

The Early Years Enriched Curriculum Evaluation Project: EYECEP

Year 5 technical supplement:

Full description of the statistical analysis

(Data gathered during the period 2000-2005)

Not to be quoted without prior agreement from CCEA

Contact: csroule@ccea.org.uk

Sproule, L¹., McGuinness, C¹., Trew, K¹., Harvey-Smith, D¹.,
Bojke, C²., Rafferty, H¹., Walsh, G³., Sheehy, N¹. and O'Neill, B¹.
Queens University Belfast¹, Pharmerit International², Stranmillis University College³

September 2006

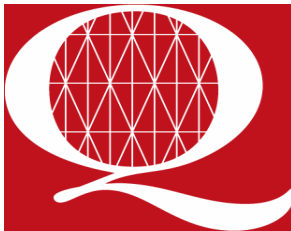


Table of Contents

	Page	
1	Scope of the supplement	4
2	Rationale for the statistical design	5
3	Experimental design and statistical technique	6
4	Further data and modelling issues	11
5	The multilevel statistical model	13
6	Model outcomes for reading and mathematics	16
7	Discussion of mathematics and reading outcomes	23
8	Other PIPS data	
	References	28

Acknowledgements

The research team would like to acknowledge:

the cooperation and contributions of the schools, teachers and pupils who have participated in the evaluation over the past five years and who have made this longitudinal study possible;

the parents of the children for their support and contributions;

the Board Officers from the Curriculum Advisory Support Service in all five Education and Library Boards who have been involved in the Enriched Curriculum;

Mr Chris Bojke, statistical analyst formerly of the National Foundation for Educational Research, who provided statistical support on the multi-level modelling aspects of the statistical analyses;

Dr Liz Sproule who manages the project on a daily basis, Ms Diane Harvey-Smith, research assistant, and Ms Jacqui Lockhart who is the project administrator;

the funding from Northern Ireland Council for the Curriculum Examinations and Assessment and particularly the support from Ms Marilyn Warren, our main contact at the Northern Ireland Council for the Curriculum Examinations and Assessment.

<p>The views expressed are those of the research team and not necessarily those of the Northern Ireland Council for the Curriculum Examinations and Assessment</p>
--

1. Scope of the supplement

The Early Years Enriched Curriculum Project and its evaluation have been running since September 2000, when the Enriched Curriculum (EC) was first introduced to six schools in the Shankill district of Belfast. This document is a supplement to the Year 5 report submitted by the evaluation team in June 2006 (Sproule, McGuinness, Trew, Rafferty, Walsh, Sheehy, and O'Neill 2006). It is concerned with the analysis of data gathered during the period 2000-2005 and combines new data gathered during the school year 2004-2005 with data gathered earlier and analysed in our report for the end of Phase 1 (Sproule, McGuinness, Trew, Rafferty, Walsh, Sheehy, and O'Neill 2005). It contains full details of the rationale for and a technical description of the multilevel statistical model of the PIPS data that is used to track the attainment of the children in reading and mathematics at the end of each year. It then moves on to describe the outcomes for children in the intervention and control groups, after controlling for known predictors such as gender and month-of-birth. This reporting period takes the pilot EC cohort from Shankill schools up to Year 5. The quantitative analysis differs in two major ways from that contained in our end of Phase 1 report (*ibid.*): The children being followed have completed Year 5 rather than Year 4 and the number of schools has been augmented from 12 to 24. The similar patterns of attainment found in the additional schools tend to further enhance our confidence that the findings are representative of all Enriched Curriculum schools.

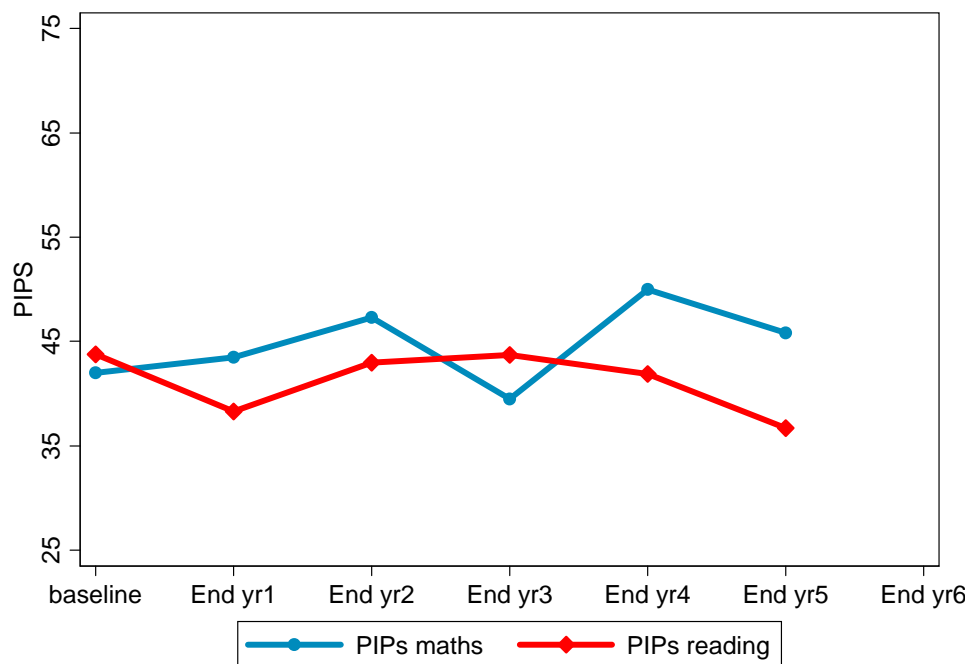
In order to get a full picture of the Enriched Curriculum and its evaluation, *this document should not be read in isolation from the annual reports*. These yearly reports are available on the website of the Northern Ireland Council for the Curriculum Examinations and Assessment (CCEA) at <http://www.ccea.org.uk/>. For those who do not wish to have details of the mathematical aspects of the analysis, the Year 5 report gives a coherent summary of the model and its outcomes.

2. Rationale for the statistical design

The impact of the EC on the pupils who are taught this new curriculum is measured by the difference in educational outcomes they achieve under this new regime and the outcomes they would have achieved under a different curriculum. That is, we measure the impact of the new curriculum as the *incremental* difference in outcomes expected for pupils being taught under different regimes.

An important component of measuring the effect of the EC is that the curriculum may have a changing impact over time. Indeed the true effect of the curriculum change may be intended to last well beyond the time frame of the evaluation and so it is important that the time-limited evaluation has an explicit attempt to incorporate a time dimension in its analysis. Figure 1 shows the progression over time of an individual in terms of their age-corrected PIPS mathematics and reading scores.

Figure 1: Academic Progression of an Individual Pupil over Time



The figure shows the results measured at time points year 0 (baseline), end of years 1, 2, 3, 4 and 5 in both mathematics and reading. This particular individual shows a

fairly stable journey over time with some variation – for example an apparent blip in mathematics results recorded at the end of Year 3. The objective of the analysis is to identify what this path would have looked like under a different curriculum regime, attempting to separate out random variation and that which is systematically related to the choice of curriculum, and hence identify what the incremental difference would be (in fact this individual experienced the EC and so the trajectory/path we would want to consider is this actual path against their estimated path under the traditional curriculum.)

This discussion highlights the essential *evaluation problem*, that is, we are interested in the incremental difference between two curriculum choices, but for any individual we may only observe the outcomes from one of those curricula. We refer to the unobserved outcomes from the alternative curriculum as the *counterfactual* outcomes. If it were possible to observe both of these outcomes, measuring the incremental effect for each individual and hence for a group of individuals or population would be a fairly trivial operation. However given the mutually exclusive nature of the data, it is a key joint requirement of both the experimental design and the statistical technique to try and ensure an unbiased estimation of the difference. In essence this is achieved by comparing results between comparable groups under different regimes.

3. Experimental Design and Statistical Technique

The analysis adopts a quasi-experimental design which ensures that we obtain observations under different regimes between comparable groups. Specifically, we compare the progress of EC intervention group children with control children attending the same school, often with exactly the same teachers. The control group children are one or two years older than the intervention children and so experienced the traditional pedagogy but in the same classroom context in all other respects. The control children were therefore tested one or two calendar years earlier than the EC children on any given test, so that children could be compared at the same stage in their school career.

In the first year of the project, a random sample of half of each intervention and control class was tested at baseline and the end of the first year. There were nine classes in six inner city schools (for each group $N \geq 84$). As the project continued, the sample size has expanded in terms of both numbers of children and numbers of schools, such that there are now 950 EC children and 696 controls, with 427 of these children being tested for the first time this year in the twelve schools new to the project. Due to the requirements of the funding body, the alternate half sample has also been tested at different times from the original sample. Thus, very few children have generated data at each wave of testing, though importantly, missing results are nearly always a function of the experimental design rather than any self-selection mechanism (with the exception of earlier results from the new schools). In these circumstances, one of the great strengths of using multilevel models estimated using maximum likelihood is that it allows us to make use of every piece of test data.

The rationale for our strategy of year-ahead and two-years-ahead control groups might be argued to suffer from creeping implementation of the Enriched Curriculum, as some of the schools were experimenting with some of the ideas before the formal implementation of the Enriched Curriculum. i.e. some of the control observations may be partly contaminated. Our response is as follows:

1. Similar issues of contamination arise in the medical literature where it is not always possible to ensure patients adhere to compliance instructions or do not access the intervention via private means after patients have been randomised into groups. However even where it is known patients deviate from their randomised allocation, analysis is still often conducted using data on these patients and analysed according to the group they were allocated. Such trials are known as *intention to treat (ITT) analyses*. The rationale that underpins the use of ITT analysis is that it incorporates in a randomised controlled trial (RCT) an element of external validity (whereas RCTs are usually designed to ensure internal validity). This is useful if we are aiming to measure the incremental difference that a new curriculum would make if rolled out nationally (i.e. effectiveness) rather than an explanatory assessment of the effect of the new curriculum under perfect laboratory conditions (i.e. efficacy). In other words “Intention to treat gives a pragmatic estimate of the benefit of a change in treatment policy rather

than of potential benefit in patients who receive treatment exactly as planned.” (Hollis and Campbell, 1999). Although the use of ITT analysis is predominantly in medical evaluations, the rationale applies in this educational context: schools and teachers do have a degree of autonomy in choosing how they educate their pupils and this may include elements which teachers have ‘cherry-picked’ from the EC even if the EC approach is not adopted as a formal policy. If the pragmatic effectiveness is the measure of interest, then ITT (i.e. analysing pupils according to their initial group control or EC) is justified.

2. This issue of our control groups was first raised in inner-city schools in the second year of the first phase of the project. For that reason, we assessed year-ahead and two-year-ahead control groups in the original 12 schools at that time in order to counter the possibility that different cohorts of pupils are systematically different from each other (in observed or unobserved characteristics) and thus confounded with the control and intervention groups. Using two years of controls, we find that there are no significant differences between the two control groups at any point in the study (including data up to Year 4), indicating no sign of systematic differences (in observed characteristics and outcomes). Statistically, this gives us greater scope in attributing any observed systematic differences between control and intervention groups (after allowing the observable and measurable differences in pupil characteristics) as being due to the intervention rather than initial differences between pupils. For example, we do not see the significant dip in performance at the end of the first and second years with either of the control groups that we see with the Enriched Curriculum groups. This is now confirmed retrospectively by the multilevel model.
3. Year 1 and Year 2 teachers are consistently telling us that their practice changes dramatically with the Enriched Curriculum, even if some of its methodology had been introduced previously. One of the most important changes appears to be the perceived permission to let the child’s ability and progress dictate the pace.
4. The children’s proxy IQ scores and where available, their baseline scores, act as separate benchmarks for any value added by the Enriched Curriculum. In a sense, each child is their own control. To date, the baseline scores and proxy IQ scores

have shown strong consistency with attainment scores at the end of the third and fourth years, both for Enriched Curriculum and control groups. Nor is there any significant trend for changes in IQ scores, the existence of which would indicate that it might be invalid to use them as a baseline for value added.

5. Our structured classroom observation instrument, the Quality of Learning Instrument, has been easily able to distinguish EC from other control classes in non-Enriched Curriculum schools, even though some of these different controls have claimed to be “doing the Enriched Curriculum anyway” and have included a lot of structured play in the curriculum (Walsh, Sproule, McGuinness, Trew, Rafferty and Sheehy, 2006).
6. A large number of parents have reported on the very different experiences of their EC children and older siblings following the traditional curriculum in the early years (Sproule, Trew et al., 2002, 2003, 2004).
7. Several researchers involved in testing the children have reported on the very different levels of confidence, verbal communication skills and vitality of EC children compared with controls.

Overall, we believe the design helps ensure we achieve our objective that we observe outcomes under the conventional curriculum and EC conditions for comparable groups of pupils. This provides the sound basis for the statistical model to disentangle the effects of different curricula from the other sources that cause individuals to have different progressions over time.

Whilst the data provide the means of estimating the effect of different curricula, the actual statistical model used is further underpinned by the application of a simple theoretical economic model, Human Capital Theory (HCT). HCT conceptually formulates education as an investment by which individuals, given their initial stock of characteristics (genetic, IQ, parental input and capital) may improve their expected outcomes over time. This simple theory thus stipulates that the observed educational outcome at any point in time will be a function of an individual’s characteristics and of the education they receive. As such, there is a rationale for gearing the statistical

model towards understanding the impact of an individual's characteristics as well as that of the different curricula. The rationale for including individual characteristics is heightened if we wish to consider how the impact of different curricula may systematically differ across different sub-groups of the population.

Thus the model we implement in this research is difference-in-difference regression model. Difference-in-difference models conceptually work by estimating a baseline progression for an individual's outcomes, which may be no change, a linear or non-linear progression over time, and estimating the incremental effect of an intervention by the expected departure an individual's outcomes would have from this baseline estimate (hence difference-in-difference). In practice this is achieved by estimating an intervention dummy, which would measure a time constant shift in outcomes (and is the standard measure of an incremental difference), and either an intervention continuous time slope or additional time-point intervention dummies (which will allow for the effect to change over time). Thus a standard model may look like:

$$y_{it} = \beta_0 + \sum_{n=3}^N \beta_n x_{nit} + \beta_1 * time_{it} + \gamma_0 * int_{it} + \gamma_1 * int * time_{it} + \varepsilon_{it}$$

The term $\beta_0 + \sum_{n=3}^N \beta_n x_{nit}$ would capture the starting point for the progression of an individual over time and β_1 models the baseline change in expected outcomes over time, a quadratic term could be added to allow for a non-linear quadratic progression. The γ terms capture the incremental difference that the intervention makes to the expected outcomes. First the term γ_0 is attached to a dummy term, set to 1 if the observation from the individual is obtained under intervention conditions, and to 0 otherwise (i.e. it will be set to zero for all individuals for the baseline observation). The estimate of this parameter thus represents the time invariant effect of the intervention i.e. a vertical shift to the baseline intercept which measures an effect of the intervention which is constant over time. The term γ_1 is attached to an interaction of time and the intervention dummy, it is set to zero if the individual is in the control group and is equal to $time$ if the individual is in the intervention group. The estimate of γ_1 thus gives the difference in the expected progression over time due to the

intervention. As with modelling the baseline/control progression a quadratic term could be included. Note that the overall effect of the intervention at time t is given by the combination of these parameters and not just the value on the intervention dummy. Thus, in terms of inference, the hypothesis test of the intervention of having no effect in either a time invariant shift or time varying progression, may be conducted by testing the null hypothesis that γ_0 and γ_1 are significantly different from zero.

Other variants of this model are possible, such as modelling time in discrete dummy variables rather than continuous time. The benefit of using discrete dummy variables is that no relationship over time is assumed (either linear or quadratic) however where there are several time periods, creating dummies for each period becomes cumbersome, lowers degrees of freedom and is not amenable to extrapolation beyond the period of data collection. The cost of using dummy variables increases if we wish to interact individual characteristics with the intervention variables to capture sub-groups differences in interaction effects. Whether continuous or discrete time is used, the basic model above captures the rationale of the difference-in-difference model.

4. Further Data and Modelling Issues

Three further data issues require additional commentary. Firstly outcomes are clustered within individuals and within schools. Secondly and in a related issue, the outcomes from an individual may be a function of individual characteristics which are not observable to the analyst (parental input at home for example). The clustering of results is often presented in the literature as a problem which requires solving – this is because the clustering implies that the regression assumption of uncorrelated error terms may be untenable. However, the multiple observations per individual allow considerable scope in addressing the possibility that outcomes are a function of unobservable individual characteristics. Thus the multiple observations per individual are potentially more of a blessing than a curse.

Both of these issues are handled by the same addition to the statistical model – the imposition of a multi-level (a.k.a. panel data) component to the basic difference-in-difference model. It is not the purpose of this section to give a detailed exposition of multi-level models, but we will identify that we use a *random-effects* (a.k.a. random intercept) model to address these two issues. The random-effects model estimates a time-consistent individual specific constant term (which is estimated under the assumption that these terms are random draws from an underlying normal distribution with a mean of zero and an unknown but constant standard deviation). Such an effect can be estimated at each level i.e. an individual specific random effect which influences all observations of that individual (and no others) and a school specific random effect which influences all observations of all individuals within that school. The random-effects model is amenable to statistical testing which may be used to determine whether the additional complexity is necessary.

The third major statistical issue is the fact that we do not observe outcomes for all individuals at all time points. That is, the data are perpetuated with missing data. The key issue for the statistical analysis is whether the mechanism by which the data are missing causes any biases or not i.e. whether the missing mechanism is *ignorable* or *non-ignorable*. Biases may be caused if the mechanism is a causal function of the value the outcome may take, so for example if an individual obtains a very poor test score and, as a result, the score is not recorded – such mechanisms are considered non-ignorable and require further consideration. However, if the mechanism is either totally random or a causal function of an observed characteristic (rather than a causal function of the unobserved outcome measure), then the missing data need not necessarily cause any additional problems. In this example if data are missing because the school at this point were not included in the data collection and we include school variables in the regression model, then it can be shown that the missing mechanism is incorporated in the model (i.e. we may successfully *condition* on the cause of missingness) and thus the mechanism is ignorable. Furthermore the fact we observe different numbers of outcomes per individual (i.e. the panel is unbalanced) causes no additional problems for a multi-level model when maximum likelihood is used in the estimation procedure (Rasbash et al., 2000).

Thus we conclude that no additional steps are required to deal with missing data in this analysis. For further details on dealing with missing data, the interested reader is directed towards (Schafer and Graham, 2002).

5. The Multilevel Statistical Model

The principle components of the model are:

- Pupils (described by Level 1 variables) are clustered within schools (described by Level 2 variables).
- Schools are clustered within school group (Level 3, Shankill or not Shankill)
- Multiple observations are obtained for many pupils over time.
- Allocation to EC or control groups is exogenous.
- Whether an observation is observed at a time point is exogenous (i.e. missingness of data is ignorable)

Thus the model chosen is the multivariate random effects (intercept) difference-in-difference with three levels. Level one identifies the time period, level two identifies the pupil and level three identifies the school. Random effects are estimated at pupil and school level.

The outcomes at time t are modelled as a function of an individual's observable characteristics (gender, month of birth, a measure of IQ, discrete time (to capture a non-linear transition over time); a random-effect for each individual which is argued to capture the effect all time-invariant unobserved characteristics; school observable characteristics (% of FSM in a school, inner-city or not); a random-effect for each school; and an EC dummy with EC time interactions. The model was further supplemented by creating IQ dummies which captured whether the individual had a mean PIPS IQ score below 45 points or above 55 points (baseline omitted category being an individual with a PIPS IQ score between these values). These dummy variables were interacted with constant terms and time variables (including EC time variables) to see whether pupils with different latent abilities progressed differently over time. A further school variable indicating the average IQ of pupils within a

school (i.e. a cluster mean) is added mainly to ensure that the school random effect is uncorrelated with the included explanatory variables. See Skrondal and Rabe-Hesketh (2004, p53) for further details.

As stated the model which treats time as discrete time points with dummy variables for each time period is estimated. The data did not appear to support a polynomial continuous time model.

The full discrete time model estimated is thus:

$$\begin{aligned}
 pips_{jkl} = & \beta_0 + \beta_1 IQL_{jkl} + \beta_2 IQH_{jkl} + \beta_3 sex_{jkl} + \beta_4 birth_month_{jkl} + \beta_5 sch_fsm_l + \\
 & \beta_6 sch_IQ_l + \beta_7 inner_l + \beta_8 IQLyr1_{jkl} + \beta_9 IQLyr2_{jkl} + \beta_{10} IQLyr2_{jkl} + \beta_{11} IQLyr3_{jkl} + \\
 & \beta_{12} IQLyr4_{jkl} + \beta_{13} IQLyr4_5_{jkl} + \beta_{14} IQLyr5_{jkl} + \beta_{15} IQLyr6_{jkl} + \beta_{16} IQLyr1EC_{jkl} + \\
 & \beta_{17} IQLyr1_5_{jkl} + \beta_{18} IQLyr2EC_{jkl} + \beta_{19} IQLyr2_5EC_{jkl} + \beta_{20} IQLyr3EC_{jkl} + \\
 & \beta_{21} IQLyr3_5EC_{jkl} + \beta_{22} IQLyr4EC_{jkl} + \beta_{23} IQLyr5EC_{jkl} + \beta_{24} IQMyr1_{jkl} + \beta_{25} IQMyr2_{jkl} \\
 & + \beta_{26} IQMyr3_{jkl} + \beta_{27} IQMyr3_5_{jkl} + \beta_{28} IQMyr4_{jkl} + \beta_{29} IQMyr4_5_{jkl} + \beta_{30} IQMyr5_{jkl} \\
 & + \beta_{31} IQMyr6_{jkl} + \beta_{32} IQMyr1EC_{jkl} + \beta_{33} IQMyr1_5EC_{jkl} + \beta_{34} IQMyr2EC_{jkl} + \\
 & \beta_{35} IQMyr2_5EC_{jkl} + \beta_{36} IQMyr3EC_{jkl} + \beta_{37} IQMyr3_5EC_{jkl} + \beta_{38} IQMyr4EC_{jkl} \\
 & + \beta_{39} IQMyr5EC_{jkl} + \beta_{40} IQHyr1_{jkl} + \beta_{41} IQHyr2_{jkl} + \beta_{42} IQHyr3_{jkl} + \beta_{43} IQHyr3_5_{jkl} \\
 & + \beta_{44} IQHyr4_{jkl} + \beta_{45} IQHyr4_5_{jkl} + \beta_{46} IQHyr5_{jkl} + \beta_{47} IQHyr6_{jkl} + \beta_{48} IQHyr1EC_{jkl} \\
 & + \beta_{49} IQHyr1_5EC_{jkl} + \beta_{50} IQHyr2EC_{jkl} + \beta_{51} IQHyr2_5EC_{jkl} + \beta_{52} IQHyr3EC_{jkl} + \\
 & \beta_{53} IQHyr3_5EC_{jkl} + \beta_{54} IQHyr4EC_{jkl} + \beta_{55} IQHyr5EC_{jkl} + \zeta_{kl} + v_l + \varepsilon_{jkl}
 \end{aligned}$$

Where $pips_{jkl}$ refers to the PIPS score of observation j from individual k in school l (separate models are done for maths and reading scores).

In terms of model variables: β_0 is the standard constant term (which will have the school and individual level random effects added to it); IQL and IQH are additional dummy variables indicating whether the individual has a mean PIPS IQ score below 45 or above 55. The estimated coefficient on these terms indicates the time-invariant shift in the constant term that an individual in either of these groups would be expected to have relative to an individual in the medium IQ group, all other things being equal. Sex is a dummy variable capturing any systematic differences between males and females after allowing for differences in IQ etc. It is set to 1 if the individual is male, 0 otherwise. $Birth_month$ is a variable capturing the month of

birth of an individual. It is centred on January (i.e. value of *birth_month* = 0) and increases by 1 for every month after this the individual is born (i.e. *birth_month* = 1, if birth month is in February) and decreases by one unit for every month before January. *Sch_fsm* and *inner* are school level variables. *Sch_fsm* is set to the percentage of pupils eligible for Free School Meals in a school (e.g. 25% is recorded as 25) and *inner* is a dummy variable set to 1 if the school is from the inner city region. This dummy is designed to pick up any difference between schools in this region and schools in the mainstream which are not captured by differences in %FSM, pupil IQ levels, etc. *Sch_IQ* is a cluster-mean of the IQ of pupils in a given school. These variables (along with the random effects, described below) are used to establish the baseline scores for individual pupils. The variables that follow are there to capture the progression over time, i.e. the deviation from the baseline, of the individuals.

The estimated coefficients on variables such as *IQLyr1* or *IQMyr3.5* show the estimated deviation from baseline at that particular time point for an individual in that particular time period (in this particular example a pupil from the lower IQ group at the end of Year 1 and a pupil from the medium IQ group after three and a half years (i.e. Feb in Year 4). Thus if an individual pupil from the medium IQ group had a baseline score of 48 and the estimated coefficient attached to *IQMyr3.5* was estimated to be -3 (i.e. $\hat{\beta}_{37} = 3$, the accent or 'hat' above the beta indicates it is an estimate) then the expected PIPS score for that individual at Feb Year 4 under the conventional curriculum would be 45.

The estimated coefficients on variables such as *IQLyr1EC* or *IQMyr3.5EC* show the estimated incremental effect of the Enriched Curriculum – the substantive focus of the estimation. These estimated coefficient represent the difference that a pupil is expected to achieve between being taught the conventional curriculum and the Enriched Curriculum. Positive values indicate that pupils have a higher expected outcome under the Enriched Curriculum; negative values indicate that a higher expectation would be achieved under the conventional curriculum. Statistical significance of these coefficients in a regression model indicate a statistically significant difference between outcomes. Note that this 'difference in difference'

interpretation only stands where there are corresponding Enriched Curriculum variables and conventional curriculum variables.

The final variables ζ_{kl} , ν_l and ε_{jkl} represent the multi-level component of the model. ζ_{kl} represents a time-consistent individual pupil specific random effect and ν_l represents a time-consistent school specific random effect. ε_{jkl} is the standard econometric error term. It is assumed that:

$$\zeta_{kl} \sim N(0, \sigma_{\zeta}^2)$$

$$\nu_l \sim N(0, \sigma_{\nu}^2)$$

$$\varepsilon_{jkl} \sim N(0, \sigma_{\varepsilon}^2)$$

and that these random effects and error terms are uncorrelated across clusters and uncorrelated with the other explanatory variables. As the expectations are known (0 in each case) the estimation procedure will estimate the relevant variances (σ_{ζ}^2 , σ_{ν}^2 and σ_{ε}^2).

6. Model outcomes for reading and mathematics

Table 1 below shows the estimated models, listing the variables with their estimated coefficients and the associated standard errors, significant coefficients being highlighted in bold. The models were estimated using the GLLAMM command in Stata 8. Graphical presentations of pupil progress are found below in Figures 2 and 3 for mathematics and literacy respectively.

Main effects at the individual pupil level

The following text is a summary. Quantitative details of coefficients are given in Table 1 below. Coefficients in bold type in Table 1 indicate significant effects. Main effects are given for various predictors; at the individual pupil level, school level (mean ability of the school, free school meals percentage), school group level (Shankill or not), followed by the significant interactions between various predictors at each time point of data collection in the child's school career (Years 1 – 5). Please

take note that some coefficients refer to Shankill schools only, as identified at the bottom of Table 1, as Year 5 data is only complete for Shankill EC and Shankill controls and Year 6 data for Shankill controls only.

Main effects at the individual level

As might be expected, IQ group (low, medium or high) is a significant predictor of outcomes in both reading and mathematics. The higher the ability of the child, the better their performance compared to national average, even after making allowance for the individual baseline.

Gender is a significant effect for reading but not for mathematics. The effect size is approximately one-third of a standard deviation in reading, considered a moderate effect.

Month-of-birth is a significant effect for both mathematics and reading. It is approximately one-quarter of a standard deviation for mathematics, a moderate effect. For reading, the effect is much smaller, about half of that for mathematics.

These effects are cumulative. In reading for example, the youngest boy in the class is expected to score more than half a standard deviation below the oldest girls. This would be considered a large effect size.

Main effects at the school level

The average developed ability *of the class*, as measured by the mean of the PIPS vocabulary and non-verbal measures in that class (class/school variable), is a significant predictor of outcomes. This is similar to a result found in Newcastle schools in the UK using PIPS tests (Tymms, Heron et al., 2005). A similar effect on attainment has also previously been reported by White (1982). This effect of the class mean score is characterised as a peer effect: A child in a class of peers who are highly advanced in attainment will do better than a child in a class of less advanced peers.

For mathematics, this average class ability effect was 0.66 points per PIPS standardised point difference in the class mean. Taking the difference between the top school in terms of ability, with a mean ability of 61.5, and the bottom school, with a mean of 43.6, the difference in outcomes related to these schools for a given child will

be 11.8 points, which is a very large effect. As seen in Table 4, when this variable was included in the analysis, the percentage of free school meals was not a significant predictor. If we excluded this class ability variable from the analysis, then the free school meals variable would become significant. The class mean ability is preferred because it explains more of the variability between scores.

This average class ability effect was 0.33 points per PIPS standardised point for reading, which is much smaller than for mathematics. Taking the difference between the top school in terms of ability, with a mean ability of 61.5, and the bottom school, with a mean of 43.6, the difference in outcomes related to these schools for a given child would be 5.9 points, which is a large effect. As seen in Table 1, when this variable was included in the analysis, the percentage of free school meals was again not a significant predictor. If we excluded this class ability variable, then the free school meals variable would become significant. The class mean ability is preferred because it explains more of the variability between scores.

Main effects at the school group level

There were no significant main effects at school group level. There was no effect of being in the Shankill group of schools, after taking account of the effects described above. It follows that a child with a given ability, and in a class with a certain average ability, is expected to make similar progress irrespective of the area in which the school he/she is attending is located, after taking the peer effect into account.

Effect of social deprivation combined with other main effects

It should be noted that all the above effects are cumulative, such that combining month-of-birth, gender and school mean IQ effects indicate that when deprivation (as evidenced by school mean IQ or percentage of free school meals) is combined with month-of-birth and gender effects, the cumulative effect can be very large indeed. In reading, the youngest boy of a given ability in the most deprived school is expected to perform 10.7 PIPS points behind the oldest girl *of equal ability* in the least deprived school. For mathematics, there is no gender effect; the youngest child of a given ability in the most deprived school is expected to perform 14.9 points behind the oldest child *of equal ability* in the least deprived school.

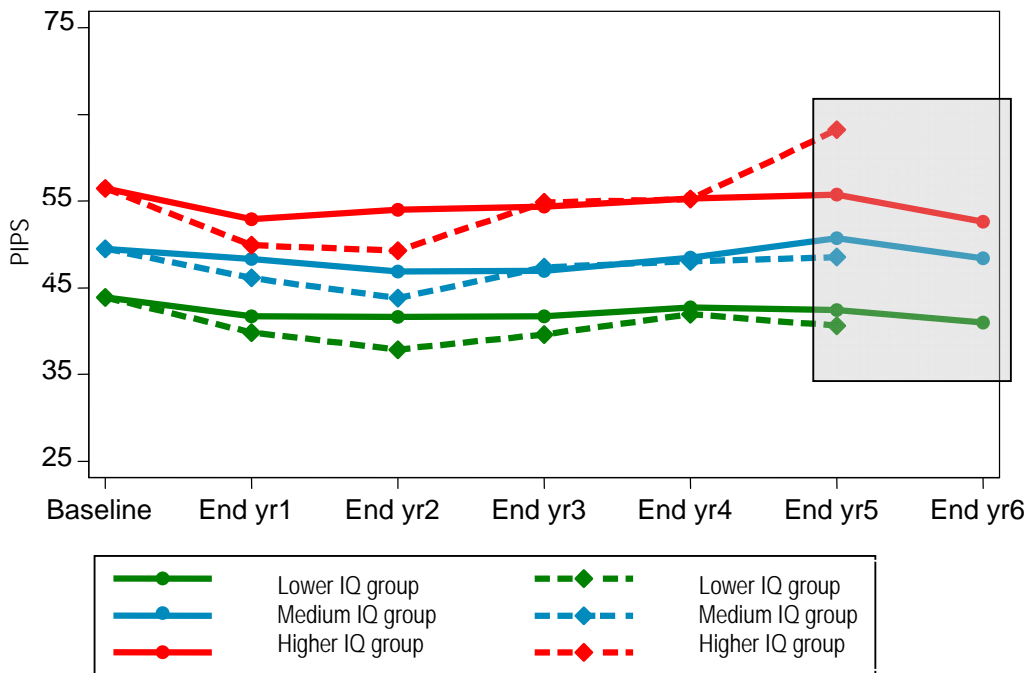
Table 1. Multi-Level Model Regression Results					
<i>Variable names containing EC denote Enriched Curriculum effects</i>		BOLD type indicates a statistically significant result			
Variable name	Variable information	Effect on PIPS Reading score		Effect on PIPS mathematics score	
		<i>Coefficient</i>	<i>Std error</i>	<i>Coefficient</i>	<i>Std error</i>
<i>Constant term</i>		36.77	7.37	16.75	6.29
Main effects on both EC and controls					
IQL (Low IQ)		-6.35	0.77	-5.61	0.68
IQH (High IQ)		5.80	1.07	6.98	1.00
Gender	<i>Boy</i>	-3.15	0.35	-0.59	0.31
Month of birth	<i>Per month compared with January</i>	±0.14	0.05	±0.25	0.04
School % FSM	<i>Per % point</i>	0.02	0.04	0.04	0.03
School mean IQ	<i>Per PIPS point</i>	0.33	0.13	0.66	0.12
Inner city		-1.20	1.79	0.88	1.40
Interactions Low IQ group					
<i>IQLyr1</i>	<i>End Year1vs baseline</i>	-4.21	0.92	-2.20	0.85
<i>IQLyr2</i>	<i>End Year2</i>	-1.94	1.17	-2.26	1.06
<i>IQLyr3</i>	<i>End Year3</i>	-2.26	0.95	-2.16	0.86
<i>IQLyr3.5</i>	<i>Mid Year4</i>	-1.33	1.88	-2.09	1.71
<i>IQLyr4</i>	<i>End Year4</i>	-1.69	0.65	-1.17	0.59
<i>IQLyr4.5</i>	<i>Mid Year5</i>	-1.34	1.18	-4.06	1.07
<i>IQLyr5</i>	<i>End Year5</i>	-0.91	1.58	-1.49	1.44
<i>IQLyr6*</i>	<i>End Year6*</i>	-6.88	1.08	-2.89	0.99
<i>IQLyr1EC</i>	<i>End Year1 EC</i>	-1.85	1.21	-1.85	1.09
<i>IQLyr1.5EC</i>	<i>Mid Year2 EC</i>	-10.18	2.51	-9.48	2.27
<i>IQLyr2EC</i>	<i>End Year2 EC</i>	-2.13	1.25	-3.76	1.13
<i>IQLyr2.5EC</i>	<i>Mid Year3 EC</i>	-8.74	1.14	-7.94	1.03
<i>IQLyr3EC</i>	<i>End Year3 EC</i>	-1.29	0.99	-2.11	0.89
<i>IQLyr3.5EC</i>	<i>Mid Year4 EC</i>	-2.16	2.11	-2.32	1.93
<i>IQLyr4EC</i>	<i>End Year4 EC</i>	-1.55	0.85	-0.80	0.76
<i>IQLyr5EC</i>	<i>End Year5 EC*</i>	-2.05	1.82	-1.79	1.65
Interactions Middle IQ group					
<i>IQMyr1</i>	<i>End Year1vs baseline</i>	-3.13	1.12	-1.16	1.02
<i>IQMyr2</i>	<i>End Year2</i>	-1.66	0.84	-2.65	0.76
<i>IQMyr3</i>	<i>End Year3</i>	-2.50	0.75	-2.56	0.69
<i>IQMyr3.5</i>	<i>Mid Year4</i>	-3.08	1.08	-3.31	1.00
<i>IQMyr4</i>	<i>End Year4</i>	-2.07	0.65	-1.03	0.59
<i>IQMyr4.5</i>	<i>Mid Year5</i>	-1.47	1.11	-3.21	1.01
<i>IQMyr5</i>	<i>End Year5</i>	0.22	0.97	1.20	0.90
<i>IQMyr6*</i>	<i>End Year6*</i>	-5.87	1.06	-1.12	0.96
<i>IQMyr1EC</i>	<i>End Year1 EC</i>	-3.42	1.28	-2.23	1.17
<i>IQMyr1.5EC</i>	<i>Mid Year2 EC</i>	-9.20	1.20	-8.64	1.09
<i>IQMyr2EC</i>	<i>End Year2 EC</i>	-5.47	0.87	-3.06	0.79
<i>IQMyr2.5EC</i>	<i>Mid Year3 EC</i>	-4.61	0.94	-3.13	0.86
<i>IQMyr3EC</i>	<i>End Year3 EC</i>	-1.66	0.75	0.42	0.67
<i>IQMyr3.5EC</i>	<i>Mid Year4 EC</i>	-1.82	1.53	-1.08	1.39
<i>IQMyr4EC</i>	<i>End Year4 EC</i>	-0.48	0.75	-0.43	0.69
<i>IQMyr5EC*</i>	<i>End Year5 EC*</i>	-3.32	1.40	-2.20	1.27
Interactions High IQ group					
<i>IQHyr1</i>	<i>End Year1vs baseline</i>	-4.07	2.22	-3.58	2.02
<i>IQHyr2</i>	<i>End Year2</i>	0.13	1.30	-2.50	1.21
<i>IQHyr3</i>	<i>End Year3</i>	0.68	1.16	-2.14	1.07
<i>IQHyr3.5</i>	<i>Mid Year4</i>	2.05	1.42	-4.28	1.32
<i>IQHyr4</i>	<i>End Year4</i>	0.57	1.16	-1.16	1.08
<i>IQHyr4.5</i>	<i>Mid Year5</i>	-1.66	2.89	-2.90	2.76
<i>IQHyr5</i>	<i>End Year5</i>	3.51	1.28	-0.72	1.18
<i>IQHyr6*</i>	<i>End Year6*</i>	-4.84	2.01	-3.89	1.88

Table 1 continued					
Variable name	Variable information	Effect on PIPS Reading scores		Effect on PIPS mathematics scores	
		<i>Coefficient</i>	<i>Std error</i>	<i>Coefficient</i>	<i>Std error</i>
Interactions High IQ group					
<i>IQHyr1EC</i>	<i>End Year1 EC</i>	-3.87	2.46	-2.99	2.24
<i>IQHyr1.5EC</i>	<i>Mid Year2 EC</i>	-7.28	1.69	-8.64	1.55
<i>IQHyr2EC</i>	<i>End Year2 EC</i>	-7.61	1.16	-4.73	1.07
<i>IQHyr2.5EC</i>	<i>Mid Year3 EC</i>	-3.46	1.39	-0.45	1.28
<i>IQHyr3EC</i>	<i>End Year3 EC</i>	-4.16	1.00	0.49	0.91
<i>IQHyr3.5EC</i>	<i>Mid Year4 EC</i>	-4.32	4.70	-1.01	4.31
<i>IQHyr4EC</i>	<i>End Year4 EC</i>	1.64	1.27	-0.12	1.17
<i>IQHyr5EC*</i>	<i>End Year5 EC *</i>	4.45	6.05	7.45	5.14
<i>Variance</i>					
σ_v^2 (school re)	<i>School level variability</i>	2.32	0.97	1.34	0.59
σ_ϵ^2 (pupil re)	<i>Pupil level variability</i>	30.44	1.64	24.78	1.38
σ_ϵ^2 (error term)	<i>Error variance</i>	26.05	0.84	21.92	0.70
<i>Model goodness of fit indicating that the model explains the data well</i>					
<i>log-likelihood</i>		-11367.72		-11053.60	
* Result applies to Shankill data only at this stage					

Interactions

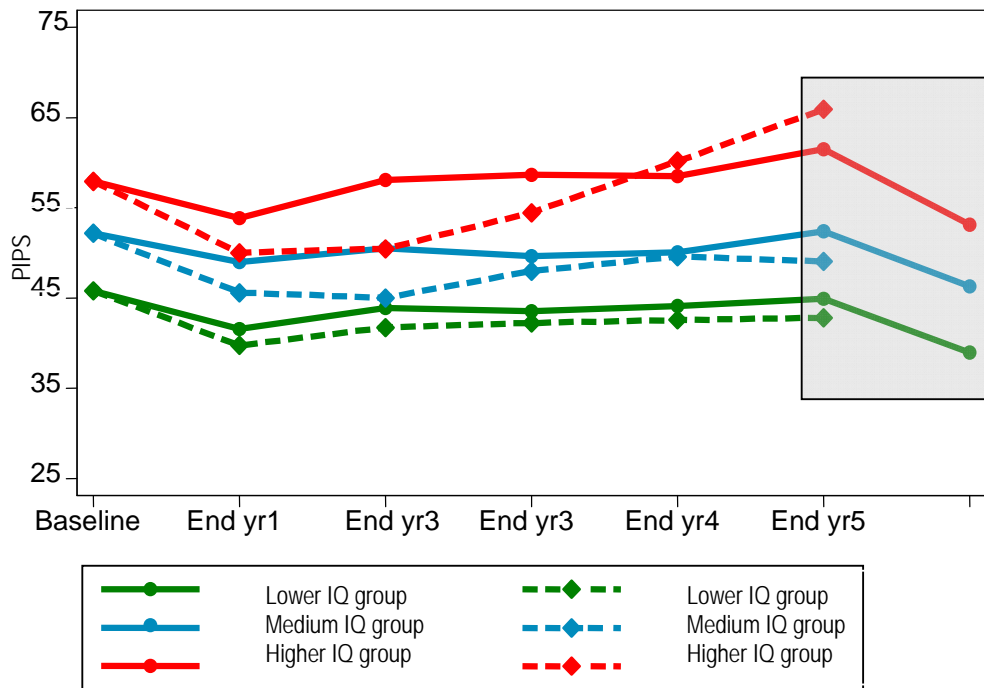
Interactions are summarised in Table 1 above. Coefficients in bold type indicate significant interactions. In all three ability groups, low, middle and high, we can see a significant effect of being in the EC on either reading or mathematics or both until the end of Year 3, with EC children having depressed scores compared with controls. For the first time since the evaluation began, and for the middle ability group in particular but to a lesser extent for other ability groups, we start to see some depression of scores for control groups over the first four years (see items labelled all groups in the explanation column of Table 1), with the effect lasting into the fourth year. Figures 2 and 3 below show the combined effects clearly for mathematics and reading respectively.

Figure 2. PIPS Mathematics Attainment over Time for EC/Control group at different levels of ability (EC = broken line; Control=solid line)



Data points in the shaded box refer to a limited sample, mostly Shankill schools only. This graph refers to a girl with a January birthday. For other subgroups the graph will move up and down in accordance with reported effects, for example, up by 0.25 points for a girl with a December birthday.

Figure 3. PIPS Reading (English) Attainment over Time for EC/Control group at different levels of ability (EC = broken line; Control= solid line)



Data points in the shaded box refer to a limited sample, mostly Shankill schools only. This graph refers to a girl with a January birthday. For other subgroups the graph will move up and down in accordance with reported effects, for example, up by 0.14points for a girl with a December birthday.

Figures 2 and 3 above show the expected progression over time for the same representative female (from the lower, medium and higher IQ groups) pupil in an average school under both the pre-existing and Enriched regimes. Main effects, such as gender, will move the graphs up or down but will not affect the shape of the child’s progress.

The figures show the pronounced EC dip at Year 2 which is eliminated by Year 4, as also demonstrated in EC interactions in Table 1. The EC variables at Years 1 and 2 show that in both subjects and across IQ groups, the EC is expected to lead to a short-

term drop in the outcome compared to what would have been expected to be gained by that individual being taught the traditional curriculum, particularly at Year 2. However both models show that over time, the incremental effect of the EC increases, such that at the end of Year 4 the outcomes achieved under EC are very similar to those expected under the conventional curriculum.

One final note about the results is the downwards trend at the end of the conventional curriculum (between end Year 5 and end Year 6) and the Enriched Curriculum (between end Year 4 and end Year 5). Both of these estimated progressions were estimated using data from the first cohort of inner-city schools. It is therefore advisable to wait till further data are available from the other schools before drawing a firm conclusion that this final dip for controls is an artefact or a genuine change in expected progression.

Of all the Year 4 and Year 5 end EC variables only one, the reading score for the mid IQ group at end of Year 5, shows a significant negative difference (and this result is subject to the same caveat as the Year 6 data, that the sample is as yet limited for these data points).

7. Discussion of reading and mathematics outcomes

The effects of the Enriched Curriculum on reading and mathematics outcomes are limited to the earliest years. Figures 2 and 3 above can summarise many of the effects described above clearly. The effects of the Enriched Curriculum across ability groups is evident, with the situation in Years 5 and 6 being still limited by incomplete data as described.

Note: It is important to remember that the coefficients on the EC variables capture the estimated effect of the intervention if, and only if, there is a corresponding estimated conventional (pre-existing) curriculum time point, otherwise the estimation is picking up the net effect of the expected progression and the Enriched Curriculum and not the incremental effect of the Enriched Curriculum alone. i.e. as there is an estimate for

$IQHyr1$, then the estimate of $IQHyrIEC$ estimates the incremental difference the Enriched Curriculum has made. On the other hand as there is no $IQHyr1.5$ then the estimate of $IQHyr1.5EC$ picks up the combined effect of the natural progression (including the expected mid-year dip) and the effect of the Enriched Curriculum. Thus estimates for the incremental effect of the Enriched Curriculum are available at: end of Year 1; end of Year 2; end of Year 3; February Year 4 (yr3.5); end of Year 4; and end of Year 5.

For both reading and mathematics models the results indicate the complexity involved in producing a multi-level is worthwhile, in particular there is considerable variation due to unobservable pupil characteristics that is captured by the individual pupil random effects in mathematics and reading, indicated by the non-zero estimates of σ_{ϵ}^2 . (24.78 in mathematics and 30.44 in reading). After allowing for the casemix of pupils, there are still significant school random effects in reading (and borderline significance in mathematics), although the variation indicates that **most variation observed between pupils is due to their own characteristics (both observed and unobserved) rather than differences between schools**. That is to say, school effects are relatively small.

In both disciplines, the patterns across time and influence of explanatory variables are broadly the same: pupils in the lower and higher IQ groups have roughly the same expected different outcomes with lower IQ groups approximately 6 PIPS points lower than the medium IQ group and pupils in the higher IQ group approximately 6 points higher. The coefficients on birth-month are significant and in the direction hypothesised, not large for a single month but showing a considerable difference between the oldest and youngest in the class. The average IQ in a school is significant and has a large positive effect, i.e. all other things being equal, pupils will have an expected higher outcome if the school average IQ is higher. However note that this cluster average is likely to be picking up the net effect of omitted variables correlated with average IQ, such as socioeconomic variables.

After allowing for differences in pupil characteristics and school characteristics the inner city dummy variable is not significant indicating that differences between Inner city schools and mainstream schools are adequately captured by the pupil and school

characteristics (including the class average ability) and random effects. When the cluster (class) mean variable is omitted from the analysis, the percentage of free school meals becomes a significant predictor. However, the former variable explains more of the variability in scores, and is thus preferred.

A 'peer' effect on outcomes?

This socioeconomic/class-mean ability finding above may be interpreted by recognition that the child's total environment; home, school, local community and wider culture, all interact to modify the child's learning experience. Let us take the school environment first. Schools have quite a lot of control over the school environment, but even here, the characteristics of the children in the class as a whole will impact on the individual. This context is most easily pictured with a consideration of language in the classroom; each child is immersed in the local forms of speech, with more formal English being supplied only by the teacher. Hence the great emphasis placed on oral language in the pilot year of the Enriched Curriculum, with schools in a disadvantaged area trying hard (quite successfully) to bring all children up to a minimum standard (Sproule, McGuinness, Trew, Rafferty, Walsh, Sheehy, 2001). Furthermore, teachers may tend to teach to the level that is ideal for the mean ability of the class. As the approach to hit the target with the greatest number of children, this is a rational strategy. In a weak class however, the most able children may find themselves little challenged under this strategy, unless the teacher makes a special effort to provide alternative experiences for this group. Similarly, weak children in a weak class may lack peer models of both good levels of learning skills and good attitudes to school and work, as captured in our structured classroom observation study with the Quality of Learning Instrument (Sproule, Trew et al., *ibid*). Secondly, let us consider the home environment. Schools may succeed in modifying that of some children, albeit with a lot of effort; there will always be parents who are less willing to comply or less confident in doing so. These parents are likely to be disproportionately represented in areas where there is social disadvantage. Finally, the school's effect on the local culture is likely to be very small indeed and there is no effect at all on the wider culture. Considering the whole context of the peer effect then, we have characterised it as such only because it is best measured by the ability of a child's peers in our analysis, but this is done with the understanding that it is mediated by the child's total environment as we have described above.

Creeping implementation of the Enriched Curriculum

For the first time, we see depression of scores for all groups, including controls, in the first four years, especially marked in the middle ability group. Although we have previously believed that there was no creeping implementation of the Enriched Curriculum in the original twelve schools, the situation is likely to have changed for the third wave of schools entering the Enriched Curriculum and represented in this analysis by 10 of the schools new to the evaluation. By the time these schools were thinking about joining the project, there was much more information in the public domain about the Enriched Curriculum. It is entirely conceivable that schools were more knowledgeable about the new curriculum before formally joining the project and had partially implemented it with the year-ahead group but quite properly not going all the way with a play-based curriculum until teachers had had training.

In terms of progression over time for both reading and mathematics, observations taken at half years tend to show negative drops below baseline as was anticipated. Half-year scores have an additional correction applied because PIPS is not being administered at the standard time (in June). This is the likely reason for the discrepancy. In any case, it does not affect the comparison between EC and control groups.

8. Other PIPS data

*PIPS picture vocabulary and non-verbal scores in the original twelve schools
(Formerly known as Shankill and Contrasting Areas schools)*

We do not have, and cannot now obtain¹, data in the new group of 12 schools showing the progress of the same group of children from Year 2 to Year 5. Data in the following section therefore refer only to the original twelve schools in the evaluation. In considering these data, it is appropriate to treat Shankill schools separately because we do not have data on schools intermediate in terms of demographics in this sample, such as percentage of free school meals. Therefore we cannot allow for

¹ We only started testing children in the 12 new schools in 2004 – 2005 and cannot obtain data on their earlier career retrospectively.

socioeconomic factors in the same way as for the main analysis above, in which the 24 schools constitute a good spectrum of school mean ability and free school meals.

For the Contrasting Areas schools (i.e. the second group of schools which came into the project in its second year), there are no significant changes in developed ability measures over the lifetime of the evaluation, either for the EC group or the control group. This finding indicates that neither the Enriched Curriculum nor the pre-existing curriculum is giving rise to changes in developed ability.

Picture vocabulary scores in Shankill schools continue to be less good than associated scores in the non-verbal test. For Year 6 control children only, the picture vocabulary scores are 2.4 PIPS standardised points below the equivalent non-verbal scores. A comparison with Year 6 EC children will not be available until next year. For Year 5 EC children, the gap between picture vocabulary and non-verbal scores has closed to 1.6 PIPS standardised points. However, we must remember that the sample size was down to 59 individuals, so we cannot conclude at this stage that this is the beginning of a trend of improved picture vocabulary performance.

Taken together, all these data suggest that oral vocabulary scores continue to be low in the Shankill schools into Key Stage 2, irrespective of the Enriched Curriculum. This deficit suggests that oral language remains relatively poor in these schools. There is a likelihood that this is adversely affecting reading skills.

References

- Basic Skills Agency (2002). *Survey into Young Children's Skills on Entry to Education*, Wales: Basic Skills Agency.
- Bodrova, E., Leong, D. J & Paynter, D. (1999). Literacy Standards for Preschool Learners. *Educational Leadership*, 57, 2, 42-46.
- Hollis, S., Campbell, F (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials, *British Medical Journal*, 670-674.
- Moyles, J., Adams, S. & Musgrove, A. (2002). SPEEL: *Study of pedagogical effectiveness in early learning*. Published by DfES, UK. Available online from the DfES website.
- National NFER (2005). *A review of Literacy Initiatives to inform advice on the Foundation Stage*. Report prepared privately for the Northern Ireland Council for Curriculum Assessment and Examinations.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., Lewis, T., (2000). *A User's Guide to MLwiN* (Centre for Multilevel Modelling, Institute of Education, University of London,
- Schafer, J.L. & Graham, J.W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods* 7, 147-177.
- Skrondal, A. & Rabe-Hesketh, S. (2004) *Generalized Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC Boca Raton, Florida
- Snow, C.E. (2001) *The Centrality of Language: A longitudinal study of language and literacy development in low-income children*, Institute of Education, University of London.
- Sproule, L., Rafferty, H., Trew, K., Dudgeon, K., Sheehy, N., M^cGuinness, C. & Walsh, G. (2001). *The Early-years Enriched Curriculum Evaluation Project: First Year Report*. Report produced for Northern Ireland Council for the Curriculum Examinations and Assessment (CCEA).
- Sproule, L., Murray, R., Spratt, V., Rafferty, H., Trew, K., Sheehy, N., M^cGuinness, C. & Walsh, G. (2001). *A Review of Recent Research Relevant to the Early Years Enriched Curriculum Project*. Report produced privately for Northern Ireland Council for Curriculum Assessment and Examinations.
- Sproule, L., M^cGuinness, C., Trew, K., Rafferty, H., Walsh, G., Sheehy, N. & O'Neill, B. (2002). *The Early Years Enriched Curriculum Evaluation Project: Year 2 report*. Report produced privately for Northern Ireland Council for Curriculum Assessment and Examinations.
- Sproule, L., M^cGuinness, C., Trew, K., Rafferty, H., Walsh, G., Sheehy, N. & O'Neill, B. (2003). *The Early Years Enriched Curriculum Evaluation Project: Year 3 report*. Report produced privately for Northern Ireland Council for Curriculum Assessment and Examinations.
- Sproule, L., M^cGuinness, C., Trew, K., Rafferty, H., Walsh, G., Sheehy, N. & O'Neill, B. (2004). *The Early Years Enriched Curriculum Evaluation Project: Year 4 report*. Report produced privately for Northern Ireland Council for Curriculum Assessment and Examinations.
- Sproule, L., M^cGuinness, C., Trew, K., Rafferty, H., Walsh, G., Sheehy, N. & O'Neill, B. (2005). *The Early Years Enriched Curriculum Evaluation Project: Final Report Phase 1*. Report produced privately for Northern Ireland Council for Curriculum Assessment and Examinations.
- Sproule, L., M^cGuinness, C., Trew, K., Rafferty, H., Walsh, G., Sheehy, N. & O'Neill, B. (2006). *The Early Years Enriched Curriculum Evaluation Project: Year 5 report*. Report produced privately for Northern Ireland Council for Curriculum Assessment and Examinations.

Tymms, P., Heron, T., Merrell, C., Jones, P., Albone, S. & Henderson, B. (2005). *The Performance of Newcastle Schools*. Private report by the CEM centre for the Newcastle commission.

White, K. R. (1982). The relationship between socioeconomic status and attainment. *Psychological Bulletin*, 91, 461 – 481.