This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document 👇

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND
Browse Books & Publications
Make a charitable contribution

For More Information
Visit RAND at www.rand.org
Explore RAND Europe
View document details

Limited Electronic Distribution Rights
This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND PDFs to a non-RAND Web site is prohibited. RAND PDFs are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see RAND Permissions.
This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.
Ten years of reform in primary mathematics education in England

A review of effectiveness

Ben A. Vollaard, Lila Rabinovich, Richard Bowman, Christian van Stolk

Prepared for the National Audit Office
The research described in this report was prepared for the National Audit Office.
Executive Summary

The UK National Audit Office (NAO) commissioned RAND Europe to look at the evidence showing the effectiveness of the reform of mathematics (herein abbreviated to ‘maths’) teaching in primary schools in England. This study looked at government-sponsored evaluations of maths reform in England, independent evaluations of the educational outcomes of maths reforms in England, and the international evidence regarding some of the main components of government policy aimed at improving educational outcomes. The main findings were:

Maths teaching has undergone significant changes in England over the last two decades
Primary school education in England has undergone major reforms in the last 20 years, including the introduction of a National Curriculum at the end of the 1980s, national testing with league tables for schools in the 1990s, and detailed guidelines for teaching laid down in the National Numeracy Strategy shortly before the new millennium. These included new guidelines on the structure and content of maths lessons in primary schools in England. Additionally, spending per primary school pupil has greatly increased, with pupil to teacher ratios consistently falling over the last 10 years, according to the Survey of Teacher Numbers and Teacher Vacancies. To speed up progress in improving attainment, the Government has announced several initiatives, including a major review of primary maths teaching, focusing on subject knowledge of teachers (the ‘Williams Review’), additional support for the lowest attaining primary school pupils (‘Every Child Counts’), and pilots to provide personalised maths tuition (‘Making Good Progress’).

Government-initiated evaluations of reforms focus primarily on their implementation
Government-initiated evaluations tend to focus on the challenges and opportunities in the implementation of the reforms. Most of the evaluations discuss progress in the structure of mathematics education and teaching practices, rather than changes in pupils’ educational attainment.

Nevertheless, this focus is still important because, when evaluating the effect of reforms in maths education, it does not suffice to look at changes in pupils’ educational attainment. A crucial issue is how these changes came about, and implementation has a central role in assessing what worked, what did not work, and why. As the evaluations indicate, classroom practice is not always in line with official policy, at least not immediately, and reforms are not always uniformly implemented (Kyriacou, 2005: 173). In other words, the ‘treatment administered’ may differ from the ‘intended treatment’. The relationship between reforms and pupils’ attainment is mediated by the way in which changes are implemented, and the
resources – in terms of finances, human capital and time – that are available to put changes in place.

The impact of the reforms on pupil attainment are not robustly assessed in the government-initiated and independent evaluations

The lessons that can be learned from both government-initiated and independent evaluations are limited for the following reasons:

- Evaluation and monitoring are often not considered at the design stage of the reform and reforms are mostly not piloted.
- When evaluating reforms that are uniformly 'prescribed' across the nation, it is difficult to include a control group.
- There is a general absence of evaluations that track performance over time, follow cohorts of pupils through the system, and methodically assess the factors that could shape this performance.
- Some of the reforms have been implemented only to a limited extent, precluding evaluation of some of the initiatives.
- The reforms often included a package of measures, which makes it difficult to see which (combination of) measures made the difference and which measures were not effective.
- It is often difficult in education evaluations to attribute impacts to individual factors such as teacher quality, class sizes, pedagogy, and types of assessment that influence educational outcomes.
- Many evaluations look solely at educational outcomes in terms of national test scores, without comparing progress in educational outcomes with independent and international tests on maths ability such as Trends in International Mathematics and Science Study (TIMSS) and PISA (Programme for International Students Assessment) or placing such test scores in the context of wider educational attainment in terms of further study or job outcomes.

Government-initiated evaluations show that while the prescribed format and structure for maths lessons have been adopted there is little evidence of ‘deeper change’ in teaching methods and improvements in pupil learning

In this study, we looked mostly at government evaluations of the Primary National Strategy (PNS) and National Numeracy Strategy (NNS). These strategies were the most important reforms of maths education over the last ten years and they therefore merit special attention. The evaluations (Earl et al 2003; Ofsted 2002; Ofsted 2005; and Ofsted

---

1 See also Chapter 5.
3 See http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html, (accessed June 2008).
4 These reforms are described in Chapter 2.
show that the changes in classroom practice tend to be in line with the NNS guidance, with most teachers using the format and structure of the prescribed daily mathematics lesson. Fundamental changes in methods of teaching – including ‘assessment for learning’ and ‘stimulating mental calculation strategies’ – have been slow in coming, even many years after the implementation. ‘Deeper’ change in teaching and learning has been hampered by poor subject knowledge, poor leadership, and limited understanding of the purposes of the NNS and PNS.

The research design of the evaluations does not allow for any definitive claims regarding the effectiveness of the reform in improving pupil attainment. The evaluations do not include control groups and do not control for other trends (such as the declining pupil to teacher ratio) that may have affected pupil test scores. As a result, the studies do not provide more than some hints as to what the impact of the reforms could have been. The results suggest that the impact of the NNS on pupils’ attainment was mostly limited to the time period immediately after its implementation – when teachers’ motivation was high and easy-to-implement changes in lesson structure were realised. In later years, gains in test scores were minimal.

High-stake testing5 was found to result in considerable test preparation in the term leading up to the national assessments. How this affects pupils’ learning is not clear, however.

**Independent evaluations show a similar range of improvement in maths and attribute at least part of this to the NNS**

The three independent evaluations (Brown et al 2003; Anghileri 2006; Basit 2003) show surprisingly similar overall gains in maths attainment up to four years after the implementation of the NNS. According to these evaluations, test scores improved by some 10 percent in the three to four years following the introduction of the NNS.

None of the studies provide hard evidence of the extent to which the gain in scores can be attributed to the NNS and later reforms. All studies are based on simple before/after comparisons, which leave the results open to alternative interpretations. Other national trends may well have played a role. However, the studies provide some indications that the NNS made at least part of the difference. Firstly, the effects of the NNS can be traced to relatively strong gains in test scores in areas that were emphasised in the NNS, including the number system and mental calculation. Secondly, the international comparative study, TIMSS, shows that gains in test scores over 1995–2003 only show up in year 5 and not in year 9, which suggests that the NNS made the difference, since the NNS is aimed at Reception through to year 6. Thirdly, lesson structure has evidently changed in line with the NNS and PNS, this being a necessary condition for the reforms to have had an effect. Nonetheless, like government-initiated evaluations, the independent evaluations find little evidence of more fundamental changes in teaching.

---

5 High stakes tests are defined as those tests that carry serious consequences for students or educators. The consequences from standardised achievement tests range from decisions on passing and failing in subjects for pupils to rewards or punitive measures for schools and school districts (see e.g. http://www.eplc.org/mpearlman.html [accessed August 2008]).
The independent evaluations show that whole class teaching is of little benefit to low attaining pupils, and appears to favour boys over girls

All studies find a greater variance in test scores pre-reform compared with post-reform, with the lowest-performing quintile of pupils continuing to lag behind. The evaluations suggest that the greater variation could be the result of the emphasis on (interactive) whole class teaching, which was one of the requirements introduced by the NNS. The needs of low-attaining pupils are not fully met in a whole class setting; on average, they made almost no gain in attainment in maths. Some high-attaining pupils felt frustrated at their progress being held back; on average, their improvement was lower than that of average-performing pupils. Based on interviews and classroom observations, the evaluations suggest that interactive whole class teaching favours boys over girls, with girls perceiving it as competitive and unpleasant. Just like government-initiated evaluations (Earl et al. 2003 and Ofsted 2002), the independent evaluations find greater gains in scores for boys than for girls. TIMSS, however, does not confirm this result.

The study looked at international evidence from a number of interventions aimed at improving educational outcomes

This study examined the evidence in the literature on the effectiveness of a number of interventions, consisting of: class size reductions; formative assessments (group teaching versus whole-class teaching); pedagogy (types of teaching); teacher quality; and standardised testing. These areas were identified in meetings between the NAO and the RAND Europe study team and were considered especially relevant for evaluating the past and present strategy of Government to improve maths outcomes. They are some identifying features of reforms of maths in the last decade.

This overview aims to give important indications of effectiveness on the basis of meta-analyses and important field studies. However, this report does not aim to offer a comprehensive overview of all the evidence and the validity of each research approach. There are some important observations about the international evidence. A first observation in most studies is that attributing the effect on educational outcomes to specific factors and controlling for others remains difficult. Second, the meanings of concepts in education often overlap or are somewhat nebulous. There are overlaps between the use of formative assessment and teacher quality as a factor in improving educational outcomes and indeed between certain types of pedagogy, formative assessment and teacher quality. Moreover, the concept of pedagogy can include a wide range of teaching interventions. Therefore, a report like this has to take care not to confuse effects or overemphasise specific effects. Thirdly, the evidence regarding the impact of many interventions is often not specific to maths skills, but rather focuses on more general educational outcomes, effect sizes, or the equivalent of additional months of education in a given year. Thus, studies that look at the role of interventions such as reduction of class sizes, formative assessment, and personalised teaching in improving maths have to take care that the evidence is specific to the area of maths or indeed generalisable.

---

6 Schools’ use of assessment data to tailor the teaching to the pupil’s need

International literature suggests that teacher quality and the use of formative assessment are important factors in improving educational outcomes, while the effect of class size reductions is more moderate.

Several studies point to teacher quality as the most important factor in affecting educational outcomes. This is an important observation, as teacher quality is often not the central component of government reforms. A study by Barber and Moursesh (2007) emphasises the importance of teacher quality and cites 1997 data from Tennessee on the difference in pupil outcome for pupils with low- and high-performing teachers over a three-year period. Other studies (Hedges et al 1994; Darling-Hammond 1998) have also found that teacher quality has a positive impact on pupil achievement. These studies raise important questions on which aspect of teacher quality (content knowledge, experience, pedagogy) contributes most to improvements in educational outcomes and also how governments can best support the development of teacher quality.

Formative assessments allow teachers to provide diagnostic feedback about their pupils’ strengths and weaknesses and their instruction is guided by the results of assessments. Several meta-analyses (for instance Black and Wiliam 1998a and Black and Wiliam 1998b) show that formative assessments can have a positive effect on educational outcomes. The effect sizes mentioned range between 0.4 and 1.8. In some cases the effect sizes studied were as large as in one-to-one instruction (Bloom 1984). From the evidence, the use of formative assessments appears an attractive proposition to policy-makers. However, the main question around the use of formative assessment is scalability. By definition formative assessment is a teacher-led intervention and studies have not looked at the effectiveness of the use of formative assessment when applied in a standardised way across an educational system.

There has been substantial research on the impact of the reduction of class size on educational outcomes. Though most studies highlight the importance of educational resources in promoting better educational outcomes, most agree that a reduction in class sizes will have only a moderate effect on pupil achievement (also distinguishing between educational outcomes and attainment). Moreover, the effect sizes differ for the various attainment groups and the intervention is potentially costly compared to other interventions for the same effect size. If class size reductions are to be part of the policy mix, these reductions should be directed at areas where the evidence shows the largest gains can be made, namely in classes containing low-achieving students and those in the first years of schooling.

There are important trade-offs in terms of cost-effectiveness, unintended outcomes and targeting of ability groups to consider when looking at the effectiveness of specific interventions.

Some studies explicitly look at trade-offs between particular interventions. Several studies have found that small group instruction tends to have a greater effect on pupil achievement than whole class teaching (see Kulik and Kulik 1987; a meta-analysis in Davidson 1985; and Johnson et al 1995). However, the studies also found that improvements in pupil

---

8 The effect size is the ratio of the average improvements in the test scores of pupils involved in an innovation over the range of scores for typical groups of pupils on these same tests (taken from Black and Wiliam 1998b).
achievement are often associated with particular teacher and pupil behaviours and interactions. Thus, the real underlying factors driving improvements in pupil achievement might not necessarily be the specific organisational arrangements in place but some of the characteristic behaviours found in interactions of smaller groups. These might also be replicated in whole class teaching.

A more recent study by Wiliam (2007) in the United Kingdom looked at trade-offs between formative assessment and reductions in class sizes in terms of their cost-effectiveness. He found that the use of formative assessment is about 20 times more cost-effective than the reduction of class sizes to achieve a similar gain in pupil achievement. Thus, the cost of interventions and the comparison of costs between interventions are areas where more research would be welcome and are important for policy-makers to consider when undertaking educational reforms.

In addition, certain interventions might have unintended consequences. For instance, class size reductions are associated with an increase of teachers within the educational system. Some observers have pointed to the effect of class size reductions on the quality of teachers, as new and not always the most qualified teachers are brought into the educational system. Such unintended effects could have consequences for pupil achievement.

Finally, certain interventions specifically affect pupil achievement in particular ability groups. Group teaching tends to have the largest effects in medium- and high-ability groups (respectively Webb 1991 and Good et al 1992). Though the evidence shows that class size reductions have a limited effect on educational outcomes, they do tend to be more effective in low-ability groups and with younger-aged pupils (Krueger 2000). Formative assessment tends to show the largest effects in low-ability pupils. These observations have to be taken into account when looking at wider trade-offs between interventions.

**There is limited evidence regarding the impact of standardised testing**

Standardised testing often provokes an emotive debate about its merits and its effects on educational outcomes. A positive aspect of such testing is that it makes it easier to track progress in educational outcomes. A negative aspect is that it could lead to the narrowing of the curriculum and ‘teaching to the test’, which might have a detrimental effect on numeracy skills. In poorly designed regimes some negative outcomes are noted from standardised testing, such as doubts over the validity and the robustness of reported test scores, and ‘teaching to the test’ is noted in some government-initiated evaluations in England. However, there is no systematic evidence that standardised testing has a limiting or negative effect on overall educational outcomes.