadsheet, which I imported it into a database. As for summer learning loss, the pre-test and post-test items were handled in the same way as the in-class test data (e.g. Student S of class Reed-Control scored 7 out of 9 on problem 4). The third examination comprised of two tests: one on algebraic and geometric topics from the first two terms and the other test focused largely on trigonometry and function topics from term 3. In fact, this second test was based primarily on work that had undergone treatment, which made it ideal for post-testing. A blend of problems were selected from this second test ranging from easy to difficult based on mean class performance.

The database. I will briefly explain the rationale and content of each table in the database.

tblStudentItemScore. This aggregate table contains the total number of items (*CountOfScore*) that a student completed as well as her total score (*SumOfScore*) for all those items. The *Percentage* column is simply *SumOfSCore*/*CountOfScore**100. Figure 10 shows what the data in this table look like. These data were used to check for correlations of in-class test performance with term grades and to follow the performance of specific students, specifically those who performed the best on the in-class tests.

qStudentItemScore

admin	CountOfscor	SumOfscore	Percentage
[illegible]	129	72.75	56.3953488372093
[illegible]	134	76.75	57.2761194029851
[illegible]	133	76.25	57.3308270676692
[illegible]	116	66.876	57.651724137931
[illegible]	136	78.876	57.9970588235294
[illegible]	124	73.126	58.9725806451613
[illegible]	129	76.5	59.3023255813954
[illegible]	125	74.876	59.9008
[illegible]	133	80.5	60.5263157894737

Figure 10: Snapshot of table *tblStudentItemScore* data

tblClean. This table stores all scores of students that are unrelated to the in-class test data; for instance, Class, Teacher, term grades (T1, T2, T3, T4), as well as problems from both third term exams e.g. 2_5 is the score that a student achieved on problem five of the second exam. These data were used in the statistical analysis of hypothesized term-to-term performance increase and the reduction of summer learning loss.

tblClassTestScore, tblTestDate, tblDescription, and tblSequencing. These four tables were used to keep track of the sequencing of in-class test items, dates, and individual student scores. Figure 11 shows how these tables are connected. Every day had a topic associated with it, for instance day 18 has the description of "Graphs: General," where the teacher discussed and posed problems on graphs like domain and range or intervals. I then generated two sets of six items similar to the problems posed during this class – one practice set and one in-class set – and sequencing these items on the expanding schedule of 1-2-4-8-16-32 days. The grading of homework on the day following its assignment was considered the first repetition or exposure, R1, of the topic; all the following exposures took the form of in-class test items. This is why *tblSequence* does not have R1 column but runs from R2 through R7.

The example given in figure 11 shows the fifth exposure of day 18's topic: this item is linked to *tblTestDate* and scheduled for September 29th, 2011. This date table, in turn, is linked to *tblStudentClasstestScore* which stores all student scores for each respective item, and its repetition, of that specific test. Finally, figure 11 shows that a student received a perfect score for the fifth exposure to topic 18, in test 31, on September 29, 2011.

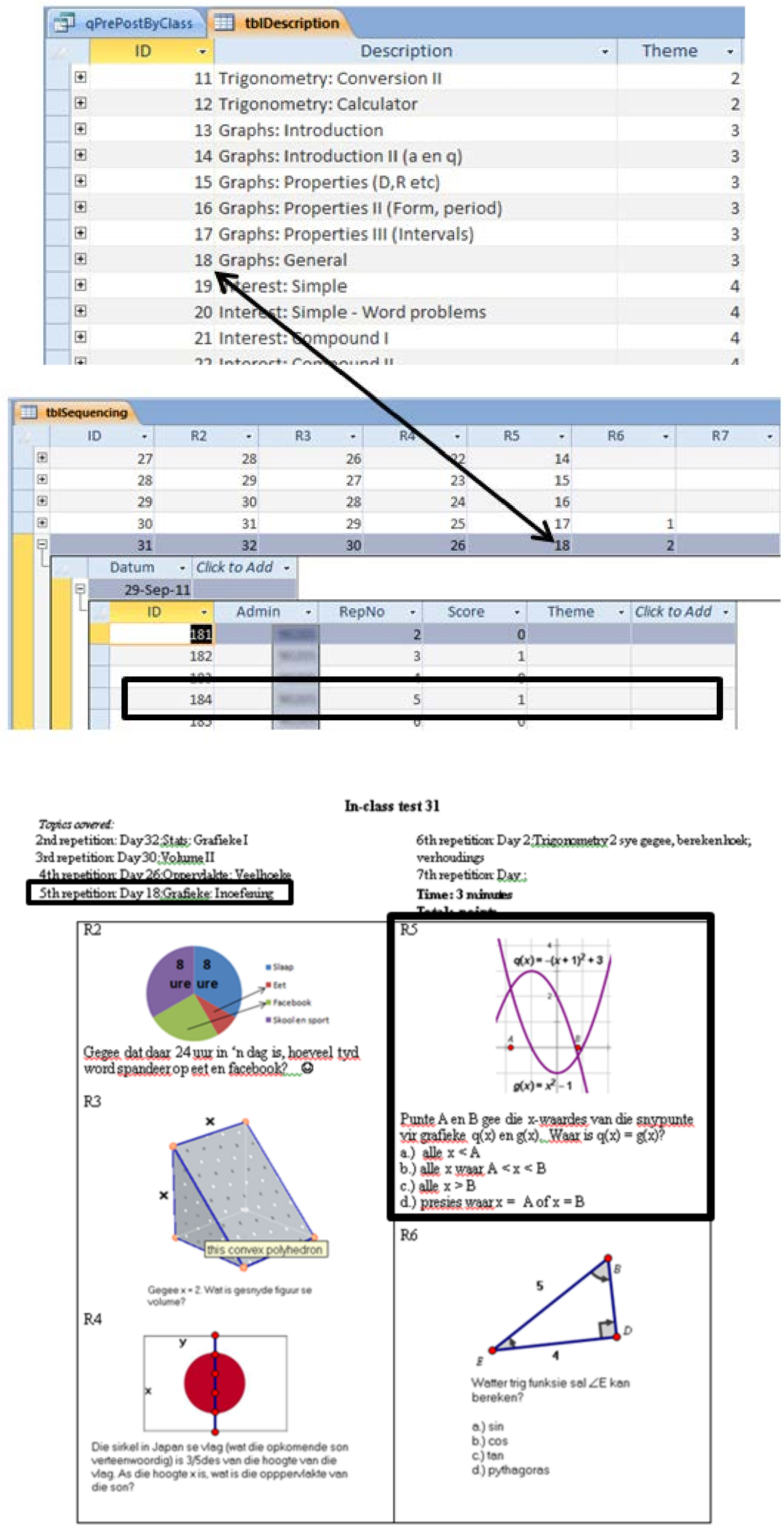


Figure 11: Coordinating items, in-class tests, and student scores.

CHAPTER 4

FINDINGS

The data used to investigate the possible effects of distributed practice on the four grade 10 mathematics classes are:

- Performance by overall term grade for each student;
- Performance on the pre- and post-test for summer learning loss;
- In-class test performance;
- Interview data; and,
- *Question 8* of the second test of the third term exam.

I made use of the in-class data to look for an overall correlation with mean term by class grades as well as individual students. I organized the interview data by themes and followed up on claims made about individual students by referring to in-class test scores as well as term grades.

Examination Performance. Finding any significant differences between the control and experiment groups was challenging. Initially, I suspected that there would be a significant difference between the treatment and control examination performance and hoped for some p-value that would substantiate, along with the interview data, my suspicions; this was not the case. Interview data also suggest that teacher differences are more obvious than any effect by treatment. The main factors that subdued comparison were between-class differences, between-test differences, and the proportion of treated to untreated topics that comprised the term grades.

Grade term means. Table 2 shows the mean across all mathematics students in grade 10. The reason for showing these data is to highlight the third term: all classes improved markedly on this grade. As fate would have it, the experiment classes were graded first and showed a marked improvement which caused some excitement among the teachers and myself, but as the other grades came in we realized that this excitement was naive: “It was just an easier test.” commented Mr. Reed. However, this examination did reveal two findings: one came from comparing standardized class means and the other from an observation by Mr. Reed concerning student responses to a particular problem, *Question 8*, of test two of this examination. Mr. Reed observed that student responses varied more than usual in terms of strategies used.

Table 2

Class means per term for the academic year by class

	Term 1	**Term 2**	**Term 3**	**Term 4**
Reed Control	56	53	**62**	59
Reed Treatment	55	57	**63**	62
Smith Control	41	38	**51**	42
Smith Treatment	49	41	**54**	49

Class performances. This section shows a figure comparing the grade means of the experiment and control classes. Figure 12 shows the overall difference between the experiment and control classes as well as the marked improvement in the third term. This posed a challenge to the study as we had decided to glean the most data for analysis from these third term exams, we were unable to do the same in the fourth quarter as access to the final exams were not part of IRB approval. Figure 12 shows class scores, and the treatment and control groups' improvement during the third term echoes what was found with all other classes; the smaller spread of the third term data also seems consistent with comments of an easier test. Treatment class means remained above the total grade mean. Comparison beyond this – especially through the lens of treatment/control – is complicated by tracking individual topics and themes from test to test; more intricacies are introduced by the weights allocated to individual portfolio and tests that comprised the final grade of a student, or gauging the relative difficulty of a problem. Based on these issues and the data collected in this study, I hesitate to suggest any causal relationships between class term means.

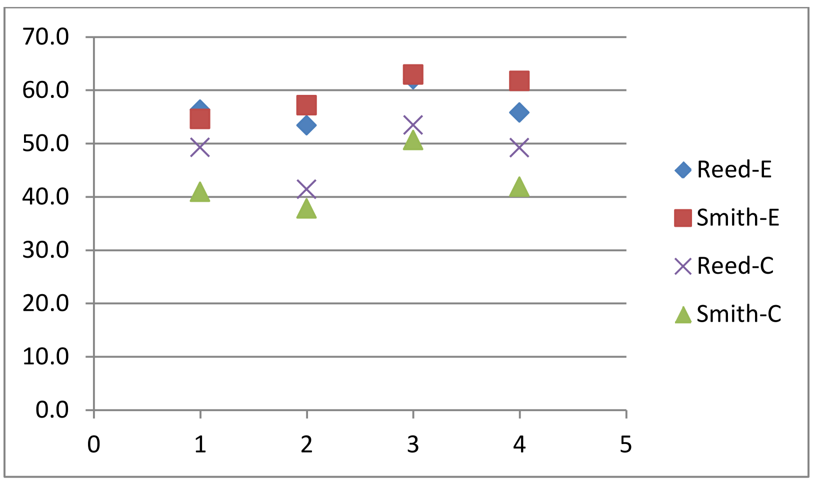


Figure 12: Treatment and control class term means

Correlations. I tested for any correlation between in-class test performance and end-of-term test performance. The motivation for checking correlation stems from an original intention of aiding weaker students: How strongly do these performances correlate? If the treatment mainly aids stronger students, is it worth a performance dip? I correlated the aggregated in-class test performance, as a percentage, with term grades. Figure 13 shows the correlation coefficient for in-class test performance (horizontal axis) versus the first semester mean (term one and term two combined -- vertical axis), where figure 14 shows the same correlation with the second semester mean (term three and term four combined). The respective correlations were r = 0.64 and r = 0.75; the higher second correlation was expected as the in-class test content focused on term three and term four content. Although there is a correlation, the treatment does not seem to aid stronger students exclusively. The first correlation shows also that while there is a connection between how well students do on in-class tests versus exams, it is not definitive.

In addition to the correlations, I categorized students into three groups based on the mean of their term one and two grades; the lower third of this high-medium-low categorization were defined to be weaker students. I used ANOVA to look for any interaction between the three high, medium, and low groups with the treatment and control groups based on third term grades; no interaction was found (p = 0.832). This test was also repeated based on a similar three way classification using the combined average of term three and four grades, and no interaction was found in this case either (p = 0.880).

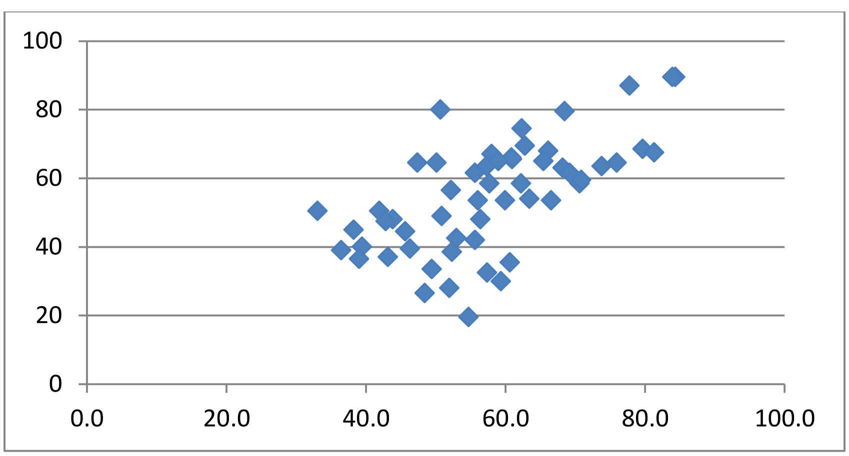


Figure 13. Treatment-Exam correlation for term one and two (r = 0.64).

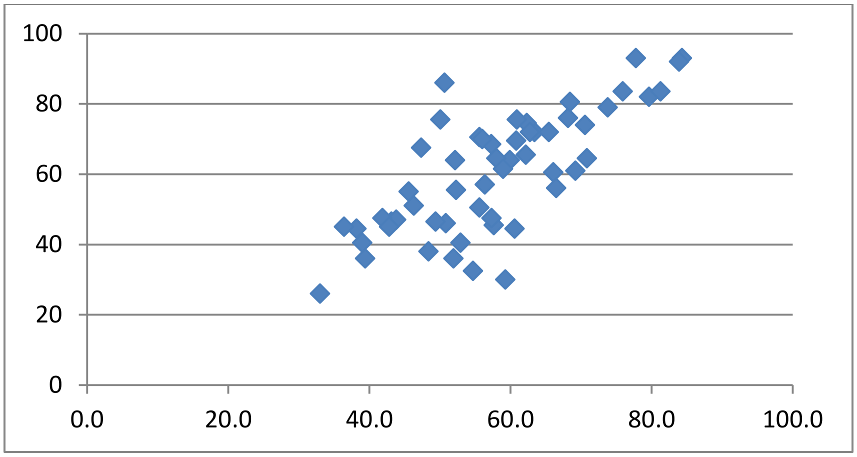


*Figure 14.*Treatment-Exam correlation for term three and four (r = 0.75).

Question 8. During one of the interviews, Mr. Reed referred to what he described as *know-how*:

I: What do you mean by *know-how*?

R: We did previously talk about what I thought was the influence of the test…sometimes you learn A and you get an answer B. But the cases where I taught children on this program I got answers B, C, and D. The children would think, not only algebra, they would apply algebraic ideas to trig ideas and they use algebra formulas (distances and things like that) and they would incorporate it into geometry. I found that quite uhm…what [would] you call that? They could work around a problem in different ways and I was glad to see that, that's the first time I've seen children work around a problem in three different ways… on my testing.

Mr. Reed was able to refer to specific student work concerning this problem, Question 8 (shown in figure 15), and the different student responses are shown in figure

16. The student was given two functions *f* and *g*, and she was asked to determine the length of AB (8.1), the length of CD, given that $\theta = 100°$ (8.2), and to solve an equation (8.3.) The question that elicited the varied responses was 8.2 where the student had to determine vertical distance between these two transformed sine functions. The responses shown in figure 16 have Afrikaans comments made in bold: *Aflees, Afstand f.*, *Substitusie*, and *Combo*; I will briefly discuss each of these. What was interesting to note was that although students were able to apply different strategies, this did not guarantee correct calculation.

Aflees. *Read off*, or*, by inspection*: The student estimated the coordinates of points C and D based on x = 90^o. The estimates were close but not accurate and the student also applied the distance formula incorrectly. The student's final solution was 3.5 units where it is actually 5.5 units. This response shows that the student possibly made use of inspection and also the distance formula.

Afstand f. *Distance formula*. This student also used the function values at x = 90^o, substituting these values into the distance formula.

Substitusie. *Substitution.* This substitution specifically refers to a student substituting the *f* and g functions into her formula "Bo minus onder", or "Top minus bottom". Here the substitution of functions and their evaluations were correct.

Combo: Subst + Afstand f. *Combination: Substitution and distance formula*. Here the student made use of the distance formula and the value 0.98 for x_d and x_c indicate that the student substituted and evaluated function $g(x) = -sin(x)$ at x = 100^o.

Discussion. These responses may indicate two subtle effects of this study: firstly, the students tried to apply procedures from one context to a different context, and secondly, they *attempted* to do this. This might suggest that the frequent, *no-stakes* testing helped students attempt problems without fear of failure or penalty. Mrs. Smith talked about confidence during her interview and made a comment about exams that I felt applicable: "*the child gets used to the tests every day and how to approach a certain problem because he sees it so often … but the exam pressure is something that always gets them…*". Even though the testing might have prompted the students to connect more topics, and to make attempts more confidently, the examination remains a challenge.

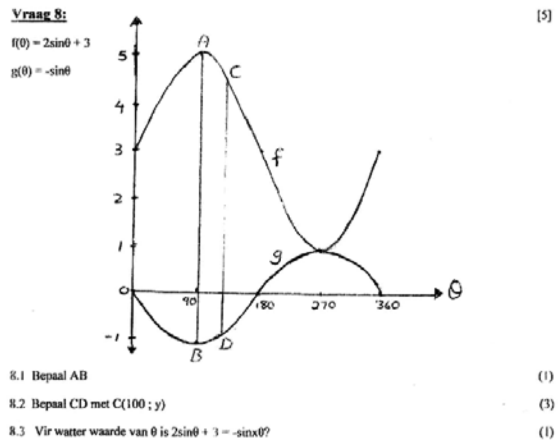


Figure 15: Question 8.

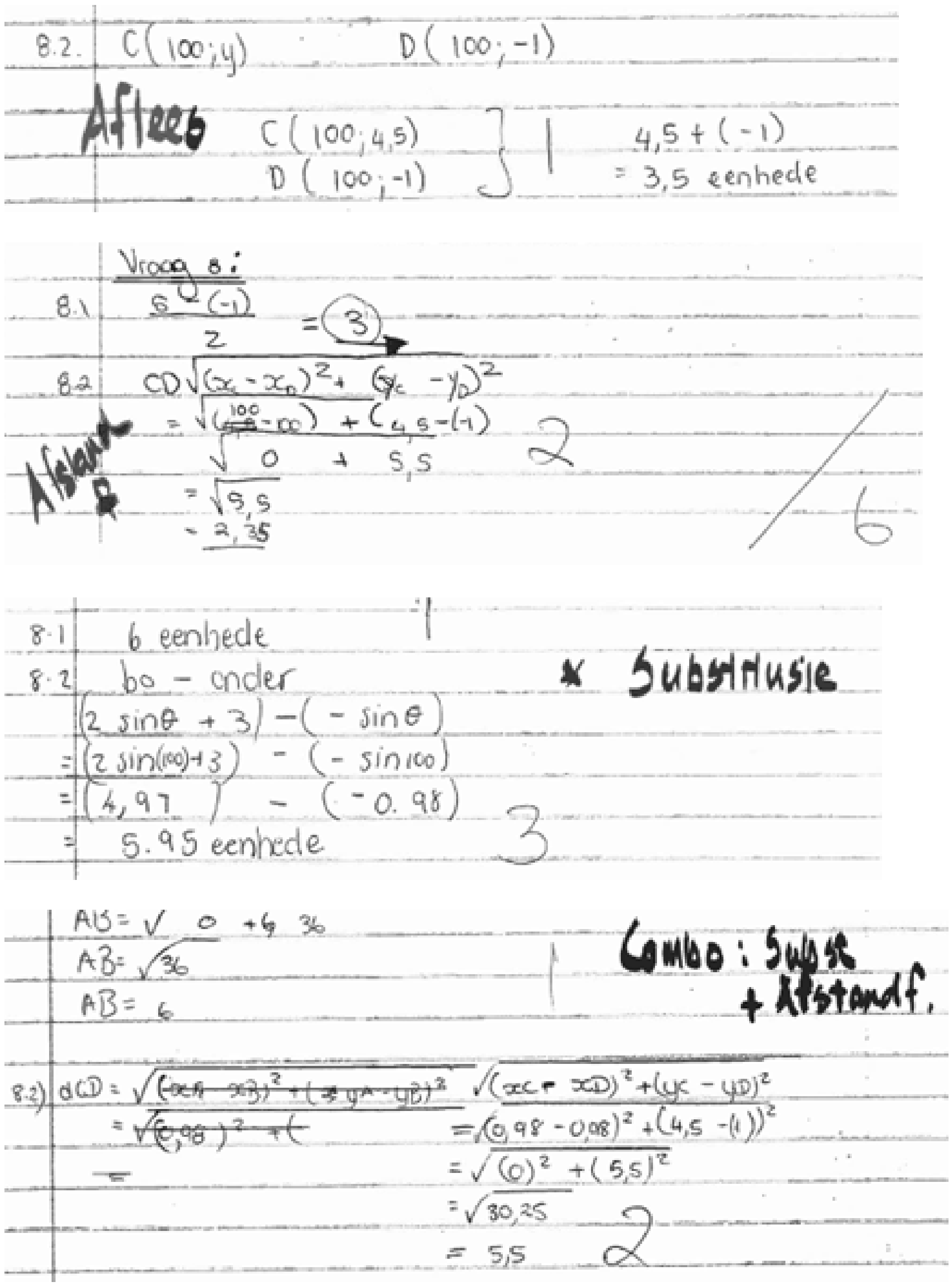
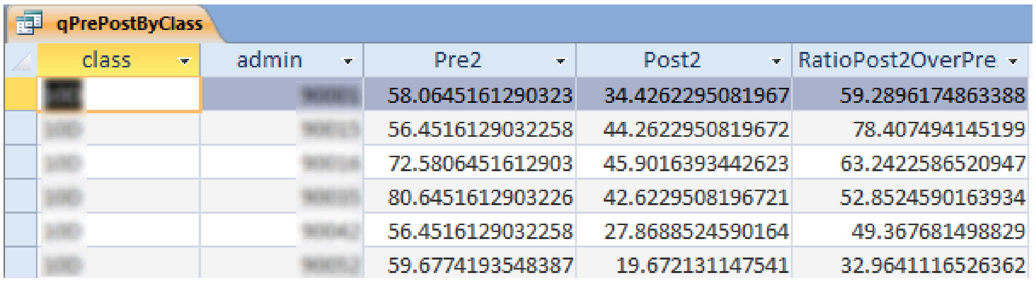


Figure 16. Student responses to question 8.

Summer Learning Loss. The South African summer runs from November through March. The post-test was written on the 18th of February in the new year. School closed for the end of the academic year on December 10 and reopened on January 12. Even

though this holiday is significantly shorter than the American summer, there were about two months between exposure to the content on the pre- and post-tests .
The first and second examination of the third term was used as pre-tests for summer learning loss as it dealt with topics used in the treatment (second exam) and not in the treatment (first exam). Knowing the time constraints on the teachers during the school day, the two exams were shortened significantly. The measure used to select specific problems was the combined means of all treatment and control classes. I selected a spread of questions, focusing mainly on the second examination where students performed well, average, and poorly. In the end, I selected four problems (or parts of these problems) from the first test and eight from the second test.

Unfortunately, the new academic year could not practically afford to give the study two periods required for administering the post-test and we decided to assign the test as homework. The purpose was to see how much students remembered from the test without studying for it again; the test did not count toward their grade. As students were generally positive about the experience the previous year and did not seem to mind the assignment, especially as it had been shortened significantly, the majority of students in each class completed and returned the test. There were, however, several complications that arose: first, the results would not be as reliable as in-class testing and this was confirmed by a few students returning identical responses; also, the tests were given to classes in grade 11 where many students had been moved to different classes and the effect of this was that one of the classes outside the study received this assignment in the stead of one of the control classes. Lastly, due to the reassignment of students in grade 11 only about half of the students from the original control and treatment had ended up taking the post-test. The coordination of tracking down the original treatment and control students was impractical at the time but I was thankful for data that we did receive: there were pre- and post-test scores for individual students on both tests (Figure 17).


qPrePostByClass

class	admin	Pre2	Post2	RatioPost2OverPre
[redacted]	[illegible]	58.0645161290323	34.4262295081967	59.2896174863388
[illegible]	[illegible]	56.4516129032258	44.2622950819672	78.407494145199
[illegible]	[illegible]	72.5806451612903	45.9016393442623	63.2422586520947
[illegible]	[illegible]	80.6451612903226	42.6229508196721	52.8524590163934
[illegible]	[illegible]	56.4516129032258	27.8688524590164	49.367681498829
[illegible]	[illegible]	59.6774193548387	19.672131147541	32.9641116526362

Figure 17 Snapshot of pre- and post-test data.

Table 3

Pre-test and post-test data by class

Class	Pre-Test 1	Post-Test 1	Pre-Test 2	Post-Test 2
Reed-Treatment	60.3	73.3	57.8	36.2
Reed-Control	52.6	78.9	51.3	30.4
Smith-Treatment	61.4	80.4	67.8	45.8
Smith-Control	48.9	--	51.4	--

Table 3 shows the available data for all four classes as well as the two main concerns: we did not have post-test data for Mrs. Smith's control group; and the selection of problems (or parts thereof) from the first examination turned out to be items that did not reflect the overall difficulty of the problem. Hence, what seems like an improvement is misleading as it is unlikely that the students had a summer learning *gain*. The second test for Mr. Reed's classes seemed to be the most likely candidate for comparison and I decided to use post2/pre2 ratio to compare the two classes. The reason for choosing division over subtraction was the substantial difference in class means. Had one control student and one treatment student both achieved half of their previous scores on the post-test, then the proportion is equal but the *difference* in scores of the treatment student would be significantly more than the control student's. A t-test comparison of Reed-C versus Reed-E on post/pre scores yielded $p>0.05$ with $p = 0.057$.

Best case summer learning loss. Assuming that summer learning loss is mitigated by this kind of treatment one might expect the students who performed the best on the in-class test to show some significant improvement. Surprisingly, the top two performing students on the in-class tests each scored about 59% of their pre-test scores on their post-tests, which is almost exactly the control group mean of the ratio mentioned above. Considering these data, it does not appear that the in-class tests had any significant effect on summer learning loss.

Interviews and individual students. This section draws on interview data (Appendix B) and performance data to generate themes and to select specific, individual students from data. Four teachers were selected for the interviews; the two teachers that administered the in-class tests, Mr. Reed and Mrs. Smith, and the two teachers who taught the treatment and control students the following year. The interview data is presented largely verbatim and claims made of specific students are compared to performance data.

During the study. I interviewed Mr. Reed and Mrs. Smith at the end of the study and I was specifically interested in how they felt during the first week of implementing the tests, how they perceived the effects of the study, and whether they had any comments about the study in general. The data below respond to those three topics.

Mrs. Smith.

Implementing the Tests.

"I was initially skeptical and a bit negative because of the time that the tests take every day, but within the first three or four tests the student had gotten used to the routine and I was able to walk around the class and take in the tests as they finished and everything went more smoothly… so time-wise it wasn't a problem anymore: the students had gotten used to it and I had gotten used to it."

Students.

"In the beginning the students didn't really want to do it, they didn't understand why they are doing it so I had to explain it to them a couple of times that they can opt out if they want to, but despite that every single student continued to participate… so everyone did ultimately have a positive attitude… later they enjoyed it."

"Every time [a student] might do a little better, if he had two answers correct then the next time he might get three right and that was actually fun for them."

" You can see that they don't like getting zeroes all the time, so they start making a bit of an effort to try and improve, and yes, their results have definitely improved, I think the child is putting more effort into it …some of them go back and look at the pre-test and others don't: some just make the effort and others don't. You can see it clearly."

"There are certain topics that children struggle with… it helps the less-confident student, who generally won't ask a question but the stronger student would, and now he has the opportunity to sit and listen to something that he also didn't know but didn't have the guts to ask. They do get a little extra feedback because I would, as it doesn't take much time, quickly do an example that everyone can take advantage of."

Comments About the Study

"I think technology would definitely help a lot to have a laptop to enter something like this would make a huge difference, or if it could be coordinated with a SmartBoard it would save a huge amount of time, teachers have so much paperwork and things to do already. But then again you also have the issue of laptops getting stolen."

Mr. Reed. A long-time colleague and class neighbor of Mr. Reed said to me "I think your experiment was actually Mr. Reed!" After instructing for 14 years, mainly making use of his blackboard, Mr. Reed has (since this study) started making use of a laptop, a projector, Geometer's SketchPad software, and a tablet. The items and subsequent data capturing seemed to have sparked a significant interest in technology and software, as well as instructional practices for Mr. Reed. All the quotations in this section are attributed to Mr. Reed.

Implementing the Tests.

"[I feel] quite positive; testing on a regular base did help in the end. I think that it did help me in the end. Other people might feel it is extra work."

Students.

"Testing on a daily base makes a child more realistic about what he knows and what he does not know and I think he learns more that way. If he only gets assessed every six weeks then he will only find out how good he is at the work on that basis. This reminds him: *Listen, I can't do this* and tomorrow he's going to see that item again. I think it makes the child more attentive to the problem."

"I think it helps discipline, they are *switched on* after the little test…would get those who are not working, to work on a more continuous basis."

"I think it just made them a little wiser."

Comments on the study.

"I think with technology one could incorporate [this] within a class… the way we did it took more time than what we initially thought it would take but I think with technology testing on a regular base from a database could be something to look into and I would apply it for sure."

After the study. Once the students moved on to the 11th grade, I interviewed the teachers who were then teaching the treatment and control classes. Those interviews suggested that the items may have had some effect on summer learning loss and classroom dynamics but the most observable influence was instruction: in this case, Mr. Reed's unit on the untreated topic, exponents, seems to have been particularly effective for both of his classes – treatment *and* control – a likely indication of instruction proving more influential than treatments of this nature.

Mrs. Post-Reed

On the Reed-Treatment Group.

"I need to point out here… all three of my classes are from different teachers, on the other hand, [Reed-Treatment is] really more focused… there is a very interesting thing here …their high marks are maybe not as high but the under-achievers… really empowered the rest… all-together their average is much higher than the other two classes, so overall as a whole group did much better than the others… I would say that the experiment had an influence on all of them in a positive way, especially I am talking about the weaker kids, they are amazing. There was a definite, noticeable difference between them and the other classes with exponents, a difficult component."

Mrs. Post-Smith.

On the Smith-Treatment Group.

"I have been teaching 11th grade for many years and I did find [Smith-Treatment] to be a strong class…[Reed-Control] is the weaker class, you need to spoon feed them a bit. [Smith-Treatment] like to be challenged, I saw that with an investigation they had to do; it can be that the frequent testing made them more used to it. I suppose they are a bit stronger, basically, academically, you know? It was as if they're background in exponents was a little weaker, [Reed-Control]'s were definitely stronger in exponents.

"In [Smith-Treatment] we just did analytical geometry and I definitely found some better recollection of the distance formula etcetera. It could be, it could be…especially number patterns (untreated) also. I don't think they differ that much… [Smith-Treatment] do work faster, [and on the first test] I would even say that the weaker ones could have benefitted, my opinion is that they could have benefitted."

Specific Students.

"Specific children that I know from grade 9, I did not teach them in grade 10, I am looking at Adele, for instance, from the Smith-Treatment class: I feel that she is doing better than expected. Karen is another one; I did find that there could be some improvement."

Specific Student Data. Table 4 shows individual term means for Adele, Karen, and Nadia[3]: as percentages and as standardized according to grade mean for all 187 mathematics students in 10th grade. The first two students: Adele and Karen, were selected because Mrs. Post-Smith referred to them in her interview as students who are doing *better than expected.* Nadia was selected because of a conversation I had with a parent that came to thank me personally for implementing the study, adding that her daughter enjoyed participating (she also happened to win one of the raffle prizes) and feels that the study has helped her in terms of performance and confidence.

Table 4

Individual cases: Percentages and standardized by grade mean.

Name	Class	T1	T2	T3	T4	zT1	zT2	zT3	zT4
Adele	S-E	30.0	15.0	40.0	35.0	-1.4	-2.1	-1.3	-1.2
Karen	S-E	30.0	30.0	40.0	35.0	-1.5	-1.1	-1.0	-1.2
Nadia	R-E	55.0	75.0	80.0	80.0	0.0	1.4	1.2	1.3

The standardized scores seem to support the idea of better performance in the third and fourth terms for Karen and Adele, but show that Karen's improved performance might already have occurred during the second, untreated, term. In Nadia's case, the data

[3]Individual in-class test performances (z-score and performance for all in-class items):
Adele: -0.23 deviation, 55/125 items correct
Karen: -0.46 deviation, 52/137 items correct
Nadia: +1.34 deviation, 74/124 items correct

also show a similar case to Karen: Nadia's best performance occurred during the second, untreated, term and continued to do well from there onward.

CHAPTER 5

CONCLUSION

The purpose of this study was to investigate the effects of distributed practice on test performance and summer learning loss in two grade 10 mathematics classes. I hoped that the treatment might significantly affect the performance and summer learning loss of students; also, I was interested in any subtle changes that this treatment may have affected. In this chapter I will briefly discuss results from the study and comment on factors from my design and methodology that may have influenced the outcomes. Finally, I comment on future implications for research as well as current assessment.

In line with previous research, definitive effects of distributed practice remain elusive. This study found that there was no significant effect of distributed practice on exam performance or summer learning loss. However, interview data suggest that students in the treatment classes showed some evidence of strategic competence on specific exam items, meaning that students who received the same instruction were able to approach an exam item – like Question 8 – using different strategies. Furthermore, the interview data from post-study teachers suggested that students from the treatment groups seemed to enjoy being challenged, whereas the control students required more spoon-feeding. In addition, teacher effects seemed more pronounced than treatment effects because of Mr. Reed’s treatment and control classes’ performance on exponents – a theme that was dealt with before the treatment began.

I hesitate to attribute strong causalities to any results in this study largely because of the magnitude of uncontrollable factors present in this naturalistic setting, one of which was the substantial difference in academic performance between the treatment and control classes. I also note that the items themselves were not designed to practice specific procedures but rather to create repeated opportunities or exposures to revisit past content. Furthermore, *Question 8* suggests that mere familiarity with topics, or the ability to employ different strategies, does not contribute significantly to the ability to calculate correctly, especially under examination conditions. Results concerning summer learning loss were disheartening as the control and treatment student data seemed almost randomly scattered despite the large differences in mean performance of treatment and control classes. It may have been a mistake to have the students take the test again in the new year without letting them prepare for it in the same way that they did for the pre-test.

These results, however, do suggest that by using technology to schedule distributed practice of curricular items on a daily basis, students can enhance strategic competence, have more opportunities to discuss or question past work, and possibly improve mathematical confidence. Also, the treatment cost roughly a week's worth of instructional time without significantly impacting academic performance. I believe this does show that it is possible to take time out of the day to focus on stimulating strands of mathematical proficiency without fear of negatively impacting test scores.

The main component that helped make this study successful was the two teachers who committed to the project. Also, clear communication with the principal, mathematics coordinator, students, and parents were pivotal. In future implementations, I would recommend more local involvement in item development from, for example, teachers, tutors, and students. This implementation of distributed practice requires a database be maintained, and six items to be developed every day for one full year. As I developed 274 items during the time of study, I am sure that sharing the workload among two or more people might make it more practical by several orders of magnitude.

Concerning future research, I feel that it was crucial that the tests did not count toward student grades during this study, but I can see this being adapted to having the students pick three out of every five items to count toward a portfolio grade. It should be emphasized that we felt it critical to the study that there were several opportunities for the students to attempt items without fear of penalty, therefore, they were free to try or discuss any strategy or procedure. Lastly, the items on an in-class test could be combined into one, rich problem that draws on the respective topics.

In closing, as the economy continues to call for innovation and creativity while policy demands accountability, I feel that there is a need for more opportunities in which students can move more freely and experiment during their learning. We are afforded these opportunities as never before because of technologies that enable mass-coordination of resources and collaboration. I feel that curriculum needs to adapt to this new environment by finding better ways to measure and target strands of mathematical proficiency that are lost in the noise of current measurement.

REFERENCES

Bakle, B. R. (2010). Summer learning loss: The influence of summer school programs on student achievement in language usage, math, and reading. *Dissertation Abstracts International*: Section A. p71.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, 89–99.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing Distributed Practice. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *56*(4), 236–246.

Cohen, N. J., & Eichenbaum, H. (1995). *Memory, Amnesia, and the Hippocampal System*. MIT Press.

Cooper, H. (2003, May). Summer Learning Loss: The Problem and Some Solutions. ERIC Digest.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology (H. A. Ruger & C. E. Bussenius. New York, NY: Teachers College, Columbia University. (Original work published 1885).

Engelmann, S., Carnine, D., & Steely, D. G. (1991). Making connections in mathematics. *Journal of Learning Disabilities*, *24*(5), 292–303.

Gold, K. M. (2002). *School's in: the history of summer education in American public schools*. New York, NY: P. Lang.

Harden, R. M. (1999). What is a spiral curriculum? *Medical Teacher*, *21*(2), 141–143.

Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition (pre-2011)*, *38*(1), 116–24.

Kilpatrick, J., Swafford, J., Findell, B., & Committee, N. R. C. (U. S.). M. L. S. (2001). *Adding It Up: Helping Children Learn Mathematics*. Washington, DC: National Academies Press.

Litman, L., & Davachi, L. (2008). Distributed learning enhances relational memory consolidation. *Learning & Memory*, *15*(9), 711.

Matsumoto, D. (2009). *The Cambridge Dictionary of Psychology*. Cambridge University Press.

Moore, C. (2010). The Effects of Summer Vacation on Mathematical Knowledge of Rural Students Transitioning from Third to Fourth Grade. *Journal for the Liberal Arts and Sciences*, *14*(2), 58.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*, *14*(2), 187–193.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101.

Rohrer, D., & Taylor, K. (2006). The Effects of Overlearning and Distributed Practice on the Retention of Mathematics Knowledge. *Applied Cognitive Psychology, 20, 1209-1224.*

Rohrer, D., & Pashler, H. (2007). Increasing Retention Without Increasing Study Time. *Current Directions in Psychological Science*, *16*(4), 183–186. doi:10.1111/j.1467-8721.2007.00500.x

Sisti, H. M., Glass, A. L., & Shors, T. J. (2007). Neurogenesis and the spacing effect: Learning over time enhances memory and the survival of new neurons. *Learning & Memory*, *14*(5), 368–375. doi:10.1101/lm.488707

APPENDIX A

SOME PRACTICE TESTS AND IN-CLASS TESTS

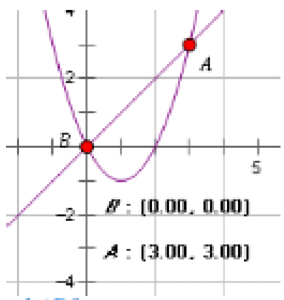
Effects of Expanded Retrieval Practice and Frequent Testing on Test Performance and Summer Learning Loss

Practice test 34

Topics covered:

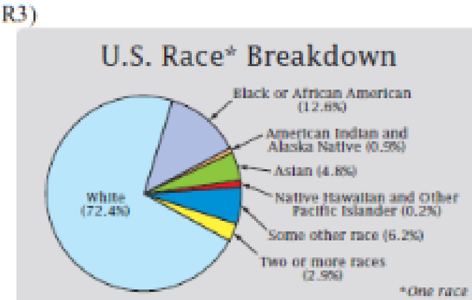
2nd repetition: Day 35:Grafieke: Hersiening I
3rd repetition: Day 33:Stats: Grafieke II
4th repetition: Day 29:Volume I
5th repetition: Day 21:Rente: Saamgesteld I
6th repetition: Trigonometrie: Woordprobleme (cos + tan)
7th repetition: Day :
Time: 3 minutes
Total: 5 points

R2)



Wat is afstand AB?

R3)



As daar 300 miljoen mense in Amerika is, hoeveel van hulle is wit?

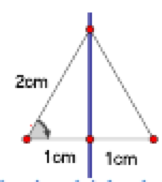
R4) As ek 'n reghoekige prisma vat en al drie dimensies se lengtes verdubbel, hoeveel keer groter sal die volume wees?

R5) Probeer die volgende sonder 'n sakrekenaar

$100(1.01)^2$ **?** $100(1.02)^1$

a.) ≥
b.) ≤
c.) =
d.) Geen een van die bogenoemde

R6)



Gegewe 'n gelykbenige driehoek, hoeveel grade is die gemerkte hoek?

R2) $\sqrt{9+9} = 4.2$eenhede R3)300 000 000 X 0.724 = 217 200 000 R4) 8 keer groter (vol X 2 X 2 X 2)
R5) a R6) 60, dis 'n gelykbenige driehoek (arccos(1/2) sal ook werk)

Case study:

Effects of Expanded Retrieval Practice and Frequent Testing on Test Performance and Summer Learning Loss

In-class test 34

Topics covered:

2nd repetition: Day 35:Grafieke: Hersiening I

3rd repetition: Day 33:Stats: Grafieke II

4th repetition: Day 29:Volume I

5th repetition: Day 21:Rente: Saamgesteld I

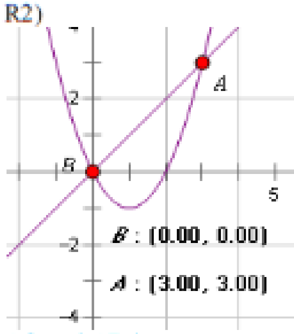
6th repetition: Trigonometrie: Woordprobleme (cos + tan)

7th repetition: Day :

Time: 3 minutes

Total: 5 points

R2)

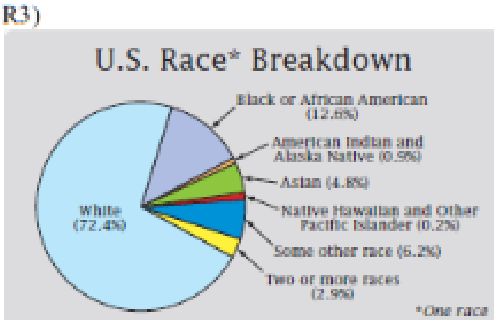


Afstand AB is:

a.) $\sqrt{(3-3)^2+(0-0)^2}$ b.) $\sqrt{(3-0)^2+(3-0)^2}$

c.) $\sqrt{(0+3)^2-(0+3)^2}$ d.) $\sqrt{(0+0)^2-(3+3)^2}$

R3)



As daar 300 miljoen mense in Amerika is, hoeveel van hulle is nie wit nie?

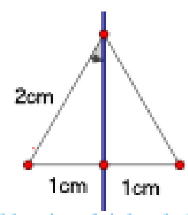
R4) As ek 'n kubus vat met sylengte 1m en al drie dimensies se lengtes verdubbel, wat sal die nuwe volume wees?

R5) Probeer die volgende sonder 'n sakrekenaar

$$100(1.03)^1 \quad ? \quad 100(1.01)^3$$

a.) ≥
b.) ≤
c.) =
d.) Geen een van die bogenoemde

R6)



Gegewe 'n gelykbenige driehoek, hoeveel grade is die gemerkte hoek?

Practice test 37

Topics covered:

2nd repetition: Day 38:Grafieke: Hersiening IV

3rd repetition: Day 36:Grafieke: Hersiening II

4th repetition: Day 32:Stats: Grafieke I

5th repetition: Day 24:Rente: Inoefening

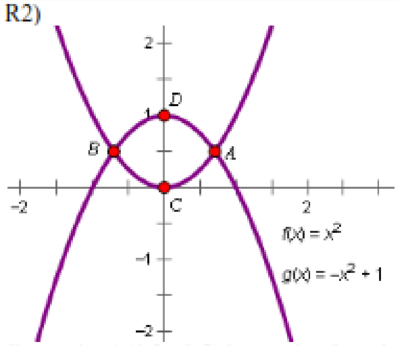
6th repetition: Day 8:Trigonometrie: Vervanging

7th repetition:

Time: 8 minutes

Total: 5 points

R2)



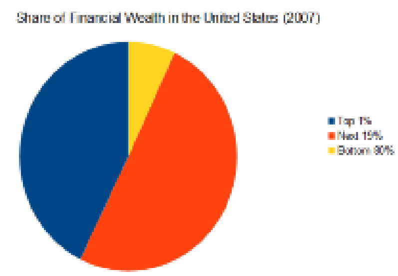
Gegee dat A(1.2 , 0.5) is, wat is afstand AD?

R3) Gegee 'n punt A met koordinate $(\frac{\sqrt{3}}{2}, \frac{1}{2})$. Wat is die afstand vanaf die oorspong tot by punt A?

R4)

Distribution of Wealth in the U.S.

click the images to go to the source with more info



Hierdie grafiek wys hoe die rykdom van Amerika tussen sy mense verdeel is. As Amerika R100 gehad het, **skat** hoeveel van dit sou die onderste 80% van die populasie van dit kry?

R5)

Skat watter belegging is meer werd?

A) (20 000)(0.05)(2)

B) (10 000)(0.1)(2)

C) Hulle is dieselfde werd

R6) As x = 54° 26', bepaal 2cos(x)

R2) Pythagoras. 1.3 R3)Presies 1, dis 'n punt op die eenheid sirkel R4) Jy wil skat wat hoeveel die klein sektor van die hele sirkel opneem – ek dink dis omtrent 'n twaalfde (een uur op 'n horlosie gesig) so dit beteken so R8 van die R100

In-class test 37

Topics covered:

2nd repetition: Day 38:Grafieke: Hersiening IV

3rd repetition: Day 36:Grafieke: Hersiening II

4th repetition: Day 32:Stats: Grafieke I

5th repetition: Day 24:Rente: Inoefening

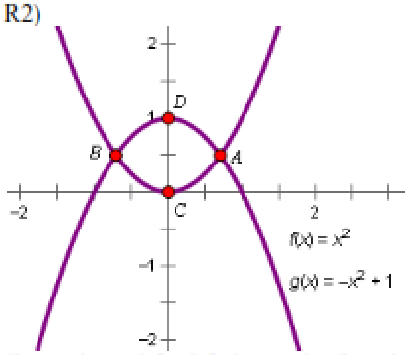
6th repetition: Day 8:Trigonometrie: Vervanging

7th repetition: Day :

Time: 8 minutes

Total: 5 points

R2)



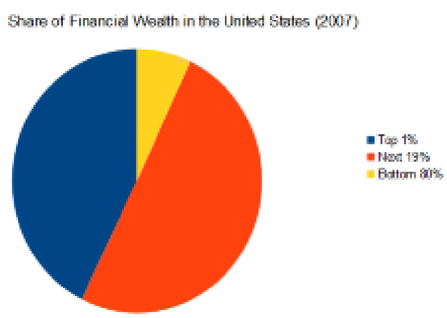
Gegee dat A(1.2 , 0.5) is, wat is afstand AC?

R3) Gegee 'n punt A met koordinate $(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Wat is die afstand vanaf die oorspong tot by punt A?

R4)

Distribution of Wealth in the U.S.

click the images to go to the source with more info



Hierdie grafiek wys hoe die rykdom van Amerika tussen sy mense verdeel is. As Amerika R100 gehad het, **skat** hoeveel van dit sou die volgende 19% van die populasie van dit kry?

R5)

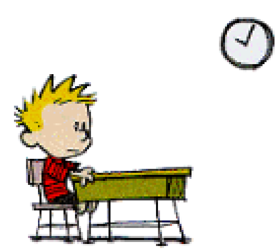
Skat watter belegging is meer werd?

A) (10 000)(0.01)(20)

B) (10 000)(0.1)(2)

C) Hulle is dieselfde werd

R6) As x = 26° 54', bepaal 2tan(x)



"Continuous effort, not strength or intelligence, is the key to unlocking our potential." Winston Churchill

APPENDIX B

INTERVIEW NOTES

Interview guide :	**Notes**
Overall, how do you feel about this project?	
What kind of reactions do you think other teachers might have to this?	
Would you say that this project had an effect on how you view technology in the classroom?	
Effects on summer learning loss?	
How do you think the students felt about the testing? How did it affect them? Or not?	
Are you still continuing this the project this year? How are you implementing it this year?	
Anything else?	
Interviewer notes:	