

The Early Years Enriched Curriculum Evaluation Project: Special fourth year interim report

*This report should be read in conjunction with
previous reports*

Sproule, L., Trew, K., Rafferty, H., Walsh, G.*, O'Neill, B.,
M^cGuinness, C. & Sheehy, N.

School of Psychology,
Queen's University Belfast &
*Stranmillis University College

March 2004

Contents

Executive summary	3
1. Introduction and summary of main findings	6
2. Method of data analysis	7
3. Detailed findings	9
4. Interpretation and discussion	26
Reference	30

Executive summary

The Early Years Enriched Curriculum Project is now around the middle of its fourth year. The first cohort of children from the Shankill group of six schools is currently in Year 4. The first cohort of children from the Contrasting Areas (CA) group of six schools is now in Year 3. This special report was commissioned separately by the Northern Ireland Council for Curriculum Assessment and Examinations (CCEA) in order to assess the progress of the Enriched Curriculum project children in time to facilitate funding decisions for 2004-05.

This task was by its nature problematic for a variety of reasons. Our first problem was that we were required to provide a judgement on the efficacy of the Enriched Curriculum by means of a 'snapshot' test of attainment. Secondly, in order to achieve this we had to make a number of unverifiable assumptions about the manner in which children are progressing (as explained in Section 2).

The interpretation of the results is complex. There would appear to be a number of results involving cohort, gender and ability, and interactions between these, some of which were unexpected. Whilst the overall sample size is good, the sample sizes involved in the interactions are too small to permit full confidence in results of statistical tests on small group differences. For these reasons, we must guard against over-interpretation of the findings and urge caution in reaching conclusions from this data-set.

Nevertheless the results are interesting and have provided some new information relevant to the evaluation of the Enriched Curriculum, even though the status of this information is, as suggested, extremely tentative.

The main findings

Effects on non-verbal ability and picture vocabulary

There are indications that the Enriched Curriculum is associated with increases in proxy IQ measures in Shankill schools. This possibility of increases in a range of verbal, spatial/pictorial abilities for those children in the Enriched Curriculum was not anticipated and so should be treated with caution. Nevertheless we believe that it is important to follow up such promising findings over the medium and longer term to see whether they hold up over a longer period of time.

Mathematics attainment

Within the results as a whole, a pattern of data is beginning to emerge which suggests that at this stage, Enriched Curriculum children are doing somewhat better in mathematics than they are in reading. Three of the four Enriched Curriculum groups performed well overall in tests of mathematics and those in Years 2 and 3 would be expected to equal or outperform the appropriate controls by the end of their third year. The exception in mathematics attainment is the first Enriched Curriculum cohort in

Shankill schools. Given their good performance in June 2003, this group had a disappointing score, although they may still catch up with controls by June. Any hypothesis about the reason for this performance is very speculative at this stage. However, boys had made somewhat more progress than girls.

Reading attainment

Three of the four Enriched Curriculum groups performed well overall in tests of reading. Those children who are in Years 2 and 3 should equal or outperform the appropriate controls at the end of their third year. Again in reading, the exception in attainment is the first Enriched Curriculum cohort in Shankill schools. Given their good performance in June 2003, this group had a disappointing score, although they may still catch up with controls by June. Any hypothesis about the reason for this lower performance is very speculative at this stage. However, the shortfall was largely explained by the relatively poor progress of high-ability girls in the first cohort only within the Shankill schools. There were no such problems with the second cohort of children.

Cohort effects

Within the results as a whole, a pattern of data is beginning to emerge which suggests that children following the Enriched Curriculum with an experienced Enriched Curriculum teacher may be doing better than children following it with a novice Enriched Curriculum teacher. The second cohort of children in the Enriched Curriculum generally outperformed the first. This is not an unexpected finding.

Gender effects

One of the reasons for introducing the Enriched Curriculum was the significantly poor performance of boys, especially in reading: Principals and teachers in the Shankill schools were convinced that the pre-existing curriculum was particularly problematic for boys. The present evaluation has found very few significant gender effects in tests of attainment to date. There are some indications that boys in the Enriched Curriculum have now closed the gap with girls. However, looking at these current test results in the context of previous testing, it is possible that boys may be starting to pull ahead of girls. There may be the beginnings of a significant trend here which we will be able to investigate statistically after the June 2004 round of testing.

Between schools effects

There are greater differences between schools' performances in mathematics than there are in reading. A school doing well in reading does not necessarily do well in mathematics and vice versa. A school's progress since the June 2003 round of testing does not necessarily correlate with its longer term performance. Some schools do consistently well in one or both subjects, indicating that performance is probably an interaction between school effects and teacher effects.

Contrasting Areas schools versus Shankill schools

As expected from their intake characteristics, the Contrasting Areas schools continue to outperform Shankill schools. However the pattern of Enriched Curriculum pupils' catching-up in maths and reading is similar in both groups of schools, Year 3 appears to be the year when they catch up with those following the traditional curriculum. The first cohort has already done so in mathematics and will hopefully do so in reading by June 2004. There may well be continued improvement in this group after June 2004. We therefore recommend that these children are followed up in the medium term.

Overall picture

The evaluation of the enriched Curriculum to date has produced a generally favourable verdict on the Enriched Curriculum by means of a wide-ranging and robust methodology which gave weight to a variety of data types, including the views of teachers and parents and a classroom observation study.

The present report is considered to be more tentative because it draws on a 'snapshot' picture and rests on a number of unverifiable assumptions. However it is considered to be consistent with previous, generally favourable findings about the progress made by pupils. Nevertheless it does suggest that the impact of the Enriched Curriculum may be more complex, more subtle, and more unexpected than previously anticipated even when the focus is only on abilities/attainments measured at one point in time.

1. Introduction

We were commissioned by the Northern Ireland Council for Curriculum Assessment and Examinations (CCEA) to investigate six groups of children, two Enriched Curriculum groups and one control group from each of the original school groups (the Shankill schools and the Contrasting Areas schools). The Enriched Curriculum groups included the first and second cohorts to follow the curriculum in a given school, thus introducing as a new variable the extent of the teacher's experience in teaching the Enriched Curriculum.

The sample tested was the alternate half sample, that is, all the children in each year not included in the original project evaluation sample, which consisted of half the population in each school (randomly selected). The children were tested using the age-appropriate version of the instrument we have consistently used throughout the evaluation, namely Performance Indicators in Primary Schools (PIPS). PIPS is a psychometrically sophisticated and highly valid and reliable instrument, which is normally administered at the end of the school year. It yields reading, mathematics, picture vocabulary and non-verbal test scores.

One Contrasting Areas school refused access the evaluation team because of the ongoing industrial action of teachers. This school has hitherto had mean scores which are close to the group mean, so this limitation is not as serious as it might have been under other circumstances.

For the most part, small group testing was employed, with typically six children in each group.

2. Method of data analysis

The data analysis is not straightforward for this round of testing. We recommend that readers try to understand this explanation thoroughly before turning to the results.

Age norms for PIPS tests are not available and PIPS is normally used at the end of the school year. A variety of alternative tests were considered but it was agreed with CCEA's representatives that Enriched Curriculum children would be disadvantaged by such tests for a variety of reasons. Some tests started with items which were too difficult for children more than one standard deviation below the mean. Since this end of the ability spectrum is overrepresented in Shankill schools, the consequent floor effects would render the results statistically invalid. Other tests used extensively formats which would be familiar to children used to doing many worksheets but not as familiar to Enriched Curriculum children. Still other tests are now somewhat old fashioned and used antiquated language which would be unfamiliar to Enriched Curriculum children whose literacy experience is still centred on modern texts. It was agreed with CCEA that PIPS was the instrument which best addressed conceptual understanding in mathematics and comprehension in reading over the whole ability range.

Using PIPS in the middle of the school year gives rise to a complex situation for analysis. Normally, one must assume a linear model of progress in this situation and use it to predict end of year scores. An assumption of this model is that a control group will be progressing at the same rate as Enriched Curriculum children: The model implies that the group mean of the standardised score for an older control group is 45 in February, then it should be 45 for Enriched Curriculum group children also, provided all the children are in Year 4 or above. Below Year 4, we have established in our previous reports, that Enriched Curriculum children do not do as well on PIPS as children following the traditional curriculum because of the mismatch between curriculum matched by PIPS¹ and the Enriched Curriculum in the early years.

Where Enriched Curriculum group children are in Year 3 or below, they have a 'catch up' phase during which they make faster progress than would be usual under the traditional curriculum, as described in our Year 3 report (Sproule, Trew, Rafferty, Walsh, O'Neill, McGuinness & Sheehy 2003). It is also well documented in the research literature that children at the 'take off' point in their reading, the point where they have grasped the basics, do not follow a linear model but begin to progress much faster in their reading skills and maintain that rate of progress for some time. For the majority of Enriched Curriculum children, this happens in Year 3, although we estimate about 20% took off in Year 2 and around 5% had still not taken off at the end of Year 3. We would therefore expect that in Year 3 or below, Enriched Curriculum would have poorer scores than controls in February, with a much more marked effect in Year 2.

¹ PIPS is very carefully matched to the English National Curriculum. There is no test yet designed to match the Enriched Curriculum in Years 1 and 2.

Originally, we had also planned to use proxy IQ measures as a benchmark for performance on reading and mathematics in this round of testing. However, it now appears that the Enriched Curriculum may be having a positive effect on proxy IQ itself, particularly in Shankill schools. This makes the comparison a little more difficult. If the Enriched Curriculum is causing a rise in proxy IQ but no rise in reading or mathematics attainment as yet, then the higher IQ scores will make the reading and mathematics scores appear poor by comparison. On the other hand, provided we keep this caveat in mind, the comparison may still prove useful.

PIPS measures proxy IQ by a combination of picture vocabulary and non-verbal scales. Although this measure would have quite a large uncertainty associated with individual measurements, it is a robust measure of group means in sample sizes of 70 and over. All sample sizes were 70 or more in this round of testing.

In addition to using proxy IQ as a benchmark, we can obtain a conservative estimate of how much children will gain by the end of the year as a result of maturation, by making use of age-corrected data. We have used data supplied to us from PIPS to estimate the number of standardised points that children will gain on average between now and the June round of testing. This varies between 1.3 and 2.5 points, depending on year group. There are larger age corrections for younger groups. It must be remembered that predicted scores based on age-corrected data do not allow for the contribution of teaching: Children should perform somewhat better in June than age-corrected scores indicate.

We are in a position to make several comparisons with controls:

- We can compare February scores of Enriched Curriculum children with February scores of control children. The latter group is one year older for the first Enriched Curriculum cohort and two years older for the second Enriched Curriculum cohort and follow the pre-existing curriculum in the same schools. To some extent, looking for equal performance with these controls may be asking Enriched Curriculum children to hit a moving target. There is some evidence that the schools are managing to improve the performance of control groups as they go up the school.
- We can compare predicted age-corrected scores of Enriched Curriculum children for June with actual scores of control groups in June. We must bear in mind that the age-corrected scores do not take into account the effect of teaching and therefore represent the *minimum* level of attainment we would expect in June.
- For reading and mathematics scores, we can compare Enriched Curriculum children's February scores with their proxy IQ scores.

PIPS is standardised on a mean score of 50 and a standard deviation of 10. Each standardised mark is one and a half times larger than a mark on the scale used for IQ tests. For example, 4 marks on this scale is the equivalent of 6.5 marks on the IQ scale.

3. Detailed findings

It is easier to understand the findings if we begin with the measures of non-verbal ability and picture vocabulary. For ease of reading tables we designate the original pilot cohort as Shankill Enriched Curriculum 1 (Shankill EC1 in the tables of results), and the second Shankill Enriched Curriculum cohort as Shankill Enriched Curriculum 2 (Shankill EC2). Similarly we have Contrasting Areas Enriched Curriculum first and second cohorts (CA EC1 and CA EC2 respectively). There are also two control cohorts for schools for which we have comparable data, namely the year-ahead and two-years-ahead cohorts (Control 1 and Control 2 respectively). There is less data available for Control 2 because they had already completed certain school years before testing began. The baseline data we have on these children comes from an earlier project.

It should be remembered that it is during Year 3 that we expect children in CA schools to catch up with controls in reading and mathematics; it is not therefore reasonable to expect children in Year 3, or indeed in Year 2 either, to have caught up at this time. In this respect, the results are fully in accordance with our expectations. This group is designated CA EC2 in tables, standing for Enriched Curriculum Cohort 2. The other Enriched Curriculum groups follow the same system of designation.

Non-verbal ability

Enriched Curriculum children had very good performance on tests of non-verbal ability. The results are summarised in Table 1. The results show that all Enriched Curriculum groups had a good performance and could reasonably be expected to equal or outperform controls in June 2004.

Table 1: Group mean scores for non-verbal ability

Shankill schools	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 4 score	45.6 (\pm 0.9)	47.2
EC2: Feb Year 3 score	48.1 (\pm 1.0)	49.9
Control 1: Feb Year 5 score	44.0 (\pm 0.9)	45.3
		June testing: Actual mean
EC1: End of Year 3 score		48.1
Control 1: End of Year 4 score		46.7
Control 2: End of Year 4 score		46.2
CA schools	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 3 score	50.1 (\pm 1.0)	51.9
EC2: Feb Year 2 score	52.6 (\pm 1.0)	54.9
Control 1: Feb Year 4 score	50.3 (\pm 1.1)	51.9
		June testing: Actual mean
Control 1: End of Year 2 score		52.4
Control 1: End of Year 3 score		55.0
Control 2: End of Year 3 score		53.9
EC1: End of Year 2 score		51.4

* Predicted from data supplied by PIPS

Note

The end of Year 4 results for the two control groups in Shankill schools and the end of Year 3 results for the two control groups in CA schools are very close. This is an indication of the validity and reliability of PIPS.

Comparing February scores for Enriched Curriculum groups with February scores for older controls

Group means were superior for all Enriched Curriculum groups compared with controls, but the differences were only significant for the second Enriched Curriculum group in Shankill schools { $t = 3.14$, $df = 147$; $p < .01$ (2-tailed)}. This group was 4.1 standardised points higher than controls, a notably large difference in terms of education interventions. Non-verbal ability is the most culture-fair measure we use and it therefore important that disadvantaged children are doing well on it.

Comparing predicted age-corrected scores with controls

In Shankill schools, the predicted age-corrected scores for both Enriched Curriculum groups are superior to measured scores for control groups at the end of Year 4. The second cohort has already exceeded control end-of-year scores. If the first cohort equals its own Year 3 score, it will also outperform controls in June. This pattern suggests that the Enriched Curriculum may be having a positive effect on non-verbal scores in Shankill schools.

The predicted age-corrected scores for both Enriched Curriculum groups in Contrasting Areas schools are superior to predicted scores for control groups at the end of Year 4. We may also conjecture about the effects of adding in another four months of learning by looking at the actual scores of control groups at the end of Year 3. If Control 1 was to equal its own performance at the end of Year 3 and if Enriched Curriculum groups were to improve by the same amount, then Enriched Curriculum groups would be equal or superior to the performance of controls. However, this is very speculative: The pattern of progress may differ in the different groups.

The effect of the teachers' experience with the Enriched Curriculum

In both groups of schools, there is a trend for the second Enriched Curriculum cohort to be already outperforming the first Enriched Curriculum cohort. In Shankill schools, the figures are $t = 1.88$, $df = 142$; $p = .06$ (2-tailed) and in Contrasting Areas schools, the figures are $t = 6.49$, $df = 151$; $p < .001$ (2-tailed). It may be that a test on the whole sample, rather than our half sample, would have yielded a significant result. If so, that would have been a definite indication that non-verbal ability would have been positively affected by the teacher's experience. As it is, we can only say that it should be further investigated.

Gender effects

One normally expects to see boys outperforming girls on tests of non-verbal ability.

There were no gender effects in Shankill schools.

In the first Enriched Curriculum group in Contrasting Areas schools, boys significantly outperformed girls by 5.0 standardised points { $t = 2.36$, $df = 74$; $p < .05$ (2-tailed)}.

There were no gender effects in the second Enriched Curriculum group in Contrasting Areas schools.

Picture vocabulary

Three out of four Enriched Curriculum groups also had a very good performance on tests of picture vocabulary. The results are summarised in Table 2. The results show that these three Enriched Curriculum groups had a good performance and would be expected to equal or outperform comparable controls in June 2004. For the fourth Enriched Curriculum group, the original pilot cohort in the Shankill schools, the prediction is a little more speculative. Nevertheless, if this group do as well as they have done in previous years, they will also equal or outperform controls in June.

Table 2: Group mean scores for picture vocabulary

Shankill schools	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 4 score*	40.2 (\pm 0.9)	41.8
EC2: Feb Year 3 score*	43.5 (\pm 0.9)	45.3
Control 1: Feb Year 5 score	42.0 (\pm 0.8)	43.3
		June testing: Actual mean
EC1: End of Year 3 score		44.5
Control 1: End of Year 4 score		42.1
S Control 2: End of Year 4 score		42.4
CA schools	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 3 score*	50.2 (\pm 1.1)	52.0
EC2: Feb Year 2 score*	47.8 (\pm 0.9)	50.1
Control 1: Feb Year 4 score	48.1 (\pm 1.0)	49.7
		June testing: Actual mean
Control 1: End of Year 3 score		52.3
Control 2: End of Year 3 score		52.4
Control 1: End of Year 2 score		51.6
EC1: End of Year 2 score		48.6

* Predicted from data supplied by PIPS

Comparing February scores for Enriched Curriculum groups with February scores for older controls

There were no significant differences between any Enriched Curriculum cohort and controls. There was a trend for the first Enriched Curriculum cohort in Contrasting Areas schools to do better than controls { $t = 1.74$, $df = 158$; $p < .09$ (2-tailed)}.

Comparing predicted age-corrected scores for June with those of controls

In Shankill schools, the predicted age-corrected scores for the second Enriched Curriculum cohort is superior to measured scores for control groups at the end of Year 4. In fact, the second cohort has already exceeded control end-of-year scores presented in Table 2. If the first cohort equals its own Year 3 score, it will also outperform controls in June. This pattern suggests that the Enriched Curriculum may be having a positive effect on picture vocabulary scores in Shankill schools. However, picture vocabulary scores in Shankill schools are still poor compared with non-verbal ability scores across all groups.

The predicted age-corrected mean score for the first Enriched Curriculum cohort in Contrasting Areas schools is superior to predicted scores for control groups at the end of Year 4. They are also equal to measured end-of-year scores for controls. We may also conjecture about the effects of adding in another four months of learning by looking at the actual scores of control groups at the end of Year 3. If Control 1 was to equal its own performance at the end of Year 3 and if Enriched Curriculum groups were to improve by the same amount, then Enriched Curriculum groups would be equal or superior to the performance of controls. However, this is very speculative: The pattern of progress may differ in the different groups.

The effect of the teachers' experience with the Enriched Curriculum

Again, in both groups of schools, the second Enriched Curriculum cohort is already outperforming the first Enriched Curriculum cohort. This difference is highly significant in Shankill schools { $t = 2.60$, $df = 142$; $p < .01$ (2-tailed)} but is only a trend in Contrasting Areas schools { $t = 4.78$, $df = 159$; $p < .001$ (2-tailed)}, thus suggesting that teachers may be having a more positive effect on picture vocabulary once they are experienced in teaching the Enriched Curriculum.

Gender effects

One normally expects to see girls outperforming boys on tests of picture vocabulary.

In the first Enriched Curriculum group in Shankill schools, boys outperformed girls by 4.4 standardised points in picture vocabulary { $t = 3.14$, $df = 69$; $p < .05$ (2-tailed)}. This is a considerable difference in scores.

There were no other significant gender effects.

Note

The end of Year 4 results for the two control groups in Shankill schools are very close and the end of Year 3 results for the two control groups in Contrasting Areas schools are very close. This is an indication of the validity and reliability of PIPS.

Mathematics

Three out of the four groups of Enriched Curriculum children had good performance on tests of mathematics. The results are summarised in Table 3. The exception was the first cohort in Shankill schools where the poor performance of high-ability girls was influential on the results. Taken together, the results show that three out of four Enriched Curriculum groups had a good performance and might reasonably be expected to equal or outperform controls, either in June 2004, or in the case of the second cohort in Contrasting Areas schools, in June 2005.

Comparing February scores for Enriched Curriculum groups with February scores for older controls

In Shankill schools, there was no significant difference between the second Enriched Curriculum cohort and controls. Although they are still in the catch-up year, this group have already caught up with controls. We also note that the February mean for the second Enriched Curriculum cohort was significantly superior to that for the first Enriched Curriculum cohort { $t = 2.80$, $df = 142$; $p < .01$ (2-tailed)}.

The control group significantly outperformed the first Enriched Curriculum cohort { $t = 3.15$, $df = 145$; $p < .05$ (2-tailed)}. This suggests that this group may have lost ground since June 2003, when they equalled the performance of this control group.

Both Enriched Curriculum cohorts in Contrasting Areas schools are still expected to be catching up with controls but we hypothesised in our last report that the first cohort would catch up during the current school year (Sproule et al. 2003). In Contrasting Areas schools, there was no significant difference between the first Enriched Curriculum cohort and controls, signifying that this group have already caught up in mathematics.

As expected in Year 2, the second Enriched Curriculum cohort in Contrasting Areas schools significantly underperformed compared with the control group {EC1: $t = 4.61$, $df = 153$; $p < .001$ (2-tailed)}.

Table 3: Group mean scores for mathematics

Shankill schools	Feb: Proxy IQ	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 4 score*	43.1 (± 0.9)	39.1 (± 0.9)	41.1
EC2: Feb Year 3 score*	45.8 (± 0.9)	42.3 (± 0.8)	44.6
Control 1: Feb Year 5 score	43.0 (± 0.8)	42.7 (± 0.8)	44.3
			June testing: Actual mean
Control 1: End of Year 3 score			45.8
Control 1: End of Year 4 score			44.5
Control 2: End of Year 4 score			44.3
EC1: End of Year 3 score			44.9
CA schools	Feb: Proxy IQ	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 3 score*	50.2 (± 1.0)	47.8 (± 1.1)	50.1
EC2: Feb Year 2 score*	50.2 (± 1.0)	41.2 (± 0.8)	43.7**
Control 1: Feb Year 4 score	49.2 (± 1.0)	46.3 (± 0.8)	48.3
			June testing: Actual mean
Control 1: End of Year 2 score			47.7
Control 1: End of Year 3 score			51.7
Control 2: End of Year 3 score			50.5
EC1: End of Year 2 score			43.7

* Predicted from data supplied by PIPS

** Not expected to equal the performance of controls until end Year 3

Comparing predicted age-corrected scores for June with those of controls

In Shankill schools, predicted age-corrected scores for the second Enriched Curriculum cohort are 0.3 standardised points higher than predicted scores for Control 1 at the end of Year 5 and 1.2 standardised points below actual scores for that control at the end of Year 3². Remembering that the predicted scores do not take account of teaching and noting that these differences in means are small, we may speculate from these data that this group is likely to equal the performance of controls in June.

For the first Enriched Curriculum cohort in Shankill schools, the mean score is predicted to be between 2.3 and 2.7 below that of controls. This suggests that this group is not quite as likely to equal the performance of controls in June as the second Enriched Curriculum group.

In Contrasting Areas schools, predicted age-corrected scores for the first Enriched Curriculum cohort are superior to predicted scores for Control 1. This group would therefore be expected to equal or outperform the Year 5 performance of these controls in June. The predicted mean score for the Enriched Curriculum group is also close to that actually measured for controls at the end of their Year 3. Given that the predicted scores do not take account of four months teaching and learning, we may speculate that this Enriched Curriculum group may also equal these scores in June 2004.

For the second Enriched Curriculum cohort in Contrasting Areas schools, the mean score is predicted to be below that of controls, in line with what the evaluation team has measured for the first Enriched Curriculum group. Their mean score is predicted to be 43.7 in June against an actual score of 43.7 in June for the first Enriched Curriculum cohort, signifying that this group is likely to be on track to equal the performance of the first cohort and catch up with controls in June of their Year 3.

Comparing February scores with proxy IQ

All groups had a shortfall compared with their proxy IQ scores except the Control 1 group in Shankill schools, where the difference was within the standard error for mathematics scores.

The second Enriched Curriculum cohort in Contrasting Areas schools had the largest shortfall. However, it is not meaningful to compare these scores in terms of present performance on mathematics because this group is still in the catch-up phase.

² For technical reasons, statistical tests between these two groups might not be strictly appropriate. The gain in mean score can be predicted but every child will not gain the same amount. The standard deviation could change without affecting the mean.

The shortfall for the first Enriched Curriculum cohort in Contrasting Areas schools was smaller than that for controls, further confirming that this group is performing well in mathematics.

The shortfall for the first Enriched Curriculum cohort in Shankill schools was larger than that for controls, further confirming that this group may be underperforming in mathematics.

The effect of the teachers' experience with the Enriched Curriculum

In Shankill schools, predicted age-corrected scores for the second cohort are virtually equal to those of all control groups. When we remember that these scores do not allow for the effect of teaching between now and June, we may speculate that the second cohort will do at least as well as the first at the end of their third year.

In Contrasting Areas schools, predicted age-corrected scores for the second cohort are virtually equal to those of the first cohort at the end of their Year 2. When we remember that these scores do not allow for the effect of teaching between now and June, we may speculate that the second cohort will do at least as well as the first.

However, we also note that the February mean for the second Enriched Curriculum cohort was significantly superior to that for the first Enriched Curriculum cohort { $t = 2.80$, $df = 142$; $p < .01$ (2-tailed)}. This suggests that a positive effect of teacher experience may be starting to emerge.

Gender and ability differences

Normally we would expect boys to do better than girls in mathematics. PIPS measures the difference as 1.3 standardised points in the national sample.

There were no gender effects in mathematics for any Enriched Curriculum group.

However, there is an interaction between ability and gender in the first Enriched Curriculum cohort in Shankill schools. Our data show that boys had done better than girls in mathematics up until the end of Year 3. The gap narrowed between the end of Year 2 and Year 3. After June 2003, the high-ability girls begin to close the gap, as shown in Figure 1, whereas the low-ability girls continue to lose out compared with boys of the same ability. The effect is under half a standard deviation at the extremes.

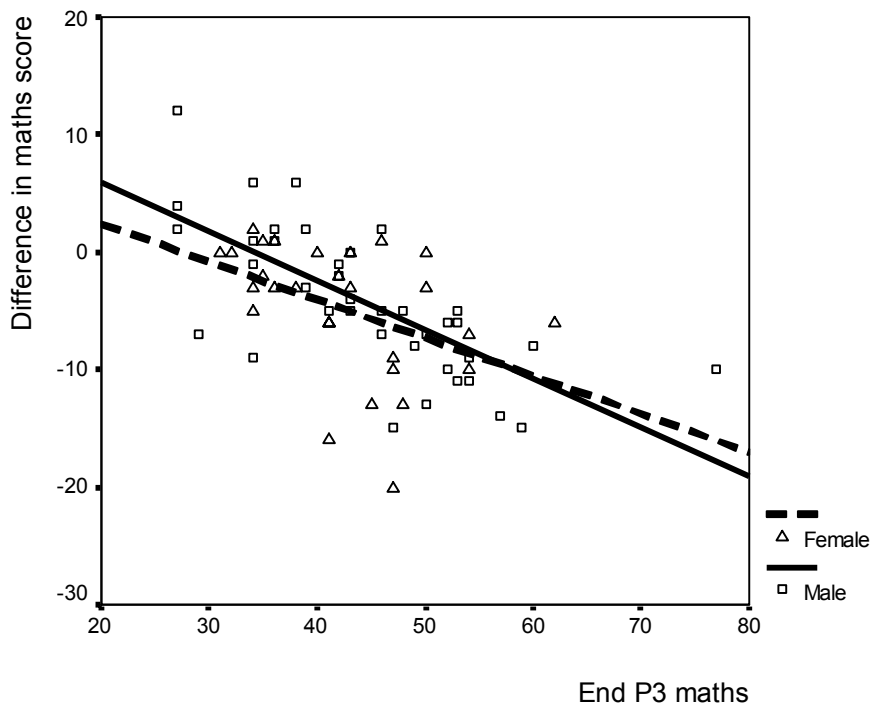


Figure 1
Progress in maths since June 2003
First Shankill Enriched Curriculum cohort

Reading

The performance of Enriched Curriculum children in reading is not quite as good as in mathematics. The results are summarised in Table 4. Taken together, the results show that three out of four Enriched Curriculum groups had a good performance and may equal or outperform controls, either in June 2004, or in the case of the second cohort in Contrasting Areas schools, in June 2005. The exception was the first cohort in Shankill schools where the poor performance of high-ability girls was influential on the results.

Table 4 Group mean scores for reading

Shankill schools	Feb: Proxy IQ	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 4 score	43.1 (± 0.9)	39.4 (± 0.9)	41.1
EC2: Feb Year 3 score	45.8 (± 0.9)	40.6 (± 0.8)	42.4
Control 1: Feb Year 5 score	43.0 (± 0.8)	44.6 (± 0.9)	45.9
			June testing: Actual mean
Control 1: End of Year 3 score			44.9
End of Year 4 score			43.7
Control 2: End of Year 4 score			43.9
EC1 end of Year 3 score			43.8
CA schools	Feb: Proxy IQ	Standardised mean (Standard error)	June: Predicted age-corrected <i>minimum</i> mean score*
EC1: Feb Year 3 score	50.2 (± 1.0)	46.2 (± 1.2)	48.0
EC2: Feb Year 2 score	50.2 (± 1.0)	40.7 (± 1.2)	42.7**
Control 1: Feb Year 4 score	49.2 (± 1.0)	49.7 (± 1.1)	51.0
			June testing: Actual mean
Control 1: End of Year 2 score			48.8
End of Year 3 score			53.2
Control 2: End of Year 3 score			51.5
EC1: End of Year 2 score			45.1

* Predicted from data supplied by PIPS

** Not expected to equal the performance of controls until end Year 3 *Comparing*

February scores for Enriched Curriculum groups with February scores for older controls

These results must be seen in the context that only the first Enriched Curriculum group in Shankill schools are expected to equal the performance of controls in February 2004.

In Shankill schools, the control group significantly outperformed the first Enriched Curriculum cohort {EC2: $t = 4.32$, $df = 145$; $p < .001$ (2-tailed)}, with controls superior by 5.2 standardised points. However, when we look at Table 4, we see that this group appear to have improved their reading, achieving the same score in the middle of the current year as they achieved at the end of Year 4, and thus expected to improve to 45.9 at the end of this school year.

The control group also significantly outperformed the second Enriched Curriculum cohort {EC1: $t = 3.28$, $df = 147$; $p < .01$ (2-tailed)}, with a somewhat lower margin here of 4.0 standardised points. This group is still in their catch-up year, not being due to catch up until June 2004, so this result was expected. In the case of reading, the difference in February scores between the first and second Enriched Curriculum cohorts is not significant.

Both Enriched Curriculum groups in Contrasting Areas schools are still expected to be catching up with controls. The control group significantly outperformed the first Enriched Curriculum cohort by 3.5 standardised points {EC1: $t = 2.10$, $df = 160$; $p < .05$ (2-tailed)}.

As expected, in their Year 2 the second Enriched Curriculum cohort significantly underperformed compared with the control group {EC1: $t = 5.55$, $df = 153$; $p < .001$ (2-tailed)}.

Comparing predicted age-corrected scores for June with those of controls

In Shankill schools, predicted age-corrected scores for the second Enriched Curriculum cohort are 3.5 standardised points lower than predicted scores for Control 1 at the end of Year 5 and 2.5 standardised points below actual scores for that group at the end of Year 3³. However, remembering that the predicted scores do not take account of teaching, or the ‘take-off’ boost which can happen in reading and noting that these differences in means is small, we may speculate that this group may still equal the performance of controls in June, especially when we note that Control 1 predicted scores would represent an improvement for that group.

For the first Enriched Curriculum cohort in Shankill schools, the mean score is predicted to be between 2.6 and 4.8 standardised points below those of the various controls. This

³ For technical reasons, statistical tests between these two groups might not be strictly appropriate. The gain in mean score can be predicted but every child will not gain the same amount. The standard deviation could change without affecting the mean and this would affect statistical tests.

suggests that this group is not quite as likely to equal the performance of controls in June as the first Enriched Curriculum group. As we have already said above, the Control 1 group appears to have improved their position. Thus the difference in means is more likely to be at the lower end of this range when we do Year 4 comparisons with control scores at the end of Year 4. See under gender differences below for more information.

In Contrasting Areas schools, predicted age-corrected scores for the first Enriched Curriculum cohort are 3.0 - 5.2 standardised points behind the scores of the various control groups. Given that many of these children are still in the 'catch-up' phase of learning to read, they may still equal the performance of controls in June.

For the second Enriched Curriculum cohort in Contrasting Areas schools, the mean score is predicted to be below that of controls, in line with what the evaluation team has found for the first Enriched Curriculum group. Their mean score is predicted to be 43.2 in June against an actual score of 43.7 in June for the first Enriched Curriculum cohort, signifying that this group is on track to equal the performance of the first cohort.

Comparing February scores with February IQ scores

It is notable that control groups in both schools performed at higher levels in reading than would be expected from their proxy IQ measure, although in the case of the Contrasting Areas schools the difference was small.

All Enriched Curriculum groups underperformed with respect to IQ in the February testing. For the second cohort in Contrasting Areas schools, the comparison is definitely not meaningful because of the mismatch between PIPS and the Enriched Curriculum at this stage. For Year 2 and 3 groups, there are shortfalls of 5.2/ 3.7 and 4.0 points for the two Shankill Enriched Curriculum cohorts and the Contrasting Areas first Enriched Curriculum cohort respectively. Given that this shortfall is partly due to the improvement in non-verbal ability and picture vocabulary already discussed, we believe these figures tend to exaggerate the shortfall somewhat.

The effect of the teachers' experience with the Enriched Curriculum

In Shankill schools, there is no significant difference between Feb Year 4 reading scores for the first cohort and Feb Year 3 reading scores for the second cohort.

In Shankill schools, the predicted age-corrected score for the second Enriched Curriculum cohort is below measured end of Year 3 scores for the first Enriched Curriculum cohort at the end of Year 3 but only by 1.4 standardised points. The effect of four months learning and teaching may be enough to close the gap.

The predicted age-corrected mean score for the first Enriched Curriculum cohort in Contrasting Areas schools is lower than that of the measured end of Year 2 scores by 2.4 standardised points.

Taking these findings together, there is no evidence yet of a positive effect of teachers' experience on reading scores.

Gender and ability differences

One would normally expect to see girls doing better in reading than boys. PIPS measure the difference as 2.5 standardised points in the national sample.

There was a trend for girls to do better than boys in the second Enriched Curriculum group in Contrasting Areas schools, { $t = 1.74$, $df = 68$; $p < .09$ (2-tailed)}.

There were no gender effects in any other Enriched Curriculum group.

However, if we look at the regression lines for the first Enriched Curriculum group scores in Shankill schools on to their end of Year 3 scores, we again see an interaction between ability and gender. We see that is predominantly high-ability girls who have not performed well. Figure 2 shows the progress since June 2003. At the extremes, we can see the effect is very large, being approximately equal to a standard deviation. Low-ability girls have much improved scores on average.

There is no ready explanation for this effect. One possibility is a failure to challenge the higher ability girls, possibly partly through a shortage of reading resources. In interviews, teachers in some schools in the current year have already drawn the researcher's attention to this lack of reading resources. Another possibility is that boys have become so confident that they dominate at the high-ability end of the class, monopolising the teacher's attention. However, the effect is so marked that it is not easy to imagine how it could have come about.

When we look at school level data, we see that all the schools except one display this effect to some degree. Some show it more markedly than others, although sample sizes in individual schools are too small to allow of any proper judgement.

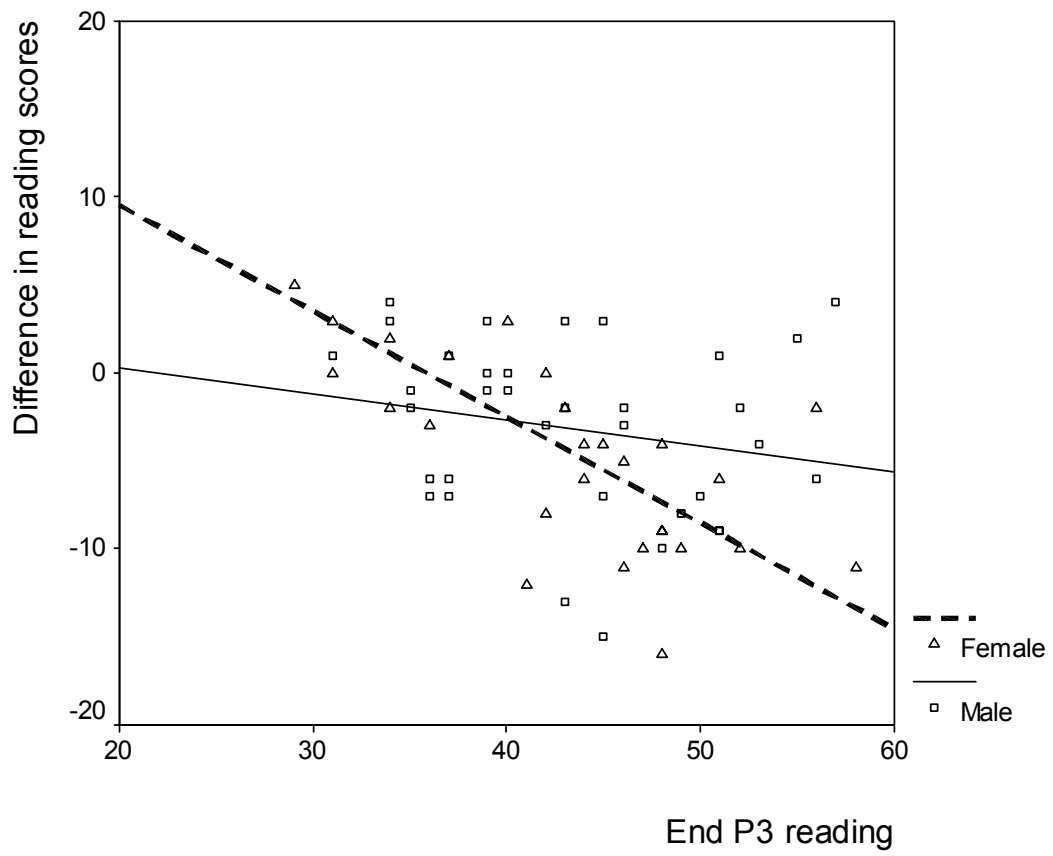


Figure 2
Progress in reading since June 2003
First Shankill Enriched Curriculum cohort

4. Interpretation and Discussion

The results are summarised in Table 5 below which lists significant group differences for the February round of testing in which Enriched Curriculum cohorts were compared with older controls. The pattern of results summarised in Table 5 is well supported by the other comparisons made in the analysis, namely predicted age-corrected minimum scores in June for Enriched Curriculum children compared with actual measurements for controls and the comparison of attainment with proxy IQ scores.

Summary of significance of results

*Table 5
February 2004 round of testing: Mean differences and associated significance level of comparisons of Enriched Curriculum cohorts with older controls*

		Non-verbal	Picture vocabulary	Reading	Mathematics
Shankill schools	EC1	ns*	ns*	↓5.5; $p < .001$	↓3.7; $p < .01$
	EC2	↑4.1; $p < .01$	ns*	ns*	ns*
CA schools	EC1	ns*	ns*	↓3.5; $p < .05^{**}$	ns*
	EC2	ns*	ns*	Not appropriate	Not appropriate

* ns means no significant difference

** Not expected to catch up until the end of the current school year

↑ *Enriched Curriculum children significantly outperformed controls by the given number of standardised marks*

↓ *Enriched Curriculum children significantly underperformed compared to controls by the given number of standardised marks*

The above table is intended to be a convenient summary. Unfortunately like all summaries, it only gives part of the picture. The reader is urged to consider the data in full before making a judgement of the current position.

It is quite clear that at present, we do not have the evidence to demonstrate that Enriched Curriculum children are consistently doing as well as controls. However, this report is very much coloured by the results for the first Enriched Curriculum cohort in Shankill schools, which are disappointing. In turn, the results for this group are coloured by the progress of the more able children, particularly high-ability girls in reading. In one sense, it is pleasing that boys are doing relatively well, since in control groups, girls outperform boys in both reading and mathematics. This difference between boys and girls is, within control groups, most marked at the high-ability end. If the relatively poor performance of high-ability girls is confirmed in June, then this is a problem which is relatively easier to fix than dealing with a group of disaffected low-ability boys.

When we look at the pattern of results in the other three Enriched Curriculum cohorts, there is nothing to give rise to serious doubt about its efficacy. Even the

reading score for the first Enriched Curriculum cohort in Contrasting Areas schools may very well recover before the end of the year because of the ‘take-off’ in reading which occurs once children grasp the basics.

In Shankill schools, we must also set the results for mathematics and reading in the context of the good non-verbal and picture vocabulary results. Although the results were not all significant, there is a clear pattern across the groups. The low picture vocabulary score in the first Enriched Curriculum cohort is entirely in keeping with the low reading scores for that group. It suggests that part of the explanation for the reading result may lie in poor oral language skills, a problem we pointed out in our Year 3 report (Sproule, Trew, Rafferty, Walsh, O’Neill, McGuinness and Sheehy 2003). Otherwise, **the results of predicted non-verbal and picture vocabulary scores as a whole suggest that there may be an improvement in proxy IQ scores in Shankill schools as a result of the Enriched Curriculum.** This is an indication that we should look for an improvement in attainment scores in the medium term.

In most groups, picture vocabulary scores were not keeping pace with non-verbal scores. The exception was the first Enriched Curriculum cohort in Contrasting Areas schools where picture vocabulary scores were not significantly different from non-verbal scores. This disparity between picture vocabulary and non-verbal scores was much more marked in all the Shankill groups, providing further support for the hypothesis in our last evaluation report that the focus on oral language may no longer be sufficient in Years 3 and 4 in Shankill schools (Sproule et al. 2003).

We do not have a complete analysis of item level data available but in reading, it appears that Enriched Curriculum children may be superior compared to controls in subtests which called for understanding of the text. On the other hand, it appears they may have performed less well in items calling for knowledge of spelling or formal grammar.

The researcher is in the process of conducting interviews with parents. Not many parents in Shankill schools have as yet been interviewed, but all who have been interviewed remarked on the children’s positive attitude to school. This finding is confirmed by teachers in ongoing interviews, including remedial teachers, who find the children much more ready to accept help⁴. Remedial teachers also believe the children have higher self-esteem. In Contrasting Areas schools, a much larger number of Year 3 parents have been interviewed. About 80% of these have remarked spontaneously on the high quality of their child’s reading, especially in reading aloud, often comparing them favourably with older siblings or with friends’ children taught in more traditional schools. Many of them also describe a ‘Suddenly, he could read fluently’ phenomenon which has happened since September 2003; these parents may be describing the take-off phenomenon we have talked about in this report. We now estimate that children will take approximately one year to catch up with controls after this phenomenon has occurred: That will not occur for

⁴ The children in Enriched Curriculum schools do not usually get intensive help until Year 3, or sometimes even Year 4.

the majority of children in Year 3 in Contrasting Areas schools until the end of the current year.

In mathematics, it appears that Enriched Curriculum children did very well in items calling for mental arithmetic and conceptual understanding but many appeared to be confused by items setting out sums in the traditional formal manner⁵: Some were able to correctly answer quite complex 2-stage word problems using mental mathematics but wrongly answered much simpler questions presented in a formal format. In some schools, it was also evident that graphs and tables had not become sufficiently familiar, as even brighter students did poorly on these items.

In the course of teacher interviews, it has come to our attention that children, in some schools at least, were tested last September in the Shankill area during the first week of term. The results were poor and some teachers believed they were being informed that they (the teachers) had not performed well. This has had a devastating effect on the confidence of those Year 4 teachers. They have felt constrained to abandon the Enriched Curriculum ethos and pressurise children to attempt work which they believe is too difficult for them. This is another possible reason for the poor results of this cohort in these tests. We understand that the tests used in this exercise were the tests which we rejected as being disadvantageous to Enriched Curriculum children for a number of reasons, and to lower ability Enriched Curriculum children in particular. These reasons were detailed in Section 2 of this report. In addition, there are a number of other considerations in connection with the procedure:

- Research has shown that children forget over the summer holidays. Testing in September is not good practice, if using for comparison norms which have been gathered over the whole year. Children will remember only really well consolidated information over the summer holidays. Children are especially unsettled during the first week of term.
- We have been careful to do small group testing with low-ability children. This ensures that they are not disadvantaged in the mathematics test because of poor reading ability. It also ensures that in reading, they understand the instructions.
- The procedure took no account of value added data. There was no baseline procedure available. This meant that teachers with a weak class, its members identified by us to be so from the beginning of their school career, felt themselves to be unjustifiably designated as failing.
- Teachers with small class sizes may have one or two children who skew the mean, again possibly identifying them as failures. Class sizes of at least 25 are required to make any reasonable judgement based on means; even this number is smaller than the ideal. That is why our evaluation employs larger sample sizes.

Patterns of differences between schools are now starting to emerge more clearly. Sample sizes are too small to permit any comparable judgement of each school's performance

⁵ Items such as $4 + * = 11$. No sums were presented in PIPS in vertical format, which may be more familiar to children.

from a single set of results. On the other hand, we can now look at progress over three years for the original half sample in the evaluation and progress since June 2003 in the alternate half sample⁶ in Shankill schools. At the moment, there is more variation in mathematics apparent than there is in reading. The range of average progress between schools varies by over one standard deviation in mathematics and by about two-thirds of a standard deviation in reading in Shankill schools. Good progress in mathematics does not guarantee good progress in reading and vice versa. The ranges are slightly smaller in Contrasting Areas schools, but again the variation in mathematics scores is larger. As we have previously indicated in this report and in previous reports, there are indications that there is a strong interaction between progress and baseline ability, meaning that schools with a lower average ability intake may be relatively disadvantaged by any straight comparison of average progress. To fully interpret the meaning of these complex interactions, we hope to use multi-level modelling in our final analysis when we will have a wide range of data covering four years on half of the children's school career (or three years in Contrasting Areas schools).

In conclusion, we reiterate that three out of the four Enriched Curriculum groups have performed well and will hopefully perform as well as controls in due course. We have generated some hypotheses about the poor results of the original pilot cohort but it may be that there are many contributory factors interacting in a complex manner to cause this unexpectedly poor performance in this group.

⁶ The alternate half sample were tested in June 2003 on PIPS for the first time and were also used in the current round of testing.

Reference:

Sproule, L., Trew, K., Rafferty, H., Walsh, G., O'Neill, B., McGuinness, C. and Sheehy, N. (2003). *The Early Years Enriched Curriculum Evaluation Project: Year 3 Report*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).