

The 1997 Education White Paper - a failure of standards

(Plewis, I. and Goldstein, H. (1998). Excellence in Schools: a failure of standards. *British journal of Curriculum and Assessment* 8: 17-20.)

by

IAN PLEWIS and HARVEY GOLDSTEIN

MATHEMATICAL SCIENCES

INSTITUTE OF EDUCATION, UNIVERSITY OF LONDON.

Introduction

February 1997 saw the publication of an influential report (Literacy Task Force, 1997a) on literacy, written by a 'task force' set up by the Labour Party and chaired by Professor Michael Barber¹. Following Labour's victory in the May 1997 election, many of the ideas from that report were incorporated into the education White Paper, 'Excellence in Schools', which was published in July. In this article, we take a critical look at the assumptions which underpin the literacy task force report, and which in turn provide support for some of the policies proposed in the White Paper. We do so from the perspective that good educational research can, and should, inform and influence policy, and with the hope that it might do so more in the near future than it has in the recent past.

Our criticisms fall into three sections. In the first section, we assess the evidence presented in the two reports (which we abbreviate to LTF and EIS) on school differences and examine how this evidence has been used. In the second section, we discuss the implications of setting national targets for levels of achievement. Finally, we consider the recommendations about the amount of school time that should be devoted to literacy at Key Stages One and Two.

Differences between schools

The conclusions in LTF are strongly influenced by an analysis of the 1995 Key Stage Two English results. The report demonstrates that there is substantial variation between schools in the percentage of pupils reaching the target level, that is level four, even after allowing for differences between schools in the percentage of pupils receiving free school meals. An almost identical diagram is found in the White Paper although it is based on the 1996 results. We do not believe that these results are presented especially well - for example, the scale for the percentage of pupils at level four or above in LTF goes down to -20%, and in both reports there is some confusing material about 'outliers', that is schools which might be unusual in the sense of having extreme

¹ An subsequent version of this report was published in September 1997 (Literacy Task Force, 1997b), but the criticisms we make of the earlier version were not addressed.

results. Our main criticism of the analysis, however, is that schools with similar proportions of pupils receiving free school meals do not necessarily have 'similar intakes' (LTF, p.11) or even 'broadly similar intakes' (EIS, p. 81), and the assumption in the White Paper that they do leads to a number of unwarranted conclusions.

There is now a substantial body of research that shows that socio-economic indicators such as the uptake of free school meals are rather poor indicators of intake achievement. Far and away the best predictor of how well a pupil will do at the end of Key Stage Two is how well they were doing when they first entered the school, as measured by suitable cognitive assessments. It is just not possible properly to control for intake without data of this kind, and the use of 'free school meals uptake', especially when measured at the school level only, is an inadequate substitute. The evidence we have from research at the primary level (e.g. Plewis, 1991) is that attainment at the beginning of the reception year accounts for about 35% of the variation in attainment at the end of Year Six, and that schools account for a further 10% or so. This is broadly supported by the study commissioned by SCAA (1997). The correlations between socio-economic variables and attainment are much lower, typically accounting for about half as much variation as measures of prior attainment (Gray, 1989). At the secondary level the picture is similar. Typically, intake achievement measures account for between half and two thirds of the differences (variation) between schools, whereas the percentage of free school meals accounts for at most a third. It is now generally accepted that we can only sensibly compare schools after controlling for intake achievement. It must also be emphasised strongly that the use of the term 'intake' in EIS is misleading since it there refers to free school meals uptake, *not to* achievement at the start of schooling.. Therefore, the diagrams in LTF (p.12) and EIS (p.81) give an exaggerated impression about how much variation in pupils' English attainment at the end of Key Stage Two can be attributed to differences between primary schools. It is most unfortunate that research evidence of the kind referred to here has been ignored and that the authors of LTF and the architects of the White Paper have chosen instead to highlight apparently wide disparities between schools.

There are three regrettable consequences of such exaggeration. The first is a tendency to blame schools and teachers for the fact that not all pupils are achieving the expected levels. Thus we read in LTF that 'expectations and overall performance need to be substantially raised' (p.11) and that there will be a 'zero tolerance of failure' (p.15). The second is to downplay the well-attested and far-reaching effects of poverty, parental unemployment and poor housing on children's educational performance. The LTF states that 'whether children learn to read well is *a lottery* (our italics) in both advantaged and disadvantaged areas.' This is not the case. Chance plays only one part in whether and when a child learns to read: there are also systematic effects of social class, income, gender and ethnicity on children's attainments.

We do not wish to imply by these criticisms that differences between schools in pupils' progress are unimportant. Indeed, research into school effectiveness has provided important insights which can be used to improve schools and hence to raise standards. Yet to focus so much on schools, and on comparisons between schools, and to ignore broader economic and social inequalities is to distort reality. Mortimore and Whitty (1997) and Robinson (1997) make a similar point.

This brings us on to another serious criticism we have of LTF and, especially, of EIS. The belief that there are substantial school effects has led to the proposal that a number of Education Action Zones should be established. EIS states that these Action Zones will be 'set up in areas with a mix

of underperforming schools and the highest levels of disadvantage' and that they 'will have first call on funds from all relevant central programmes' (p.39). In other words, scarce resources will be diverted to schools in disadvantaged areas. However, we have been down this route before. In the late 1960s, Educational Priority Areas (EPAs) were set up with very similar aims to the proposed Education Action Zones. They were not, however, a success partly because, as Barnes and Lucas (1975) showed, the majority of poor pupils do not in fact live in poor areas and attend disadvantaged schools. Data in EIS (p.81) show that there are over four times as many primary schools with fewer than 30% of pupils receiving free school meals than there are disadvantaged schools with more than 30% uptake of free school meals. If we make the reasonable assumption that the proportion of poor pupils in disadvantaged schools is three times the proportion of poor pupils in advantaged schools, then it follows that there will be one third times more poor pupils attending advantaged schools as there are poor pupils attending disadvantaged schools. Moreover, recent figures suggest that one third of children are living in poverty. Consequently, directing resources at disadvantaged schools will inadvertently give a further benefit to about ten per cent of pupils who are not poor, and who attend disadvantaged schools.

A redistributive policy of allocating relatively more resources to those groups identified as disadvantaged is sensible in principle but the difficulty lies in deciding how much more, relatively, the disadvantaged are to receive. It is perfectly possible to squander resources, as we have suggested above, unless careful consideration is given to the optimal policy. When overall resources are scarce and no new resources for education are being made available the removal of resources from some schools or areas into others may have an overall deleterious effect.

It is possible, in principle, to derive a rational allocation policy which makes the most efficient use of available resources. This has been studied in the area of health screening (Alberman and Goldstein, 1970) where even relatively simple models can provide useful information to inform policy. A similar approach to the allocation of educational resources, of whatever type, could yield important insights.

What does EIS mean by its proposals on performance tables?

It is absolutely clear from Chapters 3.7 and 3.9 of the White Paper that the Government is committed to continuing to publish league tables of test and exam results based on raw data, with the emphasis on local comparisons. It believes that these tables are useful to parents in choosing schools. It admits that proper value added tables are not yet available, and there is no certainty that these will ever be feasible on a widespread basis. The policy of publishing league tables, begun by the previous Government, has been shown to be detrimental in a number of ways and there has been a considerable public debate about this. It is clear that the Government recognises the force of this debate since it talks about the desirability of 'value added' tables. Yet the whole point of 'value added' tables is that they correct the false impressions given by 'raw' tables which take no account of intake. It is logically inconsistent therefore, to maintain that 'raw' tables are misleading and, at the same time, to recommend them to parents for choosing schools!

The White Paper nowhere provides any indication that league tables, of whatever kind, have serious drawbacks. We believe that parents and others have a democratic right to be provided with information about any potentially misleading inferences which could be drawn from published tables. A 'health warning' seems to us essential and we believe that government has a major

responsibility for ensuring that this is done (see Goldstein and Myers, 1996, for a detailed discussion). In addition to what we have already mentioned the other major factor is that any published ranking of schools is subject to large measures of uncertainty. Again, the research is very clear on this - up to two thirds of institutions cannot be separated due to sampling variability. This information considerably diminishes the usefulness of league tables of whatever variety, other than as crude initial screening instruments. To continue to promote league tables as described in the White Paper is likely further to damage schools and pupils and impede any attempts to raise standards generally.

Setting national targets

Both LTF and EIS set targets for English and mathematics although not, curiously, for science. EIS states that 80% of eleven year olds will be at level four in English by 2002. The first point to make about these targets is that we will need accurate and reliable data in order to assess whether they have been reached. The omens are not good. EIS states that 58% of 11 year olds are at level 4 or above in English, the figure given in LTF is 57% but data from SCAA indicate that the figure of 57% is based on the 94% of pupils who were allocated a level in 1996. If we allow for these missing data, then the best estimate for 1996 is 61% (see Plewis, 1997). This kind of inconsistency, if repeated in 2002, would clearly lead to opportunities for massaging the figures.

We must assume that the White Paper regards level 4 in 1996 as a standard that, in a properly functioning system, virtually all pupils should reach. If, therefore, level 4 is to be interpreted as having been set as a target, this raises the issue of how that was done and whether such a target can be maintained consistently over time.

The original TGAT report (DES, 1988, para. 108) stated that "the average expectation for an age 11 pupil will be level 4". It must be assumed that this expectation has informed the design of Key Stage Two tests and the results quoted above, therefore, simply reflect this aim.

In reality, of course, there is no absolute criterion which determines what pupils will be able to do. At any age, in every educational system, there is a large variation among pupils and to use level 4 or any other level as a 'benchmark' involves a contestable judgement about what is desirable. To quote a figure of 6 out of 10 with disapproval is both unhelpful and strictly meaningless. Likewise, it is meaningless to contrast the 82% who achieve the Key Stage One 'target' of level 2 or above in mathematics, with the 54% who achieve the Key Stage Two target of level 4 or above. Yet this is just what EIS does. Even so, it might still be argued that a specific target for improvement could be useful and this is examined in the next section.

Maintaining level 4 over time.

The White Paper lays great stress on its proposals for achieving the 80% and 75% targets by 2002: it invites the public to judge its policies in terms of its ability to meet those targets. Hence, there is clearly considerable interest in knowing precisely how it is to ensure that the target is maintained consistently over time, without 'shifting'. The White Paper gives no indication whatsoever how this is to be done and presumably the Government does not view this as problematical: the Literacy Task Force report made a reference to the task being handed over to SCAA (QCA). Yet all the evidence suggests that it is actually impossible to define such a consistent standard.

The debate in the UK over how to measure standards over time dates back at least to the early 1970s with concern over apparent declines in reading standards (Start and Wells, 1972). It was a major concern of the Assessment and Performance Unit (APU) and the debate is summarised in Gipps and Goldstein (1983). In the USA a similar debate surfaced around the comparison of reading performances between 1984 and 1986 in the National Assessment of Educational Progress (NAEP) (Beaton and Zwick, 1990). In the 1990s in particular the debate has tended to centre around trends in GCSE performance from year to year.

In all of these cases the conclusion essentially has been the same, namely that the attempt to measure absolute standards over time is doomed to fail. In effect, it is impossible to distinguish 'real' changes in the performance of pupils from changes in the (different) tests or exams that are used over time. For example, it was shown by both the APU and NAEP research that even apparently minor changes in question format, question ordering or small changes to content could greatly affect the proportion of correct responses. The only possible way to ensure that the same thing is being measured is to use precisely the same test in precisely the same way over time. This, of course, would be unacceptable, so the conclusion must be that there is literally no way in which level 4 (in 1996 or 1997) could be maintained as a standard through to 2002.

There are of course, other useful targets to aim for, some of which are mentioned in the White Paper. Setting out to reduce the gaps between ethnic groups, social classes and the sexes is both legitimate and, importantly, measurable. There is a strong case for the Government to set up rigorous evaluations of policy initiatives to see what does and doesn't work in terms of improving performance. It would also be worthwhile using highly trained experts to judge extensive random samples of pupil work in English and mathematics every year, with a view to making informed judgements about basic levels of literacy and numeracy. Such judgements, however, would not be able to provide precise indicators of 'standards': rather they would act as a mechanism which aimed to detect any large shift which might occur - possibly as a result of policy or external factors - so that further investigation could be undertaken. In other words it would have a crude, but potentially useful, screening function.

We see, therefore, that the White Paper targets are simply unachievable because there exists no way to measure what is happening. The Government needs to recognise this and to drop these targets, rather than trying to do the impossible. If it persists there is a real danger that attention will become narrowly, and entirely misleadingly, focussed on the numbers achieving the levels each year, while ignoring more important issues.

Finally, we should make it clear that the argument we are putting forward is not a new one and nor is it merely a debating point. Much of the White Paper is predicated on being able to achieve these targets: this cannot be delivered and the sooner this is recognised the better.

Literacy time in schools

Both LTF and EIS recommend that primary schools devote a structured hour a day to literacy, the implication being that this does not happen at present. Although the introduction of the National Curriculum has led to changes in the way teachers and pupils organise school time (Plewis and Veltman, 1996), the Education Reform Act of 1988 stopped short of prescribing a timetable for all schools. However, the 1994 Dearing review recommended that, on average, pupils at Key Stages One and Two should spend an hour a day, and 55 minutes a day respectively

on English. So the LTF and EIS recommendations are in line with those in the Dearing review. We do not have much research evidence about the way school time is actually divided between the different National Curriculum subjects and other activities. However, Plewis and Veltman (1996), in their study of 24 inner London Year Two classrooms, found that pupils spent 79 minutes a day on literacy activities. In other words, teachers, at least at Key Stage One, are already devoting a lot of time to literacy, perhaps rather more than Dearing and EIS recommend. It is also important to recognise that a teacher can devote a lot of time to, for example, hearing individual pupils read without any one pupil getting a lot of individual attention. Plewis and Veltman (1996) found, like Tizard et al. (1988) before them, that, on average, a teacher will listen to a seven year old reading aloud for eight minutes a week. This seems very little but aggregated over say 25 pupils in a class, this can amount to 40 minutes a day. Of course, sometimes reading aloud will take place in small groups, sometimes with a classroom helper, but the research evidence we have all points to the fact that teachers are already devoting a lot of time to literacy and that the proposed literacy hour may not in fact add very much.

Conclusions

The concerns expressed in LTF and EIS about the poor academic attainments of pupils from disadvantaged backgrounds are ones we share. Children who do not learn to read at primary school are unlikely to be able to participate fully in a democratic society. However, we believe that proposals to solve these problems should be based not only on good intentions but also on careful analysis, taking account of what has happened to similar ideas in the past, and to what we now know about differences between schools' performance, about ways of sensibly comparing schools, and about what is happening at present in primary schools. On these criteria, we find that both LTF and EIS fall short of the standards which a democratic society has a right to expect.

References

- Alberman, E. D. and Goldstein, H. (1970) The at risk register: A statistical evaluation. *British Journal of Social and Preventive Medicine* **24**: 129-135.
- Barnes, J. H. and Lucas, H. (1975) Positive Discrimination in Education: Individuals, Groups and Institutions. In Barnes, J. (Ed.) *Educational Priority, Volume 3*. London, HMSO.
- Beaton, A. and Zwick, R. (1990) *Disentangling the NAEP 1985-1986 reading anomaly*. Princeton, Educational Testing Service.
- DES (1988) *Task Group on Assessment and Testing*. Department of Education and Science. London.
- DfEE (1997) *Excellence in Schools*. London, DfEE.
- Gipps, C. and Goldstein, H. (1983) *Monitoring Children*. London, Heinemann.
- Goldstein, H. and Myers, K. (1996) Freedom of information: towards a code of ethics for

performance indicators. *Research Intelligence*(57): 12-16.

Gray, J. (1989) Multilevel models: Issues and Problems emerging from their recent application in British studies of school effectiveness. In Bock, R.D. (Ed.) *Multilevel Analysis of Educational Data*. San Diego, Academic Press.

Literacy Task Force (1997a). A reading revolution. London, The Labour Party.

Literacy Task Force (1997b). The implementation of the National literacy strategy. London, DfEE.

Mortimore, P. and Whitty, G. (1997) *Can school improvement overcome the effects of disadvantage?* London, Institute of Education.

Plewis, I. (1991) Pupils' progress in reading and mathematics during primary school: associations with ethnic group and sex. *Educational Research*, 33, 133-140.

Plewis, I. (1997) Presenting educational data: Cause for Concern? *Research Intelligence*(61): 9-10.

Plewis, I. and Veltman, M. (1996) Where Does all the Time Go? Changes in Pupils' Experiences in Year 2 Classrooms. In Hughes, M. (Ed.) *Teaching and Learning in Changing Times*. Oxford, Blackwell.

Robinson, P. (1997) *Literacy, Numeracy and Economic Performance*. London: Centre for Economic Performance, LSE.

Start, B. and Wells, K. (1972). *The trend of reading standards*. Slough, NFER.

SCAA (1997). *The value added National Report*. London, Schools Curriculum and Assessment Authority.

Tizard, B., Blatchford, P., Burke, J., Farquhar, C. and Plewis, I. (1988) *Young Children at School in the Inner City*. Hove, Lawrence Erlbaum.