The impact of study guides on "matric" performance: Evidence from a randomised experiment

STEPHEN TAYLOR AND PATRICIA WATSON

Stellenbosch Economic Working Papers: 13/15

KEYWORDS: SOUTH AFRICA, EDUCATION, EDUCATIONAL ACHIEVEMENT, STUDY GUIDES, RANDOMISED CONTROLLED TRIAL
JEL: I20, I21, I28

STEPHEN TAYLOR
DEPARTMENT OF BASIC EDUCATION AND
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: STEPHEN@SUN.AC.ZA

PATRICIA WATSON
DEPARTMENT OF BASIC EDUCATION
PRIVATE BAG X895
PRETORIA, 0001
SOUTH AFRICA
EMAIL: WATSON.P@DBE.GOV.ZA

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

The impact of study guides on "matric" performance: Evidence from a randomised experiment

STEPHEN TAYLOR AND PATRICIA WATSON*

ABSTRACT

Most international literature on the impact of textbooks on educational achievement suggests that this is a relatively cost-effective intervention. However, recent experimental evidence from developing countries has called this into question, suggesting that resources alone are unlikely to impact on performance and that changes in school organisation, pedagogical methods or incentives facing teachers are more effective. South African studies, using observational data, typically show weak associations between achievement and additional resources, though in some studies textbooks emerge as an exception. Some argue that school management is a key mediating variable.

This paper evaluates the impact of providing study guides to pupils shortly before their secondary school leaving examination (the "matric" exam). From a sampling frame of 318 schools in the Mpumalanga province, 79 schools were randomly selected to receive study guides, leaving 239 control schools. These study guides were developed by the National Department of Basic Education and distributed to treatment schools for four subjects – accounting, economics, geography and life sciences – resulting in four distinct treatments per school.

The impact of the study guides was estimated using matric results from 2011 (baseline) and 2012 (endline). The accounting and economics guides did not have a significant impact on performance. However, the geography and life sciences guides improved scores in those subjects by approximately two percentage points. Treatment heterogeneity was apparent for geography where students in better-performing schools gained more from the guides than students in low-performing schools. This may relate to other studies suggesting that additional school resources matter conditionally upon overall school functionality, particularly management. A simulation indicated that distributing the geography and life science at scale could increase the overall matric pass rate by roughly one percentage point. A cost-benefit analysis calculating the standard deviations of impact on test scores per $100 spent indicates that this intervention is amongst the most cost-effective of educational interventions internationally that have been tested using randomised experiments. Possible reasons why the guides were effective in two subjects but not the other two are discussed.

Keywords: South Africa, education, educational achievement, study guides, Randomised Controlled Trial

JEL codes: I20, I21, I28

* The authors recognise the work of the Department of Basic Education in commissioning the impact evaluation that is reported on in this paper.

1. Introduction

The South African school system is heavily geared towards the National Senior Certificate (NSC) examination at the end of grade 12, known informally as the "matric exam". Those passing this examination attain a "National Senior Certificate" and those meeting certain higher level requirements obtain a "Bachelors Pass" which qualifies one for entrance into university. Attainment of a NSC and especially of university education is strongly related to labour market performance, as the literature on returns to education shows (for example, Branson *et al*, 2012). In 2012 there were 510 173 candidates that wrote the NSC exam of whom 377 829 (73.9%) achieved a NSC and 136 047 (26.6%) also achieved a bachelors pass.

The Department of Basic Education (DBE) recently launched a series of new study guides to assist learners taking the NSC exam. The *Mind The Gap* study guide series is designed to assist learners in weakly performing schools where curriculum coverage may have been incomplete and learners thus lack exposure to the content being tested in the NSC. Incomplete curriculum coverage is a feature of many weakly performing primary and secondary schools throughout the grades thus leading to accumulated deficits by grade 12 (Taylor, 2011). The guides initially covered four subjects – Accounting, Economics, Geography, and Life Sciences in both English and Afrikaans – and have since also been developed in several other subjects.

The *Mind The Gap* guides were written by four subject teams comprising teachers, examiners, curriculum officials and academics.[1] The study guides were designed to be effective for self-study without requiring a teacher to mediate content. Item response analysis was used to identify specific concepts and skills that are most frequently incorrect and to target these in the *Mind The Gap* study guides. The study guides use simple explanations to clarify key terminology, provide references to textbooks and include learner assessment with model answers. In addition, the guides orientate learners to the examination requirements by drawing on the official Grade 12 examination guidelines and through providing examples of specific questions from previous examination papers and their memos. All of this is available on the DBE's website and is widely used in affluent schools but not sufficiently used in poor schools where connectivity may be lacking.

The *Mind The Gap* study guides were not intended to replace textbooks, but were designed to fulfil a different purpose which would be complementary to textbook use. This intervention can be seen as part of a broader policy emphasis within the DBE over the last few years on providing materials specifically targeted to schools in poor communities. A major intervention since 2011 involves the distribution of colour-printed workbooks for literacy, numeracy and life skills to all children in grades R to grade 9. Ironically, this emphasis on materials has overlapped with extensive media critique of textbook delivery, prompted by specific delivery problems experienced in the Limpopo province in 2012. However one looks at it, learning support materials have recently received a lot of attention in South Africa.

[1] The materials development process was guided by an explicit set of values and practices, which was referred to collectively as "ordinary innovation". A full discussion of this aspect of the materials development process may be of interest to some readers and can be found in the Department of Basic Education's evaluation report (DBE, 2013).

Despite the intentions of official policies and funding norms, school-based surveys indicate that access to textbooks in South African classrooms is erratic. For example, the School Monitoring Survey of 2011 indicates around 72% of Grade 6 learners having access to a mathematics textbook, with a figure of 90% for Grade 12. Previous surveys such as the SACMEQ[2] 2007 survey and the NSES[3] survey of 2009 suggest even lower access. Low access to school resources such as textbooks reflects inefficiencies in delivery as well as non-existent or poorly managed inventories and retrieval systems at the school level.

The majority of schools in poor communities can therefore be characterised as having limited access to learning materials and incomplete curriculum coverage. Furthermore, children in these schools typically lack home-based support in the forms of additional learning resources such as books and computers and of parents who can assist with school work. Therefore, an additional resource such as the *Mind The Gap* study guide can be expected to provide a valuable secondary site of learning acquisition for these children. With the high-stakes NSC close at hand, learners may well have a stronger motivation than ever to use available learning materials. This reinforces the theory of change underlying the *Mind The Gap* intervention.

However, the South African and international economics of education literature is not particularly optimistic about the impact of additional resources on educational achievement. Hanushek (2002) argues that input-based education policies around the world have failed to produce any systematic improvement in achievement. According to Hanushek, the more effective avenues for improving achievement have proven to be aspects of teacher quality that are unrelated to resources and policies that change the incentives facing teachers. Kremer, Brannen and Glennerster (2013: 297), discussing 30 experimental studies of educational interventions in developing countries, conclude that, "despite very low levels of resources, providing more-of-the-same educational inputs without changing pedagogy or accountability typically has very limited impacts on test scores."

Similarly, additional school resources have generally not been found to have substantial effects on school performance in South Africa. Van der Berg (2008) proposes that the effectiveness of resources appears to be conditional on the managerial efficiency of schools to convert resources into outcomes. Most large-scale school surveys provide weak measures of school management, thus restricting estimation of this notion. The NSES survey contained a richer than usual amount of information on school management and teacher practices. Using this data, Taylor (2011) highlights several indicators of effective management that are significantly associated with student achievement in multivariate education production functions. One of these indicators is the quality of inventories for learning support materials. Schools with inventories that were evidently in operational use performed better, *ceteris paribus*.

Yet, as Glewwe, Kremer and Moulin (2007: 1) point out, even those sceptical about the benefits of spending more on education are often optimistic about the impact of certain non-teacher inputs such as textbooks. Lockheed and Hanushek (1988), for example, in their review of four retrospective

[2] SACMEQ stands for Southern and Eastern African Consortium for Monitoring Educational Quality.
[3] NSES stands for the National School Effectiveness Study, which was undertaken by JET Education Services between 2007 and 2009.

studies focussing on textbooks, observe that significant and large effects were obtained in all four studies. Similarly, Heyneman, Farreland and Sepulveda-Stuardo (1978), Fuller (1986) and Fuller and Clarke (1994) all observed significant effects of textbooks in the majority of studies they reviewed. However, such optimism has been based mainly on estimates from retrospective studies using observational data which may be biased if textbook availability is correlated with other unobserved aspects of school quality.

Experimental or quasi-experimental studies on textbooks are less common, although two of the earliest Randomised Controlled Trials (RCTs) in education involved textbooks. Jamison *et al* (1981) conducted an RCT in Nicaragua which found that workbooks improved first grade mathematics achievement by about a third of a standard deviation. Heyneman, Jamison and Montenegro (1984) randomly allocated textbooks to children in grades 1 and 2 in the Philippines. 26 schools received one textbook per child, 26 more schools received one textbook for every two children, and 52 schools served as a control group. Both treatments improved test scores by a similar magnitude – about 0.4 standard deviations.

More recently, the results of an RCT focussing on textbooks in Kenyan primary schools have been widely disseminated. Glewwe *et al* (2007) found that schools receiving textbooks did not perform significantly better than the control group. They did, however, find that textbooks improved performance at the higher end of the student ability distribution. Glewwe *et al* contend that most students were not sufficiently fluent in English to benefit from the textbooks, which were in English, and that this explains why the textbooks only benefited more advanced students.[4] Glewwe (2002), reflecting on the RCTs in Nicaragua, Philippines and Kenya, suggests one other possible explanation that would be consistent with all three experiments: it may be that textbooks are only effective when teachers have been trained to use them.[5]

While a substantial international literature on the use of textbooks does exist, there is little evidence regarding other types of learning materials such as study guides. Horton and Lovitt (e.g. 1989) and their collaborators conducted several small scale experiments in the 1980s to investigate the impact of study guides. Horton and Lovitt (1989: 447) found that study guides, by extracting the main ideas from text material, can be an effective remedial intervention for students with poor reading ability. Their research indicated that study guides benefited students with learning disabilities, those with learning deficits and regular students.

The South African literature on the effect of textbooks is limited to analyses of observational data. Education Production Function analyses using PIRLS 2006 data have not found textbook availability to be a significant predictor of grade 5 reading achievement (Taylor, 2010; Shepherd, 2011). Similarly, Taylor's (2011) production function analysis of the NSES data indicated that textbooks were not significantly associated with literacy or numeracy in grades 4 and 5. On the other hand, analysis of SACMEQ data has suggested a positive effect of good textbook access on reading

[4] Note that the impact of the study guides should not have been substantially affected by the ability to read English since grade 12 students would have been exposed to English instruction and materials for much longer.
[5] The *Mind The Gap* study guide intervention was not accompanied by teacher training and their use would almost certainly not have been mediated by teachers.

achievement in grade 6 but not on mathematics achievement. This finding was obtained in analyses of the 2000 SACMEQ dataset (Gustafsson, 2007; Van der Berg, 2008) and of the 2007 SACMEQ dataset (Spaull, 2011). One possible explanation for the emergence of a mixed picture from these studies is that the available selection of control variables may influence whether a significant effect is observed – a key limitation with observational data. The estimates in these studies may therefore be biased if textbook availability in South Africa is endogenously determined: important unobserved school characteristics such as management efficiency may influence textbook access. Therefore, observed correlations between textbook availability and test scores may reflect unmeasured school efficiency rather than the causal impact of textbooks.

No rigorous impact evaluation of study guides has previously been conducted in South Africa. One study randomly allocated mathematics workbooks (a hybrid between a textbook and an exercise book) that were developed by an NGO to 21 primary schools and conventional textbooks to another 21 schools (Fleisch *et al*, 2010). Both groups improved test scores by a similar amount over the treatment period, indicating that the two interventions could not be distinguished. However, the lack of a pure control group (receiving neither of the books) means that the overall impact could not be measured.

The *Mind The Gap* study guides were developed during 2012 and were therefore not physically distributed to all schools in the country. The guides were made available electronically on the DBE website and were physically distributed to certain underperforming districts in the Limpopo and Eastern Cape provinces and to a lesser extent in the Western Cape, Northern Cape and Free State. The province of Mpumalanga was used to conduct a randomised control trial in which 79 schools received *Mind The Gap* study guides based on a computerised lottery, leaving a control group of 239 schools. Each school selected into the treatment group received the guides for all four subjects: Accounting, Economics, Geography and Life Sciences.

This design allows one to estimate the causal impact of the *Mind The Gap* study guides in each subject. While the primary interest in this experiment was to evaluate the impact of this new government programme, it also has bearing on several other interesting research questions. Can a relatively low-cost support material that is not mediated by a teacher improve performance in the NSC exam? Is the impact of study guides the same across subjects? To the extent that we have information on pupil and school characteristics, can we identify who benefits most from such an intervention?

The next section explains the methodology of the experiment in greater detail and introduces the data that was used for the evaluation. Section 3 reports the results from a number of estimation strategies employed to evaluate the impact of each of the four study guides. This section presents analysis of overall impact, an investigation of whether the impact varied across different types of pupils and schools ("heterogeneous effects"), a simulation to estimate what the overall effect of the study guides might be on the NSC pass rate if all candidates across the country received the guides, and various placebo tests to assess how likely it is that the observed results are a peculiar artefact of the data rather than a reflection of a true impact. Section 4 discusses the significance of the findings for policy-making and for research on the determinants of educational performance. Section 5 concludes.

2. Data and Methodology

The available budget (funds for transport costs and the number of guides already printed) dictated that a maximum of 79 schools could be chosen to receive *Mind The Gap* study guides within the impact evaluation component of the programme. The province of Mpumalanga was chosen as the site for an RCT because it is relatively large and predominantly poor province (thus strengthening external validity in applying the findings to the largest and most policy-relevant part of the South African school system) and because prior commitments to distribute the study guides to some school districts in several other provinces excluded those provinces from being suitable.

NSC registrations data collected by the DBE by June 2012 provided a sampling frame from which to select treatment and control schools. The sampling frame was restricted to schools in Mpumalanga that had registered to write the exams in English. The latter restriction excluded a small proportion of schools writing in Afrikaans, most of which would be more affluent and better resourced schools. The sampling frame was then further restricted to include only schools in which all four subjects under evaluation were offered. Schools in which fewer than five candidates were registered for any of the four subjects were then excluded. Schools in which more than 170 pupils were registered for any of the four subjects were also excluded. These restrictions were imposed to improve the ratio of statistical information to the cost of distribution. This process yielded a final sampling frame of 318 schools.

Using a computerised lottery, 79 schools were then randomly selected from these 318 schools to receive the *Mind The Gap* study guides. This left a control group of 239 schools.[6] Importantly, given the randomisation, there is no reason to expect the 79 treatment schools to be systematically different in any from the control schools. Therefore, after allocating the study guides to the treatment schools, the control schools provide a legitimate estimate of the counterfactual – how the treatment schools would have performed had they not received the study guides.

Delays in the printing process meant that it was only possible to deliver the study guides between the 24th and the 28th of September 2012. This was just less than one month before the first examination amongst the four *Mind The Gap* subjects (Accounting, 23 October) and about two months before the last (Economics, 21 November). The short period in which to use the study guides may have limited their potential impact, although it is precisely during this period that the guides might be expected to have maximum impact and during which there is the strongest incentive to use them.

[6] Using all the remaining schools in the sampling frame affords extra statistical power when estimating whether performance in the treatment group was statistically significantly different from performance in the control group. Since there was no cost associated with increasing the number of control schools, all 239 were included so as to maximise precision.

The distribution of the study guides was successful. A private company couriered the books to treatment schools and proof of delivery notes were returned to the DBE for all schools. However, it is not possible to know to what extent school staff distributed the study guides to pupils or to what extent pupils actually used the material.

A minimal amount of treatment spillover into the control group might have occurred since electronic versions of the study guides were available on the DBE website from September. However, in a poor and largely rural province such as Mpumalanga where most households do not have internet connectivity it is unlikely that many students would have downloaded the guides. Moreover, there was not a high degree of public awareness around this new material.

No assessment costs were incurred in the evaluation since outcome measures at the learner level for all schools and for all subjects could be obtained from the NSC database. This contributed to the extremely light "footprint" of this experiment in schools: there was no additional testing, no training accompanying the materials and no visits or monitoring by officials or fieldworkers. Therefore, many of the external validity concerns that typically plague RCTs due to the special experiment conditions involved were circumvented in this project. One can confidently rule out any potential Hawthorne effects (where the treatment group changes its behaviour due to the extra attention rather than the intervention itself) and there was absolutely no intrusion into control schools and therefore no reason to expect any John Henry effects (where the control group changes its behaviour in response to the experiment).

NSC data for 2011 and 2012 was used to measure impact. While the schools are the same across the two years, different pupils wrote the exam in each year. Therefore, two estimation strategies were used; one using data aggregated at the school level and treating it as a panel, and one treating the data as two cross-sections at the pupil-level. Table 1 shows the numbers of pupils in treatment and control schools in each subject in 2011 and 2012. Note that individual pupils are duplicated as each writes several subjects, including some that are not on this list. The 10 subjects shown here include the four *Mind The Gap* subjects and six other popular subjects that are used in some of the analysis as a control and in several placebo tests. For both years and all subjects these pupils were grouped in 79 treatment schools and 239 control schools (though one control school was not found in the 2011 data). Note that "treatment schools 2011" is used when referring to 2011 characteristics of those schools that were treated in 2012 – no treatment occurred in 2011.

Table 1. Numbers of pupils in treatment and control schools by subject and year

	Control schools 2011	Control schools 2012	Treatment schools 2011	Treatment schools 2012
Accounting	6 374	5 989	2 229	2 029
Economics	7 310	7 337	2 422	2 297
Geography	12 130	12 252	3 674	3 765
Life Sciences	12 931	12 694	4 206	4 271
Business Studies	8 446	8 780	2 799	2 805
English First Additional Language	23 964	24 145	7 811	7 660
Mathematics	9 587	8 976	3 169	2 928
Mathematical Literacy	15 085	15 753	4 817	4 890
Physical Science	8 240	7 822	2 748	2 604
Tourism	6 645	8 350	1 637	1 763

Due to the randomisation the identification of the causal impact of treatment is straightforward. Firstly, for each of the four *Mind The Gap* subjects the following regression model is estimated using Ordinary Least Squares:

$$Y_j = \beta_0 + \beta_1 T_j + \beta_2 X_j + \varepsilon_j \qquad (1)$$

where Y_j is the mean score in school *j* in 2012 (endline) expressed in percentage scores, T_j is a dummy equal to one if the school was in the treatment group receiving study guides, and X_j is a vector of school level control variables. In the fully specified model X_j includes the mean score in school *j* in 2011 (baseline) and the official poverty quintile of the school.[7] In one model the treatment dummy is interacted with the baseline score in order to estimate whether the impact of study guides differed according to the overall level of school performance. In other models the outcome (Y_j) is not the mean score but performance at various percentiles of scores within the school. In another model the outcome (Y_j) is the school pass rate in the NSC (not for a specific subject but overall).

Secondly, a triple-difference model is estimated using pupil-level data for pupils in treatment and control schools, in 2011 and 2012, and for treated subjects and non-treated subjects. In this model, the subjects that were not affected by *Mind The Gap* form another control group. Although there is no reason to expect it, this strategy is robust to the possibility that treatment schools experienced overall improvement in all subjects relative to control schools. As a sensitivity check this model was also estimated. The performance of pupil *i* in school *j* in year *y* for subject *s* is modelled as a function of being in a treatment school (*T*), the subject (*S*, which equals one for the *Mind The Gap* subjects and zero for other subjects), the year(*Y*, which equals one for 2012 and zero for 2011), interactions amongst all of these variables, and a vector of pupil and school characteristics (X_{ij}):

$$Y_{ijys} = \beta_0 + \beta_1 dT + \beta_2 dY + \beta_3 dS + \delta_0 dT.dY + \delta_1 dT.dS + \delta_2 dS.dY + \delta_3 dT.dY.dS + \beta_4 X_{ij} + u \qquad (2)$$

[7] For the purposes of funding norms, schools in South Africa are categorised into five poverty quintiles based on community-level measures of socio-economic status.

The OLS estimate $\hat{\delta}_3$ is now the coefficient of interest since it expresses the impact of being in a treatment school in 2012 for a *Mind The Gap* subject, over and above any overall year-, subject- or group-specific trends, and any year-subject-specific trends and any group-year-specific trends. $\hat{\delta}_3$ can also be expressed as follows, making clear the intuition of the triple difference (or "difference-in-difference-in-difference"):

$$\hat{\delta}_3 = \left(\bar{y}_{T,Sx,2012} - \bar{y}_{T,Sx,2011}\right) - \left(\bar{y}_{T,Sz,2012} - \bar{y}_{T,Sz,2011}\right) - \left(\bar{y}_{C,Sx,2012} - \bar{y}_{C,Sx,2011}\right) \quad (3)$$

where $\bar{y}$ denotes the mean score, Sx denotes subjects affected by *Mind The Gap*, Sz denotes subjects not affected by *Mind The Gap*, *T* denotes treatment schools and *C* denotes control schools.

3. Results

3.1 Baseline Differences

Despite the strictly random allocation of schools into treatment and control groups, the treatment group scored slightly better on baseline achievement in all four subjects. These differences were not statistically significant. In order to check whether the observed differences between the treatment and control groups were unusual, ten new draws of treatment and control groups were conducted using the same sampling frame and randomisation procedure. Table 2 reports the differences (in percentage points) between treatment and control groups for the actual draw as well as for the ten "placebo" draws. Note that the numbers reported in the table are the coefficients on the "treatment" dummy in an OLS regression predicting 2011 percentage scores based only on a treatment dummy. Evidently the differences between the actual treatment and control groups are not unusually large. This provides assurance that the treatment and control groups were well balanced.

Table 2. Typical baseline differences between treatment and control groups

	Accounting	**Economics**	**Geography**	**Life Sciences**
Actual	0.54	0.78	0.49	1.36
Placebo draw 1	-0.9	0.4	0.8	-1.1
Placebo draw 2	-0.63	1.36*	1.29	1.74**
Placebo draw 3	1.93**	0.95	2.55**	0.62
Placebo draw 4	1	-0.45	-0.45	0.04
Placebo draw 5	1.68*	0.78	-0.9	0.35
Placebo draw 6	0.27	-0.59	0.52	0.15
Placebo draw 7	-1.09	-0.19	-0.95	-1.62*
Placebo draw 8	-0.8	0.76	-1.34	-0.78
Placebo draw 9	0.99	0.79	0.87	2.94***
Placebo draw 10	-0.51	0.22	-1.02	0.34

Note: One, two and three stars denote statistical significance at the 90%, 95% and 99% levels of confidence respectively (* p<0.1, ** p<0.05, *** p<0.01).

3.2 Overall impact of the study guides on subject performance

Using kernel density curves, Figure 1 provides a descriptive view of the endline scores for treatment and control groups in each subject. These curves depict the proportions of pupils (along the vertical axis) who achieved specific scores (along the horizontal axis). A cursory glance at these graphs suggests that treatment and control schools had very similar distributions of economics scores; treatment schools had a marginally superior distribution of accounting scores; and treatment schools had noticeably superior distributions of geography and life sciences scores.

Figure 1. Kernel Density curves of subject performance in 2012 (endline)

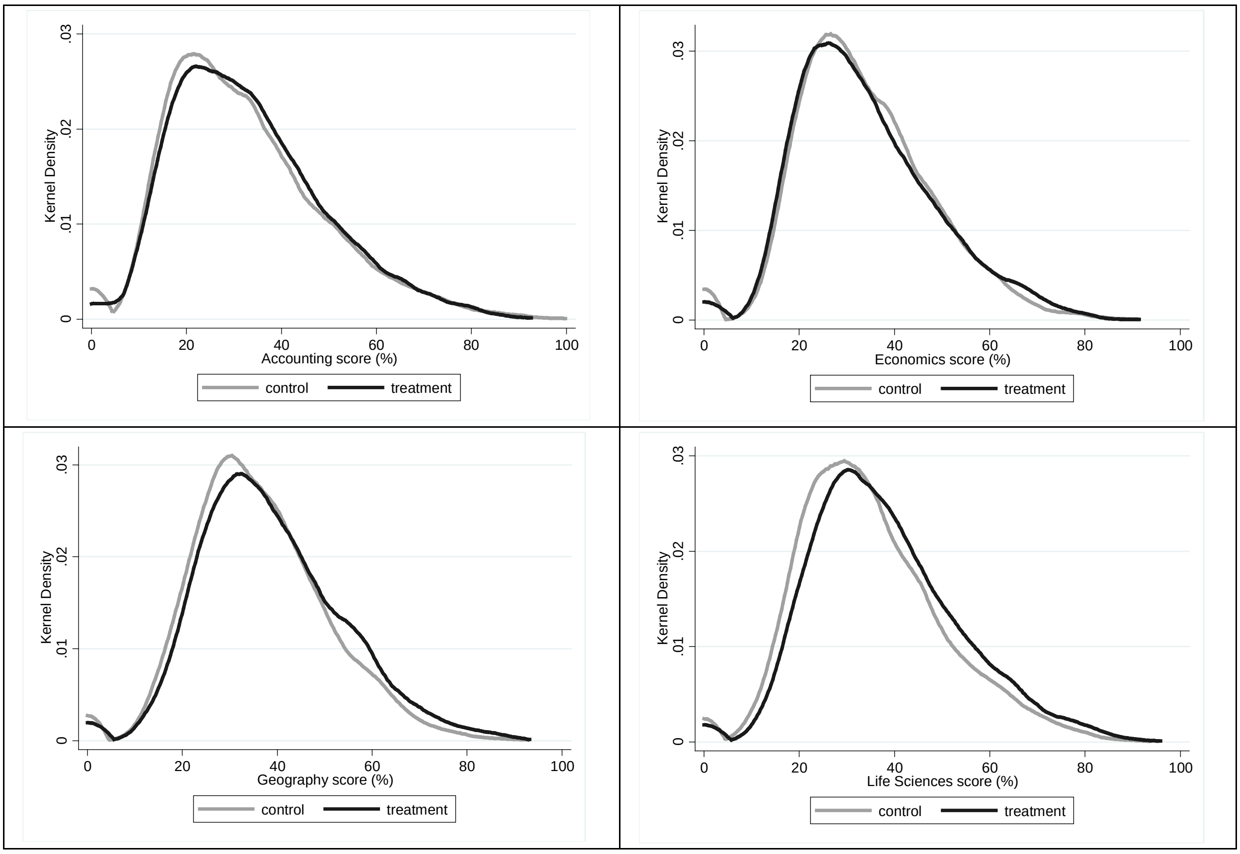


Table 3 reports a first set of “unconditional” regression estimates based on equation (1) above, predicting school mean achievement in 2012 controlling only for the treatment dummy. In the cases of accounting and economics treatment schools performed slightly better than non-treatment schools on average, but this difference was not statistically significant. However, the average scores in geography and life sciences were significantly higher for treatment schools than for control schools.

Table 3. "Unconditional" OLS regressions

	Accounting	Economics	Geography	Life Sciences
Treatment	0.84	0.17	2.39***	3.01***
	(1.12)	(0.88)	(0.89)	(0.81)
Quintile	No	No	No	No
R-square	0.0018	0.0001	0.0223	0.0421
Observations	318	318	318	318

Notes: In all school-level regressions, schools were weighted by the number of pupils that wrote that specific subject in 2012. Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Controlling for the school poverty classification hardly changes the estimated coefficients on the treatment dummy. This is to be expected given the randomised allocation into treatment and control groups. Table 4 reports these results.

Table 4. OLS regressions controlling for school poverty quintile

	Accounting	Economics	Geography	Life Sciences
Treatment	0.46	0.09	2.49***	2.93***
	(1.12)	(0.88)	(0.89)	(0.81)
Quintile	Yes	Yes	Yes	Yes
R-square	0.0481	0.0249	0.0488	0.0739
Observations	318	318	318	318

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 5 presents the main regressions for estimating the overall impact of the study guides. It is based on equation (1) and controls for the poverty quintile as well as the school mean achievement in 2011 (baseline). The estimates on the treatment dummy are slightly lower than in Tables 3 and 4. This is because the treatment schools scored marginally better on baseline achievement. Controlling for baseline therefore reduces the size of the estimated treatment coefficient slightly. Nevertheless, the overall picture is consistent with Tables 3 and 4: The impact of study guides was not significantly different from zero for accounting and economics, but was significantly positive for geography and life sciences. For both these subjects a positive impact was obtained at the 99% level of confidence. The size of the impact for both subjects was roughly two percentage points.

Table 5. Main OLS regression models controlling for baseline school performance

	Accounting	Economics	Geography	Life Sciences
Treatment	0.29	-0.37	1.94***	2.17***
	(0.89)	(0.76)	(0.63)	(0.63)
Mean 2011 score	0.74***	0.59***	0.64***	0.59***
	(0.05)	(0.05)	(0.04)	(0.04)
Quintile	Yes	Yes	Yes	Yes
R-square	0.4182	0.2892	0.5342	0.4344
Observations	315	317	315	317

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

The results of the triple-difference model, as explained using equations (2) and (3), are presented in Table 6. The positive coefficient on the "Year 2" dummy indicates that on average, over all subjects and schools, NSC scores were higher in 2012 than in 2011. Overall, those in treatment schools were not significantly better-performing than those in control schools. All four of the *Mind The Gap* subjects were harder than the average of the other six subjects combined. In 2011, treatment schools did not enjoy subject-specific advantages over control schools in any of the four *Mind The Gap* subjects, as indicated by the non-significant coefficients on the treatment-subject interactions. Accounting and geography appear to have been at a similar level of difficulty in 2011 and 2012, whereas performance in economics was better for all schools in 2012 and performance in life sciences was lower for all schools in 2012.

The triple interaction terms combining year, treatment school, and subject are the coefficients of interest. These indicate that there was no particular effect for being in a treatment school in 2012 in accounting or economics. However, for those in treatment schools in 2012 there was a relative advantage in geography and in life sciences. This confirms the earlier analysis demonstrating that study guides did not significantly impact on performance in accounting and economics and that the guides did have a positive impact on performance in geography and life sciences. The size of the impact is roughly in the same order of magnitude as in Table 5.

Table 6. Triple Difference OLS regressions across Year, Subject and Treatment

Dependent variable: **Score for student *i* in school *s* in time *t* in subject *z* (Y_{istz})**	
Year2	1.46***
Treatment school	0.95
ACCN	-9.24***
ECON	-8.98***
GEOG	-4.90***
LFSC	-3.34***
Treatment*ACCN	0.17
Treatment*ECON	-0.38
Treatment*GEOG	-0.58
Treatment*LFSC	0.51
Year2*ACCN	0.75
Year2*ECON	1.88***
Year2*GEOG	0.02
Year2*LFSC	-2.78***
Year2*treatment*ACCN	-0.45
Year2*treatment*ECON	-0.48
Year2*treatment*GEOG	2.04**
Year2*treatment*LFSC	1.58**
Quintile	Yes
Gender	Yes
R-square	0.0513
Observations	293334
Number of schools	317

Notes: Standard errors clustered at the school level; * p<0.1, ** p<0.05, *** p<0.01

3.3 Heterogeneous effects

Two possible types of heterogeneity in the impact of the study guides were explored. Firstly, it was established using a variety of models and for all four subjects that the impact of the guides did not differ by pupil gender. Secondly, the impact was allowed to vary with the baseline level of school performance. Table 7 shows results from models where the treatment variable was interacted with the school mean score in 2011 (baseline). The interpretation of the various coefficients is aided by the graphs in Figure 2 which show the predicted values from the models in Table 7.

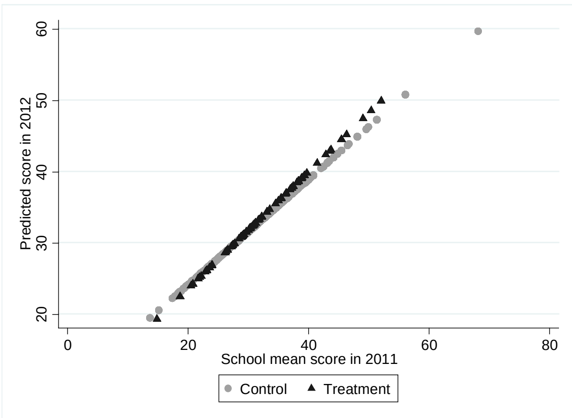
As before, the results confirm that geography and life sciences guides had an overall impact on performance while accounting and economics did not. However, the geography guides appear to have had a greater impact within schools with relatively higher baseline achievement. For accounting and economics the estimated impact of the guides, though not statistically significant, also appears to have been slightly higher at higher levels of baseline scores.

Table 7. OLS regressions interacting treatment with baseline measures

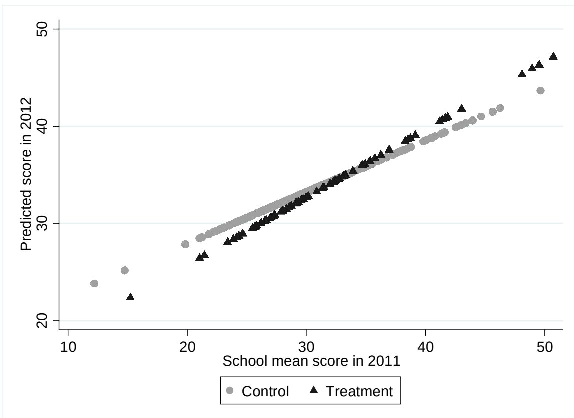
	Accounting	Economics	Geography	Life Sciences
Treatment	-2.17	-5.62	-3.61	3.42
	(4.11)	(3.87)	(2.90)	(3.57)
Mean 2011 score	0.74***	0.53***	0.59***	0.61***
	(0.06)	(0.07)	(0.04)	(0.05)
Interaction: Treatment & 2011	0.08	0.17	0.16**	-0.03
	(0.13)	(0.12)	(0.08)	(0.09)
Quintile	No	No	No	No
R-square	0.4038	0.2651	0.5332	0.4306
Observations	315	317	315	317
Treatment & Interaction Jointly significant	No	No	Yes***	Yes***

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

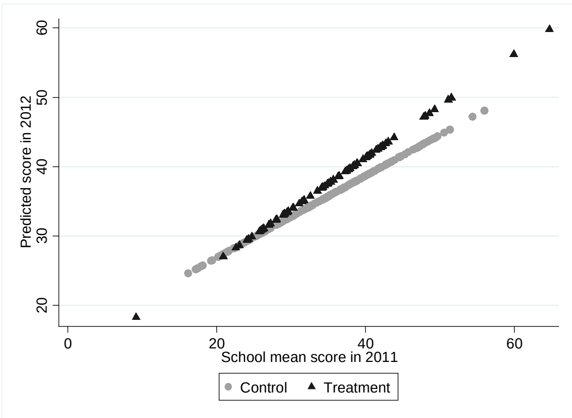
Figure 2. Predicted scores for treatment and control schools based on Table 5 estimates



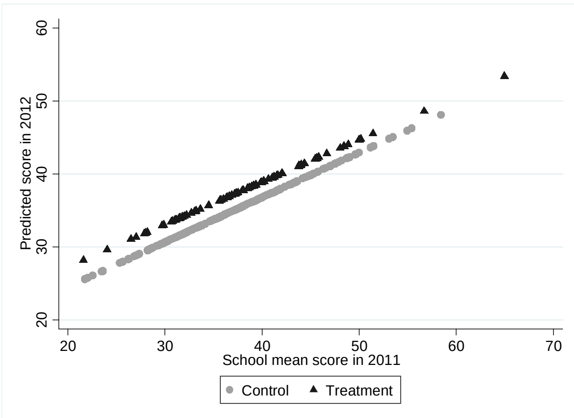
Accounting



Economics



Geography



Life Sciences

Taken together, these results suggest that better-performing schools in the sample were able to benefit more from the study guides. However, due to the type of schools in Mpumalanga chosen for this experiment, these "better-performing" schools had mean scores in the region of only 40% to 60%. Therefore, it is perhaps more meaningful to say that very poorly functioning schools were less likely to benefit from the guides. However, this finding applies only to geography. In life sciences the impact of the study guides was uniform throughout the spectrum of school performance.

It would also be interesting to know whether the study guides impacted equally on pupils at different points in the within-class distribution of performance. Since there was no pupil-level baseline measure, this is not possible to accurately estimate. However, one way to produce suggestive evidence is to change the outcome of interest from the school mean score to the school score at specific percentiles in the performance distribution. Table 8 reports the estimated coefficients on the treatment dummy for value-added regressions with the outcome variable being the 10^{th} percentile of achievement within a school, the 25^{th} percentile, the median, the 75^{th} percentile and the 90^{th} percentile, for each subject.

As before, no significant impact of the study guides was observed in the cases of accounting and economics, and this applied irrespective of the outcome measure used. Positive impacts were estimated for all models for geography and life sciences. Interestingly, in both of these subjects the size of the effect increased systematically as the outcome variable was set at higher percentiles in the performance distribution. This would suggest that the study guides had a greater impact for the stronger pupils within each class.

Table 8. OLS regressions with the outcome score set at various percentiles

	10th percentile	**25th percentile**	**Median**	**75th percentile**	**90th percentile**
Accounting	-0.04	0.55	0.22	0.59	0.2
	(0.86)	(0.84)	(0.97)	(1.17)	(1.39)
Economics	0.57	-0.07	-0.62	-0.35	-0.24
	(0.84)	(0.78)	(0.83)	(0.97)	(1.11)
Geography	1.33**	1.57**	1.82***	2.46***	2.57***
	(0.67)	(0.61)	(0.66)	(0.77)	(0.94)
Life Sciences	1.46**	1.79***	1.86***	2.92***	3.49***
	(0.65)	(0.65)	(0.69)	(0.80)	(0.93)
Quintile	Yes	Yes	Yes	Yes	Yes
2011 score at relevant percentile	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

3.4 Effects on passing the matric examination

If *Mind The Gap* study guides impacted on performance in geography and life sciences, then this should translate into more NSC passes overall. For some pupils an extra few percentage points in the geography or life sciences paper would have allowed them to pass that subject, and in some of these cases passing that subject would have been the difference between attaining an NSC and not.

Estimating the effect of the study guides on passing matric is not straightforward since there will be spillover across the subjects as some pupils would have taken more than one of these four subjects. For instance, although the economics study guides had no impact on economics scores one might observe a higher pass rate amongst economics pupils in treatment schools owing to the impact of geography or life sciences guides amongst those who took these subjects as well as economics.

Therefore, a more appropriate model predicts the school pass rate only counting those pupils who took both geography and life sciences. This model thus reflects the combined impact of receiving effective study guides in two NSC subjects. Table 9 reports the results. Only 260 schools had pupils taking both geography and life sciences. Amongst pupils taking both subjects in these schools the impact of receiving two effective study guides on attaining a NSC is estimated to be 6.5 percentage points. This large effect is not unsurprising given the distribution of scores. As Figure 1 illustrates, the bulk of pupils have scores between about 25% and 40%. An increment of 2 percentage points in two subjects is therefore likely to be the difference between passing and failing for many of these candidates.

Table 9. OLS regressions predicting pass rate amongst those taking geography and life sciences

	Pass rate 2012
Treatment	6.50**
	(2.69)
2011 pass rate	Yes
Quintile	Yes
R-square	0.2726
Observations	260

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

A different way of assessing the impact of the study guides on NSC passes is to consider what the effect would be on the national pass rate if guides were distributed at scale. To this end a simulation exercise was conducted using NSC data for 2010, since this was a year in which no pupils received *Mind The Gap*. We assumed that only the Geography and Life Sciences study guides impact on performance, and that those impacts are two percentage points in both cases. We then imputed an extra two percentage points to each candidate in the 2010 dataset for each of these two subjects, as long as their original mark was less than 80%.

According to our calculations (which differ slightly from official NSC 2010 figures due to excluding certain unusual candidates), 367 417 candidates passed the NSC in 2010 (a pass rate of 65.54%). Had pupils who took Geography and Life Sciences achieved two percentage points more than they did (as we predict the *Mind The Gap* study guides would have caused) then 373 026 pupils would have passed the NSC (a pass rate of 66.54%). This means that 5 609 children who did not pass the NSC in 2010 would have passed had the Geography and Life Sciences study guides been distributed nationally.

Placebo tests

Several placebo tests were carried out in order to check the likelihood that the observed impacts of the study guides were somehow an artefact of the data. The first set of placebo tests involved randomly selecting another 79 treatment schools and 239 control schools from the same sampling frame and then performing the same regression models on these groups. The schools that actually did receive study guides would have been randomly distributed across the new “fake” treatment and control groups and there is therefore no reason to expect the treatment dummy to be associated with performance. This procedure was repeated ten times and the results for the main value-added models for each subject are reported in Table 10. Of the 40 regressions that were carried out only one significantly positive effect was obtained, although there were three significant negative effects. This is broadly in line with the expected frequency of type I errors at given levels of statistical significance.

Table 10. OLS regressions on placebo draws of treatment and control schools

Placebo dif in dif	**Accounting**	**Economics**	**Geography**	**Life Sciences**
Placebo draw 1	0.8	-0.15	-1.11*	-0.58
Placebo draw 2	0.19	-0.42	-1.32**	-0.25
Placebo draw 3	0.3	-0.09	0.94	0.92
Placebo draw 4	-0.17	-0.26	0.5	0.21
Placebo draw 5	1.16	0.97	1.27**	0.36
Placebo draw 6	0.18	0.46	-0.56	-0.38
Placebo draw 7	0.52	0.07	-0.3	-0.1
Placebo draw 8	-2.66***	-1.17	-0.01	0.06
Placebo draw 9	-0.44	0.32	0.29	-0.65
Placebo draw 10	1.23	1.04	0.94	-0.44
Baseline	Yes	Yes	Yes	Yes
Quintile	Yes	Yes	Yes	Yes

Notes: Only the coefficients on the treatment dummy variable are shown.
* p<0.1, ** p<0.05, *** p<0.01

A second set of placebo tests used the actual treatment and control groups but estimated the regression models for other subjects not covered by *Mind The Gap*. Model 1 replicates Table 3, Model 2 replicates Table 4 and Model 3 replicates the main value-added models as reported in Table 5. In the two models that do not control for baseline achievement a spurious treatment effect is observed for two subjects. However, once the baseline school achievement is controlled for there are no statistically significant “impacts” of being in a treatment school for non-treated subjects.

Table 11. Three main regression models applied to other subjects

	Model 1	Model 2	Model 3
Mathematics	2.04*	2.41**	0.80
Physical Science	1.48	1.57	0.64
English FAL	0.66	0.51	0.13
Business Studies	0.98	0.73	0.35
Maths Literacy	1.87**	1.74**	0.73
Tourism	0.95	0.61	0.22
Quintile	No	Yes	Yes
Baseline	No	No	Yes

Notes: Only the coefficients on the treatment dummy variable are shown.
* p<0.1, ** p<0.05, *** p<0.01

A final set of placebo tests predicted 2011 performance using 2010 performance as the baseline. Six models were estimated for each *Mind The Gap* subject: The standard model predicting school mean score controlling for quintile and baseline mean score as well as the models with the outcome being school performance at various percentiles of within-school achievement. In none of these models was a statistically significant coefficient on treatment obtained.

Table 12. Predicting 2011 outcomes with 2010 as a baseline

	Mean	10th pctile	25th pctile	median	75th pctile	90th pctile
Accounting	0.09	0.15	-0.16	-0.36	0.25	0.73
Economics	0.24	0.28	0.11	0.2	0.34	0.48
Geography	0.9	1.35	1.12	1	0.29	0.61
Life Sciences	0.87	0.89	0.43	0.77	1.09	1.38
Quintile	Yes	Yes	Yes	Yes	Yes	Yes
Baseline (2010)	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Only the coefficients on the treatment dummy variable are shown.
* p<0.1, ** p<0.05, *** p<0.01

The various placebo tests therefore provide assurance that the observed impacts of study guides on geography and life sciences performance were not the result of a spurious data generating process.

4. Discussion

The main findings of the study can be summarised as follows. *Mind The Gap* study guides had no observable impact on accounting and economics performance, but did positively and significantly impact on performance in geography and life sciences. There was some evidence in the case of geography that the impact was lower amongst seriously underperforming schools. Improvements in geography and life sciences achievement led to an increased probability of attaining a National Senior Certificate.

The fact that there was an impact, albeit in only two of the four subjects, is perhaps surprising given how light an intervention this was: No training or awareness campaign accompanied the delivery of the study guides to treatment schools; moreover, pupils only had between one and two months to use the guides before writing their final exams. On the other hand, the timing of delivery may have been fortuitous in one sense. It is well known that the management of learning materials leaves much to be desired in many South African schools, with reported cases of boxes remaining unopened in store rooms and books getting lost throughout the course of a year. Therefore, the arrival of a new resource at exactly the time in the school calendar when all energies are targeted toward the high-stakes NSC examination may have provided the right incentive to make proper use of the guides. This relates to an important interaction between policy interventions and the existing incentives facing stakeholders: when the incentives facing stakeholders are strongly aligned to the goals of interventions these interventions will be most effective.

The findings in this study relate to several broad debates of policy and theoretical interest. Glewwe *et al* (2007: 27) suggest that the textbooks experiment in Kenya pointed to a curriculum mismatch in that country: textbooks reflecting ambitious curriculum requirements were only appropriate for the most talented individuals. Similarly, Banerjee and Duflo (2011) argue that in many developing countries an unintended negative consequence of ambitious curricula is that the majority of pupils fall behind. Despite recent improvements and simplifications in South Africa's Curriculum and Assessment Policy Statements (CAPS), national assessments demonstrate that across the grades the majority of pupils are performing substantially below curriculum requirements. One possible implication of this study is that well designed study guides, which extract the core subject material, can partly mitigate the unintended consequences of an ambitious curriculum in weakly performing schools. However, the fact that the study guides benefitted the stronger learners within each school more than the weaker learners could reflect that strong learners had been delayed by teaching to a lower level in the class but were able to fulfil more of their potential through a new exposure to the entire curriculum through the study guides and at an individual-specific pacing. In many developing countries and in South Africa, there are many classrooms with a wide range of cognitive ability present, with many learners several years behind the grade-specific curriculum (Pritchett and Beatty, 2012; Spaull and Kotze, 2015). Interventions that offer learners an alternative site for learning allowing for a differentiated pacing, such as through study guides or computer assisted learning programmes, hold potential for improving learning outcomes.

This study also suggests that incremental improvements in educational performance are possible through interventions that do not involve teacher training or teacher mediation. Given the strong deterministic influence of pupil socio-economic status and the central role of teachers (e.g. Hanushek, 2010), some are pessimistic about the impact of isolated interventions when these are not well integrated within deeper instructional and organisational reforms. Richard Elmore (1996: 25) argues that, "getting to scale with good educational practice requires nothing less than deliberately creating and reproducing alternatives to the existing flawed institutional arrangements and incentives structures." While fundamental and substantial systemic change no doubt does require this level of institutional reform, other research indicates that specific remedial interventions that to some extent bypass existing institutional arrangements can lead to incremental improvements. Banerjee *et al* (2007) report on an RCT in India where remedial classes administered by young women in the community significantly improved basic literacy and numeracy amongst primary school children. Similarly, Banerjee *et al* (2010) found that remedial reading camps run by volunteer youth in India improved reading acquisition. The *Mind The Gap* RCT provides another example of a specific intervention that improved performance without fundamentally reforming instructional practice and institutional arrangements.

The finding that, at least for geography, the study guides had a larger impact on better-performing schools may relate to the finding that additional resources in South African schools matter conditionally upon the quality of school management and overall functionality (Van der Berg, 2008). It is often found in education production function analyses that interventions and resources which may be intended to be pro-poor are more effective in more affluent schools, perhaps because the success of these interventions is dependent on other capabilities and the accountability structures within schools. It may well be that well managed schools with a greater academic emphasis effectively distributed the study guides to pupils and more actively promoted their use.

The fact that two of the study guides had an observable impact and two did not, despite the same delivery procedures, should caution against simplistic views on whether resources such as textbooks matter. For example, the findings of the influential textbooks RCT in Kenya (Glewwe *et al*, 2007) should not be categorically interpreted as evidence that textbooks do not improve performance in African countries. The contextual caveats regarding language proficiency of pupils and curriculum mismatching must be given due consideration. Moreover, as Bold *et al* (2012) demonstrate, the quality of treatment implementation in RCTs can play a strong role in determining whether an impact is observed. This study demonstrates that even when implementation is exactly the same, impact can vary across subjects.

Yet the burning question remains: why did the geography and life sciences guides have an impact but not the accounting or economics guides? This study cannot conclusively answer this question, but one can think of several plausible hypotheses. The first possible explanation is that the quality may have differed across the guides. There is no prior reason to expect this given the similar and extensive processes of material development by teams of teachers, examiners, curriculum officials and academics. However, separate teams of experts were involved in each subject so the possibility cannot be ruled out.

A second possibility is that contextual factors predisposed geography and life sciences to benefit from study guides but were different for accounting and economics. For example, study guides may have different effects depending on the availability and quality of existing materials in the specific subjects. For example, the geography study guides contained map-work materials, something which is crucially important yet particularly scarce in many poorly resourced schools. Secondly, accounting and economics are arguably more skills-intensive and less content-based than geography and life sciences. Sudden improvements through intensive studying are perhaps less likely in accounting and economics than in content-based subjects.

It is important for policy-makers to realize that interventions such as this, and similarly the DBE workbooks intervention which provides workbooks to all children in grades R to 9, should not be viewed as blanket interventions. Rather, certain grade or subject-specific materials may be more or less effective than others due either to quality differences or contextual factors. In particular, materials should continuously be revised to ensure relevant and high quality products. This is especially relevant when materials are centrally developed rather than in the private sector.

What can be concluded about the magnitude of the observed impacts and the expectations in scaling up this intervention? Amongst the 318 schools in Mpumalanga used for this study, one standard deviation of pupil performance was 15.1 percentage points for life sciences and 14.4 percentage points for geography.[8] Therefore, the coefficient in the main regression for geography of 1.94 translates into an impact of 13.5% of a standard deviation, while the coefficient for life sciences of 2.17 translates into an impact of 14.4% of a standard deviation. One rule of thumb is that a year of learning amounts to between 40% and 50% of a standard deviation in test scores (e.g. Filmer *et al*, 2006). Viewed in this light, the impact of the geography and life sciences study guides was large: Providing pupils with a study guide lead to an increase in learning roughly equivalent to a third of a year of schooling. While this casts a favourable light on the *Mind The Gap* intervention, the fact that this sort of “slack in the rope” exists points to serious deficiencies in teaching and learning in South African schools.

Since we do not know how many pupils actually received and used the study guides, we are only able to measure the so-called Intent-to-treat (ITT) effect. The true impact on those pupils who actually used the guides was probably larger than the average ITT of about 2 percentage points. If, for argument’s sake, 75% of study guides were actually handed out to pupils and 50% of learners ultimately used the books, the true impact on those using the books would be about 4 percentage points. If taken to scale and given wide systemic support there may be better use of the guides and therefore a larger effect than what was estimated in the impact evaluation. On the other hand, delivery was seamless in the RCT since a private courier company was used. Inefficiencies in delivery using standard procedures relying on distribution through district offices may contribute to some attrition of the ITT impact.

The NSC result simulation demonstrated that the expected increase in the national NSC pass rate associated with a national distribution of geography and life sciences guides is approximately one percentage point. However, amongst those *individuals* taking both these subjects the increase in the pass rate is estimated to be about 6.5 percentage points.

This is an enormous impact given the low cost of this intervention. The unit cost per study guide (reflecting material development, printing and distribution) is estimated to be R41.82. Using a technique from Kremer, Brannen and Glennerster (2013), it is possible to estimate how many standard deviations of test score improvement could be generated per $100 invested in the study guide intervention. This comes to 3.04 standard deviations per $100.[9] This places *Mind The Gap*

[8] Nationally, the standard deviation is only slightly larger. For geography it is about 15 percentage points and for life sciences it is about 18 percentage points.

[9] A rand dollar exchange rate of 0.11 (as at 8 May 2013) was used. As per the earlier discussion, it was estimated that each geography and life sciences study guide led to a test score improvement of 0.14 standard deviations.

study guides amongst the five most cost-effective interventions out of the 17 interventions for which Kremer *et al* (2013: 298) calculated this value.

One potential criticism of this intervention is that it may not produce a sustained improvement in skills but only help pupils prepare for a specific test. There may be an element of truth in this, but even so, the ability to prepare for an examination is an important skill and one which is no doubt better taught in South Africa's top-performing schools, which serve more affluent children. Therefore, one might argue that a study guide can partly offset the artificial difference in NSC results reflective of better guidance rather than underlying skill. Moreover, through increasing the probability of passing the NSC, the study guides helped pupils in what is an important step towards labour market success.

5. Conclusion

This impact evaluation of *Mind The Gap* study guides represents an example of how governments can conduct rigorous impact evaluations of specific programmes at fairly low cost. This requires coordination between programme managers and evaluation experts, preferably prior to programme roll out. Randomised Controlled Trials of course provide high internal validity and are a fair way to distribute resources when there is a budget constraint to operating at scale or when doing piloting activities. They can be expensive but when outcomes data, such as national examinations data, exists the cost is substantially lower and the artificial experimental conditions, which are often present in an RCT, are reduced. Even when an RCT is not possible, such as when a programme is in full scale operation, other quasi-experimental techniques can sometimes be employed with a little creativity. These include the setting of a clear policy rule to determine programme allocation (allowing Regression Discontinuity Techniques to be used), encouraging or incentivising certain randomly chosen groups to participate in an available intervention (thus constructing an Instrumental Variable), or merely keeping accurate administrative records of who received interventions at which time (allowing difference-in-difference and matching techniques to be used).

The primary conclusion of this paper is that the *Mind The Gap* programme can be expected to significantly impact on the performance of NSC candidates. The fact that two study guides did improve performance indicates that the basic intervention design can accomplish its intended purposes. However, the fact that two guides did not noticeably improve performance illustrates that even well-designed interventions might sometimes not have the intended impact when certain contextual factors prevent this. Rigorous impact evaluation can help dispel the myth that ostensibly well designed interventions must surely be effective and can alert policy-makers to cases where barriers to effectiveness may exist.

This experiment using study guides also has bearing on the literature on textbooks and school resources more generally. It indicates that additional learning support materials can be a cost-effective means to improve pupil achievement. The finding of greater impact within relatively well-performing schools is consistent with other literature arguing that resources matter conditionally upon school management (Van der Berg, 2008). The finding that study guides shifted achievement most at the high end of the within-class distribution is consistent with other findings that textbooks help high ability pupils most (Glewwe *et al*, 2007).

Finally, this study illustrates that even when the treatment implementation is essentially identical, there can be different results across subjects. Therefore, influential RCTs on textbooks must be interpreted with sensitivity to their context rather than taken as simplistic answers to questions such as, “do textbooks matter?” Future studies should aim to illuminate the conditions necessary for interventions, such as additional materials, to impact on learning.

6. List of References

Banerjee, A., & Duflo, E. (2011). *Poor Economics.* London: Penguin Books.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy, 2*(1), 1-30.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics, 122*(3), 1235-64.

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2012). *Interventions and Institutions: Experimental Evidence on Scaling up Education Reforms in Kenya.* Draft Paper.

Branson, N., Garlick, J., Lam, D., & Leibbrandt, M. (2012). Education and Inequality: The South African Case. *SALDRU Working Paper*(75).

Department of Basic Education. (2013). *Developing and evaluating the first phase of the grade 12 Mind The Gap study guide series.* Pretoria.

Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Educational Review, 66*(1), 1-16.

Filmer, D., Hasan, A., & Pritchett, L. (2006). A millenium development goal: Measuring real progress in education. *Center for Global Development Working Paper, 97*.

Fleisch, B., Taylor, N., Du Toit, R., & Sapire, I. (2010). Can workbooks improve learner performance? Findings of the randomised control trial of the Primary Mathematics Research Project. *Wits School of Education Research Seminar.* Johannesburg.

Fuller, B. (1986). Raising school quality in developing countries: What investments boost learning? *The World Bank*(Washington D.C.).

Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classrooms, tools, rules and pedagogy. *Review of Educational Research, 64*(1), 119-157.

Glewwe, P. (2002). Schools and skills in developing countries. *Journal of Economic Literature, XL*, 436-482.

Glewwe, P., Kremer, M., & Moulin, S. (2007). Many children left behind? Textbooks and test scores in kenya. *NBER Working Paper13300*.

Gustafsson, M. (2007). Using the Hierarchical Linear Model to Understand School Production in South Africa. *South African Journal of Economics, 75*(1), 84-98.

Hanushek, E. (2010). The economic value of higher teacher quality. *National Center for Analysis of Longitudinal Data in Education Research, Working Paper No.56*.

Hanushek, E. A. (2002). The Failure of Input-based Schooling Policies. *National Bureau of Economic Research Working Paper*(9040).

Heyneman, S., Farrell, J., & Sepulveda-Stuardo, M. (1978). Textbooks and achievement: What we know. *World Bank Staffing Paper, No.298*.

Heyneman, S., Jamison, D., & Montenegro, X. (1984). Textbooks in the Philippines: Evaluation of the pedagogical impact of nationwide investment. *Educational Evaluation and Policy Analysis, 6*(2), 139-150.

Horton, S., & Lovitt, T. (1989). Using study guides with three classifications of secondary students. *The Journal of Special Education, 22*(4), 447-462.

Jamison, D., Searle, B., Galda, K., & Heyneman, S. (1981). Improving elementary mathematics education in Nicaragua: An experimental study of textbooks and radio on achievement. *Journal of Educational Psychology, 73*(4), 556-567.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The Challenge of Education and Learning in the Developing World. *Science, 340*, 297-300.

Lockheed, M., & Hanushek, E. (1988). Improving educational efficiency in developing countries: What do we know? *Compare, 18*(1), 21-38.

Pritchett, L., & Beatty, A. (2012). The negative consequences of overambitious curricula in developing countries. *Center for Global Development Working Paper, 293*.

Shepherd, D. (2011). Constraints to school effectiveness: What prevents poor schools from delivering results? Stellenbosch.

Spaull, N. (2011). A preliminary analysis of SACMEQ III South Africa. *Stellenbosch Economic Working Papers*(11/11).

Spaull, N., & Kotze, J. (2015). Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics. *International Journal of Educational Development, 41*, 13-24.

Taylor, S. (2010). *The performance of South African Schools: Implications for Economic Development.* Unpublished PhD Thesis, University of Stellenbosch, December 2010.

Taylor, S. (2011). Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study. *Stellenbosch Economic Working Papers*(10/11).

Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation, 34*, 145-154.