South African Journal of Education

Vol 31:488-504

Evaluation of *Back to Basics* mathematics workbooks: a randomised control trial of the Primary Mathematics Research Project[1]

Brahm Fleisch, Nick Taylor, Roelien Herholdt and Ingrid Sapire
Brahm.Fleisch@wits.ac.za

Can providing learner support materials, particularly custom-designed workbooks, improve primary mathematics achievement more cost effectively than providing conventional textbooks? To contribute to this debate, this paper reports on the findings of a study conducted in 2010 by a consortium of educational researchers at JET Education Services and University of the Witwatersrand. Between January and June 2010, the consortium undertook a randomised control trial of learning support materials in Grade 6 classes in 44 primary schools serving low income communities in Gauteng province. The study shows that Grade 6 learners exhibit the same degree of learning improvement in mathematics, whether they use a conventional textbook or a workbook specifically customised to address the problems exhibited by South African learners. From a policy perspective, no warrant can be issued on the relative effectiveness of the specific workbooks considered in the study.

Introduction

In 2011 the Department of Basic Education (DBE), recognising the centrality of learning support materials in the learning process, allocated substantial expenditure to provide all Grade 1 to 6 learners in public schools with literacy/language and numeracy/mathematics workbooks. In 2012, the Department is intending to extend the workbook provision to include books for Grades 7 to 9 in Mathematics and First Additional Language. The DBE workbooks that have been provided consist of 128 discrete worksheets, four worksheets per week and eight weeks per term (Department of Basic Education, 2011). Given the scale and cost of this intervention, the provision of workbooks is clearly central to the South African government's strategy to improve learning outcomes.

But what is known about the relative cost-effectiveness of the provisioning of this kind of learning resource? To contribute to this debate, this article reports on a workbook study conducted in 2010 by a consortium of educational researchers at JET Education Services and University of the Witwatersrand. Two sets of quality learner (a workbook and a textbook) materials were tested. Learners in the experimental group were all issued with a 'workbook' consisting of a set of carefully sequenced worksheets covering the curriculum. The workbook emphasises basic skill proficiency and the four basic operations: addition, subtraction, multiplication and division. Learners in the control group of schools were all given an approved 'textbook' that is widely used in South African primary schools. The research question framing the study is: does a workbook, customised to address the lack of proficiency in basic mathematical operations by poor South African learners, produce improved learning gains over a conventional textbook?

This next section begins with a description of the background to the study, specifically exploring the rationale for the selection of Primary Mathematics Research Project workbook,

the theory that informs its design and existing research evidence of its effectiveness. The third section provides details of the evaluation study itself: the rationale for the design features, the sampling frame, and study instruments and measurements. The fourth section provides a summary of the major findings of the main study. This section provides evidence of the gains associated with the control and experimental groups, as well as important trends in relation to curriculum coverage. The paper concludes with comments on the research and policy implications of the study findings.

Background to the Primary Mathematics Research Project

To appraise the relative merits of 'workbooks' as a cost-effective intervention to improve primary school achievement, the study team determined that it would be prudent to select case study materials that had been developed and field tested in South Africa and had existing empirical evidence to demonstrate effectiveness. With these criteria in mind, we reviewed possible options and selected the Primary Mathematics Research Programme (PMRP) workbooks designed and produced by Eric Schollar and Associates.

The PMRP workbooks were originally field tested as part of a wider intervention programme in primary mathematics (Schollar, 2007). The intervention programme, with workbooks at its centre, included the implementation of 70 mathematics workbook lessons over 14 weeks. The assumption underlying the original PMRP workbook intervention is that mathematics performance would only be enhanced if certain 'bedrock' skills were taught systematically. These bedrock skills are associated with the capacity to perform mental calculations quickly and accurately. As Schollar notes:

> The application of these algorithms allows the solving of extremely complex calculations in simple steps through an understanding and knowledge of basic number bonds, the multiplication (and division) tables and, above all, an understanding of place value in the base-10 number system. Conversely the failure of learners to understand the number system and to master arithmetic operations beyond the reach of simple counting of single units renders learners incapable of developing any degree of mathematics proficiency. (Schollar, 2007, 19).

To take account of the great variation in performance in every class, the PMRP approach begins by diagnosing which level each child is on in terms of the relevant topic: diagnostic tests are a key part of the programme. The approach assumes learners would progress more rapidly towards expected levels of competency the longer they use the programme. The PMRP workbook is graded and, following the diagnostic test, each child is placed at the appropriate level in the book (Cow, Lion, Elephant, Goat) which s/he then follows through the book. The lesson plans for Day 1–4 of each week, have a simple logical structure, while Day 5 lesson plans prescribe review, assessment, enrichment and remediation exercises. The assumption in the workbook materials is that the prescribed lessons and exercises ensure that teaching proceeds at the correct pace required to cover the curriculum and, most importantly, that children work from the text, reading and writing every day.

The PMRP approach, with the workbooks at the core, was field tested in 40 schools (over 3,000 learners) in one district in Limpopo in 2007 (Schollar, 2007). In the original study, schools were randomly assigned to two groups: an intervention group and a control group, with 20 schools in each. Grade 4 and 6 teachers in the intervention schools attended a one and a half day training workshop at which they received a teacher manual including lesson plans for 14 weeks, learner workbooks with daily worksheets corresponding to the planned lessons and

related assessment materials. The control group received no intervention and in these schools teachers and learners continued working as they normally did. Teachers in the intervention group were visited twice during the 14 weeks of implementation, when their compliance to the PMRP method was monitored. The same pre- and post-tests were administered to both groups. The original study (Schollar, 2007) found that schools which used the workbooks as planned made very significant gains (126.7% increase for the workbook group compared to an increase of 15.6% for the control) compared to the control schools.[2] The original study concluded that the improvements in learner performance in the intervention schools can be attributed to the use of the PMRP workbooks and that using a direct instruction approach associated with the workbook is a cost-effective option for improving South African learner achievement. The gain in scores achieved by the programme using the workbooks is in the order of twice to three times the kinds of learning gains effected by donor-funded school intervention programmes in South Africa in the last decade (Taylor, 2007).

Literature review

Lockheed & Verspoor (1991) and more recently Abadzi (2006) have reviewed empirical evidence from a range of countries on the relative importance of learner materials for improving learning outcomes. Research from a wide range of contexts including Nicaragua, the Philippines, Brazil, Fiji, Ghana, Guinea, have consistently shown improvement in learner performance when sufficient textbooks were supplied. Advocates of textbooks argue that these resources work as they reduce wasted instructional time (Lockheed, 1991). Advocates of textbooks do acknowledge, however, that the effectiveness of these learning resources is conditional. They work if the materials are pedagogically sound, culturally appropriate and durable. They also require teachers to be trained in the use of the materials, and learners need to be able to take textbooks home.

The emphasis on provisioning of textbooks as a key part of the improvement strategy, closely associated with the World Bank approaches of the 1980s, is not without critics. Fuller (1991), for example, drawing on his field work in Botswana and Malawi, expressed scepticism about the use of textbooks in developing countries. This scepticism has recently been supported by robust empirical evidence. Glewwe *et al*.'s (2007) study revealed the failure of textbooks to increase student performance. Their research shows that learners, who were academically strong regardless of textbook availability, increased their performance levels substantially with the use of textbooks, but that those students who were average or weak showed no substantial gains through textbook use.

Absent from the developing country context literature are analyses of the relative merits of different types of learner materials. In particular, little is known of the relative cost-effectiveness of approach that employs 'workbooks', i.e. soft-cover books consisting of a series of sequenced worksheets designed to be written directly in compared to conventional and generally more expensive standard textbooks. Developers of the newer 'workbooks' assume that these materials reduce wasted instructional time associated with teachers writing exercises on the chalkboard. They also assume that workbooks are lighter and can and will be taken home. It is also assumed that as workbooks are designed to be replaced annually, they eliminate the difficult management of textbook retrieval (Department of Basic Education, 2011).

Research design and methodology

Within the areas of school improvement, JET Education Services has been working systematically over the past 16 years to develop a comprehensive knowledge base on school effects and school improvement (see Taylor, 2007 for a synthesis). Other initiatives in the country have focused on knowledge development through analysis of data produced by regional, national and cross-national studies of quality (Van der Berg, 2008). In addition, there continues to be a large grouping of researchers undertaking qualitative research in the field. (For an overview of education research on learning achievement in South Africa, see Fleisch, 2008). While these are important developments, what has been missing in South Africa is evidence-based research generated from randomised experiments. A randomised control trial is an experimental design in which the researcher randomly assigns subjects to either the control or the intervention group (McMillan & Schumbacher, 2006). The random assignment of subjects, when done within a large enough sample, enables the researcher to ensure that there are no major differences between the control and intervention groups prior to the intervention as well as to conclude that any differences after the intervention are not due to the characteristics of the subjects or extraneous events, but to the intervention or lack thereof (McMillan & Schumbacher, 2006). Quasi-experimental designs mimic ramdomised control trials in the direct manipulation of conditions, but lack random assignment, often due to practical constraints in applied research (McMillan & Schumbacher, 2006). Thus, in quasi-experimental designs it cannot be assumed that the groups were equal in all measurable and non-measurable characteristics before the intervention, or that differences between the groups after the intervention were not influenced by extraneous factors. Although evidence generated from matched quasi-experimental studies (of which there have been a number) has the potential to make substantial contributions to the knowledge base, there are significant advantages to using randomised experiments and, as a consequence, this is a rapidly growing area of research, particularly in the US (Raudensbush, 2005; Slavin, 2008).

The development of new statistical techniques now makes randomised control trials both economically and practically feasible (Gilmore & Goe, 2007). Consequently, there is a growing acceptance amongst education researchers that randomised control trials have an important role to play, given the high-stakes consequences of their findings. It is however necessary to expand the number of studies using these approaches in order to compare findings.

The PMRP Study used a modified standard treatment/control approach. For ethical and policy reasons, it was decided that rather than simply comparing the relative gains of intervention schools with schools who received no intervention, the control schools would receive a complete set of materials, representing not what currently exists, but what could be standard practice if the schools were properly provisioned with existing approved materials and used these regularly. This is what we refer to as 'enhanced standard practice'.

Research site

Gauteng province was chosen as the research site as it has the advantage of containing a range of poor schools of different types (rural, urban, informal and formal) in relative proximity, which reduced travel costs.

Sample design, frame and size

Particular care was taken in selecting the most appropriate sample design and sample size for the study. The original study used a simple two group comparative design model, with 20

schools in the experimental group and twenty in the control group. On closer examination, it was found that the decision made in the original study to place the number of schools at 20 for control and experimental groups, respectively, was not based on scientific sampling procedures, but rather on budget constraints. Working closely with the University of Michigan's Capacity Building for Group Level Interventions, the present study initially considered using a multisite cluster randomised field trial design (Borman *et al.*, 2008). This is a novel approach, which essentially consists of working in relatively large schools and randomly assigning all classes *within* each school to the treatment or control group respectively, rather than using only one class per school for either treatment or control, as Schollar had done in his initial evaluation of the PMRP. The multisite cluster design has a number of advantages, chief of which is that fewer schools are needed to achieve the required number of classes; hence this design holds the promise of reducing costs significantly.

We proceeded to select schools according to the multisite cluster design, using all primary schools in Gauteng with four or more Grade 6 classes as the target population. However, when we began to approach schools, we realised that it is common for one teacher to be given responsibility for all Grade 6 maths classes in the school, thus nullifying the advantage of selecting more than one unique class in each school. We decided therefore to work in schools with a maximum of two Grade 6 classes, and adopted the whole school rather than classes within a school as the unit for random selection. The advantage of this design is that it reduces the possibility of spill-over or contamination between treatment and control classes, a serious problem in a pilot study of the Singapore Mathematics Materials in Alexandra Township (Du Toit, 2010).

For ethical and practical reasons, all learners in a particular grade in a selected school were included in the study. The ethical reason is that sampling classes within schools would have meant that some children within a single school and grade would have received the benefits of the treatment or control, while others would not have. The practical reason is that if the study had had a sub-sample for treatment or control within a school, the mathematics teacher would have been required to teach two different methods simultaneously, which would have substantially added to his/her workload, and increased the risk of spill-over effects.

With a random assignment to intervention schools and enhanced standard practice schools, the variance estimates are large because the schools were the unit of interest rather than the classes (if classes had been the unit of analysis, individual schools could have had more than one unit). One of the vexing questions the researchers grappled with was the number of schools required to ensure that the study had adequate statistical power.

Given these assumptions, according to Schochet (2008), a sample size of 44 (22 treatment and 22 control) schools would be adequate, assuming a two-tail test, a value of 0.20 for the Inter-Class Correlation (ICC), a balanced allocation of the research groups and no sub-sample within units.

A list of all Gauteng public primary schools was drawn from the provincial database. All schools with more than two Grade 6 classes, fewer than 30 learners in Grade 6[3] and all schools in quintile 5 (i.e. the most affluent schools) were excluded. The remaining schools were ordered randomly and intervention and control schools assigned alternately off the random list. Starting at the top of this list, schools were then called and informed of the study. When a school could not be contacted after repeated efforts, it was deleted from the list: of the first 53 schools on the list, nine could not be contacted.[4]

Learner attrition

Learner attrition of 11% occurred in both groups, because of learners not being present for one of the two tests, as shown in Table 1.

Table 1 Learner attrition

CM*	Learners excluded	153
	Learners present both pre and post	1 374
PM**	Learners excluded	125
	Learners present both pre and post	1 141

*CM (Classroom Mathematics) refers to group that received the enhanced standard treatment.
**PM refers to the group that received the Primary Mathematics Research Programme materials.

The control group

In studies of this kind it is generally desirable to have a counterfactual: this is what the subjects of the experimental evaluation would have received if the intervention had not taken place. In the majority of South African schools, 'normal schooling' consists of very slow pacing of instruction and very little reading and writing (Taylor, 2007; Schollar, 2007) used this as the control group in his original study. However, it seems obvious that any group of classes which is subject to an intervention consisting of a set of materials which facilitates appropriate pacing and daily reading and writing would show improved learning, even if the materials were of a relatively mediocre quality. Therefore it was decided to provide control schools with additional standard materials, thus making the study a comparison of the intervention schools with control schools using 'enhanced standard practice'.

Audit of learner materials

In the last week of May 2009, the study team undertook a comprehensive audit of all learner materials for the teaching of Grade 6 mathematics in the proposed control schools. The results are shown in Table 2. What emerged from the audit is that the randomly selected primary schools were using 11 different approved titles including *Successful Mathematics*, *Mathematics Plus*, *Spot On*, *Mathematics Matters*, *Classroom Mathematics*, *Maths for All*, *Maths Today*, *Magic Maths*, *My Clever Mathematics*, *Day by Day Mathematics*, and *Maths for Fun*. The most frequently used were *Successful Mathematics* (Oxford) and Classroom Mathematics (Heinemann).

The other major finding from the audit was that all but two of the schools had textbook shortages. Half of the schools had only a teacher's copy of the book. The learners in these schools did not have access to a single copy of the book. In the rest of the schools, most learners had to share textbooks.

Choice of materials for the Control group

The results of the materials audit in control schools revealed that the schools did not have sufficient materials for the study of mathematics in the Intermediate Phase (and by implication for all levels of the primary school). Thus, implementation of the research design described above (PMRP vs. 'enhanced standard practice') would not be possible without the provision of books to all learners in control schools. The *Classroom Mathematics* textbook was chosen for the control group, on the grounds that it is widely regarded as a 'good' textbook and is very

Table 2 Audit results of the learner support materials in control group schools

Name of school	Number of classes	Number of learners	Textbook(s) used	No. of books available	Shortage
S1	3	125	Successful Mathematics		
S2	1	72	Mathematics Plus	22	50
S3	2	64	Spot On Classroom Mathematics	1	64
S4	2	78	Mathematics Plus Mathematics Matters	3	They use photo copies (59)
S5	2	59	Classroom Mathematics	1	59
S6	1	12	Maths Today	12	0
S7	2	123	Maths for All	40	83
S8	2	78	Successful Mathematics		
S9	2	61	Successful Mathematics Spot On	Teacher's copy	60
S10	1	48	Maths Today Magic Maths	5	43
S11	2	55	Successful Mathematics Spot On	0	55
S12	1	44	Spot On Understanding Maths	Teacher's copy	44
S13	2	97	Classroom Mathematics Successful Mathematics	20	77
S14	2	83	Successful Mathematics	0	83
S15	1	51	Successful Mathematics Spot On	51	0
S16	1	51	My Clever Mathematics Classroom Mathematics	Teacher's copy	51
S17	1	46	Spot On	Teacher's copy	46
S18	1	41	Classroom Mathematics	35	6
S19	2	75	My Clever Mathematics	38	37
S20	2	88	Classroom Mathematics Day by Day Mathematics My Clever Mathematics	7	81
S21	2	90	Magic Maths Spot On	Teacher's copy	90
S22	3	120	Spot On Maths for fun	0	120

widely used throughout the country. A comprehensive set of books was provided to schools during training in the second week of 2010.

The two sets of materials used by the teachers and their learners are designed to cover LO1 (Number and operations) of the NCS, but in different ways. The following review gives some indication of the similarities and differences between the material sets.

Classroom Mathematics (CM)

The book used by learners in this group was Classroom Mathematics: Grade 6 (Scheiber et al.,

2009). The CM material set is a standard textbook approved for use in South African schools. It was developed by an experienced mathematics textbook writing team of seven writers. CM is generally divided into two-page 'chunks' of work, each of which is structured as follows: introduction of terminology, explanation of a concept or technique, a number of worked examples and a set of activities (generally in the form of an exercise) which are graded from easy to more difficult. The activities are designed to be done in separate exercise books. Although the teacher guide contains supplementary material and solutions to all problems, the learner book is self-sufficient and therefore suitable for self-study by learners.

The learner book is designed to cover the entire NCS mathematics curriculum in a format that can be used by a teacher over the period of one year. A work schedule which could be followed by teachers is provided but no emphasis is placed on the use of this work schedule. The learner book also contains expositions of alternative algorithms, such as using expanded notation to 'decompress' 'long' multiplication and division: some of these examples are long and unwieldy, and while they can be an aid to conceptual understanding, it would be inefficient to use them to perform operations, particularly for large numbers. CM strongly promotes problem solving and does encourage the drilling of multiplication tables and addition/subtraction bonds.

Primary Mathematics Research Project (PM)

The book used by learners in this group was *Back to Basics! Getting Learning Outcome One Right Intermediate Phase* (Schollar, 2003). The PM material set comprises a project learner workbook and teacher manual that has been used in several South African schools. The book was explicitly designed to address the problems observed during extensive research in primary schools, specifically to move beyond the 'unit counting' methods so ubiquitous in schools serving poor children and to develop a greater degree of automaticity in executing the four arithmetic operations. PM's approach is not to prioritise basic operations at the exclusion of problem solving, but assumes that proficiency in the four operations is a prerequisite for problem solving. In order to achieve these aims, the workbooks consist of exercises which are mostly repetitive, numeric activities designed to consolidate concepts that have been taught. Each day starts with a 10 minute mental maths exercise. Contextualised word problems are also provided at the end of the fourth and the ninth weeks of using the materials and the formal assessment tasks include some word problems.

PM provides fairly thorough curriculum coverage of LO1 (Number and operations) (it is not a full "textbook" that covers the entire curriculum). The materials require that teachers test their learners and then place them into strands according to their achievement on the diagnostic test. Each level provides the learners with the same mathematical content but on different levels, ranging from Grade 3 level work to Grade 6 level work. This feature of the book may make it complex to use and certainly increases its size: this is a very bulky book.

The teacher guide and learner book are designed to be used in tandem, and the learner book is therefore not amenable to self-study, containing mainly mental arithmetic exercises, rule summaries and sets of written exercises designed to be done directly in the workbook. The teacher guide contains full lesson plans, with conceptual explanations, definitions and worked examples and follows the learner book closely, even containing advice to the teacher on what to write on the board during each lesson.

Test construction

A test covering number and operations (NCS LO1 Grade 6) was constructed by the research team. The first draft of the test consisted of 60 items covering eight skill categories — number concepts (place value, comparing numbers), fractions, addition, subtraction, multiplication, division, problem-solving and mental operations — all specified in the NCS for Grades 3 to 6. This test was administered to both project and control groups as a pre- and post-test. Item difficulty and discrimination indices were calculated per group for each item on the pre- and post-test. Using the results of the post-test, 40 items were selected with item means between 0.15 and 0.91 and item discrimination indices between 0.20 and 0.80.

Furthermore, the reliability estimates for the whole sample as well as each group were calculated on both the pre-and post-test. The measure of reliability gives an indication of the likelihood of obtaining the same results should the test be administered again under similar conditions. It is usually measured by the Alpha coefficient. The overall reliability of the test, as indicated by the Alpha coefficient, was 0.84 (CM 0.84 and PM 0.85) at the pre-test and 0.87 (CM 0.87 and PM 0.87) at the post-test. By convention, a lenient cut-off of 0.60 is common in exploratory research; Alpha should be at least 0.70 or higher to retain an item in an "adequate" scale, and many researchers require a cut-off of 0.80 for a "good scale" (Gronlund, 1998). Therefore, with an upper level of 1, the coefficient of 0.84 and 0.87 is high and would suggest that the test had good reliability at both the pre-and post-test levels, as shown in Table 3.

Table 3 Reliability statistics for the pre and post tests

	Cronbach's Alpha	
Group	Pre-test	Post-test
CM	0.842	0.871
PM	0.845	0.870

As indicated in Table 4, the 40 items that remained in the test for analysis covered the eight skills categories for Grades 3, 4, 5, and 6. Grade level does not necessarily indicate difficulty level, since, for example, a Grade 6 item might be pitched at the minimum requirement level for that grade, while a Grade 5 item might be at the maximum level. The skills categories can be further grouped into:

- operations, which were predominantly procedural in nature, i.e.:
  - Addition
  - Subtraction
  - Multiplication
  - Division
- number concept and understanding, which were predominantly conceptual, i.e.:
  - Number
  - Fractions
- strategic (problem solving), which tested the ability of learners to interpret questions that applied their number or operational skills in context.

Items can further be categorised as those:

- containing only symbolic information (numbers and mathematical signs) (10),
- involving a reading instruction with numeric information (24),
- problem type questions with the numeric information embedded in the text (6).

Achievement on these items could be compared to investigate the contention that learners struggle with reading and problem solving activities.

Table 4 Test items grouped according to Skill and Grade level*

	Number of items				
Skill	Grade 3	Grade 4	Grade 5	Grade 6	Total
Number concepts	3	3	1	1	8
Fractions	3	3	1	1	8
Problem solving	4	2	0	0	6
Addition	1	2	0	1	4
Subtraction	1	1	2	0	4
Multiplication	3	0	0	0	3
Division	2	1	0	1	4
Mental Operations	1	1	0	1	3
Total	18	13	4	5	40

* Grade levels follow the NCS specification and do not necessarily indicate the difficulty of an item.

The test items were scored according to a key which included some items for which there was only one correct answer and some items that were partially scored. The partial scoring gave recognition to the correct answer (of which there was one) but also to answers which indicated a partial understanding of or solution to the question. Tracking the improvement from incorrect, to partially and fully correct could give a deeper understanding of learner gains over the 14 week teaching period. There were 33 items with one correct answer only and 7 items with partial scoring.

Training of teachers

Training took place over two weekends in January. One of the researchers, who is an experienced teacher educator, trained teachers for both the PM and the CM groups. This was planned so that teachers had a few days to complete the necessary administrative tasks required of them at the beginning of the academic year, before beginning teaching according to the 14 week schedule. It was also hoped that by the second week of term most of the learners would have registered at their schools for the year and numbers of learners in each class would be known, since a key aspect of this project was the provision of a textbook to each learner in every class. Both groups received the same number of hours of training, although the content of their training differed according to differences in the materials being used by the group. Teachers were told that attendance at the training sessions was compulsory, but there were still certain teachers who were unable to attend for various reasons. These teachers were visited at their schools during the week between the first and the second training sessions. The purpose of the visits was to deliver the essential materials that had been handed out at the first session

and to brief the teachers on what they had missed. Teachers who were unable to attend the second training session were also visited by a fieldworker and their materials delivered to them.

Findings

Despite the differences in both form (workbook vs. textbook) and internal logic of these materials, there was no significant difference in the gains made by the two groups, each group achieving a mean improvement of over 8 percentage points (Table 5).

Table 5 Mean scores in the pre- and post-tests

	Number of learners		Mean for test: Mean % correct		Mean for test: Std. Error		Std. Deviation	
	CM	PM	CM	PM	CM	PM	CM	PM
Pre-test	1374	1141	48.07*	46.74*	0.45	0.50	16.89	17.091
Post-test	1374	1141	56.69*	54.87*	0.49	0.54	18.25	18.29
Gain			8.62**	8.13**				

* The difference between the pre- and post-test mean for each group is statistically significant, as is the gain score within each group (p value of 0 .00)

** The difference in gains between CM and PM is not statistically significant (p value of 0.29).

Although the schools were selected at random, the PM group performed less well on the pre-test. However the percentage point gains were almost identical. The respective frequency distributions of the two groups in both pre- and post-test are shown in Figure 1.

Both the CM and PM groups made substantial gains in terms of the proportions of learners who achieved scores of 50% or more (Table 6). Whether assessed on the percentage point gains, on the percentage gains from the pre-test baseline or the increased proportion of learners that 'passed' the test, there is little evidence that either set of materials is more advantageous. Both made significant gains of roughly the same magnitude.

Slavin and Lake's (2008) best evidence synthesis of elementary mathematics programmes revealed that, while the introduction of new textbooks alone did not seem to impact positively on learner outcomes, what they refer to as 'instructional process strategies' do. We would suggest that context is key as most disadvantaged South African schools do not have complete sets of textbooks or workbooks and as such, the provision and use of such learner materials is likely, at least in part, to explain the change in prevailing instructional practices.

Gain scores by coverage of the materials

Both sets of materials are designed to cover all LO1 (Number and operations) topics specified for Grade 6 in the NCS in 14 weeks. The time allocated is proportionate to the space LO1 occupies in the curriculum. The study tracked coverage of the materials through visits to classrooms on three separate occasions. Fieldworkers examined learner books and noted how far the class had progressed and whether all exercises had been completed up to that point. From these data, percentage coverage could be calculated for each class. The two interventions diverge in terms of the gain scores by degree of coverage (Table 7) at a minimum of 79% and below 79%.

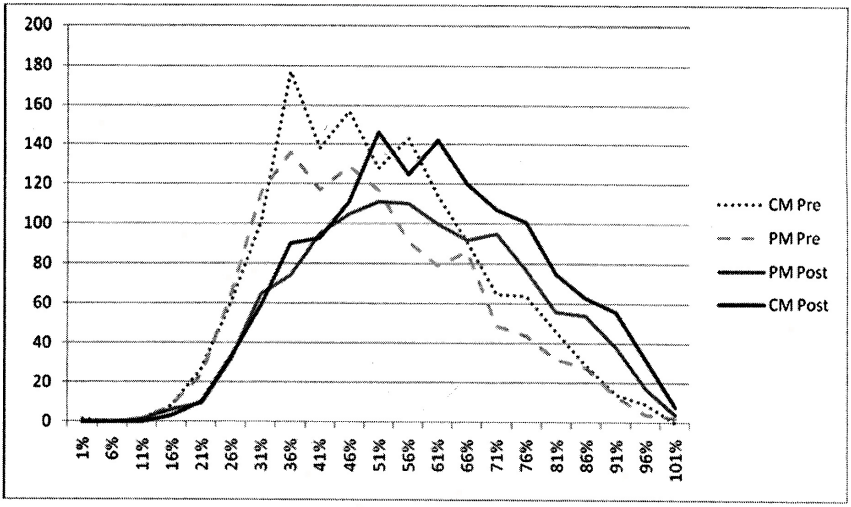

Figure 1 Pre- and post-test score distribution by Group

Table 6 Proportion of learners scoring at or above 50%

Group	No. of learners	Pre-test		Post-test		Increase		
		No. ≥50%	Perc ≥50%	No. ≥50%	Perc ≥50%	No. ≥50%	Perc ≥50%	Perc on BL*
CM	1 374	604	43.96	866	63.03	262	19.07	43.4%
PM	1 141	458	40.14	667	58.46	209	18.32	45.6%

* Base-line

Within each coverage category, each group's mean improved significantly from the pre-test to the post-test suggesting that providing programme materials makes a difference. As would be expected, the improvement is higher when 79% or more of the material is covered. The first point to note is that, whereas 607 (44%) learners in the CM group achieved 79% coverage or more, only 420 (37%) of the PM group achieved this. This may be because CM is more easily readable than PM and therefore classes are able to progress faster, although this hypothesis will have to be investigated by means of case studies. Also, those CM learners who covered 79% or more achieved higher gain scores (12 percentage points) than their PM counterparts (8.85). On the other hand, those PM learners who covered less that 79% of the material showed higher gains (7.72) than the equivalent fraction of CM learners (6.61). These differences are striking, but their origin is unclear and a great deal more qualitative work is required to gain a better understanding of this phenomenon. In general, for both groups, gain scores increased with increasing coverage, an obvious and expected development, which

provides strong evidence in support of the hypothesis that it is use of the materials that causes learning gains.

Table 7 Learner mean totals and gains per coverage at 79% level

Coverage				Mean score (%)		Std. Deviation	
		CM	PM	CM	PM	CM	PM
>79%	Pre-test	607	420	51.29**	44.07**	17.98	16.50
	Post-test	607	420	63.29**	52.91**	17.57	19.01
<79%	Gain	607	420	12.00***	8.85***	12.94	10.56
	Percentage on Pre-test	622	721	46.97**	48.30**	15.71	17.25
	Percentage on Post-test	622	721	53.58**	56.02**	17.05	17.78
Missing	Gain	622	721	6.61****	7.72****	11.61	10.16
	Percentage on Pre-test	145		39.35**		13.10	
	Percentage on Post-test	145		42.45**		13.92	
	Gain	145		3.09		9.87	

** The differences between the pre-and post-test mean for the >79%, <79% and the missing coverage categories for each group are significant (p value of .00)

*** The difference between the gain score for the PM and CM groups in the >79% coverage category is significant (p value of .00)

**** The difference between the gain score for the PM and CM groups in the <79% coverage category is not significant (p value of 0.06)

Costs

To determine the unit costs of the respective materials, the study compared the costs of the Classroom Mathematics and Primary Mathematics materials as paid to the publishers in January 2010 (see Table 8).

Table 8 Unit cost of materials (R)

	CM	PM
Teacher Guide	R 145.00	R 136.00
Learner Book	R 87.00	R 83.00
Teacher Guide Cost Per Learner*	R 3.40	R 3.63
Average Cost of Learner Book per Year**	R 43.50	R 55.33
Total Annual Unit Costs	R 46.90	R 58.96

* Per Learner Costs of Teachers Guide: we divided the full costs of the teacher's guide by 40.

** To establish the average cost of the learner book per year, we assumed that the CM book would be used for at least two years (this is a very conservative estimate that incorporates substantial loss). PM materials are designed to be used for three years in Grades 4, 5, and 6 for Learning Outcome One. However they cover half of the years work, and are designed to be used by only one learner. We therefore multiplied the cost by two.

Given that Classroom Mathematics is a textbook, we assumed multi-year use, but given the difficulties with retrieval systems, we assumed that the books would need to be replaced every two years. With regard to the annual costing of the Primary Mathematics book, we assumed that this is a workbook in which learners would be writing directly in, would be used over three years and each copy would be used by only one learner. Moreover, the book is designed to cover only half of the NCS curriculum for Grade 6 mathematics (and other years) and as such, we assumed that a full year of this kind of materials would cost twice what we paid per book. Within the constraints of our assumptions, which we consider to be realistic, the PM materials are more expensive than the CM textbook and accompanying teacher guide.

Conclusion and implications of the study

Research

The advantage of random control trial (RCT) designs as used in the present study is that they are able to generalise research findings to the target population, in this case Grade 6 classes in quintile 1–4 schools in Gauteng. Further, the answer to our main research question — does the PMRP, as a customised workbook, show an advantage over a conventional textbook — is clearly in the negative.

However, the conclusions that can be drawn from RCTs are constrained by the elements of their design. In the case of our study, the absence of a control group — for both ethical and practical reasons — means that we cannot compare the results of the present study with what normally happens in Gauteng schools. We can show relatively large learning gains, but we do not know how much of this might have happened without the materials provided by the research study. However, we can make an educated guess, based on typical gains made by school development programmes over the last 20 years, which indicate that all learners in our study gained significant advantage through using the materials.

A second constraint on RCT research designs is that they are blind to the mechanisms which drive the changes observed. In other words, they can confidently show a change accompanying an intervention, but, on their own, RCT programmes cannot explain why or how these changes happen (Maxwell, 2009). This latter area is the domain of intensive descriptive studies, which in research on materials would include classroom observations of materials in use and discussions with teachers and learners. By their intensive nature, such ‘qualitative’ studies are necessarily small in scale and therefore cannot reach general conclusions. Ideally, intensive descriptive studies should accompany RCTs to couple predictions for the population with an understanding of change mechanisms and hence of their implications for policy and practice.

Policy

The research problem that animated this study focused on the relative effectiveness of PMRP resources in comparison to conventional textbooks. We found that not only did both the Primary Mathematics and Classroom Mathematics groups make statistically significant gains between the pre-test and the post-test, but the magnitude of gains was substantially higher than most interventions in South Africa typically achieve (Taylor, 2007). The present study shows that Grade 6 learners exhibit the same degree of learning improvement in mathematics, whether they use a conventional textbook or a workbook specifically customised to address the

problems exhibited by poor learners. From a policy perspective, no warrant can be issued on the relative effectiveness of the Primary Mathematics Research Project workbooks.

Our study shows that, in Gauteng at least, children cannot use books, suitable or otherwise, because they are simply not available in schools: in the 22 control schools audited, only two had sufficient Grade 6 mathematics textbooks for all learners, and around half had no copies. Further study is needed to examine the reasons for this situation, with possible explanations including corruption in the procurement of books, poor school management systems resulting in books not being cared for and reluctance on the part of teachers to use books and the consequent low demand from schools. The first task in getting teachers to teach reading, writing, and mathematics is to get books to schools and not necessarily to expend resources on designing new workbooks.

Notes

1. This project is a joint undertaking of colleagues at the Wits School of Education and JET Education Services. Funding for the study has been provided by the DG Murray Trust and JET. We acknowledge a number of individuals who have contributed their time and expertise. Linda Zuze has played an invaluable role in helping determine the sampling frame and providing statistical training. Helen Perry and Prof. Peter Fridjhon assisted with some of the data analysis. Talia de Chaisemartin has provided guidance in the statistics and presentation of data. Special thanks to Andres Martinez at the University of Michigan's Building Capacity to Evaluate Group-Level Interventions and Stephen Raudenbusch at the University of Chicago for assistance with study design and sample size. We also thank the Gauteng Department of Education for permission to undertake the study and the Non-Medical Ethics Committee of the University of the Witwatersrand who reviewed the proposal and issued an ethics clearance certificate for the study. The findings, interpretation, and conclusion expressed in this paper however are entirely those of the authors. We welcome comments and criticism.
2. The full programme starts at Grade 4 level and ends at Grade 6 level. Therefore, an evaluation should track the Grade 4 class to Grade 6 to allow the full effect of the intervention to be measured.
3. Some school sizes decreased substantially since selection to actual data collection.
4. In the 2009 pilot, we had 23 schools in each group. In the final study in 2010, one school from each group dropped out of the study (mainly due to pressure from the district office to use the Foundations for Learning lesson plans rather than the lesson plans in the study).

References

Abadzi H 2006. *Efficient Learning for the Poor: Insights from the Frontier of Cognitive Neuroscience*. Washington: World Bank.

Borman GD, Maritza Dowling N & Schneck C 2008. A multisite cluster randomized field trial of Open Court Reading. *Education Evaluation and Policy Analysis*, 30:389-407.

Borman GD 2005. National efforts to bring reform to scale in high poverty schools: Outcomes and implications. *Review of Research in Education*, 29:1-28.

Fleisch B 2010. *Singapore Mathematics Materials study*. Johannesburg: JET Education Services (manuscript).

Fleisch B 2008. *Primary Education in Crisis: Why South African Children Underperform in Reading and Mathematics*. Cape Town: Juta.

Fuller B & Snyder C 1991. Vocal teachers, silent pupils? Life in Botswana classrooms. *Comparative Education Review*, 35:274-294.

Department of Basic Education 2011. *The Rainbow Workbooks*. http://www.ar-t.co.za/DOEFlyer/Workbooks%20flyer%201706.pdf. Accessed 25 July 2011.

Gitomer G & Goe L 2007. Editors note. *Educational Evaluation and Policy Analysis*, 29:4-5.

Glewwe P, Kremer M & Moulin S 2007. *Many children left behind? Textbooks and text scores in Kenya*. www.econ.yale.edu/~egcenter/infoconf/kremer. Accessed 7 April 2010.

Gronlund N 1998. Assembling, administering and evaluating the test. In NE Gronlund. *Assessment of student achievement*, 6th edn. Boston: Allyn & Bacon.

Lockheed M & Verspoor A 1991. *Improving primary education in developing countries*. World Bank Publication, Oxford University Press.

Maxwell J 2009. Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, 33:3-11.

McMillan JH & Schumbacher S 2006. *Research in Education: Evidence-based inquiry*. 6th edn. Boston: Pearson.

Raudenbush S 2005. Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher*, 34:25-31.

Scheiber J, Brown M, Lombard S, Markides M, Mbatha K, Randall I & van Noort D 2009. *Classroom mathematics: Grade 6*. Johannesburg: Heinemann.

Schochet PZ 2008. Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioural Statistics*, 33:62-87.

Schollar E 2007. *Towards evidence based education development in South Africa* (manuscript).

Schollar E & Kemp A 2003. *Back to Basics! Getting Learning Outcome One Right Intermediate Phase* (Learner Workbook and Teacher Manual).

Schreuder B 2008. *Improving literacy and mathematics in the primary school: The country's most important priority*. Paper presented to the Conference What's Working in School Development, JET Education Services. Available at www.jet.org.za .

Slavin R & Lake C 2008. Effective programs in elementary mathematics: a Best-evidence synthesis. *Review of Educational Research,* 78:427-516.

Slavin R 2008. What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37:5-14.

Taylor N 2007. Equity, efficiency and development in South African Schools. In T Townsend (ed.). *International handbook of school effectiveness and improvement*. Dordrecht, Netherlands: Springer.

Van der Berg S & Louw M 2008. South African student performance in regional context. In Chisholm L, Fleisch B & Bloch G (eds). *Investment Choices for South African Education*. Johannesburg: Wits University Press.

Authors

Brahm Fleisch is Professor of Education Policy at the University of the Witwatersrand. His current research interests include whole systems change, randomised control trials in education, and education law.

Nick Taylor is Visiting Researcher, University of the Witwatersrand and a member of Umalusi's Assessment Standards Committee. His interests are in the fields of assessment and evaluation of school improvement programmes.

Roelien Herholdt is a Senior Research Manager at JET Education Services and an executive committee member of the Association for Evaluation and Assessment in Southern African (ASEASA). Her current research interests include assessment and randomised control trials in education.

Ingrid Sapire is a Research Project Coordinator at the University of the Witwatersrand. Her focus is mathematics teaching and teacher development. She has designed and presented materials for mathematics teacher education.