Julie, C. et al. (2016). The effect size of an intervention focusing on the use of previous national senior certificate mathematics examination papers.
Journal of Educational Studies, 15(1): 1 – 22
http://hdl.handle.net/10520/EJC199584

The effect size of an intervention focusing on the use of previous national senior certificate mathematics examination papers

Cyril Julie, Monde Mbekwa and Marius Simon

Abstract

This study reports on an intervention that emanated from a concern a mathematics teacher had about the unsatisfactory performance of grade 12 learners in the school-based mid-year examination. The intervention was based on distributed practice and the effect size of the intervention was determined as an indicator of the effectiveness of the intervention. Different effect sizes are reported and the reasons for their acceptance or non-acceptance are presented. Overall the results indicate that if all the different effect sizes are taken into account, the intervention had a medium to high effect. Given that the intervention was driven by aspects of teaching and was done at school, it is recommended that more attention be accorded to those aspects of teaching that enhance achievement in Mathematics.

Introduction

Interventions for the improvement of achievement in the National Senior Certificate Mathematics examination abound. Rarely are the interventions designed with inputs from teachers who teach Mathematics in grades 10 to 12. In addition the interventions are generally not school-based and normally target learners who are identified as having the potential to be successful. Rarer is the reporting on the effectiveness of the interventions other than descriptive percentage data on the success of learner participants. Furthermore, if effect sizes are reported, the decisions involved for the adoption of a particular effect size indicator are seldom made explicit. This study reports on the use of previous examination with a particular strategy, spiral revision, and uses effect sizes to determine whether the implementation of the strategy was significant.

The use of previous examination papers for preparation for high-stakes examinations such as the National Senior Certificate (NSC) examination is a well-grounded practice in schools. In fact the Department of Basic Education (DBE) in South Africa makes previous question papers available and encourages teachers and learners to use them for revision purposes. It states on its website:

Old examination papers are a great way to revise and prepare for upcoming NSC examinations. This way you can find out what you already know and what you don't know.

They also help you manage your time better and be familiar with the terminology and vocabulary used in the actual exam. (www.education.gov.za).

With respect to Mathematics in grade 12, the use of previous examination papers is implemented in various ways. Schools normally complete the curriculum by the third term. In fact it is suggested in the Curriculum, Assessment and Policy Statement (CAPS) (DBE, 2011) for grade 12 Mathematics that the 3 weeks comprising the allocated class teaching time for the fourth term be devoted to revision. These 3 weeks are normally devoted to the use of previous examination papers.

Observations and informal discussions with teachers rendered that at least three different ways of use of previous examination papers for revision and preparation of learners for the NSC Mathematics examination are employed. One is that learners work through an entire paper over a few periods with teachers giving hints and assistance as they monitor and assess the learners' ways of dealing with the problems. This is normally accompanied by learners having the memorandum of marking of the paper at hand and they can check their responses by referring to this document. Another method being used is that learners work through some relevant questions of a particular topic. These questions are selected by teachers from questions of previous examinations on the selected topic. After this period of engagement, normally a week, with questions related to the topic, the learners are given a test on the curriculum topic with the questions selected from previous examinations but which were not part of the practice set of previous examination questions learners were engaged with. The third way is that teachers use their knowledge of learners' performance on particular topics in particularly the mock, externally- set September examination to select particular problems from previous examinations in which the performance of the learners fall below a particular threshold. Normally, the teacher will, in an expository manner, work through one of the questions selected from a previous examination and then allow learners to work individually or in groups through some problems, selected by the teacher, on the same topic from other previous examinations. Characteristic of the last two approaches is that after learners have engaged with the topics, these topics are only dealt with by the learners during their self-study sessions during which they normally do what is done in the first approach. The first approach can be viewed as a distributed approach to using previous examination papers since learners engage with a topic more than once in a concentrated manner. The last two approaches can be deemed as a massed approach in the sense that learners practice, also in a concentrated manner, a topic once by doing problems only related to that topic. This study reports on the achievement outcomes when a distributed approach to revision is used as an intervention to address unsatisfactory achievement in the NSC Mathematics examination. The next section discusses the two constructs, massed and distributed practice.

Massed and distributed practice

Massed practice is concentrated doing of activities of a topic after the topic has been taught. The practice activities focus on what has been taught immediately and the next engagement the learners will have with the topic will be during revision sessions normally before tests or assessments. Mathematics textbooks are normally structured to facilitate mass practice. So, for example, when dealing with the simplification of fractions with different denominators there will be a few worked out examples followed by a set of exercises dealing with the simplification of fractions with different denominators. Teachers normally follow this organisation of the textbook for their teaching. The next time practice on the simplification of fractions with different denominators will be practiced will either be during the cumulative review of the section dealing with fractions or, at the discretion of the teacher, before a test or examination. Saxon (1982, p. 484) depicts this kind of textbook structure (and by implication teaching) as an organisation "in discrete units or chapters [which] present all facets of a particular concept in a single chapter." In a similar sense Rohrer & Taylor (2007, p. 482) describe mass practice as that within which "most or all of the problems relating to a given lesson are concentrated or massed into the immediately following practice set…" A criticism of mass practice is that the retention of taught and learned work is compromised. In colloquial parlance this is articulated by learners after taking a test as "when we did the work in class I could do it. In the exam I just went blank." In a recent discussion with teachers about the achievement of their learners in a Mathematics examination, one teacher articulated the non-retention phenomenon as, "When I was teaching the learners could do the work. But when I marked their exam scripts, I was wondering whether I was teaching them at all!"

In order to address the problem of retention, different teaching approaches and organisation of textbooks evolved. A common attribute of these approaches and textbook organisations is that the presentation, practice of taught mathematical constructs and ideas and formative assessment of these constructs is spread over time and not as a once-off occurrence. A variety of terms emerged to describe these approaches. Some of these terms are "distributed practice" (Seabrook, Brown, & Solity, 2005, Smith & Rothkopf, 1984, Johnson & Smith, 1987), "snappies" (Cramp & Nardi, 2000), "incremental approach" (Saxon, 1982, Klingele & Reed, 1984), "shuffling of mathematics problems" (Rohrer & Taylor, 2007), "spiral testing" (Wineland & Stephens, 1995), "deliberate practice" (Ericsson, Krampe & Tesch-Romer, 1993) and the "spacing effect" (Dempster, 1988). Such approaches are deemed to "reinforce previous learning and encourage retention of material" (Wineland and Stephens, 1995, p. 228). The core of these approaches is that [Of the] problems contained within each problem set, only a few deal with the most recently presented topic; the remaining problems are review problems of previously learned material. The frequency of exposure to examples specific to the original types is never completely withdrawn. It is [the] intent to provide, within each problem set, elements of all previously introduced topics, either through direct example problems, or by incorporating a number of previously learned functions within a more complex problem. *(Johnson & Smith, 1987, p. 98)*

Of the studies referred to above all but one (Johnson & Smith, 1987) report that distributed approaches improved achievement compared to massed approaches. Hattie (2009) meta-analysed the effectiveness of 2 meta-analyses dealing with 63 studies and found that distributed practice approaches was highly effective compared to massed approaches for enhancing achievement.

Distributed practice is closely linked to mastery learning. Mastery learning was a central component Escalante's teaching approach to an Advanced Placement calculus course which caused much discussion and debate in the early 1980s. None of the Latino students in the school previously attained success in the course but as a result of Escalante's teaching programme, 14 students scored so well in the placement test that eyebrows were raised implying cheating with the examination board being suspicious of the success of the students. Twelve students sat for a second round of the examination and repeated their success. Escalante's teaching and ways of dealing with students are captured in the film "Stand and Deliver." Escalante's programme demanded "practice, practice, and more practice is…from each student" (Escalante & Dirmann, 1990, p. 411).

The purpose and aim of this article is to discuss an intervention in the form of a revision programme using past examination papers where a particular variant of a distributed practice approach was used. The article discusses the effect of this intervention in a grade 12 class.

The Study

The study originated when one of the schools participating in the Local Evidence-Driven Improvement of Mathematics Teaching and Learning Initiative (LEDIMTALI) expressed concern about the performance of their Grade 12 learners' unsatisfactory performance in the June examination. The overall aim of the project is to improve the quality of teaching of Mathematics in secondary schools. It is underpinned by the notion that the development of teaching can substantively contribute towards increasing of the number of learners offering Mathematics, the number who passes Mathematics in the National Senior Certificate Mathematics examination and the quality of the passes achieved. In the project, teachers, Mathematics Education specialists, Mathematicians, curriculum studies specialists and curriculum advisors from the Western Cape Education Department (WCED) work collectively to develop strategies to address the unsatisfactory mathematics achievement of learners in the NSC Mathematics in schools in socially and economically depressed environments in Cape Peninsula in the Western Cape Province.

Discussions with the teacher of the school indicated that some learners in low achieving bands have given up focusing on Mathematics and decided to rather concentrate their energies on their other subjects to achieve a satisfactory pass in the overall NSC examination.

As is the case with all projects of this nature, participation is voluntary. Since the inception of the project teachers participate in continuing professional development dealing with the development of teaching. These activities comprise 2 after-school workshops of 2 hours and 1 residential weekend institute (±16 hours) per quarter for the first three quarters of the school year. Amongst the activities addressed during the workshops and institutes is the discussion of dilemmas teachers experience in the teaching of mathematics and the development of practically implementable strategies to address the articulated dilemmas.

One of the concerns teachers had was that learners did not do homework. Of the various purposes homework serve the consolidation and practising of work that was completed, was deemed as a major factor contributing to learners' low achievement in examinations. Early in the project an entire workshop was dedicated to the seeking of strategies to address the issue of consolidation of work covered. Since the majority of learners, for various reasons, did not do this after school, a strategy of doing regular revision by them during normal mathematics classes was developed. This strategy was labelled "spiral revision".Julie (2013, p. 93) describes "spiral revision" as follows Spiral revision is the repeated practising of work previously covered. It is underpinned by the notion that through repeated practice learners will develop familiarity with solution strategies of mathematical problem types that they will come across in high-stakes examinations. Productive practising has to do with allowing learners to develop general ways of working in school mathematics through "deepening thinking"-like problems whilst practising. An example of such a problems is "Factorise $ak - (k + a) + a^2$ in more than one way." The procedure suggested for implementing and sustaining spiral revision and productive practising is that 2 or 3 problems on work previously done are presented to learners at the start of a period. This should preferably not take more than 10 minutes of the time allocated for the lesson period.

"Spiral revision" is thus a variant of distributed practice as discussed above. The term "spiral revision" was coined by the project participants. It does not necessarily deal with work which was completed immediately before the lesson at hand. At workshops subsequent to the development of the "spiral revision" strategy, teachers reported that they used the strategy but its appropriation was differential and not as regularly implemented due to "the curriculum being overloaded and we must cover it". However, there were always positive reports about learners' views about the use of the strategy. As one teacher reported at a workshop "The learners actually liked it. One learner asked 'Miss, can't we use this more because I forgot the work we did in the first term.'" This was after the teacher did a "spiral revision" activity in the second term on work that was done in the first term.

As referred to in the afore-mentioned paragraph, the "spiral revision" strategy is deemed by teachers as a viable strategy to address a concern they had but its implementation is as yet, not cemented in their daily teaching practice. With this insight, it was decided to use "spiral revision" in a regular and targeted manner with selected topics for grade 12 from the start of the third term. Given that the teacher still had to complete the curriculum which for grade 12 must be completed during the 3rd school term, an additional period was sought to be dedicated to "spiral" revision. Negotiations with the principal for this additional period resulted in him agreeing that the school's assembly period be used. Learners who achieved in the band 15% to 45% in the mid-year examination were particularly invited to attend the "spiral" revision period. This period was conducted by a fieldworker and the subject teacher. The fieldworker selected problems at levels 1 and 2 of cognitive demand from past examinations, exemplar grade 12 examinations from the Department of Basic Education and exemplar examinations that appeared in the media. The topics focused on were algebraic manipulative quadratic theory (graphs of quadratic functions and nature of roots excluded), straight line analytic geometry (circles excluded) and manipulative trigonometry with specific or general unknown given angles (trigonometric graphs, solution of trigonometric equations and distance and height problems excluded). The general procedure was that learners would be presented with a worksheet with two to three problems on these topics. If they experienced difficulties, the underlying concepts would be explained and learners had to practice problems related to the concepts until the fieldworker and the teacher were satisfied that the majority of learners have gained sufficient mastery in the solution of the problems. Only after they were assured that such mastery was attained would they move on to another mathematical idea. In order to maintain momentum, the teacher would take some of her normal teaching time to further consolidate learners' fluency of dealing with examination-like problems. An example of a revision card is given below.

Grade 12 Spiral Revision Exercises Card 1

Solve for x

1. $3x(x-2)=4$
2. $3x^2-5x=2$
3. $(2x+3)(3-x)>4$

4. ***Solve Simultaneouslyfor x and y.***

$2+y=-2x$

$-2x^2+8xy+42=y$

Figure 1: Example of a spiral revision activity

Approximately 14 hours of tutorial sessions, including a 6-hour Saturday session, were used for this intervention.

The research interest was in whether this intervention would result in a positive increase in achievement between the June and September

Research Approach

Interventions in school mathematics education are seen as important in facilitating improvement in learner performance globally. In order to assess the viability of such interventions, it is important to test whether a particular intervention achieves what it sets out to achieve. An important statistical measure utilized to ascertain whether an intervention has been effective or not is that of effect size. Ferguson (2009, p.532) states that "Effect size estimates the magnitude or effect or association between two or more variables." The interest in this study is whether the use of revision based on questions in the final mathematics examination question papers using the "spiral revision" approach will have an effect on the level achievement of learners in mathematics examinations. The specific research question pursued was, “Is there a difference in achievement scores between the June and September scores on selected mathematical topics after an intervention driven by “spiral revision” was implemented?” A quasi- experimental design was selected since the learners were not randomly selected. The sample was therefore an opportunistic one. As stated above the selected mathematical topics were algebraic manipulative quadratic theory (graphs of quadratic functions and nature of roots excluded), straight line analytic geometry (circles excluded) and manipulative trigonometry with specific or general unknown given angles (trigonometric graphs, solution of trigonometric equations and distance and height problems excluded). As is customary, the examinations consisted of two papers. Algebraic manipulative quadratic theory was part of the first paper. Straight line analytic geometry and manipulative trigonometry were part of the second paper. The data were the combined scores for the questions for the topics the learners obtained in the June and September examinations, with the June and September scores the pre- and post-treatment scores respectively.

The June examination was set by the Mathematics teacher of the school. As is common practice, these examinations closely follow the patterns of previous examinations and any exemplar examinations provided by the DBE. The September examination was the mock, preparatory examination for the upcoming 2014 examination which was externally set by the WCED. An example from the equations, inequalities and algebraic manipulation question in the WCED-set September examination is given below.

Question 1

1.1 Solve for x (to two decimal places if necessary)

1.1.1 $(x-1)(2x+5)=0$ (2)

1.1.2 $\frac{1}{2}x^2+3x-10=0$ (4)

1.1.3 $x^2 \geq 20+x$ (4)

1.1.4 $96=3x^{\frac{5}{4}}$ (3)

1.2 Solve for x and y simultaneously:

$2^{x+y}=4$

$x^2=52-y^2$ (7)

1.3 Calculate, without using a calculator, the value of a and b if a and b are integers and:

$\frac{14}{\sqrt{63}-\sqrt{28}}=a\sqrt{b}$ (4)

Figure 2: September 2014 examination question for the Paper 1 topic

Both examinations were marked by the teacher and these marks were accepted. The marking can be assumed to be reliable since it was the same teacher who marked both examinations and teachers have their own particular habits of marking assignments according to their interpretation of the marking memorandum.

The marks obtained by the individual learners for the sections of import for this article were transferred to an Excel sheet to make it amenable for treatment by SPSS 22.

The marks were inspected for possible anomalies. It was found that some learners did not write both examination papers. These learners were excluded and rendered 42 learners who wrote both examination papers.

Data Analysis

For the determination of the effect size of an intervention various indicators can be used. The choice for selecting a particular route is dependent on the data fulfilling specific requirements.

Hence data need to be examined prior to selecting a particular route to follow. For the parametric route a crucial requirement is that the data fulfil the criterion of normality. Real data are generally skewed and leptokurtic.

The skewness and kurtosis of data can be determined. If they are within pre-defined ranges then the data can be considered as approximating normality and parametric statistical procedures can be applied to the data. If the data are outside these recommended ranges the suggestion is that non-parametric statistical procedures be used to arrive at effect sizes.

In order to decide whether to go the parametric or non-parametric route, the general appearance of the histogram of the data with an overlay of the normal curve, tests linked to skewness and kurtosis can be conducted to arrive at the decision.

The histograms with associated normal curves for the data sets of this study are given the Figures 3 and 4 below.

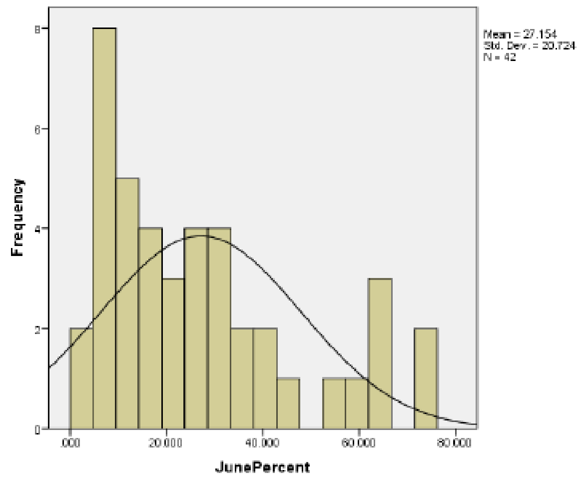

Figure 3: Histogram with normal curve for June Percentage scores

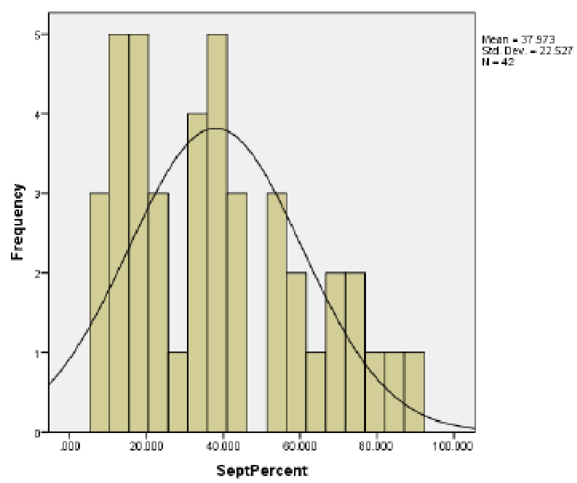

Figure 4: Histogram with normal curve for September Percentage scores

An inspection of the histograms indicates that both the June and September scores are right-skewed.

The Kolmogorov-Smirnov and Shapiro-Wilk tests are recommended tests to ascertain whether a data sample is normal. The latter test is suggested for sample sizes less than 2000. It tests the following null (H_0) and alternate (H_A) hypotheses:
H_0: the data is normally distributed
H_A: the data is not normally distributed
Table 1 provides the descriptives of the data under scrutiny and the positive values for the skewness of the data confirm that the data is skewed to the right.

Table 1: Descriptive statistics for the June and September data sets

	June Percent Score		September Percent Score
Mean	27.15420		37.97314
95% Confidence Interval for Mean	Lower Bound	20.69623	30.95338
	Upper Bound	33.61216	44.99290
5% Trimmed Mean	25.95112		37.07774
Median	21.42857		34.61538
Variance	429.473		507.445
Std. Deviation	20.723725		22.526534
Minimum	2.381		7.692
Maximum	73.810		87.179
Range	71.429		79.487
Interquartile Range	27.381		33.974
Skewness	.910		.593
Kurtosis	-.152		-.686

Table 2: Results of the normality tests

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Jun Total	.139	42	.039	.894	42	.001
Sept Total	.127	42	.088	.934	42	.017

a. Lilliefors Significance Correction

The statistic, normally indicated by D, for the Shapiro-Wilk test is the reported value and the significance (sig.), called the P-value, is the significance level for the calculated D. For the null hypothesis to be accepted the P-value must be greater than the predetermined significance level, which in this case is .05. The P-values for both June (.001) and September (.017) are less than the predetermined significance level and hence the null hypothesis is rejected. This implies that the scores are not normally distributed.

The most commonly used (see, for example Kelley, 2005) effect size indicator for normally distributed data is Cohen's (1988) d, given as:

$$Effect\ size = d = \frac{Mean_{end\ of\ tratment} - Mean_{start\ of\ treatment}}{Pooled\ sample\ standard\ deviation}.$$

For the data under consideration Cohen's d = 0.52 which would have been the effect size if the data satisfied the normality criterion. Many analysts use Cohen's d as an indicator even if the data under consideration do not comply with the normality criterion (Kang, Harring & Li, 2015). This is particularly the case when a sample is the entire population as is the case with an opportunity sample of all learners in a school who sat for a particular examination. A possible reason offered for this is that the outcome of the analysis is of importance to the particular teachers involved in the intervention to use the outcome as feedback on the effectiveness of an employed teaching strategy to reach their goals of enhancing achievement (Hattie, 2012).

McGraw & Wong (1992) suggest that effect sizes should be reported in a more understandable format than that found by the aforementioned formula. They developed an effect size formula called the Common Language (CL) effect size indicator. CL converts an "effect into a probability" (McGraw & Wong, 1992, p. 361) using the difference of the means and a standard deviation found from the variances of the two samples. The difference of the means is converted to the standardized score with 0 as the mean. This standardized score is given by: =

$$\frac{0-difference\ between\ the\ means\ of\ the\ scores\ before\ and\ after\ treatment}{\sqrt{variance\ of\ scores\ before\ treatment+variance\ of\ scores\ after\ treatment}}.$$

CL is the probability associated with z as the standardized score for a normal distribution. Thus if $z = a$ and the associated probability for this standardised value is b then CL is the probability converted to a percentage, $100b\%$ in this case, and is interpreted as a $100b$ out of 100 treatments the requisite effect will be obtained. The CL effect size indicator also requires that the requirement of normality be fulfilled. McGraw & Wong (1992) tested the degree of violation if data are not normally distributed. They concluded "that violations of the normality assumption alone do not severely comprise the practice of using the unit normal curve to estimate CL" [and] "the absolute maximum error was never greater than about .10." (pp. 364, 365). Thus, although using parametric tests for the determining the effect size is seemingly not a viable route to follow, at least the CL effect size indicator does not lead to sizable under- or over-estimations.

For violations of the normality criterion, the use of non-parametric procedures is suggested with the Spearman rho correlation coefficient, rs, as the recommended statistic for the effect size. Furthermore, the non-parametric procedure recommended for testing significance is the Wilcoxon signed rank test. As for the Spearman rho coefficient the calculations are based on the ranks of the data.

Kerby (2014) offers a difference formula for the effect size of paired ranked data. It is essentially the difference between the sums of the positive (favourable) and negative (unfavourable) ranks for the differences of the ranks of the two sets of data.

The effect size is the difference between the percentage of the sum of favourable (positive) and unfavourable (negative) ranks. Kerby (2014) argues that reporting the difference in these proportions is a viable way to represent the "common language effect size" indicator proposed by McGraw & Wong (1992). The results emanating from the application of the four procedures—Wilcoxon signed rank test for significance, the Spearman rho coefficient, the differences in proportions of favourable and unfavourable rankings and the CL effect size indicator —are presented in the next section.

The above discussion dealt with various effect size indicators, the conditions under which they can be used and an argument for selection the particular indicators to present the results. The Wilcoxon signed rank test for significance and the Spearman rho coefficient are further discussed in the section that follows.

Results and Discussion

Figure 3 and its accompanying frequency distribution confirm the concern of the school. Only 32% of the learners obtained a pass level of 30% and higher for the topics focussed on in the June examination. This percentage increased to 60% after the intervention as indicated by figure 4 and its accompanying frequency distribution.

The mean percentage score for the June examination was 27.15 and it increased to 37.97 for the September examination.

The result from the application of the Wilcoxon signed rank test is presented in Table 3.

It indicates that the difference between the medians of the two scores is significant. The implication of this is that the intervention contributed significantly towards to the improvement in scores in the September examination.

Table 3: Results from the Wilcoxon Signed Rank Test

Null Hypothesis	Test	Sig.	Decision
The median difference between JunePercent and SeptPercent equals 0.	Related Samples Wilcoxon Signed Rank Test	.000	Reject the null hypothesis

Asymptotic significances are displayed. The significance level is .05. The interest of this research is whether the intervention was effective. As indicated by many researchers, significance tests do not provide information about the effect size of an intervention. Olejnik & Algina (2000, p. 241) cautions that "Statistical significance testing does not imply meaningfulness". Coe (2002) suggests that statistical significance does not convey the size of an effect and hence the suggestion that some effect size indicator accompanies the significant consideration.Table 4 gives the Spearman's rho, rs, correlation coefficient of .649 and it indicates that the association of the ranks of June and September scores is high (Hopkins, 1997) at the .01 level of significance. Keeping in mind that the significance of the Wilcoxon Signed Rank Test reported in Table 3, it can be stated the high association resulted from the intervention.

Table 4: Spearman Rho correlation coefficient between June and September percentage scores

			June Percent	Sept Percent
Spear-man's rho	JunePercent	Correlation Coefficient	1.000	.649**
		Sig. (2-tailed)	.	.000
		N	42	42
	SeptPercent	Correlation Coefficient	.649**	1.000
		Sig. (2-tailed)	.000	.
		N	42	42
** Correlation is significant at the 0.01 level (2-tailed).				

The Spearman rho, r_S, correlation coefficient is obtained from ranks and not the actual scores. The effect size is reported as the "percentage variation accounted for". Since the rankings and not the actual scores are use the effect can only be viewed as the "variation accounted for" in the rankings. This percentage of variance explained is obtained by squaring the Spearmen rho, r_S, coefficient.

Table 5 presents the outcome of the sum of ranks generated by SPSS 22. The percentage of the sum of positive ranks and the sum of negative is 83% and 17% respectively.

Table 5: Sum of ranks

Ranks				
		N	Mean Rank	Sum of Ranks
SeptPercent - JunePercent	Negative Ranks	10[a]	15.40	154.00
	Positive Ranks	32[b]	23.41	749.00
	Total	42		903.00
a. SeptPercent < JunePercent b. SeptPercent > JunePercent				

Table 6 summarises the different effect sizes obtained from the application of the three effect size procedures deemed appropriate for non-normalised data.

Table 6: Different effect sizes

Effect size procedure	Percentage variance explained (r_s^2)	Difference in sum of ranks	CL effect size indicator
Effect size	42% ($r_s = 0.65$)	66%	64% ($z = -0.35$)

With the caveat alluded to above about to which aspect the variance refers in mind, 42% (r_s^2) of the variance between the June and September ranks is accounted for by the intervention. House, Spangler & Woycke (1991) argue that anything from 20% to 66% of variance explained indicates a strong effect. However, Ableson (1985) cautions against summarily dismissing small effect sizes in terms of variance explained. He contends "that percent variance explanation is a misleading index of the influence of systematic factors in cases where there are processes by which individually tiny influences cumulate to produce meaningful outcomes." (p. 129). For the case at hand the influence cannot be considered "tiny" and hence the variance explanation signals a reasonable highly positive effect size. The "difference in sum of ranks" indicator is customarily interpreted in terms of a weighing scale and the direction in which the scale is tipped.

In the case under scrutiny the scale is tipped by 66% in favour of the positive sum of ranks. The CL effect size indicator denotes that in 64 out of a 100 implementations of interventions of the nature described in this study the achievement scores will increase. In addition, if the effect size indicated by Cohen's d = 0.52 is taken into consideration, then it falls within Hattie's (1992) benchmark of 0.40 for judging the effect of an innovation. Furthermore, Hattie's (2012) recommendation is that Cohen's d be used to ascertain the impact of teaching a class. Hence the effect sizes indicate that the impact of the intervention based on spiral revision can be deemed to be in the medium to high range. At another level the principal, teacher and fieldworker reported that they observed that the learners' motivational levels for Mathematics had increased. The fieldworker exemplified this by stating that a learner who said "I am not going to focus on Mathematics and put my energies in my other subjects" saw it fit to change this view to "I am beginning to see that I can make it in Mathematics." These are indicative of the intervention addressing affective issues related to school mathematics.

Conclusion

The success or not of various interventions to improve achievement in the high-stakes NSC Mathematics examination is receiving much attention in the country (see for example, Padayachee, Boshoff, Olivier & Harding, 2011). The reported successes of such interventions are in many instances used to offer proposals for curriculum implementation policy. Much of the interventions are premised on an apparent weak subject content knowledge base of teachers (Taylor, Van der Berg & Mabogoane, 2013) and hence significant attention is accorded to interventions which focus on enhancing teachers' subject content knowledge.

The connection of this apparent enhanced content knowledge to improved achievement scores in the high-stakes NSC is rarely addressed. Other interventions report on achievement enhancement but these are done in summative manner and normally reported in percentage terms. This makes it hard to fathom what the effect of the reported improvement is. What, however, runs through reports of this nature, is that the interventions rarely address teaching in classrooms and interventions focussing on issues identified by teachers. The above study focussed on an intervention of a particular aspect of teaching to address a problem identified by a teacher. It is contended that more attention be accorded to aspects of teaching to improve achievement scores in school mathematics. In South Africa there is a tendency to propose (and implement) improvement interventions based on singular studies and take the outcomes of such studies as evidence that the interventions will work. What this study is opening up is that interventions based on strong meta-analytic evidence focussing on teaching have the possibility of enhancing achievement. An implication of this is that policy should reconsider its inordinate attention on teachers' subject matter knowledge and shift towards the development of teaching of mathematics in classrooms. It also implies that textbook authors and evaluators of textbooks for schools should consider the forms of re-organisation which take into account distributive practice.

Acknowledgement
This research is supported by the national research foundation under grant number 77941. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the national research foundation.

References

Abelson, R. (1985). A variance explanation paradox: When a little explains a lot. *Psychological Bulletin*, 97(1),129-133.

Coe, R., (2002). *It's the Effect Size, Stupid. What effect size is and why it is important*. Paper presented at the Annual Conference of the British Educational Research Association, England 2002. Retrieved on September 8, 2014, from http://www.leeds.ac.uk/educol/documents/00002182.html

Cramp, S. & Nardi, E. (2000). A snappy start to a mathematics lesson. *Mathematics Teaching, 17, 46-51.*

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

Dempster, F. N. (1988). The Spacing Effect: *A Case Study in the Failure to Apply the Results of Psychological Research.* American Psychologist, 43(8), 627-634

Department of Basic Education, (2011). *Curriculum and Assessment Policy Statement (grades 10-12): Mathematics.* Pretoria: Department of Basic Education.

Ericsson, K. A., Krampe, R. Th. & Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review*, 100(3), 363-406.

Escalante, J., & Dirmann, J. (1990). The Jaime Escalante Math Program. *Journal of Negro Education, 59*, 407-423.

Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional psychology: Research and practice,* 40(5), 532-538.

Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education*, 36(1), 5 -13.

Hattie, J. (2009). *Visible learning: A synthesis of 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J. (2012). *Visible learning for teachers: Maximising impact on learning*. London: Routledge.

Hopkins, W. G. (1997). *New view of statistics*. Retrieved on May 14, 2014, from http://www.sportsci.org/resource/stats/effectmag.html.

House, R.J., Spangler, W.D., & Woycke, J. (1991). Personality and charisma in the U.S. presidency: A psychological theory of leader effectiveness. *Administrative Science Quarterly*, 36, 364 - 396.

Johnson, D. M. & Smith, B. (1987). An Evaluation of Saxon's Algebra. *Journal of Educational Research,* 81(2), 97-102.

Julie, C. (2013). Towards a model for intentional teaching for improving achievement in high-stakes mathematics examinations. In Davis, Z. & Jaffer, S. (Eds.), *Proceedings of the 19th Annual Congress of the Association for Mathematics Education of South Africa*, *Vol. 1* (pp. 86 – 96). Cape Town: AMESA.

Kang, Y., Harring, J. R. & Li, M (2015). Reexamining the impact of nonnormality in two-group comparison Procedures. *The Journal of Experimental Education,* 83(2), 147-174.

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement*, 65(1), 51-69

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

Klingele, W. E. & Reed, B. W. (1984). An Examination of an Incremental Approach to Mathematics, *The Phi Delta Kappan,* 65(10), 712-713.

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

Olejnik, S. & Algina, J. (2000). Measures of effect size for comparative studies:Applications, interpretations, and limitations. *Contemporary Educational Psychology,* 25, 241–286.

Padayachee, P., Boshoff, H., Olivier, W., & Harding, A. (2011). A blended learning Grade 12 intervention using DVD technology to enhance the teaching and learning of mathematics. *Pythagoras, 32*(1), 1 – 5.

Rohrer, D. & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science,* 35: 481–498.

Saxon, J. (1982). Incremental Development: A Breakthrough in Mathematics. *Phi Delta Kappan,* 63(7), 482-484.

Seabrook, R., Brown, G.D.A. & Solity, J. E. (2005). Distributed and Massed Practice: From Laboratory to Classroom. *Applied Cognitive Psychology*, 19, 107–122 .

Smith, S. M. & Ernst Z. Rothkopf, E. Z. (1984). Contextual Enrichment and Distribution of Practice in the Classroom. *Cognition and Instruction*, 1(3), 341-358.

Taylor, N., Van der Berg, S., & Mabogoane, T. (2013). *What makes schools effective? Report of the National Schools Effectiveness Study*. Cape Town: Pearson.

Wineland, & Stephens, 1995. Effects of spiral testing and review on retention and mathematical achievement for below-average eigth- and ninth-grade students. *International Journal of Mathematical Education in Science and Technology*, 26(2), 227 – 232.