

**The Early Years Enriched Curriculum
Evaluation Project (EYECEP)
End-of-Phase 2, Report 4**

**The Enriched Curriculum:
Outcomes for Pupils over Time**

September 2009

Sproule, L.¹, McGuinness, C.¹ Trew, K.¹, Rafferty, H.¹ & Walsh, G.²

¹School of Psychology, Queen's University Belfast

²Stranmillis University College

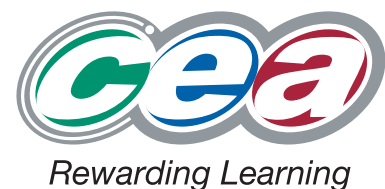
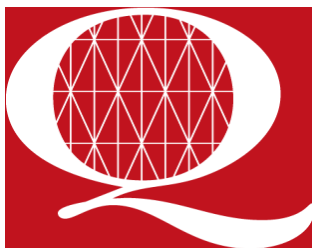


Table of Contents

	Page
Membership of the EYECEP Research Team 2000-2009	3
Acknowledgements	4
Statement on Ethics	4
1 Introduction	5
2 Design and Methodology	10
3 Outcomes for Pupils over Time: Mathematics, Reading and Science	19
4 Outcomes for Pupils over Time: Writing	36
5 Outcomes for Pupils over Time: Learning Dispositions and Attitudes	43
6 Overview of Findings and Conclusions	66
References	73
Appendix: Figure 3 reproduced from End-of-Phase 2, Report 1	75
Technical Appendices	Submitted separately

Membership of the EYCEP research team 2000-2009

Professor Noel Sheehy	Lead Investigator 2000-2001, Research Team 2000-2004
Dr Harry Rafferty	Lead Investigator 2001-2004, Research Team 2000-2009
Dr Karen Trew	Lead Investigator 2004-2006, Research Team 2000-2009
Professor Carol McGuinness	Lead Investigator 2005-2009, Research Team 2000-2009
Ms Bridgeen O'Neill	Research Team 2001-2005
Dr Glenda Walsh	Research Team 2000-2009
Dr Katrin Dudgeon	Research Assistant 2000-2001
Dr Liz Sproule	Senior Research Fellow and Project Co-ordinator 2001-2009
Dr Diane Harvey-Smith	Research Assistant 2005-2008
Ms Jacqui Lockhart	Project Administrator 2003-2008

Statistical Consultant and Analyst

Dr Chris Bojke, Pharmerit International (formerly NfER) 2004-2009

Additional help from

Mr Geoffrey Caves	Online survey
Ms Michelle Drain	Scoring protocols
Dr Alissa Lange	Scoring protocols and data entry
Ms Grainne McKenna	Video analysis
Dr Gary McKeown	Statistics, graphics
Dr Victoria Simms	Scoring protocols and data entry
Mr Andrew Wilkinson	Data entry
Team of Fieldworkers:	A group of 40 professionals, mostly retired teachers, who were trained to administer our psychometric tests and questionnaires in the schools. Most of them remained active in the project over several years and greatly facilitated the management of data collection across Northern Ireland.

The views expressed are those of the research team and not necessarily those of the Northern Ireland Council for the Curriculum Examinations and Assessment

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 9 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 were involved in the Enriched Curriculum;

Dr Chris Bojke, statistical analyst now at Pharmerit International, who provided statistical advice and support on the multi-level modelling aspects of the statistical analyses;

the Northern Ireland Council for the Curriculum Examinations and Assessment for funding the project.

Statement on Ethics

The research was conducted following the British Psychological Society's Code of Ethics and each phase of the research was granted ethical approval by the Queen's University School of Psychology Ethics Committee.

1. Introduction

1.1 The Enriched Curriculum and the Evaluation: Brief Background¹

This is the last of four End-of-Phase 2 reports on the evaluation of the Enriched Curriculum (EC) in Northern Ireland schools.

The Enriched Curriculum began as a local initiative in the Belfast Education and Library Board. In September 2000, it was introduced by six primary schools in a disadvantaged area of Belfast. One of the purposes behind the pilot project was to ease the transition of 4-5 year old children from pre-schools, playgroups, or home — to statutory schooling. This was particularly important in Northern Ireland, which has the youngest statutory starting school age in Europe. Moreover, for the Belfast schools that initiated the project, the transition issue was acute, given that their school entry children tended to have poor oral language skills and were not always well prepared for school routines. The work was influenced by the experiences of principals, teachers and Curriculum Advisory Support Service (CASS) officers in this disadvantaged area of Belfast, who reported that the traditional curriculum was not meeting the needs of the children and some schools were already exploring alternative approaches. In addition, an evaluation of a pre-school project in the area, the Greater Shankill Early Years Project (Sheehy, Trew, Rafferty, McShane, Quiry and Curran, 2000) had drawn attention to the difficulties faced by children as they progressed through the previous first year primary curriculum.

Thus, by being responsive to the developmental stage of individual children, the aim of the Belfast schools was to remove the early experience of persistent failure and to promote children's sense of self-competence and self-esteem. The teaching methods included a greater emphasis on play and activity-based learning rather than desk-work, in order to stimulate children's curiosity, creativity, social development and engagement with learning. In literacy, the methods involved more emphasis on developing oral language skills and on emergent literacy (phonological awareness), and less on formal methods such as reading schemes. In mathematics, it involved laying foundations in number through sorting, matching and counting rather than formal number recording. There was a new emphasis on the importance of outdoor play and activities. A more complete description of the characteristics of the curriculum and how it was organised can be found in Report 1 (Section 4.1, The Enriched Curriculum as originally conceived) and Report 2 (Section 2.3.1, Broad Overview of the Day)

¹ A full description of the Enriched Curriculum and the rationale behind the evaluation strategy can be found in End-of-Phase 2 Report 1 Overview: Evaluation Strategy and Curriculum Implementation (McGuinness et al., 2009).

After the first year, when information about the EC project in Belfast was disseminated, it provoked an enthusiastic response in other Education and Library Boards (ELBs) and from some individual schools, where many were already thinking along similar lines. Each ELB decided to introduce the project into a group of their own schools in September 2001, and additional schools joined in September 2002. Eventually, over 120 schools across Northern Ireland participated.

The evaluation of the project, which was initially funded for just one year, was extended for a further three years, and then for another four years. It was recognised that the impact of the EC could only be well evaluated in the longer term, hence the commitment to a longitudinal design. In the first phase of the evaluation, only 12 schools participated in the evaluation. Following an external review at the End-of-Phase 1 (NFER, 2004), an additional 12 schools, and a second cohort of EC children in each school, joined the evaluation, and substantially increased the sample size.

Thus, the evaluation was conducted over eight school years from 2000-01 to 2007-08 and was carried out in 24 EC schools. Three cohorts of children within each school were studied: the class of children who followed the curriculum that existed immediately before the introduction of the EC (called the control children), the class of children who followed the EC in the first year of its introduction (called EC1 children) and the class of children who followed the EC in the second year of its introduction (called EC2 children). As the numbers of schools and children participating in the EC increased, the evaluation strategy expanded to meet the complexities of the project.

1.2 Purpose of this Strand of the Evaluation

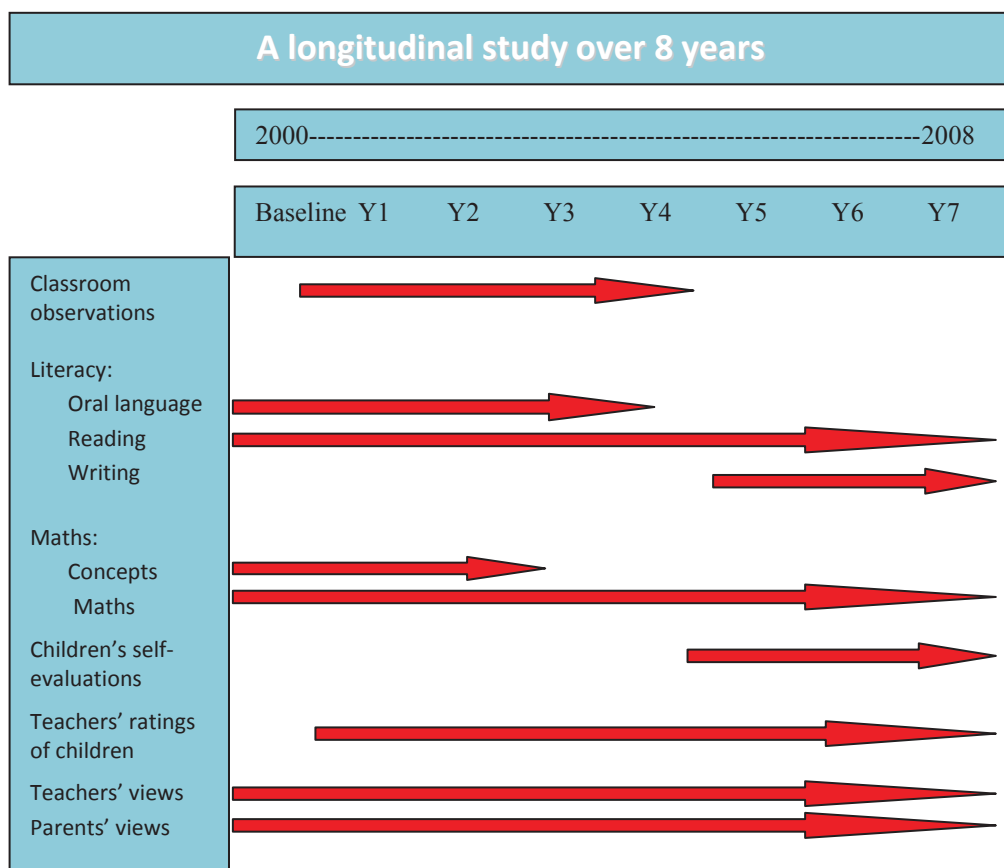
The evaluation strategy consisted of four strands and is fully described in End-of-Phase 2 Report 1. The findings from the other three strands are reported in End-of-Phase 2 Report 2 (the classroom observations, the teachers' and school principals' perspectives) and in End-of-Phase 2 Report 3 (the parents' survey). The aim of this strand of the evaluation was to assess the short- and longer-term impact of the EC on children's learning orientations and attitudes, and on their progress in literacy and numeracy.

The general purpose of the EC was to introduce a more informal approach in the early years of primary school — one that was more appropriate for young children — thus easing the transition for children from pre-school into statutory schooling. Findings from the other strands of the evaluation confirmed that the shift to a more informal approach in Years 1 and 2 did occur. For example, the classroom observations showed distinctively different patterns in the allocation of time

between play-based and desk-based activities compared with the previous curriculum; there was a more even balance between adult-initiated and child-initiated activities, and children experienced higher levels of social-emotional, motivational and cognitive experiences than they would have had in the previous curriculum. Despite some reservations, teachers, school principals and parents all valued the more informal and practical orientation of the new curriculum, particularly for Year 1.

The purpose of the longitudinal design then was to monitor the shorter and longer term effects of the curriculum shift and to evaluate the impact on a range of educational outcomes for the children over time. Standardised psychometric tests and rating scales were used to evaluate the outcomes. The testing programme incorporated different types of tests for children of different ages, as these became appropriate (see Figure 4.1).

Figure 4.1 Overview of the longitudinal design (red arrows refer to data analysed for this report)



The current report deals mainly with outcomes for children not covered in the End-of-Phase 1 Report (Sproule et al., 2005). The report will follow the children through Key Stage 2, whereas the previous End-of Phase 1 Report (and previous annual reports), reported outcomes for children up to the end of Key Stage 1. These included reading and mathematic outcomes using standardised tests

(PIPS²) as well as age appropriate tests of oral language and basic concepts in mathematics (See Figure 4.1). The current report will continue the story for reading and mathematics outcomes, but will introduce a new literacy measure appropriate for Key Stage 2 — writing composition. Furthermore, for the first time, the children were asked to rate themselves on cognitive and motivational indicators of learning and teachers' ratings of the social and behavioural competence of the children in the classroom are also reported. These ratings quantitatively evaluate the impact of the EC on children's self-perceptions and attitudes to learning, as well as the teachers' perceptions of the children in the classroom. They complement the findings from the suite of standardised tests that assesses scholastic attainments in literacy and mathematics. Also, because of the addition of new schools and a second cohort of EC children in all schools, we will be able to see if the previous patterns of findings, reported at the End-of-Phase 1, are replicated in this bigger sample of schools and children.

1.3 Issues related to the interpretation of the pupil outcomes over time

The first important issue relates to the aims of the EC. It is important to remember that the EC had broad aims with regard to its impact on children's educational development and was not primarily a reading and mathematics intervention. Nevertheless, adopting a more informal pedagogical approach in Years 1 and 2 did change how reading and mathematics were taught in the first two years and beyond. For example, interviews with the school principals (End-of-Phase 2 Report 2) confirmed that teaching reading, in particular, was subject to several revisions in the first years of implementation. In addition, interviews with the teachers (End-of-Phase 2 Report 2 and previous annual reports) showed that many Year 3 teachers did have concerns about the children's reading competence when they arrived in their classrooms, although children progressed very well when formal approaches to reading were adopted. Consequently, it is important to track the longer term literacy and mathematics development of children who experienced the EC. Literacy outcomes in particular have become increasingly salient for schools since the report for the House of Commons Public Accounts Committee on Reading Standards in 2006.

Notwithstanding the importance of literacy and numeracy as gateways to learning across the curriculum, the general thrust of curriculum reform both in Northern Ireland and in other parts of the UK (and elsewhere) has been to focus on the development of the whole child — on the social, emotional and motivated child — as well as the literate and numerate child (e.g., Northern Ireland

² See Section 2.3

Revised Curriculum³, Scottish Curriculum for Excellence⁴, The Independent Review of the Primary Curriculum in England, Rose, 2009). Recent research has pointed to the reciprocal relationships between cognitive, social and contextual factors in children's education that suggest that thoughtful education must address all three domains in seeking the maximum progress for children in school (EPPE 3-11, Technical Paper 11, 2004 — with specific reference to early years education).

The second important issue relates to the factors that influenced the implementation of the EC and that might have affected the outcomes for children. These factors are dealt with more fully in the End-of-Phase 2 Reports 1 and 2. For example, we draw attention to

- the different conditions under which the EC was implemented and resourced in the high deprivation Belfast schools compared to the schools in the other ELBs (the Contrasting Area Schools);
- the perceptions and responses of teachers and school principals to the training and support provided;
- whether schools implemented the EC as a whole-school initiative or primarily as an early years approach;
- the professional knowledge and ideas about a play-based curriculum and developmentally appropriate practice that the teachers had; and
- whether substantial changes and revisions were made to the implementation of the EC as it became more embedded in the schools.

In one sense then, the EC was not one curriculum but many. Yet this is true of any curriculum that is flexible, that is not highly prescribed and scripted, and is open to professional interpretation by teachers and schools. And, in the absence of a written curriculum document at the beginning of the project, there was probably more room for interpretation within the EC than for many other flexible curricula. Yet, there is evidence from the Key Stage 1 patterns of attainment, as well as from the teachers' reports and classroom observations, that almost all teachers moved their practice in a developmentally appropriate direction. So, despite the potential differences between groups of schools, individual schools and teachers, we will treat the EC as a 'unitary' curriculum and analyse its impact on the total sample of children in the 24 schools — at least in the first instance. Where differences emerged between subgroups of schools and children, we will report these and, as far as we can, we will relate them to differences in the conditions of implementation. To anticipate the results, the most important differences seemed to occur between the high deprivation schools and

³ Information available on <http://www.nicurriculum.org.uk/>

⁴ Information available on <http://www.ltscotland.org.uk/curriculumforexcellence/>

the remaining group of schools which joined in the 2nd and 3rd years of the project — the Contrasting Areas (CA) schools.

1.4 Structure of the Report

This report is divided into four sections — in addition to this introduction.

- Section 2 details the design and methodology of the longitudinal evaluation, including choice of instruments, sampling issues and outline of statistical analysis techniques.
- Section 3 will report on the pupil outcomes over time for reading and mathematics and for science in Year 7 only.
- Section 4 will report on pupil outcomes over time for writing.
- Section 5 will report on pupil outcomes over time for children's self-perceptions, dispositions and attitudes to learning, as well of the teachers' ratings of children's social and behavioural competence in the classroom.
- Section 6 will summarise the overall results, place them in the context of the factors that might have influenced the implementation of the EC, and draw some conclusions.

The reader is reminded that the evaluation team have produced annual reports, supplementary reports, as well as the End-of-Phase 1 (end of the Fourth Year) in March 2005. All these reports can be found at <http://www.nicurriculum.org.uk/>. There will be some overlap in the data and commentary between these previous reports and this report.

2. Design and Methodology

2.1 Longitudinal Design

With regard to the pupil outcomes which are the focus of this report, the general design of the evaluation was a longitudinal quasi-experimental⁵ design. Children in classes one year ahead in the same school, who had experienced the pre-existing more formal curriculum, acted as controls. Two cohorts of children who experienced the EC curriculum in the first and second years of implementation in each school were studied, allowing the stability or change in EC practices and

⁵ When conducting an evaluation of the effects of any intervention, it is crucial that the children in the intervention and control groups are similar, so that the effects of the intervention can be adequately evaluated. The ideal method is to assign children/classes/schools randomly to either the control or intervention groups. A **quasi-experimental** research design is one in which the comparison groups are formed by some method other than random assignment. Alternative strategies are then adopted to ensure that the two groups are as similar as possible. In the EC evaluation, the similarity of the comparison groups was ensured by using the year-ahead class in the same school as the control group.

outcomes over time to be evaluated. Use of year-ahead control groups minimises the influence of potential confounding effects, such as differences in school intake, differences in ethos between schools, and so on. The quasi-experimental longitudinal design allows a range of questions about the impact of the curriculum on the outcomes for children to be addressed, and complex statistical modelling can partial out the effects of different variables. As the Enriched Curriculum was implemented more widely and became more important, so the evaluation expanded alongside it. The evolution of the longitudinal design is described in detail in the End-of-Phase 2, Report 1: Overview (McGuinness et al., 2009), and Figure 3 from that report is reproduced in Appendix A.

2.2 Sample of Schools and Cohorts of Children

There were 24 schools in the evaluation in three school groups; the Belfast high deprivation group (6 schools), Contrasting Area 1 group (CA1 group, 6 schools) and Contrasting Area 2 group (CA2, 12 schools). These groupings reflect the time at which the schools joined the evaluation. The latter two groups of schools were divided between the Western (N=2), North Eastern (N=7), Southern (N=5) and South Eastern (N=4) Education and Library Boards. This breakdown is a reasonably good reflection of the relative population distributions of the Education and Library Boards. The schools volunteered to participate in the evaluation, and thus were not randomly selected. However, a balance between school types in terms of locality was sought as detailed in Table 4.2. Although the sample was not randomly selected, it is representative of the population of schools in Northern Ireland, in terms of geographical distribution, urban/town/rural, and social advantage/disadvantage as indicated by percentage of free school meals (FSM). In addition, a balance was sought between integrated (N=4), controlled (N=13) and maintained (N=7) primary schools⁶. Demographic statistics clearly distinguish the high deprivation group in Belfast from the rest in terms of FSM and intake ability characteristics.

There were 3414 children who contributed to the evaluation at some point but for reasons of cost and the rolling nature of the evaluation, not every one of these children took part in every aspect of the evaluation. There were three successive cohorts of children in each school; Control, EC1 and EC2. Actual sample sizes for the main measure of attainment, PIPS, in each year group are given in Table 4.2. For other measures, sample sizes were of the same order and will be given in detail as those measures are reported. It should be noted that the most complete samples were available for Years 3-6. For the CA2 group of schools, few outcomes data were available for the early years; children had already progressed further by the time these schools joined the evaluation. In addition,

⁶ Controlled schools are government controlled and admit mainly Protestant and non-affiliated children, maintained schools are controlled by the Council for Catholic Maintained Schools and Integrated have specific quotas for children from the different Northern Ireland communities.

there were relatively fewer Y7 children compared to Y6 because EC2 children in CA2 schools had not reached Y7 when the evaluation ended. Although the statistical models try to correct for any deficiencies in sample sizes by estimation, in the context of this study, Y6 outcomes should always be considered as more reliable than Y7 as the exit point from primary school, especially in relation to the EC2 cohort performance.

Table 4.1 Characteristics of the school groups in the Enriched Curriculum evaluation

Group: Cohort	Ability profile⁷ <i>Mean (SD)</i>	Range of % of FSM in school group	Year schools started EC	Type of locality of schools
6 Belfast high deprivation schools:				
Control cohort	45.5 (6.4)	50 - 76	All in 2000	6 city centre
EC first cohort	44.8 (5.2)			
EC second cohort	45.6 (6.4)			
Joined the <i>evaluation</i> in 2000				
6 CA1 schools outside Belfast:				
Control cohort		2 - 27	All in 2001	1 city centre
EC first cohort	52.6 (7.4)			3 city or suburban
EC second cohort	51.8 (7.1)			1 medium or large town
Joined the <i>evaluation</i> in 2001	51.1 (6.6)			1 rural or small town
12 CA2 schools outside Belfast:				
Control cohort	50.5 (7.0)	0 – 30	2001 (2) 2002 (10)	1 city or suburban
EC first cohort	51.5 (7.0)			4 medium or large town
EC second cohort	50.7 (6.8)			7 rural or small town
Joined the <i>evaluation</i> in 2004				

Despite these caveats, it is clear from Table 4.2 that the addition of the 12 schools and the second cohort of children in each school (EC2) for Phase 2 has considerably increased the sample size and will allow for robust statistical conclusions.

⁷ Taken from the PIPS test and explained in Section 3.1. The population mean is 50, SD is 10.

Table 4.2 Sample sizes by year group for PIPS literacy and numeracy testing⁸

PIPS test	N (EC)	N (Controls)	N (Total)
Baseline	314	278	592
End Year 1	177	188	365
End Year 2	447	156	603
End Year 3	732	307	1039
End Year 4*	738	422	1160
End Year 5	778	287	1065
End Year 6	635	369	1004
Mid Year 7**	340	364	705

* End of KS1 in Northern Ireland

** There are relatively fewer Y7 children compared with Y6 because EC2 children in CA2 schools had not reached Y7 when the project finished.

Maintaining the size of the sample in a longitudinal design: At this stage, we would like to consider the sampling issues that are encountered when running a longitudinal study — the problems of **drop-out and missing data**. These issues are inevitable in any longitudinal sample and occur for many understandable reasons — children move away from the area, they change school, they are absent at the period of testing, they are withdrawn from the study either deliberately or inadvertently through not complying with explicit consent forms. In addition, new samples of children joined the study in Phase 2 at Years 3 and 4, and, as we have explained above, the study ended before some children reached Year 7 (See the appendix for a helpful illustration). Consequently, the study was not a fully 7-year longitudinal sample for all children. Very few children generated data at each year of testing; there are 3-4 data points on average per child, with a range of between one and seven data points. Provided children’s absent data points are not systematically related to the type of curriculum in any way, that is, not related to any factor that might influence outcomes, use of an appropriate multilevel statistical model negates any need to impute missing data. Such a model therefore allowed us to make use of every piece of test data and to make reliable estimates for co-efficients and thus maintain the benefits of the sample size over time. Nevertheless, the *reliability* of any estimate does depend on the sample size at any point in time and we will make reference to that several times throughout this report.

2.3 Choice of Instruments for Key Stage 2 Assessment

Choosing measures to evaluate the pupil outcomes in any educational intervention over time presents particular challenges. First, it is important that measures are sensitive to the learning goals

⁸ Sample sizes for other tests were different. Exact numbers will be reported in the appropriate section.

of the intervention, and second, it is important that measures have well established reliability and validity. These two demands can be at odds with one another, especially if an intervention is innovative and trying to enhance types of learning for which there are limited available measures. In addition, in a longitudinal design, children are growing older, and there is a need for age appropriate instruments that measure the same construct (e.g., literacy) but different manifestations of it at different age groups (e.g., decoding skills, vocabulary, comprehension, writing).

For the EC evaluation, both in the earlier years and for Key Stage 2, measures were selected against a number of criteria: (i) the sensitivity of the measure to the learning goals of the EC; (ii) the psychometric properties of the instruments with regard to validity and reliability; (iii) the appropriateness of the measure to the developmental stage of the children over the course of the evaluation; (iv) the practicalities of testing children, both individually or in groups; (v) the burden on teachers; (vi) the cost of the measures in terms of value for money; (vii) specialist knowledge required to administer them (viii) fieldworkers' time, etc. Inevitably, compromises and trade-offs were made between these criteria in selecting a suite of instruments for any phase of the evaluation.

A variety of tests and instruments were selected for Key Stage 2 outcomes against these criteria.

- At the core is a robust measure of basic attainment, the Performance Indicators in Primary Schools (PIPS, 2001).
- This is amplified by a well-established and age-appropriate measure assessing written composition, from the Wechsler Objective Language Dimensions (WOLD, 1996).
- In addition, a new suite of self-rating instruments was introduced, called the Assessment of Learner-Centred Practices, ALCPs (McCombs, 1997) which are derived from the American Psychological Association's Learner-Centered Principles. They ask the children to assess themselves on learning and motivational indicators that are linked to be the longer term outcomes for the EC.
- A short teacher rating scale, evaluating children's social and behavioural competence in the classroom, was also used.

These latter three instruments are more fully described in the results sections below.

The PIPS suite of age-appropriate tests was chosen as a cost-effective measure of known high reliability and validity, which had been carefully designed to meet the criterion of sampling a wide

range of domains pertinent to the English National Curriculum. The PIPS database⁹ was first established in 1993 and now contains data on hundreds of thousands of children in the United Kingdom within more than 4000 schools. One of the main benefits of the PIPS suite is that it provides standardised, and comparable, end-of-year assessments across the primary school years, from baseline entry to final year in primary school. Furthermore, the tests were not widely used in Northern Ireland at that time¹⁰.

Despite the psychometric sophistication of PIPS, it should be pointed out that the measures were not well attuned to the learning goals of the EC, especially in Years 1 and 2, where an informal approach to teaching reading and mathematics was adopted. Nevertheless, the baseline PIPS measure assesses aspects of literacy and numeracy that predict later achievement. For example, the correlation of PIPS baseline with children's performance at the end of their third year at school in England is 0.65 for mathematics and 0.70 for reading (Tymms, Merrell and Henderson, 1998). More recent work shows that it is also predictive of outcomes at the end of primary school (Y6 in England), predicting up to 41% of the variance at the individual level and 53% at the school level (Tymms, Merrell, Henderson, Albone and Jones, 2007). So, even if the tests are not sufficiently sensitive to the EC curriculum in the early years, they are likely to pick up on important aspects of children's achievement in numeracy and literacy in the later years.

2.4 Analytical Method

For most analyses, the method used was multi-level modelling. Like all regression analyses, multi-level regression models allow us to estimate the effect of the main variable of interest — participating in the EC versus the pre-existing curriculum — while controlling for other effects that might impact on outcomes such as gender, age-in-class, individual difference in ability, and levels of social deprivation. In addition, multi-level modelling permits us to take into account the effects of clustering in the data — the fact that the pupils are grouped into classes in particular schools, and that schools are grouped into areas (inner city versus contrasting areas). The longitudinal design also allows us to examine the children's progress, not just at one point in time, but also the growth curves for pupils who are participating in the EC compared to the children being taught under the conventional curriculum. A full technical description of the difference-in-difference approach which was used can be found in the Technical Appendices to this Report.

⁹ The PIPS database is managed by the Curriculum, Evaluation and Management (CEM) centre for the University of Durham.

¹⁰ This has changed dramatically since InCAS testing was introduced as part of diagnostic testing in Northern Ireland schools. PIPS is a paper and pencil test and is a predecessor of InCAS.

We would just offer a word of caution about modelling. Modelling data is complex and requires various assumptions to be made about the nature of the variables, their likely effects and how they change over time. The same data can be modelled in various different ways and be plausibly interpreted. We are often looking for the best fit between the model and the data, or even a good fit. Also, because of the longitudinal nature of the sample, and the use of methods to estimate missing data points, we will often refer to the reliability of the estimates about a specific sub-set of the data at a given point in time, where the original sample was small.

We had several research questions to answer using multi-level modelling methods. The major research questions asked were:

- **What are the differences between the EC and the previous curriculum on pupils' outcomes over time?**
- **Does the EC have a similar impact on pupils' outcomes over time for high, middle and low ability children?**
- **Are the differences the same for the first cohort compared to the second cohort (EC1 versus EC2) — how does the curriculum bed down within a school?**
- **Given the differences in pupil intake, implementation and resources associated with different school groups in the evaluation, does the EC have similar effects over time for high deprivation schools, compared the 'contrasting area' group of schools?**

These are the main research questions to be investigated in the statistical analyses. But there are other variables that we know predict school progress and orientations to learning in school.

- There is continuing concern about gender differences in attainment, and boys' underachievement, so the effect of gender on the pupils' outcomes was also assessed.
- Children's age in class can vary by up to 12 months and there is now considerable evidence that month-of-birth can impact on children's progress in school, not just in the early stages but right into performance at secondary school (NFER, 2009).
- We know that children differ substantially in terms of the academic, social and personal resources that they bring to school. This can be conceived and measured in terms of
 - the individual pupil's 'developed' ability — the language and reasoning abilities that can reflect innate capacity, home learning as well as school learning and which impact on school progress;

- the average levels of ‘developed ability’ in a class — this can affect the levels of language and discourse patterns that are available for promoting learning in the classroom — the effects of the peer group on learning in the classroom; and
- economic social disadvantage, as indicated by the percentage of children receiving free school meals in a class or in a school.

The effects of these additional variables on children’s outcomes were assessed in all multi-level models, and are referred to as the *main effects* in the model. These effects are then statistically controlled for when estimating the effects of the EC versus the pre-existing curriculum, thus giving us a better understanding of the ‘true’ impact of the curriculum, all other things being equal.

Predictor variables used in our analyses fall into two categories, variables operating at the level of the individual and variables operating at the level of the school. They are summarised and explained in Table 4.3.

2.5 Reporting Conventions

There are a large number of statistical analyses described in this report. It should be remembered that statistically significant effects are based on probabilities, and that there is a degree of chance associated with each finding. We have adopted the usual convention of reporting only statistically significant effects at $p < .05$. This means that, on 5/100 occasions, a reported significant finding is likely to be due to chance rather than to the effects of the intervention. A second important point to note is that statistically significant effects are more likely to be reported in larger than in smaller samples; what looks like a trend in a smaller sample, can reach statistical significance in a larger sample, or when sub-samples are combined.

Because statistical significance can be dependent on sample size, it has become the norm in scientific research to get an estimate of the ‘size’ of a statistical effect, called an **Effect Size**. An effect size is simply a way of quantifying the size of the difference between two groups. In the case of an intervention versus a control group, the effect size can be thought of as the change that is brought about by an intervention. It is expressed in units of standard deviations (SDs) and thus allows for comparison across different scales and different samples. Effects sizes in the range of 0.2 to 0.3 SDs are normally considered as ‘small’; around 0.5 SDs is normally considered ‘medium’; 0.8 SDs or above is considered ‘large’ (Cohen, 1969). Effects sizes for educational interventions tend to be in the small-medium range.

Table 4.3 Summary of predictor variables used in the statistical analyses

Individual level variables

1. Gender – boy or girl.
2. Month of birth – distinguishes between children of different ages in a class.
3. Individual mean developed ability – as measured by the PIPS suite of tests averaged over several testing points. The measure assesses verbal and non-verbal components. (See Section 3.1)
4. Individual developed ability grouped into categories – low, middle or high ability (low ability means more than half a standard deviation below the mean [approximately 25% of the sample]; high ability means more than half a standard deviation above the mean [approximately 25% of the sample]; about 50% of the sample falls into the middle ability group.
5. Intervention or control –Enriched Curriculum or pre-existing curriculum (control).
6. Cohort group – control, EC1 or EC2 cohort (pre-existing curriculum, first cohort to follow the EC, second cohort follow the EC; all cohorts are in the same school)

School level variables – these are all indirect measures of deprivation

1. Percentage of free school meals (FSM) in the school. These data were gathered from principals, and where possible, checked against official statistics, for example, school inspection reports and FSM individually reported by teachers.
2. School mean developed ability – the mean of all PIPS individual ability scores for each school. Shows the profile of developed ability within the class in a school; likely to effect the peer interactions and classroom talk. It is highly correlated with the FSM variable (Pearson's $r = -.79, p < .001$).
3. Inner city – a variable which distinguishes the high deprivation group of schools from the rest. It is included to detect whether there is any additional effect attached to being in this group of schools *after* having controlled for the deprivation in the sample as a whole by one of the above school variables.

For accessibility and brevity, the findings will be reported in the following way:

- The focus for reporting and interpretation will be Key Stage 2, Years 5, 6, 7. The findings for Years 1-4 were reported in the End-of-Phase 1 Report (Sproule et al., 2005), and they will be commented on only when they differ markedly from previous findings.
- The estimated means from each of the multi-level models are plotted on graphs to show comparisons between the EC versus control groups over time. 'Stars' above or below graph points indicate that differences are significant. The colour of the star corresponds to the colour of the subgroup with a score that is significantly different from that of controls.
- When reporting statistically significant findings at $p < .05$ or less, the effect sizes will also be reported in the text.
- All other statistical tables are included in the Technical Appendices to Report 4.

3. Outcomes for Pupils over Time: Mathematics, Reading, and Science

3.1 The Performance Indicators in Primary Schools (PIPS) suite of tests

End-of-Year PIPS tests consist of measures of reading, mathematics, verbal reasoning (picture vocabulary) and non-verbal reasoning (e.g., identifying dots in patterns, fitting jigsaw pieces). Data on these measures were collected for each group from Y1 to Y7. The PIPS science test is available for Y7 only. Baseline and Y1 PIPS were individually administered. From Y2 to Y7, the tests were administered as group tests, with the group size being determined by the children's age. All testing was conducted by trained staff, who were either graduate psychologists or qualified teachers and who were extensively briefed on the administration of the tests by the research staff.

The reading tests in the earlier years consisted of asking the pupils to read simple text and to complete a sentence choosing an appropriate word from three options (a cloze procedure which can test knowledge of grammar and punctuation as well as word meaning and decoding), plus simple multiple choice comprehension passages. From Y5 onwards, there was a substantial shift in the level of difficulty demanded in the comprehension passages, where pupils had to read and extract information from various types of text (narrative, tables, adverts, flyers) and answer both literal and inferential types of questions in multiple choice formats. Time limits were also more stringent in KS2.

Similarly, the mathematics test shifted from computation and simple word problems to a wide range of topics such as estimations, graph and table work and more complex word problems. The support for children with reading problems, available for those who needed it in the early years of the mathematics test, was no longer allowed in mathematics at KS2. Poor, or even just slow, reading becomes a disadvantage as the test was also taken under stricter time limit conditions than before.

The Y7 Science test consisted of 47 questions on a wide range of topics such as biological cycles, electrical circuits, the solar system and metals versus non-metals. Most of it is fact-based but there is a question about a fair test and some inferences are required in a few questions.

In addition to measures of reading, mathematics and science, PIPS end-of-year assessments (after Year 1) provided a measure of verbal and non-verbal reasoning which, when combined, can be used as a proxy score for IQ or a measure of 'developed' ability, to use the PIPS terminology. This measure was used to group the sample into low, middle and high ability groups; low ability means more than half a standard deviation below the mean [approximately 25% of the sample]; high ability

means more than half a standard deviation above the mean [approximately 25% of the sample]; about 50% of the sample fall into the middle ability group.

All PIPS scores are standardised on a population mean of 50 and a standard deviation of 10, unlike many other standardised tests where the equivalent figures are 100 and 15 respectively. This means that approximately two-thirds of all children will score between 40 and 60 PIPS standardised points in a normal population. All PIPS analyses used standardised scores.

For each outcome, the data will be reported in the following way:

- The effects of the EC versus the control for the total sample of schools;
- The effects of the EC versus the control for the high deprivation sample of schools;
- The effects of the EC versus the control for the contrasting areas schools;
- Where sample sizes allow for reliable estimates, the data is reported for high, middle and low ability groups;
- Where there are consistent differences between EC1 versus EC2 cohorts, these are reported separately; if the differences between the cohorts are negligible or inconsistent across time, the data from the two EC cohorts are combined into a single EC group.
- Finally, an overview of the main effects of the predictor variables which are true for both the EC and the control classes are presented.

3.2 Mathematics: EC Pupils versus Control Pupils: Total Sample, 24 Schools

This initial multi-level model evaluates the outcomes for mathematics for the total sample. It combines the findings from all 24 schools, irrespective of any differences between them in terms of training, resources and the implementation of the EC. It benefits from the large sample size (See Section 2.2) and thus increases the reliability of the estimates from the multi-level modelling. Figure 4.2 shows the graphical representation of the regression corrected mean scores for nine groups of children (3 cohorts x 3 ability levels) over 7 years. It compares the effects of the EC1 (blue line) and EC2 (red line) outcomes relative to the control group (broken grey line) of year-ahead classes, for high, middle and low ability groups of children.

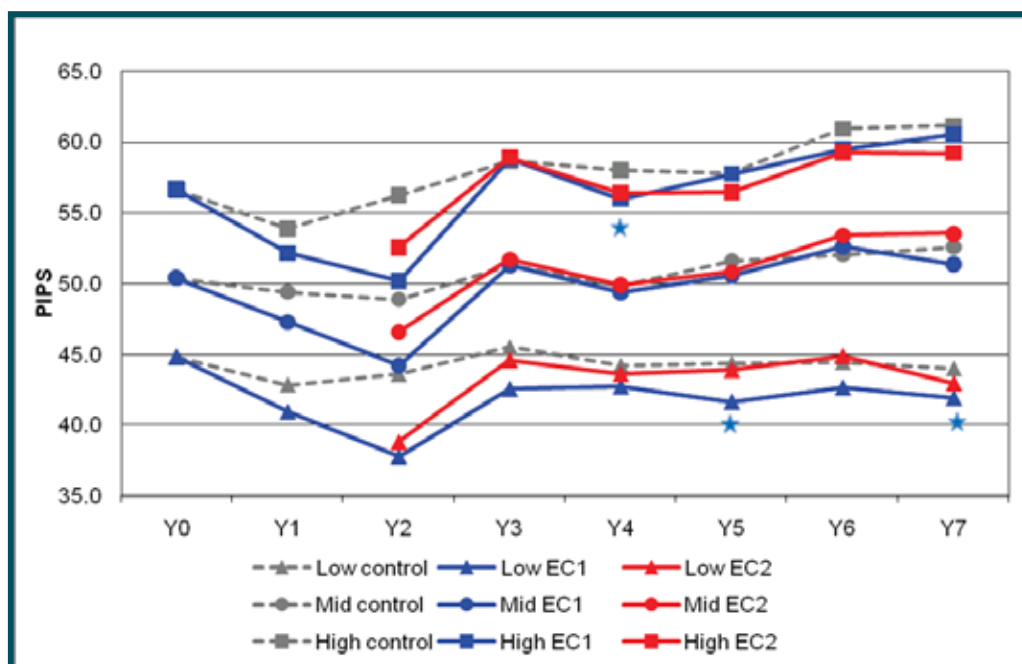
General trends: In terms of the general trends in the children's performance over time, note that the differences between the ability groups (irrespective of cohort) increased over the 7 years of primary school. There was a greater gap between the high and lower ability children when they left school compared to when they entered school. Also, the trajectory over time for high and middle

ability children was in an upwards direction, while for lower ability children the trajectory was slightly downward. This shows that middle and higher ability children were likely to achieve beyond their baseline expectations in mathematics, while the performance of lower ability children was slightly below baseline expectations.

Years 1-4 EC versus Control: This is an augmented dataset in Years 2, 3 and 4 compared with what was reported in the End-of-Phase 1 Report. Nevertheless, the first point to notice is that the pattern up to Y4 is similar to that described and discussed in previous reports. EC children did not begin formal recorded mathematics as early as in the pre-existing curriculum; when they did move on to formal work in Y2, their rapid progress was apparent in the results for the end of Y3. By Y4, there were no statistically significant differences between 5 of the 6 EC groups and the control group; for one group, the EC1 high ability group, the control group continued to perform marginally better; the effect size was 0.20SDs.

Years 5-7 EC versus Control: As the children progressed into Key Stage 2, the first point to notice is that the EC2 cohort (red line) performed better than the EC1 cohort (blue line), especially the lower ability group of children. Of the eighteen possible comparisons between subgroups of EC children and comparable controls in Ys 5 – 7, sixteen comparisons showed that there were no significant differences between the EC children and the control children. In two comparisons, the low ability EC1 group compared with controls in Y5 and Y7, performed significantly lower than the controls (blue stars). The effect sizes were 0.27SDs and 0.20SDs respectively. At three points of comparison, the means for the EC2 cohort were higher than those of controls but not significantly so.

Figure 4.2 Mathematics: PIPS regression corrected¹¹ mean scores for high, middle and low ability groups for three successive cohorts, control, EC1 and EC2: Total Sample, 24 Schools



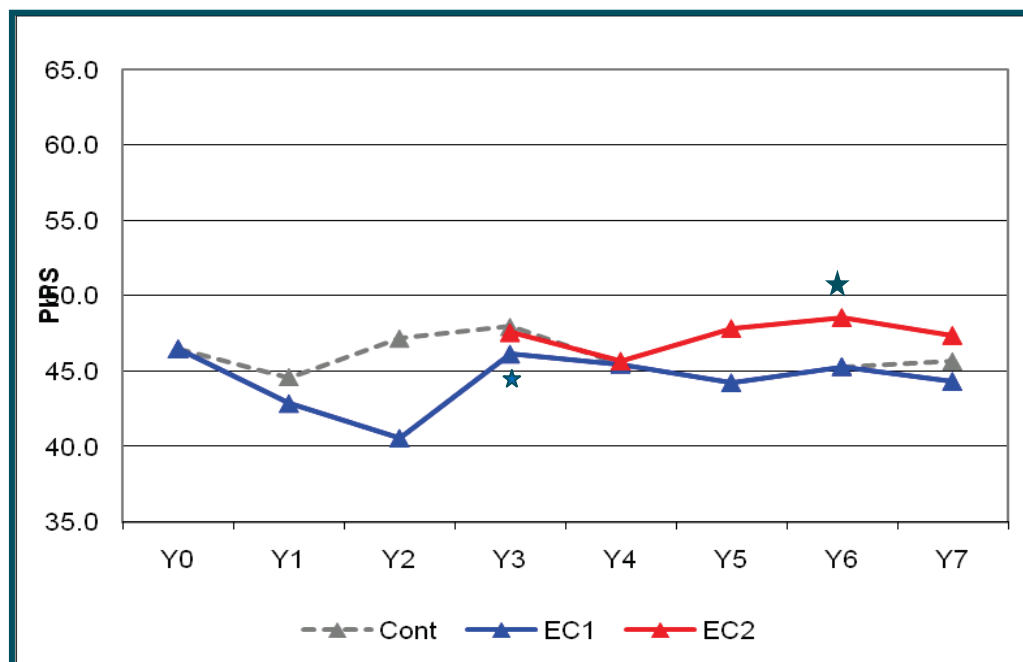
3.3 Mathematics: EC Pupils versus Control Pupils: High Deprivation Sample, 6 Schools

Because of the distinctiveness of the pupil intake into the high deprivation schools, and the differences in the training, resources, and implementation of the EC curriculum in these schools, compared to the other schools, separate multi-level analyses were conducted for the high deprivation sample alone. Given the smaller sample size, it was not possible to include a separate grouping for high, middle and low ability children in the analysis. However, according to the PIPS tests of 'developed' ability, the majority intake into these schools is in the lower-middle ability range, and there are high levels of social deprivation as shown by percent FSM.

Figure 4.3 shows the graphical representation of the regression corrected mean scores for three cohorts of children over 7 years. The analysis compares the effects of the EC1 (blue line) and EC2 (red line) outcomes relative to the control group (broken grey line) of year-ahead classes for all children.

¹¹ These scores are calculated using the statistical model that corrects for any imbalance in the sample. For example, if one year group has more scores from high deprivation schools than another, the model allows for it.

Figure 4.3 Mathematics: PIPS regression corrected mean scores for three successive cohorts, control, EC1 and EC2: High deprivation sample only, 6 schools



The most striking finding for this high deprivation sample was that, from Year 5 onwards, the second EC cohort (red line) was performing substantially better in mathematics than both the EC1 cohort and the control sample. The EC2 cohort means were superior to EC1 means in Ys 5, 6 and 7, and superior to the controls in Years 6 and 7¹². In Year 6, the EC2 performance in mathematics was significantly better than the control cohort (red star); the effects size was 0.33 SDs, a small effect. The superior performance of these high deprivation EC2 children have contributed substantially to the improvements for the EC2 lower ability children that were observed in the analysis for the total sample (Figure 4.2).

The general trajectory in mathematics for these children over time, confirms that, with the exception of the EC2 cohort, they were performing slightly below baseline expectations by the time they left school. This is consistent with the finding reported for the total sample.

3.4 Mathematics: EC Pupils versus Control Pupils: Contrasting Areas Sample, 18 Schools

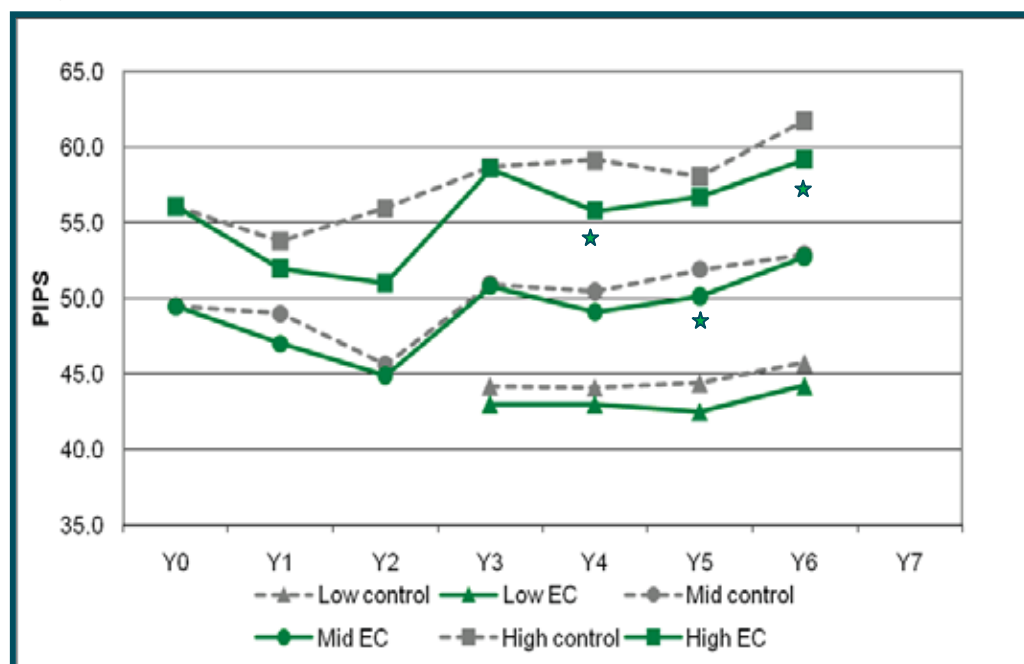
A separate multi-level modelling analysis was conducted for the Contrasting Areas sample of schools. This sample comprises the CA1 schools (6 schools), which joined the evaluation in 2001, and the CA2 schools (12 schools), which joined the evaluation in 2004. However, both groups of

¹² No control data are available for the high deprivation schools in Year 5, due to uncertainty at that time about the continuation of the longitudinal study into Key Stage 2.

schools had begun the EC in their schools in either 2001 or 2002, and they experienced similar levels of support and training. There were no high deprivation schools in either CA1 or CA2 schools: the range was moderate to low deprivation (%FSM 0-30). For the purposes of analyses, they were combined to maximise the sample size.

Preliminary comparisons of the effects of EC1 and EC2 cohorts showed that the differences between the two cohorts were not as consistent as for the high deprivation schools. Sometimes the EC2 cohort was marginally better than the EC1 cohort and/or the controls, sometimes it was marginally worse. For this reason, the EC cohorts were combined and compared directly with the control cohort. Figure 4.4 shows the regressions corrected mathematics scores for high, middle and low ability children for two cohorts, the control cohort (broken grey line) and the combined EC1/EC2 cohort (green line), for all CA schools, over six years. Y7 data were not included in this analysis because no EC Y7 data were available from the CA2 schools and only the first EC cohort was available from the CA1 schools. The imbalance in the sample sizes between the control and EC samples for Y7 reduces the reliability of the estimates in the model. In addition, the means for the low ability group for baseline (Y0), Y1 and Y2 in this sample are not shown in Figure 4.4. The sample size was small and too unbalanced between EC and control for these year groups and the comparison was not reliable.

Figure 4.4 Mathematics: PIPS regression corrected mean scores for high, middle and low ability groups for two successive cohorts, control, EC1/EC2 combined: Contrasting Areas Sample (CA1 and CA2), 18 schools



General trends: In terms of the general trends in the children's performance over time, the trajectory for high and middle ability children was in an upwards direction, while for lower ability children the trajectory was flatter, though still in an upward direction. This shows that middle and higher ability children were likely to achieve beyond their baseline expectations in mathematics, while the performance of lower ability children was more consistent with baseline expectations.

Years 1-4 EC versus Control: This is an augmented dataset in Years 2, 3, 4 compared with what was reported in the End-of-Phase 1 Report. Nevertheless, the first point to notice is that the pattern up to Y4 is similar to that described and discussed in previous reports. EC children did not begin formal recorded mathematics as early as in the pre-existing curriculum; when they did move on to formal work in Y2, their rapid progress was apparent in the results for the end of Y3. There were no statistically significant differences between the control and EC groups in Y3 across all three ability groups; in the Y4 high ability group, the EC children's performance was significantly behind the controls, although the performance of the two groups merged again in Year 5.

Years 5-6 EC versus Control: For 4/6 comparisons between subgroups of EC children and comparable controls in Key Stage 2 Ys 5-6, there were no significant differences between the groups; two comparisons were significantly different, the EC middle ability group in Y5 and the EC high ability group in Y6 were underperforming compared with controls with effect sizes of 0.17 SDs and 0.25 SDs respectively. These are small effect sizes.

3.5 Reading: EC Pupils versus Control Pupils: Total Sample, 24 Schools

This initial multi-level model evaluates the outcomes for reading for the total sample. As for the mathematics analysis, it combines the findings from all 24 schools, irrespective of any differences between them in terms of training, resources and the implementation of the EC. It benefits from the large sample size (See Section 2.2) and thus increases the reliability of the estimates from the multi-level modelling. It compares the effects of the EC1 (blue line) and EC2 (red line) outcomes relative to the control group (broken grey line) of year-ahead classes, for high, middle and low ability groups of children. Figure 4.5 shows the graphical representation of the regression corrected mean scores for these nine groups of children (3 cohorts x 3 ability levels) over 7 years.

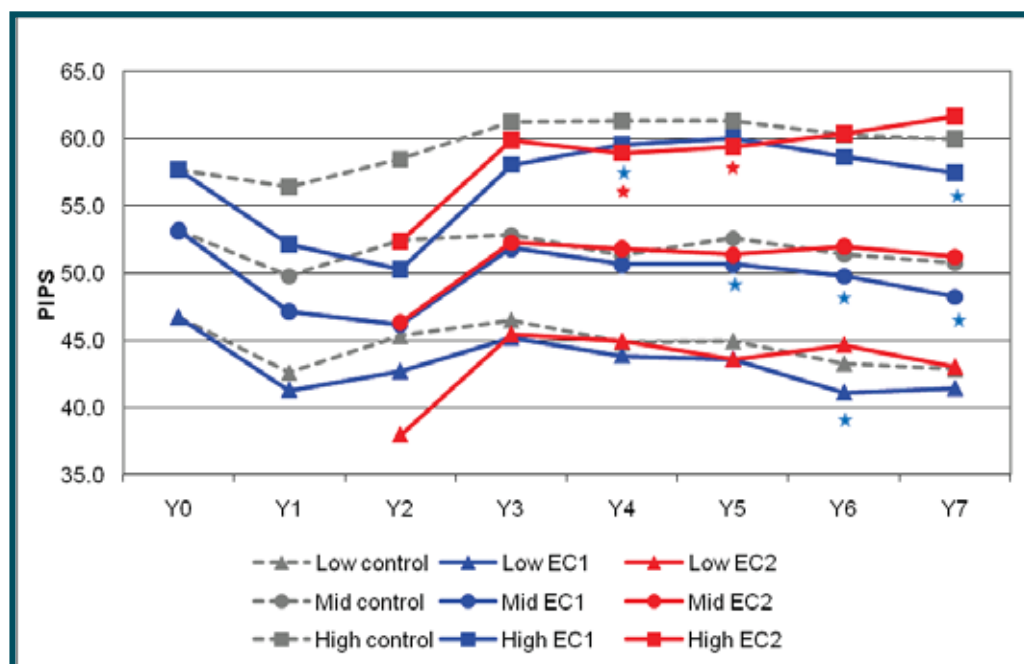
General trends: In terms of the general trends in the children's reading performance over time, note that the differences between the ability groups (irrespective of cohort) increased over the 7 years of primary school. There was a greater gap between the high and lower ability children when they left

school compared to when they entered school. In contrast to progress in mathematics, the trajectory over time is in the upwards direction only for the high ability group. For the middle and lower ability groups, the trajectory is downwards. This shows that only the higher ability children were likely to achieve beyond their baseline expectations in reading, while the performance of the middle and lower ability children tended to be below baseline expectations, particularly so for the lower ability groups.

Years 1-4 EC versus Control: This is an augmented dataset in Years 2, 3 and 4 compared with what was reported in the End-of-Phase 1 Report. Nevertheless, the first point to notice is that the pattern **up to Y3** is similar to that described and discussed in previous reports. EC children were not beginning formal reading as early as in the pre-existing curriculum; when they did move on to formal work in Y2, their rapid progress was apparent in the results for the end of Y3. By Y4, there was a statistically significant difference only between the EC1 high ability group compared with the marginally better control; the effect size was 0.20SDs. For middle and lower ability groups, all three cohorts performed very similarly and there were no significant differences between them.

Years 5-7 EC versus Control: As the children progressed into Key Stage 2, the first point to notice is that the EC2 cohort (red line) performed better than the EC1 cohort (blue line) consistently in all groups of children. Of the eighteen comparisons between subgroups of EC children and comparable controls in Ys 5 – 7, six EC groups performed significantly more poorly compared with controls; 5/6 were from the EC1 cohort. The effect sizes ranged from 0.19 SDs to 0.24 SDs, all small effects. At seven out of nine points of comparison, the means for the EC2 cohort were higher than those of controls but not significantly so.

Figure 4.5 Reading: PIPS regression corrected¹³ reading scores for high, middle and low ability groups for three successive cohorts, control, EC1 and EC2: Total Sample, 24 Schools



3.6 Reading: EC Pupils versus Control Pupils: High Deprivation Sample, 6 Schools

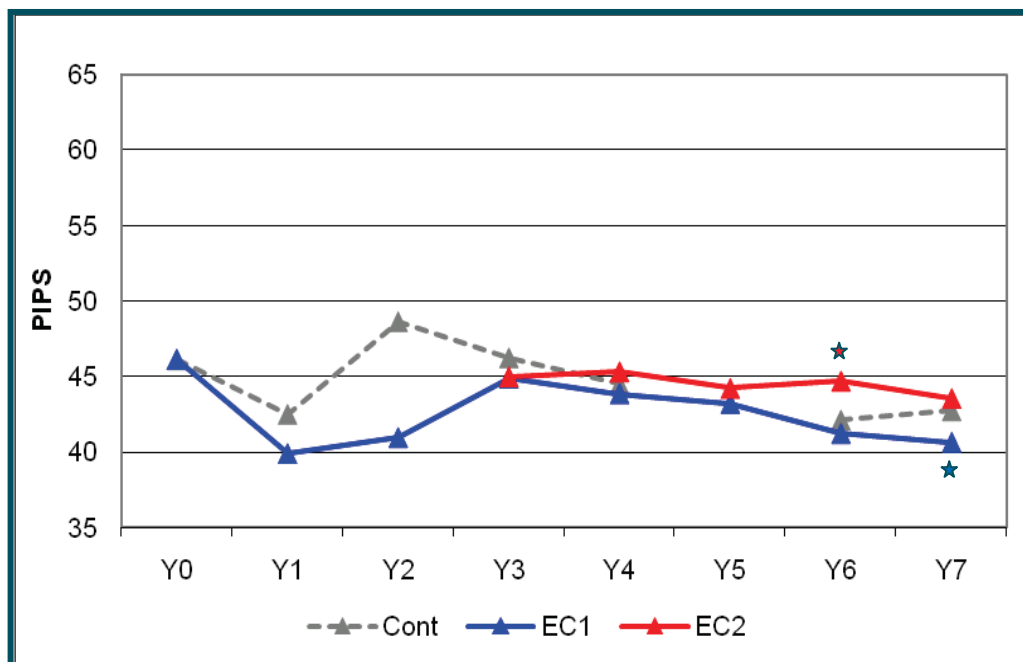
Because of the distinctiveness of the pupil intake into the high deprivation schools, and the differences in the training, resources, and implementation of the EC curriculum in these schools, compared to the other schools, separate multi-level analyses were conducted for the high deprivation sample alone. Given the smaller sample size, it was not possible to include a separate grouping for high, middle and low ability children in the analysis. However, according to the PIPS tests of 'developed' ability, the majority intake into these schools is in the lower-middle ability range, and there are high levels of social deprivation as shown by percent FSM.

The analysis compares the effects of the EC1 (blue line) and EC2 (red line) outcomes relative to the control group (broken grey line) of year-ahead classes for all children. Figure 4.6 shows the graphical representation of the regression corrected mean scores for these three cohorts of children over 7 years. Again it was clear that the EC2 (red line) was performing better than either the EC1 cohort (blue line) or the control, from Y4 onwards. The EC2 cohort was significantly better than controls in Y6; the effect size was 0.25 SDs, a small effect. The EC1 cohort was significantly lower than controls

¹³ These scores are calculated using the statistical model that corrects for any imbalance in the sample. For example, if one year group has more scores from high deprivation schools than another, the model allows for it.

in Y7; the effect size was 0.21 SDs, a small effect. Despite the positive impact of the EC in the second cohort, the general trajectory for this high deprivation group was downwards.

Figure 4.6 Reading: PIPS regression corrected mean scores for three successive cohorts, control, EC1 and EC2: High Deprivation Sample Only, 6 Schools



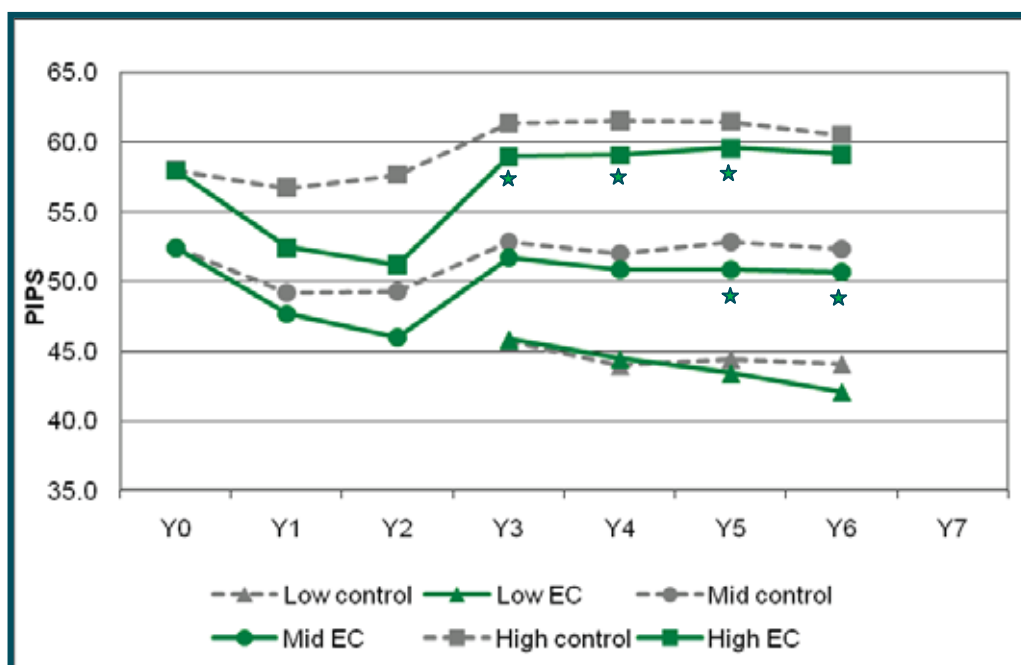
3.7 Reading: EC Pupils versus Control Pupils: Contrasting Areas Sample, 18 Schools

A separate multi-level modelling analysis was conducted for the Contrasting Areas sample of schools. This sample comprises the CA1 schools (6 schools), which joined the evaluation in 2001, and the CA2 schools (12 schools), which joined the evaluation in 2004. However, both groups of schools had begun the EC in their schools in either 2001 or 2002, and they experienced similar levels of support and training. There were no high deprivation schools in either CA1 or CA2 schools: the range was moderate to low deprivation (%FSM 0-30). For the purposes of analyses they were combined to maximise the sample size.

Preliminary comparisons of the effects of EC1 and EC2 cohorts showed that the differences between the two cohorts were not as consistent as for the high deprivation schools. Sometimes the EC2 cohort was marginally better than the EC1 cohort and/or the controls sometimes it was marginally worse. For this reason, the EC cohorts were combined and compared directly with the control cohort.

Figure 4.7 shows the regression corrected reading scores for high, middle and low ability children for two cohorts, the control cohort (broken grey line) and the combined EC1/EC2 cohort (green line), for all CA schools, over six years. Y7 data were not included in this analysis because no EC Y7 data were available from the CA2 schools and only the first EC cohort was available from the CA1 schools. The imbalance in the sample sizes between the control and EC samples for Y7 reduces the reliability of the estimates in the model. In addition, the means for the low ability group for baseline (Y0), Y1 and Y2 in this sample are not shown in Figure 4.7. The sample size was small and too unbalanced between EC and control groups and the comparison was not reliable.

Figure 4.7 Reading: PIPS regression corrected mean scores for high, middle and low ability groups for two successive cohorts, control, EC1/EC2 combined: Contrasting Areas Sample (CA1 and CA2), 18 schools



General trends: In terms of the general trends in the children’s reading progress over time, the trajectory is marginally upwards for the higher ability children, and downwards for the other ability groups – reflecting the pattern in the overall sample. This shows that the higher ability children were likely to achieve beyond their baseline expectations in reading, while the performance of middle ability children was more consistent with baseline expectations and the lower ability children were likely to achieve below expectations.

Years 1-4 EC versus Control: This is an augmented dataset in Years 2, 3, 4 compared with what was reported in the End-of-Phase 1 Report and there are slight differences in the higher ability groups at Years 3 and 4 compared to previous samples. EC children did not begin formal reading as early as in

the pre-existing curriculum; when they did move on to formal work in Y2, their progress was apparent in the results for the end of Y3 and Y4 for the middle and lower ability groups. However, for Y3 and Y4 high ability groups, the EC children's performance continued to be significantly behind the controls, with small effect sizes of 0.31SDs and 0.24SDs respectively.

Years 5-6 EC versus Control: In Key Stage 2, 3 of 6 comparisons between EC and control children showed significant differences; the EC middle ability group in Y5 and Y6 and the EC high ability group in Y5 underperformed compared with controls, with effect sizes of 0.20 SDs, 0.17 SDs and 0.19 SDs respectively. These were small differences.

3.8 Science: EC Pupils versus Control Pupils: Year 7 only

The PIPS science test was available for Year 7 only. Three separate multi-level model regression analyses were conducted to evaluate the effects of the EC on science outcomes at one point in time at Y7, for the total sample and for the high deprivation sample and for the CA sample. Table 4.4 shows the standardised regression means for EC and control samples for the high deprivation schools, and for CA1 and CA2 schools.

Table 4.4 Science: PIPS regression corrected means for Y7 for the different samples

Sample	Control Mean(SD)	EC Mean (SD)
High Deprivation sample	43 (5.45)	45 (4.88)
CA sample	51 (7.17)	51 (6.72) (8 schools only)*
Total sample	50 (7.30)	49(6.81) (14 schools only)*

*Year 7 EC2 children who joined the project in 2002 had not reached Year 7 at the end of the evaluation

EC versus control, total sample: After controlling for socioeconomic status, month-of-birth and gender, there was no significant effect of the intervention on science outcomes in the whole sample. The total sample sizes were N (EC) = 341 and N (control) = 365.

EC versus control, high deprivation sample, 6 schools: In this group, EC children showed superior performance on the science test compared with control children after controlling for socioeconomic status, month-of-birth and gender. The effect size was 0.22SDs. The sample sizes were N (EC) = 187 and N (control) = 93.

EC versus control, CA sample: In this group, there were no significant differences between EC children and control children. The sample sizes were N (EC) = 154 and N (control) = 272.

3.9 Reading, Mathematics and Science: Main Effects of Predictor Variables

This section provides a summary overview of the main effects of the predictor variables for all nine multi-level regression analyses that have just been reported. The main effects in the statistical analyses operate *irrespective* of whether the child belongs to the EC or control groups, and they have been statistically controlled when examining the effects of the intervention. The effects of individual differences in ability (high, middle, and low) have already been reported. The variables reported here are month-of-birth, gender, and two measures of social deprivation. The pattern of findings for these outcomes has not changed appreciably since the Sixth Report (Sproule et al., 2008), where they were discussed in detail. Table 4.5 shows the strength of their effects on the pupils' outcomes in reading, mathematics and science for the total sample, and for the high deprivation sample only, and for the CA sample only. Where the variables have a statistically significant or marginally significant trend effect, they are reported, together with the effect sizes. Effect sizes are particularly useful in this table because they allow the relative strength of the effects of different variables to be scrutinised.

Table 4.5 Main Effects of Predictor Variables for All Analyses: Effects Sizes for Significant Effects ($p < .05$) and for Significant Trends ($p < .1$)

Predictor	Sample	Reading	Mathematics	Science (Year 7 only)
Month-of-birth over 12 months	Whole sample	.18	.24	ns
	CA group	.19	.23	.07 trend
	High deprivation	.19	.39	ns
Gender	Whole sample	.30	ns	.13
	CA group	.32	ns	.14
	High deprivation	.38	ns	ns
		(girls are better)		(boys are better)
School mean ability over the full range of school mean ability, high to low	Whole sample	.36 trend	.45	.50
	CA group	.35 trend	.47 trend	.64
	High deprivation	ns	ns	ns
School's percentage of free school meals over the full range of % FSM	Whole sample	ns	ns	-.45
	CA group	ns	ns	ns
	High deprivation	ns	-.69 trend	-.21 trend

The key findings related to the Main Effects are presented below:

Month-of-birth: The ages of children within the same class can vary by 12 months. 'Month-of-birth' refers to the age of a child relative to the other children in their class. This variable had a statistically significant effect for both reading and mathematics outcomes and in all analyses. Children who were old-in-class performed better than those who were young-in-class. The effect was bigger for mathematics than for reading. In terms of differences in standardised PIPS points, there was a difference of between 2-4 standardised points between the youngest and the oldest in the class for mathematics and just less than two points of a difference in reading. The effects were larger when the children were younger, and they got smaller over time. For science learning in Year 7, only a small effect was detected in the CA school sample.

Gender: Gender had a substantial effect on reading performance where the girls' performance was consistently better than the boys in all the analyses by about 3 standardised PIPS points. There were no gender differences in mathematics. The boys were better than the girls in science in Y7.

Social deprivation: Two measures of social deprivation were included as main effect predictors. The mean ability of the school gives some indication of the school's pupil intake and is likely to affect the language and the classroom talk that is available to promote learning in the classroom. It should be noted that this variable had effects on children's progress *beyond* the effects of the individual ability of the child. %FSM refers more to the economic circumstances of the child's family. These two variables were highly correlated in the sample ($r=.79$) and it is difficult to statistically distinguish the effects of one variable versus another. That is why the variable *school mean ability* had a significant effect in some analyses, while *FSM* had an effect in other analyses.

Both variables negatively affect school attainment in reading, mathematics and science; schools with higher mean ability intakes and/or with few children eligible for free school meals had better outcomes for children. In those samples where the differences between the schools were greatest — the whole sample and the CA sample — school mean ability had the greater effect. For example, in mathematics, a child in the school with the highest mean ability is expected to outperform a child of similar characteristics in the school with the lowest mean ability by 4.5 PIPS points. School mean ability had no statistical effect in the high deprivation schools simply because there was very little difference between the schools in that sample on the measure. But even within the high levels of FSM reported by the high deprivation schools, those with the very highest levels of FSM tended to

do worse than those with slightly lower levels. Deprivation effects tended to be greater for mathematics than for reading. There were surprising big effects of social deprivation on science learning in Y7.

At risk children: It should be noted that, irrespective of the curriculum that the children were following, some children experienced an accumulation of factors that predict poorer progress in school — poorer baseline scores at school entry, being a boy, being young in class, and various contextual factors related to social deprivation, inside and outside school. Issues related to the detection of at-risk children and proposed interventions were extensively discussed in the Sixth Annual Report.

3.10 Summary of Key Findings for Mathematics, Reading and Science

Preliminary points

- The main focus for reporting these outcomes was on the pupils as they moved into Key Stage 2. Nevertheless, the sample was considerably increased for Phase 2, and included children from 12 new schools and a second EC cohort in all 24 schools. As well as following children into KS2, additional data were collected for Years 2, 3 and 4 pupils; so the performance of the augmented sample in those years needed to be evaluated to check if the patterns of the children's performance at the earlier year groups were similar to those that had been previously reported.
- Extensive use was made of the 'effect size' to quantify the strength of statistically significant effects and to allow for comparisons to be made across variables and across different measures. All the effects associated with the impact of the EC intervention, whether in a positive or a negative direction, can be characterised as small effects, varying from 0.15 – 0.33 SDs. This latter effect is approximately of the same size as the difference between boys and girls. The biggest effect sizes were associated with social deprivation predictor variables which had an impact on the children's progress, irrespective of the curriculum.

Key Findings EC versus Control groups

- **Year 1 and Year 2:** The general pattern of findings for Y1 and Y2 remains largely the same as previously reported. EC children did not begin formal recorded mathematics or formal reading as early as in the pre-existing curriculum, so their performance was poorer in those first two years when compared to the control groups. When they did move on to formal work towards the end of Y2, they made rapid progress, such that, with certain exceptions

(see below), there were no statistical differences between the EC and control groups in Year 3 and Year 4.

- **A broad sweep from Year 3 to Year 7:** Taking a broad sweep across the data from Year 3 to Year 7, the vast majority of statistical comparisons between EC versus control subgroups found no statistically significant differences between EC and control classes. For example, 107 statistical comparisons were made across the different samples (total sample, high deprivation sample, CA sample), different ability levels (high, middle and low ability) and across the different measures (mathematics, reading, science) from Year 3 to Year 7: 77% (83/107) showed no statistical differences between the two groups, 20% showed statistical differences in favour of the control classes, and 3% showed statistical differences in favour of the EC classes.
- **Mathematics versus reading:** In general, the EC classes performed better in mathematics than in reading. For example, of the 52 statistical comparisons made in mathematics and in reading from Year 3 to Year 7, 85% (44/52) showed no significant differences between EC and control classes in mathematics compared to 71% (37/52) in reading. In 13% of statistical comparisons (7/52), the EC class underperformed relative to the control class in mathematics; the comparable figure for reading was 27% (14/52). For both mathematics and reading, the EC class outperformed the control class in 2% (1/52) of comparisons.
- **Science:** Three comparisons were made in Science as it was assessed only in Year 7. There were no differences between the EC and control groups in 2/3 comparisons (67%) and in the remaining case, the EC performed better than the control class (33%).
- **EC1 versus EC2:** There was a general trend for the EC2 cohort to perform better than the EC1 cohort, suggesting that the curriculum had taken some time to 'bed' down. The effect was more apparent for the lower and middle ability groups overall and particularly marked in the high deprivation sample. The improved pattern for the EC2 cohort was not so obvious for the CA schools; sometimes the EC1 cohort did better than the EC2 cohort and sometimes the EC2 cohort did better. This inconsistent pattern did not always show statistically significant group differences.
- **High deprivation schools:** For the high deprivation schools, the EC began to have a distinctly positive effect on outcomes for the EC2 cohort as the pupils progressed into KS2. For both **mathematics and reading**, the EC2 cohort was better than the control groups from Year 4 to Year 7, and showed statistically significant differences in Year 6 on both measures. The EC Year 7 classes also performed better than the control classes in **science**. This finding coincides with the general reports from the school principals that the EC took some time to

bed down, but also with several other factors that may have had a positive influence on the outcomes for the EC2 in those schools. The area had experienced severe political unrest when the EC1 cohort began school and their transition to school was disrupted. In addition, the training and resources were more cohesive in the high deprivation schools, and some EC2 children in the group experienced Linguistics Phonics¹⁴ training which may have had a positive impact on reading.

- **Against the trend:** The general pattern indicated that children's reading progress was more negatively affected by the EC than their progress in mathematics. However, the achievements of the EC2 cohort in the high deprivation schools did not follow this general pattern, and showed that both reading and mathematics attainment were equally enhanced by the EC.
- **Ability levels:** In terms of ability levels, it was lower ability EC children who were most likely to achieve outcomes that were better, or as good as, the previous curriculum, especially as the curriculum bedded down (i. e. the EC2 cohorts). Higher ability children who were more highly represented in the CA schools, tended to do less well compared with controls.

Key Findings: General Patterns from the Main Effects

- **A widening gap:** There was a widening gap between higher and lower ability children as they progressed through primary school.
 - Higher ability children in the evaluation schools were progressively performing beyond their baseline expectations in both reading and mathematics, but particularly in mathematics.
 - For middle ability children, their progress was more in line with baseline expectation, but with an upward trajectory in mathematics and a marginally downward trajectory in reading.
 - Lower ability children were progressively performing below their baseline expectations, and the effect was more pronounced in reading than in mathematics.
- **Age-in-class:** Children who were young for their age in class were disadvantaged compared to their older peers. The effect was stronger for mathematics than for reading.
- **Gender:** Girls consistently outperformed boys in reading. There were no gender differences in mathematics. Boys were better than girls in science.
- **Social deprivation:** Social deprivation had a negative effect on pupil outcomes in several different ways. The effect was seen at entry into school (baseline measures) and children

¹⁴ See http://www.belb.org.uk/parents/literacy_linguistic.asp?sm=37

continued to lose ground compared with those in schools in more advantaged areas. One indication against the trend was the positive impact of the EC curriculum on mathematics performance in the second cohort of children in high deprivation schools. More generally, children in high deprivation schools were not making good progress relative to the total sample of schools. However, once deprivation was quantified, the children from the high deprivation schools were making as good progress as could be expected, given the general effects of deprivation on the pupil outcomes in the total sample. After statistically taking individual ability into account, social deprivation (measured by either the % FSM or school mean ability) was the second most important predictor of pupil outcomes. The effect of social disadvantage was stronger for mathematics and science than for reading.

- **At risk:** The effects of individual ability, social deprivation, gender and month of birth are cumulative and highly predictive of pupils' outcomes when taken together, indicating that it is possible to predict quite early when children are likely to be at risk of failure to attain basic skills in reading and mathematics in primary schools.

To conclude, as pupils progressed into Key Stage 2, the EC had little longer-term positive or negative impact on their reading and mathematics attainments for over 75% of the comparisons conducted. In the remaining comparisons where there were statistically significant differences between EC and control classes, the impact was in the negative direction, particularly for reading. In general, lower ability children seemed to benefit more than higher ability children. For the second EC cohort in the high deprivation schools the EC had a statistically significant positive impact for both mathematics and reading in Year 6 and for science in Year 7. The strength of the statistically significant effects (effect sizes) associated with the EC were generally small. Social deprivation variables had the most powerful negative influence on pupil outcomes, irrespective of the curriculum being followed and these were stronger for mathematics and science than for reading.

4. Outcomes for Pupils over Time: Writing

4.1 Brief comment on learning to write

As children progress through primary school, what counts as competence in literacy changes. As they grow older, a large part of children's language development is concerned with learning written forms of language. This involves more than learning the mechanics of production, it also involves learning syntactic structures, using vocabulary, organisation, and becoming familiar with text types

that are perhaps only found in written forms of language. Learning written language also involves learning to communicate with an audience and to structure texts in such a way that they will be perceived by their readers as coherent, and from which readers can gain a message similar to that intended by the author. Coherence is ultimately not a property inherent in the words on the page. Nevertheless, there are ways of using language which are likely to be helpful to readers in achieving coherence. It is these aspects of language use that emerging young writers have to learn.

The EC had a particular focus on developing oral language and creativity in the early years. Evaluating children's writing provided an opportunity to examine the impact of these earlier EC activities on the later development of children's writing.

4.2 Compositional Writing: The Wechsler Objective Language Dimension Test (WOLD), Test of Written Expression

The main psychometric literacy measure used in the evaluation, PIPS, did not include a measure of children's writing. Thus, in later years as children became able to write independently for reasonable periods, it was appropriate to introduce a test of written composition skills. The WOLD Test of Written Expression (one of the Wechsler Objective Language Dimensions, 1996) was selected as a suitable robust measure. It is age appropriate from the age of eight years up to sixteen years. For the test, the child is given a title, and a specified time to write (20 minutes). The child's free writing is then analysed and scored against criteria on six different dimensions. Each piece of writing receives a total writing score, and a separate score for each of the six sub-scales.

- Ideas and development: Scored on the number of ideas and the extent to which these are extended and elaborated.
- Organisation, unity and coherence: Scored on the degree to which the writing follows a plan and the different parts are logically related.
- Vocabulary: Scored on the appropriateness, precision and vividness of the words used.
- Sentence structure and variety: Scored on correct sentence construction and variety of sentence formats.
- Grammar and word usage: Scored on correct grammar and word usage but allowing colloquial expressions where these are appropriate in the context of the intended audience.
- Capitalisation and punctuation: Scored on correctness and sophistication.

Spelling is not directly a focus of the test and this is explained to the children in the instructions for the test. However, certain spelling errors such as 'where' for 'were' may be penalised under word usage.

Administration and Scoring: The administration was slightly different from the standardised approach in the test manual, and the changes were made following advice from teachers, in order to be consistent with the classroom practices that the children were used to. Instead of just providing the title without any elaboration, a ten-minute scripted discussion fleshed out the topic a little and suggested what things might be considered in order to get started. Children were then given 20 minutes to write on the given topic. The test has two alternative scenarios, 'Inviting a friend for a day out' and 'Planning the design and redecoration of the child's bedroom'. These two titles were used with both the EC and the Control pupils. All children were group tested in their normal classrooms.

The children's free writing was then scored following a strict scoring procedure outlined in the test manual. Each dimension of the writing was scored from 1-6, giving an overall score that can range from 6-24. Scoring was completed by ten markers who had been extensively briefed on the scoring protocol. Because there is some element of subjectivity in the scoring, reliability checks were conducted on a percentage of each marker's scorings by an expert marker, to ensure consistency, and adjustments were made when necessary.

Because of the changes in the administration, it will not be possible to use standardised scores to report the results. Table 4.6 gives examples of low-, middle- and high-scoring children's writing so that the reader can have some understanding of the standards of the writing that are being referred to.

Table 4.6 Examples of Children's Writing from the WOLD Test of Expressive Writing
{Spelling and punctuation preserved as in the original (names changed)}.

High scoring example – Y7 boy, 22.5/24

Dear Brandon

I have some amazing news to tell you! Two weeks ago I entered a competition that could change my life forever and guess what, I won! The prize was a once in a lifetime opportunity. I can go anywhere I want for a whole day. Also, I have some good news for you! I am allowed to bring one friend with me and I have chosen you! I am sure you are wondering where we are going, the answer is SPACE. I had better tell you all about our trip. Firstly, don't worry about the cost. Everything is being payed for us, you'll be happy to know. We will be going on our breath-taking journey in one of the world's newest space-crafts. Also, we will be the first children to venture into space! We'll be a Part of history! Unbelievable! I have made out a plan for our wonderful adventure! We will meet at the N.A.S.A. headquarters at 5.00pm on Saturday, 28th June. We will board the magnificent shuttle at 6,00am. The shuttle will have everything we could possibly need, food, water, oxygen and of course TV! Don't worry a trained astronaut, Neil Armstrong, will accompany us. Whne we come back from space we will be attending a presentation that will be shown on TV worldwide. Isn't that spectacular! There is absolutely no doubt that you will enjoy this trip, remember it for the rest of your life! Well, I hope to see you there.

Darren

Middle/high scoring example – Y5 boy, scoring 16/24

Dear Peter

You'll never believe what happened. I won a competition to go to Barrys (in Portrush) for a day. You are allowed to bring one friend and I have picked YOU! We will have the place all to ourselves and the rides will be free. You don't have to worry about getting left down. I have hired a bus all to ourselves. I will pick you up at 8.30am (look out for the bus). I am planing to go on the ghost train first (to get you ready for the scary rides). After that we will go on the Big Apple and that is some roller coaster. Next we will get a drink or whatever you want. After that you can do whatever you want. We could even go on the bumping cars. Oh, one more thing there will be a big, I mean BIG surprise when we get there. (Even I don't know what it is). One last thing BBC news will be there ar 5 o'clock see you tomorrow.

David

Low scoring example – Y4 girl, 7/24

Dear Aoife

I want you to come to Canada whith me and go skeing with me. Go swimming with me. and we can have a big ice-cream toghether. We will go shopping toghether. We will go to The park toghether. and buy friendship bracelets.

Meabh

Sample: Table 4.7 shows the sample sizes for the WOLD Test. The smaller size of the Year 7 sample simply reflects the fact that some children had not reached Y7 at the end of the evaluation period.

Table 4.7 Total sample size for the WOLD across 24 schools

	<i>Year 4</i>	<i>Year 5</i>	<i>Year 6</i>	<i>Year 7</i>
Control	325	287	340	454
EC	683	819	671	325
Total	1008	1106	1011	779

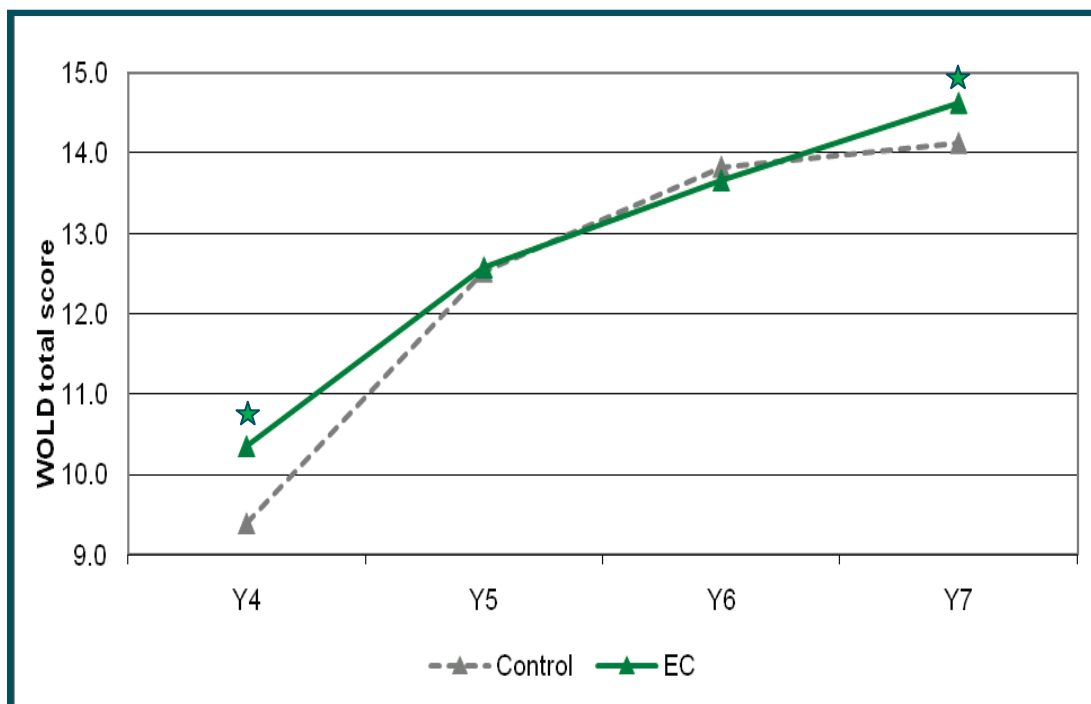
Although the size of the total sample was large, it was not possible to conduct separate multi-level modelling analyses for the high deprivation sample and for the CA sample, due to incomplete data sets in a small number of schools, which gave rise to unreliable estimates for some subgroups. However, simple multiple regression models were computed for the separate school samples and we make reference to these where appropriate.

4.3 Writing: EC Pupils versus. Control Pupils: Total Sample, 24 Schools

A multi-level modelling analysis was conducted for the whole sample for the 24 schools on the **raw** scores (not standardised scores) from the WOLD Test of Expressive Writing. Figure 4.8 shows the regression corrected reading scores for two cohorts, the control cohort (broken grey line), and the combined EC1/EC2 cohort (green line) over **four** years, from Y4 to Y7, for the total sample.

General trends: The first point to note is the overall level and quality of the writing. The overall mean for the sample across all the ages was 11.6. The mean score for the Y4 group was 10 rising to a mean of just over 14 for Y7s. Comparing this with the examples given in Table 4.7, the quality of the average Y7 child's writing was below the middle/high-scoring example in the table. In terms of the general trends in the children's performance over time, the trajectory for both groups of children was consistently upwards. As these are not standardised scores, this does not mean that children were performing above expectations; it merely shows that the children's writing was improving with age as would be expected. The scale cannot be confirmed as an equal interval scale so we cannot draw any firm conclusion about the change in gradient after Y4; it may just that it is easier to improve from a lower baseline than it is from a higher level.

Figure 4.8 Writing: PIPS regression corrected WOLD writing raw scores for two successive cohorts, control group and combined EC1/EC2 cohorts: Total Sample, 24 schools



EC versus Control children: Figure 4.8 shows that EC children did significantly better in the WOLD compared with controls in Y4 and Y7, with effect sizes of 0.42SDs and 0.18SDs respectively. These are both small effects, although the Y4 effect size was the biggest positive EC effect that emerged in any analyses. Analysis with simple multiple regression models (not a multi-level model) suggested that these gains may be attributed disproportionately to the CA schools, with the EC group in high deprivation schools showing smaller gains. However, it was not technically possible to confirm this with a multilevel model.

Analysis of individual scales on the WOLD: There was no consistent pattern of differences on individual WOLD scales across the sample as a whole. In high deprivation schools, the data suggested that EC children performed better than the control children on the more technical aspects of writing; — sentence structure and variety, grammar and word usage, and capitalisation and punctuation. The more creative aspects of writing were very similar for EC and control groups. In the CA schools, the situation was reversed, with the data suggesting that the EC children were performing better on ideas and development and on vocabulary. These were patterns in the data and could not be statistically confirmed.

Main effects in the analysis of the WOLD total score. There were strong main effects of predictor variables which apply equally to EC and Control children.

Month-of-birth: Age in class was a significant predictor of outcomes, with the oldest child in the year group scoring on average more than the youngest child with effect sizes in Years 4, 5, 6 and 7¹⁵ being 0.25SDs, 0.23SDs, 0.21SDs and 0.21SDs respectively. These were small effects.

Gender: Gender was a significant predictor of outcomes, with girls scoring on average more than boys in the year group, with effect sizes in Years 4, 5, 6 and 7 being 0.42SDs, 0.40SDs, 0.35SDs and 0.36SDs respectively. These were small effects, but larger than the gender effects reported for reading.

Social deprivation: Social deprivation, as measured by the school's mean ability score, was a significant predictor of outcomes, with children in the school with the highest mean ability scoring on average more than those in the school with the lowest mean ability, with effect sizes in Years 4, 5, 6 and 7 being 0.74SDs, 0.71SDs, 0.63SDs and 0.64SDs respectively. These are moderate to large effects. In addition, for this test only, there was an additional penalty for children in the high deprivation group schools, with effect sizes in Years 4, 5, 6 and 7 being 0.67SDs, 0.64SDs, 0.57SDs and 0.58SDs respectively. This is also a moderate to large effect. This means that social deprivation had a higher negative effect on writing skills than on any other academic outcome. About one child in seven in the highly deprived group of schools had a score of nine points or less on the WOLD (see Table 4.7) irrespective of the curriculum. This shows how poor the standard of writing was for children in high deprivation schools, in particular.

4.4 Summary of Key Findings for Writing

General quality of the children's writing

- Because of changes to how the test was administered to the children, it was not possible to use standardised scores; so we do not have a norm against which to judge the overall quality of the children's writing for the different age groups (unlike the PIPS test, where there is a substantially sized set of UK age-based norms). Nevertheless, the raw scores indicate that the standard of writing overall was poorer than the norms, given that children were given some help to get started. The average score across the age groups was 11.6, showing that the children's writing scored just under 2 points on each of the 6 subscales. Some Year 7

¹⁵ Because the scores were not standardised, it is important to give effects sizes per year group, even for main effects.

children wrote very well and scored well above normative expectations, given that the test can be used with pupils up to sixteen years. On the other hand, there were sizeable numbers of children at each year group who had very poor skills.

EC versus control groups

- The EC had a positive effect on children's writing in Year 4 and in Year 7, indicating that when the children first began to write independently that they had an advantage over the children who were following the previous curriculum. However, as they progressed into Key Stage 2, this advantage disappeared but re-appeared again in Year 7. It was not possible to conduct separate multi-level model analyses for the high deprivation schools and the CA schools. Simple regressions showed that the effect held for both samples, but was stronger in the CA schools.
- Contrary to expectations, there were no consistent patterns between the EC and the control classes on the different subscales of the writing test. It was expected that the EC might be better at the more creative or compositional aspects of writing (e.g., ideas and development, unity and coherence, vocabulary) versus the technical aspects. This was not the case. Whenever the EC children scored higher than the controls, their improved performance was evident across all the subscales.

Effects of other predictors

- In general, the effects sizes of the other predictors — month-of-birth, gender and social-deprivation — on writing were larger than those observed in reading and mathematics. Social deprivation, as measured by school mean ability, had a large and consistent effect at each age group; and there was an additional effect of being in the high deprivation schools. A substantial minority of the children in the high deprivation schools had a very poor standard of writing for their age, irrespective of the curriculum they were following.

5. Outcomes for Pupils over Time: Learning Dispositions and Attitudes

5.1 The Broader Goals of the EC: Positive Learning Dispositions and Attitudes

Although the focus of this report so far has been on pupils' literacy and numeracy outcomes, it is important to remember that the EC had broad educational goals and intentions for young children. By adopting a play-based and informal approach to learning and teaching, the intention was to provide early school experiences that were not only appropriate for younger children in the 'here

and now', but that would also have positive and long-lasting effects on their learning as they progressed through school. The more exploratory and self-directed learning opportunities provided by a play-based curriculum, compared with the more formal pre-existing curriculum, were expected to stimulate children's curiosity, to increase their motivation and their willingness to learn, and to help them become more confident and independent learners in the long run. In addition, the initial designers of the EC were keen that the children's positive views of themselves as learners would have time to develop and would not be undermined by experiences of persistent failure by being presented with developmentally inappropriate tasks and/or learning situations.

Findings from the classroom observations, and from interviews with school principals, confirmed that these broader goals were being achieved to a large extent (End-of-Phase 2 Report 2). For example, the children were evaluated as having higher quality levels of social-emotional and cognitive classroom experiences, compared to more traditional classes. Furthermore, school principals were generally agreed that the EC had positively influenced the children's social, emotional and motivational development, and for many schools, had a more general positive influence on the school climate. However, these findings were at the level of the classroom observations and general perceptions about the school, and were not derived from information about *individual* pupils. The current section turns attention directly to monitor the impact of the EC on what have been variously described as pupils' learning dispositions, learning orientations or attitudes to learning as they progressed through their primary schooling, and to compare them with the year-ahead control classes who had experienced the previous curriculum.

These attributes were assessed from two perspectives: through the **children's self-evaluations** of their learning dispositions and attitudes to learning and through **teachers' ratings** of the children's social and behavioural competence in the classrooms. For the children's self-evaluations, a suite of a self-rating instruments, called the Assessment of Learner-Centred Practices, ALCPs (McCombs and Lauer, 1997), which are based the American Psychological Association's Learner-Centered Principles, was used. They ask the children to assess themselves on learning and motivational indicators that are linked to be the longer term outcomes of the EC. Although the ALCPS suite was developed in the US, it had previously been adapted and used in research in Northern Ireland on a similar age group of children (McGuinness, 2006) and had established psychometric properties. Children's self-evaluations were collected only from children in Key Stage 2, in Y5, Y6 and Y7¹⁶.

¹⁶ On verbally presented scales of this kind, younger children's ratings are not always reliable and have limited predictive power. They tend to rate themselves very positively, resulting in a very narrow distribution of scores.

During the first year of the project, teachers were asked to complete the Adaptive Social Behavioural Inventory (ASBI) for each child (Hogan, Scott and Hauer, 1992). This inventory was lengthy and its completion was a major burden for teachers. Consequently, the response rate was low. A shorter scale with just six items was constructed by the research team. The six items on which the teachers were asked to rate the children were: makes an effort with his or her work; is confident about his or her work; gets on well with other children; seems happy and well adjusted; has problems maintaining attention; has behaviour problems. Factor analysis of the items yielded a single factor that we have called a **pro-learning disposition** and the internal reliability of the scale is high (Cronbach's Alpha=.81). The scale focuses directly on attributes of the child that might enable him/her to take advantage of learning opportunities in the classroom and correlates moderately highly with educational attainment (see later sections). The measure was used at the end of Y2, Y3, Y4, Y5, Y6 and Y7 for both EC and control classes.

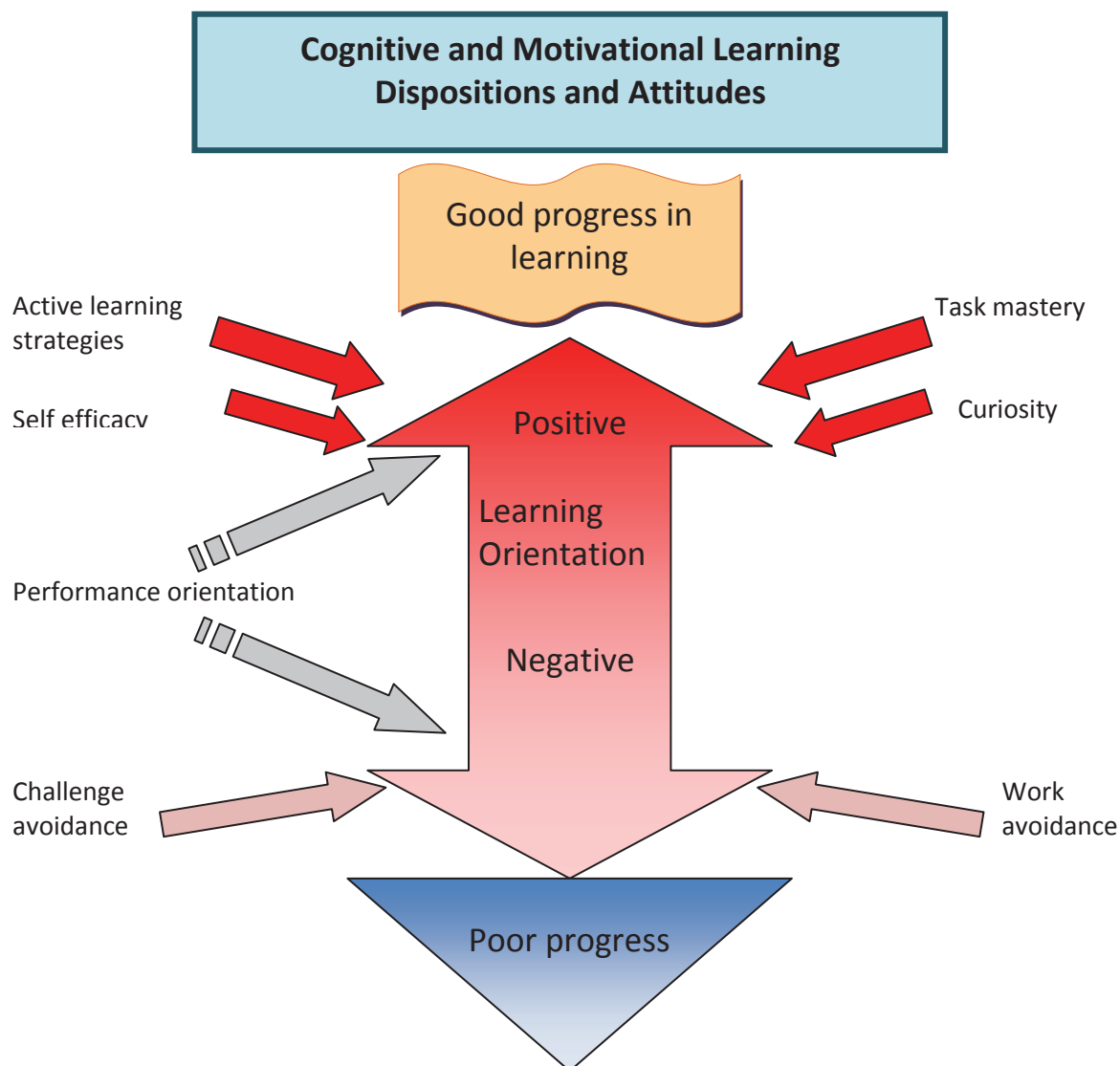
5.2 Learning Dispositions and Attitudes: Children's Self-Evaluations

The Assessment of Learner-Centred Practices (McCombs and Lauer, 1997), ALCPs, draws on an extensive research base that has identified cognitive and motivational dispositions and attitudes that are associated with a positive orientation to learning, and ultimately with positive progress in school (Alexander and Murphy, 1998). Seven scales in the ALCPs suite were used for the children's self-evaluations. Figure 4.9 gives a schematic outline of how the dimensions that underpin the scales are theoretically interrelated.

Two of the scales refer to how actively involved the children are in their learning, the strategies that they use to acquire knowledge, how they monitor and check their learning, and how open and curious they are about finding out new things. These scales are called **Active Learning Strategies** and **Curiosity**. High scores on these scales are predicted to be associated with a positive orientation to learning.

Two scales have explicitly negative connotations and refer to whether children go out of their way to avoid a mental challenge and to avoid making an effort, skip over difficult work, and try to finish as quickly as possible. These scales are called **Challenge Avoidance** and **Work Avoidance**. Low scores on these scales are predicted to be associated with a positive orientation to learning and higher scores with a more negative learning orientation.

Figure 4.9. A schematic outline of the ALCPs model of Positive Learning Orientation



The final three scales are related to motivation and beliefs about the self. Two of the motivational scales refer to the reasons why pupils might make an effort to learn about a topic and to persevere in the face of difficulty. **Mastery Orientation** refers to the desire to ‘master’ the understanding of a topic or the performance of a skill because they want to get better at it for its own sake, and is sometimes called intrinsic motivation. **Performance Orientation** refers to the desire to learn and to achieve a high standard because of the consequential rewards, status or recognition; this is sometimes called extrinsic motivation. These two types of motivational orientations are not polar opposites, and people can be motivated for both types of reasons. While mastery orientation is predicted to be associated with a positive orientation to learning, there is some debate in the research literature about the position of performance orientation; sometimes it predicts a positive orientation and sometimes less so. To anticipate the results — performance orientation also

produced some perplexing findings in this study. The final scale — **Self-Efficacy** — refers to the beliefs that pupils hold that, through their own efforts, they can influence the outcomes of their future learning. High self-efficacy scores are predicted to be associated with a positive learning orientation.

Although each scale can be considered separately, the seven scales do cluster into two groups:

- self-efficacy, mastery orientation, active learning strategies and curiosity are all predicted to be pro-learning; and
- challenge avoidance, work avoidance, and — to a lesser extent — performance orientation, are predicted to be negatively associated with learning.

Table 4.8 gives a summary of each scale and a sample of the associated items from the questionnaire.

The children were asked to rate themselves on a four-point scale on how frequently they do the activity, or have the experience, that is described in the questionnaire item: almost never (1), sometimes (2), often (3), or almost always (4). Lower scores indicate a positive learning orientation on the Work Avoidance, Challenge Avoidance and the Performance Orientation scales and vice versa for the remaining scales.

The questionnaires were completed in class and, for the younger classes, the items were read out so that children who had poor reading skills would not be disadvantaged or take too long to complete the questionnaire.

Table 4.8 The ALCPS Scales

Scale	Scale interpretation	Sample of the items to be rated by the pupils
<i>Motivational scales</i>		
Self-efficacy (6 items)	The belief that one's own actions can affect one's future learning outcomes.	<i>"I am sure I can do even the hardest work in this class if I try." "I am sure that I will do well in this class"</i>
Performance orientation (6 items)	Wanting to do well to gain recognition, approval or rewards.	<i>"An important reason why I do my class work is to get better marks than the other pupils." "I want to do well in this class so that the teacher will think I am smart."</i>
Mastery orientation (6 items)	Wanting to master a task or problem in order to get better at it, or out of interest.	<i>"I do the work in this class because it is interesting." "I do want to do my work because it really makes me think."</i>
<i>Attitudes to work scales</i>		
Work avoidance (6 items)	Avoidance of work completion, guessing or asking friends.	<i>"I feel I have done well in this class if I can do my work without much effort." "When I do work in this class, I just want to get it done as quickly as possible."</i>
Challenge avoidance (8 items)	Avoidance of putting in effort and of doing difficult work.	<i>"When I have a difficult piece of work to do in this class, I skip the hard parts." "When I have trouble with a piece of work, I give up."</i>
<i>Learning strategies and seeking out knowledge</i>		
Curiosity (7 items)	Curiosity, knowledge-seeking behaviours, seeking out knowledge because it is interesting or new.	<i>"School work is very interesting to me." "I enjoy learning schoolwork that is new to me."</i>
Active learning strategies (8 items)	Having proactive, metacognitive and independent learning strategies.	<i>"I try to figure out how new work fits with what I have learned before." "When I make mistakes, I try to figure out why."</i>

Sample: Table 4.9 shows the number of children who completed the ALCPS for each year group in the control and the EC classes. The smaller size of the Year 7 EC sample simply reflects the fact that some children had not reached Y7 at the end of the evaluation period.

Table 4.9 Sample size for the ALCPS

	Year 5	Year 6	Year 7
Control	221	365	468
EC	850	679	322
Total	1071	1044	790

Inter-correlations between the ALCPs scales: Before beginning the multi-level analyses to evaluate the effects of the EC children versus controls, simple bivariate correlations were conducted to examine the relationships between the seven ALCPs scales, to check if the predicted relationships between the scales were confirmed in this data set.

Table 4.10 shows the inter-correlations between the scales for Year 6 pupils (Y5 and Y7 show essentially the same results). The correlation coefficients in the table have been colour coded to help interpretation. The inter-correlations between the cluster of four positive learning scales — Self-Efficacy, Mastery Orientation, Active Learning Strategies and Curiosity — were all relatively large and were in the same positive direction (identified by the numbers in red). Those children who had high scores on one of the scales in this cluster, also tended to have high scores on the other scales, and those who had low scores on one of the scales also tended to have low scores on the related scales. Similarly, the cluster of three negatively oriented scales — Work Avoidance, Challenge Avoidance and Performance Orientation — were also positively correlated (identified by green numbers). Moreover, the positively learning scales were negatively correlated with the negative learning scales (identified by blue numbers). One scale stands out as anomalous — Performance Orientation. As predicted, this scale was positive correlated with the negative cluster of scales, but it was also positively correlated (though less strongly) with 3/4 positive learning scales (identified by purple numbers). Thus, the general theoretical relationships as predicted by the ALCPs were confirmed with this sample of Northern Ireland primary school children. Even the apparent contradictory correlations of performance orientation was not surprising, as the status of this type of motivation as pro or anti learning has been previously disputed (e.g., Pintrich, 2003).

The expectation is that the EC should have a positive impact on pupils' learning dispositions, that is, EC pupils should rate themselves higher than the control classes on the positive learning scales and lower than the controls on the negative learning scales. The predictions about performance orientation are less clear cut and are likely to be difficult to interpret.

Table 4.10 Correlations between the ALCPs seven scales: Example for Year 6

Scale Name	Self-Efficacy	Mastery Orient	Active Learning	Curiosity	Performance Orient	Work Avoid	Challenge Avoid
Self-Efficacy	1.00						
Mastery Orientation	.473**	1.00					
Active Learning	.451**	.606**	1.00				
Curiosity	.481**	.698**	.529**	1.00			
Performance Orientation	.165**	.177**	.197**	-.032	1.00		
Work Avoidance	-.117**	-.248**	-.149**	-.378**	.378**	1.00	
Challenge Avoidance	-.301**	-.294**	-.265**	-.400**	.278**	.567**	1.00

5.3 Learning Dispositions and Attitudes: EC Pupils versus Control Pupils: Total Sample, 24 Schools

Seven multi-level modelling analyses were conducted for the whole sample of 24 schools, one for each of the ALCPs scales. The sample comprises the Belfast high deprivation schools (6 schools,) CA1 schools (6 schools), which joined the evaluation in 2001, and the CA2 schools (12 schools), which joined the evaluation in 2004. Preliminary analyses indicated that there were no substantial differences in the patterns of findings between the schools groups or between EC1 and EC2 cohorts, so statistical analyses for the whole sample only are reported, to gain the full benefits of the sample size.

Preliminary comments about the interpreting the data:

Figures 4.10 (a, b, c), 4.11 (a, b) and 4.12 (a, b) show the pupils' mean ratings for the EC and the control classes, for Year 5, Year 6 and Year 7. The mean ratings refer to raw scores that can vary between 1 (almost never) to 4 (almost always). The first point to note is the absolute level of rating for the different scales. The rank order based on the mean ratings for the scales is Challenge Avoidance (3.22 when reverse scored), Self-Efficacy (2.97), Mastery Orientation (2.94), Work Avoidance (2.83 when reverse scored), Active Learning Strategies (2.69), Curiosity (2.67) and Performance Orientation (2.52 when reverse scored). This rank order remains the same for the different age groups and for EC and control groups. Overall, this group of children evaluate themselves as generally pro-learning, as all the ratings are above 2.5, which is the mid-point on the

2-point rating scale. Statistical differences between groups must be considered from this overall perspective. Note that the Y-axis for each figure is drawn to a slightly different scale, to draw attention to the area of the scale where the differences between the groups can be best illustrated.

The second point to note is that all that the mean rating for each scale shows a downward trajectory as the children progress through school, from Year 5 to Year 7. This age/Year/schooling effect will be cautiously interpreted as it may simply mean that the children calibrate themselves differently as they get older and thus use the rating scale differently. Thus, for the purpose of the study, the main interpretation will be on the differences between the EC and control groups, *within* each Year.

The findings from the seven scales will be reported in three groups that fall naturally together; the motivational scales, the attitudes to work scales and the attitudes to learning and seeking out knowledge scales.

5.3.1 Motivational scales

Figures 4.10a, 4.10b and 4.10c shows the graphical representation for the mean ratings for the three motivational scales — self-efficacy, performance orientation and mastery orientation. Low scores on performance orientation are associated with a positive orientation to learning, while high scores on the other two scales show a positive orientation.

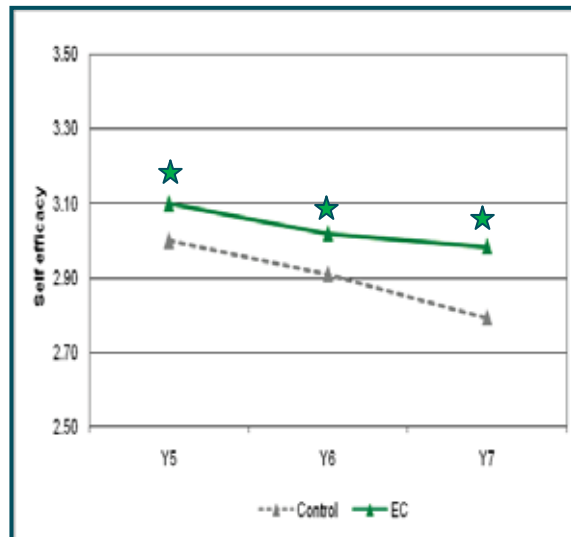
Self-Efficacy, displayed in Fig. 4.10a, shows that EC children rated themselves significantly higher than the controls in Ys 5, 6 and 7, with effect sizes of 0.16SDs, 0.17SDs and 0.31SDs respectively. The mean ratings for subgroups varied from 2.8 to 3.1.

Performance Orientation, displayed in Fig. 4.10b, shows that EC children rated themselves significantly higher than the controls in Ys 5, 6 and 7, with effect sizes of 0.25SDs, 0.12SDs and 0.24SDs respectively. The mean ratings for subgroups varied between 2.2 and 2.75.

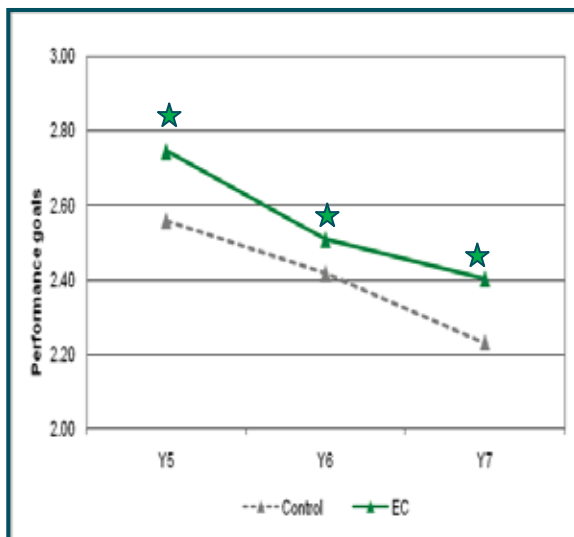
Mastery Orientation, displayed in Fig. 4.10c, shows that EC children rated themselves significantly higher than the controls in Ys 5 and 7 with effect sizes of 0.19SDs and 0.30SDs respectively. There was almost no difference between the groups in Y6. The mean ratings for subgroups varied between 2.6 and 3.2.

Figure 4.10 Learning Dispositions and Attitudes: Pupils' Mean Ratings on the Motivational Scales: EC versus Control classes, Total Sample

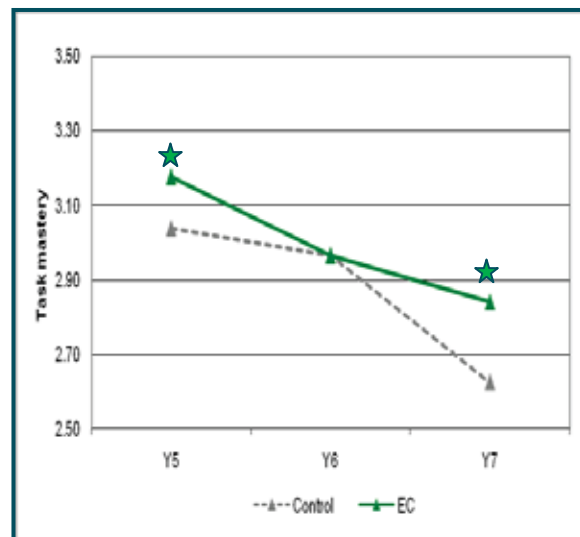
a. Self-Efficacy



b. Performance Orientation



c. Mastery Orientation



5.3.2 Attitudes to work scales

Figures 4.11a and 4.11b show the graphical representation for the mean ratings for the two attitude-to-work scales — work avoidance and challenge avoidance. Low scores on these scales predict a positive orientation to learning.

Work avoidance, displayed in Fig. 4.11a, shows that there were no significant differences between the mean ratings of the EC children compared with controls in Ys 5, 6 or 7. The mean rating for subgroups varied from 2 to 2.3.

Challenge avoidance, displayed in Fig. 4.11b, shows that EC children rated themselves performed significantly higher than the controls in Y 7 only, with an effect size of 29SDs. This is a low effect size. The mean ratings for the subgroups varied from 1.5 to 1.8.

5.3.3 Learning strategies and seeking out knowledge scales

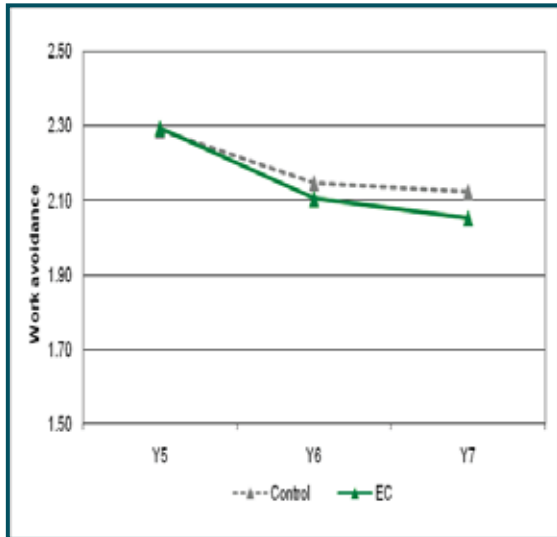
Figures 4.12a and 4.12b show the graphical representation for the mean ratings for the two scales on learning strategies, knowledge acquisition, and seeking out new knowledge — active learning strategies and curiosity. High scores on these scales are associated with positive learning dispositions and orientations.

Active learning strategies, displayed in Figure 4.12a, shows that there were no differences between the mean ratings for EC children and the control children. The mean ratings for subgroups varied from 2.5 to 2.8.

Curiosity, displayed in Fig. 4.12b, shows that the EC children rated themselves significantly above controls in Y 7 with an effect sizes of 0.19SDs. This is a low effect size. There were no differences in Y5 and Y6. The mean ratings for the subgroups varied from 2.3 to 2.75.

Figure 4.11 Learning Dispositions and Attitudes: Pupils' Mean Ratings on the Attitude to Work Scales: EC versus Control classes, Total Sample (lower ratings show better attitudes)

a. Work Avoidance



b. Challenge Avoidance

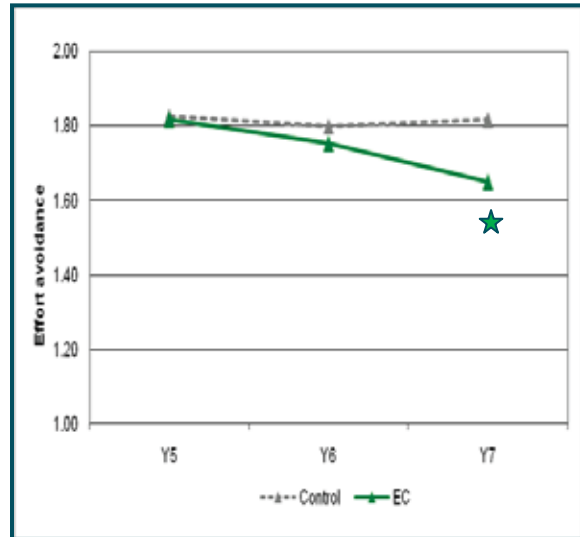
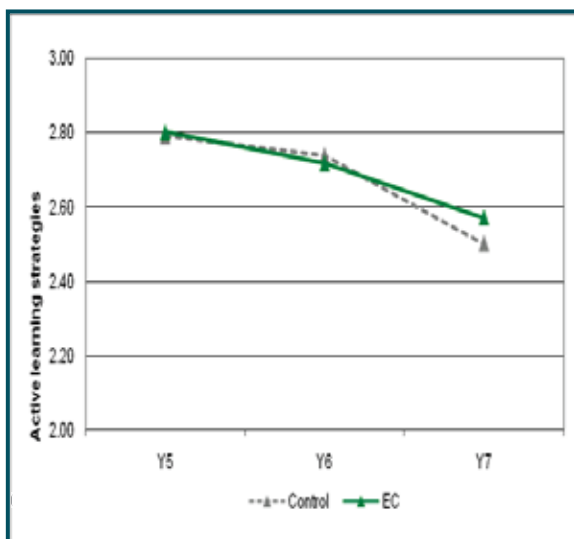
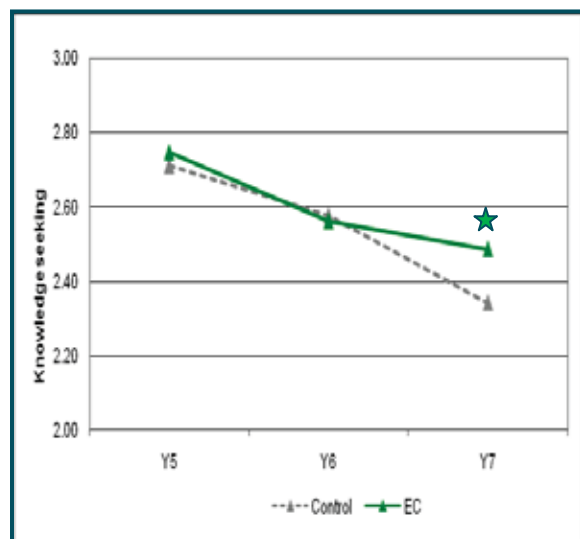


Figure 4.12 Learning Dispositions and Attitudes: Pupils' Mean Ratings on the Attitudes to Learning Scales: EC versus Control classes, Total Sample

a. Active Learning Strategies



b. Curiosity



5.4 Learning Dispositions and Attitudes: Main Effects of Predictor Variables

This section provides a summary overview of the main effects of the predictor variables for all seven multi-level regression analyses that have just been reported. The main effects in the statistical analyses operate irrespective of whether the children belongs to the EC or the control groups, and they have been statistically controlled when examining the effects of the intervention. The predictor variables included in the analyses were identical to those used in the analyses of the literacy and numeracy outcomes — gender, month-of-birth (age in class), % free school meals, school mean ability, and the effects of attending the high deprivation schools versus the rest. The more general effects of these variables on learning dispositions and attitudes have not been well researched, compared with their effects on school attainment.

Table 4.11 shows the strength of their effects — effect sizes — on the seven ALCPs scales for the whole sample only. The most striking finding is the absence of statistical main effects of these predictor variables, compared to their consistent effects on reading, mathematics and science (Table 4.5). Of the 35 effects analysed, only five were statistically significant, and there was no strong pattern related to any single variable. When effects were detected, they tended to be on the motivational indicators (self-efficacy, performance orientation, mastery orientation) rather than on the cognitive and work/challenge indicators. The biggest effects were associated with the social disadvantage predictors, though the pattern was not easy to interpret. For example, children in classes with high mean ability scores tended to rate themselves as more curious than those in classes with lower mean ability scores, and children in higher %FSM schools were more performance oriented than children in lower %FSM schools. At the same time, children in the high deprivation schools had higher mastery orientation ratings when compared to the rest of the sample. These opposing tendencies between extrinsic and intrinsic motivations occur frequently in this data set.

Surprisingly, there were no main effects of gender in the whole sample analysis. However, the preliminary analyses with the high deprivation sample showed that girls were more positively oriented to learning than boys in that sample. This effect did not emerge in the separate analysis for the CA schools. When the school samples were combined, no effect of gender was detected.

There were small effects associated with age-in-class. Older children in the class tended to have higher levels of self-efficacy than younger children, and younger children in the class tended to be more performance oriented, that is, more motivated by rewards, recognition and status, than the

older children in the class. Overall, the findings about the main effects provide hypotheses and questions for future research rather than confirm any strongly held expectations for this data set.

Table 4.11 Main Effects of Predictor Variables for All Analyses of Children’s Self-Ratings: Effects Sizes for Significant Effects ($p < .05$) Whole Sample ONLY

Predictor	Self Efficacy	Performance Orientation	Mastery Orientation	Work Avoid	Challenge Avoid	Active Learning Strategies	Curiosity
Month-of-birth over 12 months	.20 Older children rated themselves higher	.16 Younger children rated themselves more performance oriented	ns	ns	ns	ns	ns
Gender	ns	ns	ns	ns	ns	ns	ns
School mean ability over the full range of school mean ability, high to low	ns	ns	ns	ns	ns	ns	.73 Higher school mean ability rated themselves more curious
School’s percentage of free school meals over the full range of % FSM	ns	.64 Higher %FSM rated themselves more performance oriented	ns	ns	ns	ns	ns
High Deprivation Schools versus the remaining schools	ns	ns	.53 High deprivation schools rated themselves more mastery oriented	ns	ns	ns	ns

5.5 Interim Summary on Children’s Self-Evaluations

- The inter-correlations between the seven ALCPs scales confirmed that the empirical relationships between the scales were as theoretically predicted. The positive oriented

learning scales were positively inter-correlated, and negatively correlated with the negatively oriented learning cluster, which were themselves positively inter-correlated. Performance orientation stood out as not entirely aligned but it was more strongly associated with the negative cluster than with the positive cluster.

- Overall, the children's self-evaluations showed that they held relatively positive learning dispositions and attitudes to work, and this was the case whether they were in EC classes or in the control classes. For example, on average, the pupils tended to say that they engaged in positive learning activities - such as using active learning strategies, seeking out new things to learn, being motivated by interest — 'often' (3) or 'sometimes' (2); and they tended to say that they avoided challenge or making an effort by guessing or skipping difficult work only 'sometimes'(2) or 'almost never' (1). In general, the pupils showed a consistent pro-learning profile — with the exception of their rating on the performance oriented scale.
- The impact of being in an EC class in the earlier years had statistically significant positive effects on the pupils' learning dispositions and attitudes as they progressed into Key Stage 2; this was particularly true as the children got older. In Y7, they rated themselves as more pro-learning than the control classes on 4/7 scales, and on two other scales, their mean scores were higher, but not significantly so. For example, Y7 EC pupils had higher self-efficacy, high mastery orientation, were more curious and accepted more mental challenge than the control groups. The effects sizes were small but consistently in the positive direction.
- Positive effects emerged on 2/7 scales for Y5 EC pupils, who had higher self-efficacy and higher mastery orientation than the control classes. Self-efficacy remained higher for the EC children compared to the control as the children progressed through Key Stage 2. However, the positive effect of the EC on mastery orientation — wanting to learn out of interest and or to improve — disappeared in Y6 but re-emerged again in Y7, where the difference was mainly due to a severe downturn in the ratings in the control classes.
- EC pupils also had significantly higher scores than control pupils on one of the negative learning oriented scales — performance orientation. This shows that the reasons the EC children said that they wanted to do their work well was to achieve good marks and be thought 'smart' by their teachers, their friends and their family. It is easy to see why apparently more motivated children would rate themselves high on this scale. However, high scores on this scale were also positively related to avoiding challenge, avoiding making

an effort, and just 'getting the work done'. Higher scores on performance orientation is a somewhat mixed blessing.

- The predictor variables that were included in all the analyses had a more limited impact on learning dispositions and attitudes than they had on attainment, indicating that there is not likely to be straightforward correlations between the children's self-evaluations and the more traditional psychometric measures of performance and attainment. There were no overall effects of gender on any of the scales, though girls tended to be more pro-learning in the high deprivation sample. In addition, the effects of social disadvantage were difficult to interpret as it appeared to be associated with both a positive and a more negative orientation to learning — higher mastery orientation and higher performance orientation.

5.6 Pro-Learning Dispositions: Teachers' Ratings of the Children's Social and Behavioural Competence the Classroom

As outlined in Section 5.1, one of the primary goals of the EC was to improve young children's motivation through removing the experience of persistent failure for the child in the early years. This was expected to improve children's confidence and, consequently, their eagerness to succeed in their work as they progressed through school. The interactional pedagogy associated with the play-based curriculum was intended to have a positive effect on children's social and emotional development, thus improving their skills in peer relations and in social adjustment more generally. Finally, attention and behaviour problems are known to adversely affect learning (e.g., Duncan, Dowsett et al., 2007) and thus improving children's attention was also one of the goals of the EC.

Early attempts to use a well-known but lengthy inventory on children's adaptive and social behaviour in the classroom proved to be unsuccessful because of the burden it put on teachers to complete. Hence a short 6-item scale was constructed by the research team to cover the main areas of social and behavioural competence which were deemed to be associated with the goals of the EC. Every year since 2001-02, teachers were asked to rate each child in their class on these 6 items, using the four point scale, as outlined in Table 4.12. Different teachers rated the children each year.

Scores on the scale can vary from 24 (4 x 6 items) to 6 (1 x 6 items); the attention and behaviour items are reverse scored. Analysis of outcomes has shown that these six items form a highly cohesive scale; children scoring high on one of the first four scales in Table 4.11 tend to score high on the others and also tend to show fewer behaviour and attention problems. Thus we are justified

in calculating a composite score, and we have labelled the scale the **pro-learning dispositions** teacher rating scale (Cronbach's alpha = 0.81). High scores indicate a higher pro-learning disposition.

Table 4.12 Pro-Learning Dispositions: Items used for the Teachers' Rating Scale

Scale	Very seldom applies to this child (1)	Applies a little to this child (2)	Applies to this child to a moderate degree (3)	Almost always applies to this child (4)
Makes an effort with work				
Has confidence in work				
Gets on well with other children				
Seems happy and well adjusted				
Has attention problems				
Has behaviour problems				

Preliminary analyses with the ratings:

Preliminary analyses showed that the teachers tended to rate most children very highly, that the scores were not normally distributed, and were positively skewed, thus limiting their statistical analyses. In order to make the most of the data, it was decided to analyse the ratings in the following way:

- To calculate an average of the teacher ratings that were available for each child. Although the longitudinal perspective was forfeited, the average rating did give a 'global' perception of the pro-learning dispositions of the children over their primary school years. These were called the composite teacher ratings for children's pro-learning dispositions.
- Because of the skewness¹⁷ in the data, it was not valid to use mean scores for group comparisons. Instead the scores were categorised into four groups based on the following, somewhat arbitrary, cut-off scores: 6-12.5 points, 12.6-15 points; 15.1-20 points, 20+ points. Then the percentages of children with scores within each band were calculated. Characterizations of the social and behavioural competence of these different groups of children were then constructed and how their learning dispositions and attitudes might manifest themselves in the classroom. We created 'labels' for each characterization, merely

¹⁷ This means that the data are not normally distributed. Children tend to cluster towards the high rating end of the scale.

as shorthand to describe the composite ratings. It is important to note that these labels were created by the research team, not by the teachers.

The 'problem' child — rated between 6 - 12.5 on the composite teacher ratings. These children were rated from 1 to 2 out of 4 points on each scale, and teachers' ratings across the years would have to be consistently in this range if the composite score was below 12.5. Consistent ratings at this level indicate that the children have substantial attention problems, show disruptive behaviour, do not get on well with the other children, and show little interest in, or effort with, school work. Such children are likely to take up a lot of the teachers' time. Examples of behaviours of this kind which were observed in early years classrooms during the structured classroom observations included leaving the classroom without permission and refusing to sit on the carpet with the other children.

The 'difficult' child — rated between 12.6 - 15.0 on the composite teacher ratings. These children were rated mostly 2, with some ratings of 3, out of 4 points on each scale. While not as consistently disruptive or worrying as the 'problem' child, the 'difficult' child would still take up a disproportionate amount of the teacher's time and would have the potential to affect class discipline and learning adversely.

The 'good' child — rated between 15.1 - 20 on the composite teacher ratings. These children scored a mixture of 3 and 4 points on each scale on average and would be expected to need extra attention from the teacher relatively rarely.

The 'ideal child' — rated above 20+points on the composite teacher ratings. These children were rated predominantly 4 on each scale on average, were interested and confident about their work, made an effort, got on well with other children, and showed very rare behaviour or attention problems.

Separate profiles were plotted for the high deprivation schools and for the CA, as different patterns were discovered in each group. Comparisons were made between EC and control classes within each group of schools. Figure 4.13 illustrates the profile for EC (N=1059) and control (N=671) children in CA schools. About 92-93% of the children were classified as 'good' or 'ideal' and the remaining 7-8% as 'difficult' or 'problem'. Within the 93%, there were twice as many 'ideal' children versus 'good' children; the ratio was similar for EC (64.3% ideal versus 28.7% good) and Control (61.2% ideal versus 31.1% good) samples. In general, the vast majority of the children in the CA schools were consistently viewed by their teachers as being well motivated, taking an interest in their work, making an effort and having few behaviour and attention problems. Nevertheless, Figure

4.13 also shows that EC children's ratings were distributed more towards the extremes of the scale compared with those of controls. For example, there were more 'problem' children and more 'ideal' children in the EC classes than in the control classes, and fewer EC children compared to controls in the other two categories. The differences between the curriculum groups were very small, yet they do show that the EC had a slight polarising effect on teachers' ratings of the children in their classrooms. In these CA schools, EC classrooms had more children who had very high pro-learning dispositions ('ideal') but also more children who were more disruptive and hard to teach ('problem'), albeit a very small number.

Figure 4.13 Composite teacher ratings of pro-learning dispositions in children in CA schools: EC children versus controls

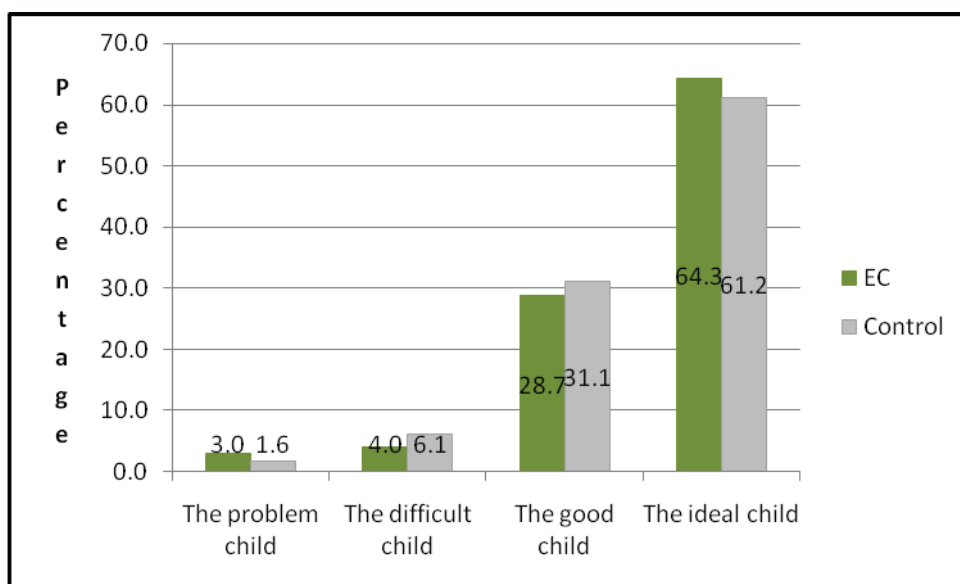
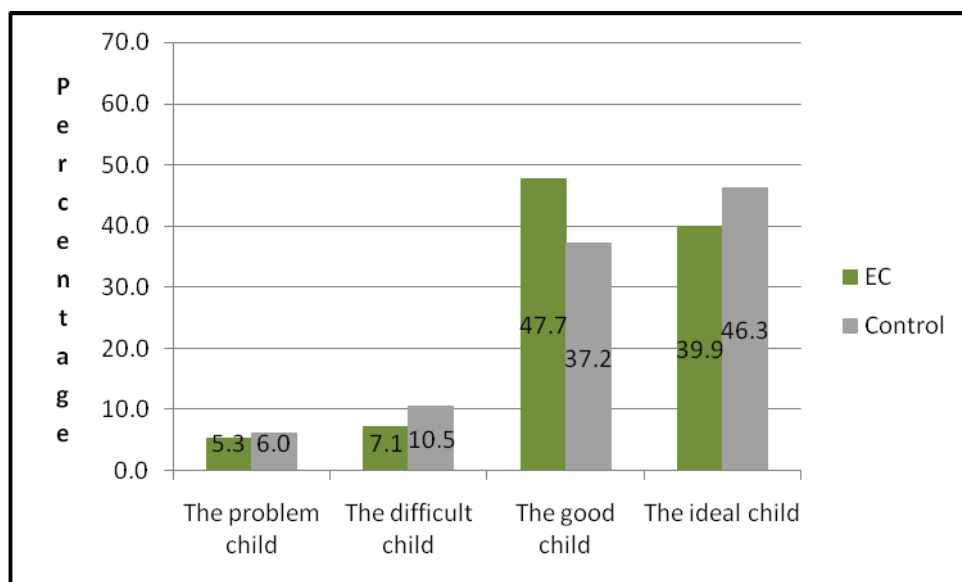


Figure 4.14 illustrates the profile for EC (N=283) and control (N=315) children in high deprivation schools. The first point to notice is that the teachers' ratings of the children were lower than in the CA schools, irrespective of the curriculum being followed: 85-86% were categorised as 'good' or 'ideal', and the remaining 14-15% were rated as 'difficult' or 'problem'. Within the 85-86%, the ratio of 'ideal' to 'good' children was approximately 50/50. Similar to the CA schools, the vast majority of the children were rated consistently by their teachers as being well motivated, making an effort, interested in their work and with few behaviour and attention problems. But a more substantial minority (14-15%) than in the CA schools (7-8%) were evaluated as having more school-related motivational, attention and behaviour problems. However, Figure 4.14 also shows that the EC seemed to have an uplifting effect on the frequency of the teachers' ratings of more difficult children. For example, the teachers rated fewer EC children as 'problem' or 'difficult' compared to

the control classes; taken together the percentage was reduced from 16.5% (10.5% 'difficult' + 6.0% 'problem' in control classes) to 12.4% (7.1% 'difficult' + 5.3% 'problem') — an important shift in terms of the numbers of likely 'hard to teach' children in such classrooms. However, there were also fewer EC children rated as 'ideal' compared to controls (39.9% versus 46.3%). In contrast to the teachers in the CA schools where the EC seemed to have a slight polarising effect, in the high deprivation schools, the EC seemed to have a more centralising effect, with greater numbers of children being rated as 'good' and fewer children being rated as 'problem', 'difficult' or 'ideal'.

Figure 4.14 Composite teacher ratings of pro-learning dispositions in children in high deprivation schools: EC children versus controls



Brief Summary of the Findings for the Composite Teachers' Rating

- The composite teachers' rating of the children's pro-learning dispositions showed that the vast majority of the children were characterised as 'good' or ideal' children (between 84%-93%). This was true both for EC and control classes, and for CA schools and high deprivation schools.
- In high deprivation schools, almost twice the numbers of children (14-15%) were categorised as 'problem' or difficult' compared to CA schools (7-8%). In addition, fewer children were in the 'ideal' category compared to the CA sample (~43% versus 62%), signifying that deprivation had a negative effect on pro-learning dispositions, as rated by teachers.
- There were very small differences between the percentages of children in the different categories for EC and control classes. Nevertheless, the patterns revealed a complex

dynamic between teachers' perceptions of children's pro-learning orientations, participation in EC versus control classes, and school groups. In the high deprivation schools, compared to the control classes, the teachers' ratings indicated that there were more 'good' children in EC classrooms, fewer 'problem' and difficult' children, but also fewer 'ideal' children. In the CA schools, the teachers' ratings seemed to say that there were more 'problem' children in EC classrooms but also more 'ideal' children.

5.7 Learning Dispositions and Attitudes: Relationship between Children's Self-Evaluations and Composite Teachers' Ratings

In order to check the validity of the ratings' measures and to see if there was agreement between how the teachers rated the pupils and how the pupils rated themselves, simple correlations were run between the composite teachers' rating of the children's pro-learning dispositions and the seven ALCPs children's self-rating scales at Year 6. Table 4.13 shows the correlations.

Table 4.13 Correlations between the Children's Self-Evaluations of their Learning Dispositions and Attitudes in Year 6 and the Teachers' Composite Ratings of their Pro-Learning Dispositions (N=1110)

Variable Name	Self-Efficacy Y6	Performance Orientation Y6	Mastery Orientation Y6	Work Avoidance Y6	Challenge Avoidance Y6	Active Learning Y6	Curiosity Y6
Composite Teacher Rating Pro-Learning Disposition	.193**	-.130**	.063*	-.204**	-.339**	.089**	.084**

* $p < .05$ ** $p < .01$ (very small correlations achieve significance because of the large sample size)

The correlations were all in the expected directions and were statistically significant. Very small correlations tend to achieve statistical significance when sample sizes are very large, so only the four largest correlations will be interpreted. The teachers' pro-learning dispositional ratings were positively correlated with the children's ratings of their own self-efficacy, although the strength of the correlation was low. Those children who were rated as more confident, who made an effort with their work, and who were able to get on with other children, also held beliefs that they could positively affect their future learning through their own efforts.

Teachers' ratings were negatively correlated with Challenge Avoidance, Work Avoidance and Performance Orientation. This again confirms that there was agreement between the teacher and

child perspectives. Those children who scored high on the composite teacher ratings, and who were viewed by the teacher as more pro-learning, also rated themselves as less likely to avoid difficult work, less likely to avoid making an effort, to skip work or just to get work done as quickly as possible. The correlation between the teachers' rating and Challenge Avoidance was remarkably high for these types of measures. Note also that scores on Performance Orientation were negatively correlated with high teacher ratings.

In conclusion, the pattern of correlations show a moderate degree of congruence between the teachers' and the children's perspectives and indicate that the ratings have some validity in terms of interpersonal and intrapersonal perceptions.

5.8 Relationships between Learning Dispositions/Attitudes and Scholastic Attainment

To further the analyses about the predictive validity of the teacher and child ratings, simple correlations were conducted between the composite teacher ratings, the seven ALCPs scales for Y6, and the PIPS mean scores for Y6 Mathematics, Y6 Reading and Y7 Science. Table 4.14 shows the correlations.

The teachers' ratings were more highly correlated with the school attainment measures than were the children's self-ratings. Nevertheless, 4/7 ALCPs scales predicted attainment in consistent ways. The highest negative correlation was between Challenge Avoidance and the attainment measures. Those pupils who rated themselves as likely to avoid mental challenges, to avoid difficult work, and to 'give up', were likely to have low scores on the school-related tests. Performance Orientation as a motivational style was negatively correlated with attainment. Despite the desire to get 'good' marks and to appear 'smarter' than other children in the class, the children who scored high on the scale did not appear to have developed the work habits to achieve their goals.

Of the four positively oriented learning scales, only Self-Efficacy was positively correlated with attainment; the remaining scales, Mastery Orientation, Active Learning Strategies and Curiosity showed correlations that were close to zero. This does not mean that these motivational styles and learning strategies are unimportant — they may acquire stronger motivational force at a future date. Beliefs about motivation, and the reasons for putting in an effort, are often developed and held before they become sufficiently powerful to translate into actions. Also, the educational environment shapes motivations and creates, or fails to create, learning opportunities for developing alternative motivational pathways.

Table 4.14 Correlations between the Teachers' Composite Rating, the Children's Self-Evaluations of their Learning Dispositions and Attitudes Y6, and PIPS mean scores in Mathematics Y6, Reading Y6 and Science Y7

Variable Name	PIPS Mathematics Y6	PIPS Reading Y6	PIPS Science Y7
Teacher Composite Rating			
Pro-Learning Disposition	+.494 ** (N=1033)	+.482** (n=1031)	+.393** (n=759)
Children's Self-Evaluations			
Self Efficacy Y6	.203** (N=859)	.174** (N=856)	.174** (N=548)
Performance Orientation Y6	-.179** (N=859)	-.203** (N=856)	-.172** (N=548)
Mastery Orientation Y6	-.022 (N=859)	.003 (N=856)	-.065 (N=548)
Work Avoidance Y6	-.141 ** (N=859)	-.145** (N=856)	-.121 ** (N=548)
Challenge Avoidance Y6	-.293** (N=858)	-.301** (N=855)	-.225** (N=548)
Active Learning Y6	.038 (N=859)	.013 (N=856)	.000 (N=548)
Curiosity Y6	.067 (N=859)	.071* (N=856)	.018 (N=548)

* $p < .05$ ** $p < .01$ (very small correlations achieve significance because of the large sample size)

It is striking that the scales that most predicted the children's school attainment were about avoiding negative states (guessing, skipping hard work, not thinking too hard) rather than by approaching positive states (learning new things, doing work that is interesting, improving skills and understanding). It was these more positive orientations and learning dispositions that the EC seemed to increase rather than decreasing the children's dependence on avoidance motivations and work habits.

6. Overview of Findings and Conclusions

6.1 Preliminary points to help interpretation

- The EC had broad aims and expectations not only for the immediate learning experiences of the children in the first years of primary school but also for creating positive learning foundations that would sustain the children's progress in school over the longer term. Thus, a longitudinal study was designed to follow the children's educational development at least up to the end of their primary schooling. For the second phase of the evaluation, the size of the longitudinal sample was substantially increased through the addition of 12 new schools, and a second cohort of children who had experienced the EC. Thus, the findings described in this report confirm (or do not confirm) outcomes that were previously reported, extend substantially the time frame of the analyses, as well as describe new outcomes that have not been previously reported.
- Although a substantial focus of this report was on literacy and numeracy outcomes for pupils, it is important to remember that the EC was not primarily a reading and mathematics intervention. Nonetheless, as noted earlier in Section 1.3, adopting a more informal and play-based approach in the early year did change how reading and mathematics were taught in Year 1, Year 2 and beyond. In addition, we know from previous PIPS testing at the end of Y1 and Y2, that the children's reading and mathematics scores were poorer than the control group — which was expected given the changes in the pedagogical approach. We also know that the children progressed very well when more formal approaches to reading and mathematics were adopted in Y3 and their PIPS scores matched the controls by the end of Y4, at least in the samples that were reported at the End-of-Phase 1. Nevertheless, it is important to track the longer term literacy and mathematics development of children who experienced in the EC in their early years, as these skills are considered gateways to learning across-the-curriculum as the children get older.
- More central to the aims and goals of a play-based curriculum was the development of children's learning dispositions, their attitudes and orientations to learning. The more exploratory and self-directed learning opportunities provided by a play-based curriculum were expected to stimulate children's curiosity, to increase their motivation and eagerness to learn, and to help them become more confident and independent learners in the long run. Observations at the classroom level in the early years confirmed that these attributes were more obviously part of the children's classroom experience than they had been in the pre-existing curriculum. It is important to evaluate the extent to which these early classroom experiences

have positively influenced children's learning orientations, their motivations for learning, their work habits and learning strategies, and the beliefs that they held about themselves as learners. For the first time, pupils' self-evaluations as they moved through Key Stage 2 were included as outcome measures. These were augmented by teachers' ratings of the pro-learning dispositions of the children which had been collected throughout the life of the project.

- Because of how the EC developed and how different school groups became involved, schools had different levels of funding, different training, and different levels of in-school and out-of-school support — all of which influenced the implementation of the EC in the schools, and may have affected the outcomes of the pupils. In the introduction to this report (Section 1.3), the question was raised whether it was appropriate to treat the EC as a unitary curriculum and conduct the analyses on the total sample of 24 schools, or where it was more appropriate to treat the school groups as complete different samples. As the analyses progressed, it became clear that the comparison between the high deprivation schools versus the remaining sample of schools outside Belfast proved to be the most useful for interpreting the pupil outcomes, especially for literacy and numeracy. Differences between the school groups were not so obvious for the analyses of the children's learning dispositions and attitudes.
- Interviews with the school principals confirmed that the EC took time to 'settle' in their schools and revisions were made in the first few years of implementation. Also, teachers became more experienced with the new pedagogy and became more confident about when to 'move children on', which had been a source of uncertainty for some teachers during the first year of implementation. The design of the study gave the opportunity to examine these 'settling in' processes by comparing the cohorts of EC pupils during the first and second years of implementation. Differences between cohorts proved useful for interpreting the pupils' outcomes, particularly in the high deprivation schools. It should be noted that revisions and adaptations to the curriculum continued beyond the first 2 years, and 'settled' versions of the EC may not have emerged in some schools for a number of years.

6.2 The impact of the EC on Mathematics, Reading, Science and Writing over Time

- At the most general level, over all the statistical analyses, the EC had no long term¹⁸ impact — positive or negative — on pupils' mathematical and reading attainments in 77% of comparisons that were made between EC classes and control classes. In 20% of the comparisons, the control classes were better than the EC classes, and in 3% of the comparisons, the EC classes were better than the control classes. Reading was more negatively affected by the EC than mathematics. The EC had a positive effect on writing in 50% of the comparisons conducted (in Years 4 and 7 only). The significant effects were all statistically 'small' (in terms of effect sizes), approximately the same size as the difference between boys and girls on the same variables.
- For the six High Deprivation schools, the outcomes were more positive than the general patterns outlined above, with 20% of their comparisons showing statistically significant positive effects of the EC, 8% showing negative effects, and 72% showing no effects. EC classes performed significantly better than the controls on mathematics (EC2 Year 6), reading (EC2 Year 6), science (Year 7), and writing (Years 4 and 7). The second cohort performed consistently better than the first cohort, and several local factors may have had a positive influence. These schools stood out from the rest in terms of the high levels of disadvantage of the schools' intakes (50-76%FSM); their baseline scores were well below average and they were very similar in terms of the low mean ability of their pupil intake. The curriculum had originated from these schools; they had a coherent view of its purpose and had a high sense of ownership. They had higher levels of funding and support than the other schools. Their training was intensive; school principals and the teachers expressed general satisfaction with the support that they received.
- For the 18 Contrasting Areas schools (CA1 and CA2 combined), the outcomes were less positive than the general patterns outlined above. Sixty-five percent of the comparisons between EC and control classes showed no significant differences between them; in 28% of the comparisons, the control classes were better than the EC classes, and in 7% of the comparisons the EC classes were better. Reading was more negatively affected than mathematics. Higher/middle ability groups were more negatively affected than lower ability children. The EC had a positive effect on writing in Years 4 and 7. There were inconsistent differences between the first and second EC cohorts, sometimes the second cohort was better than the first, and sometimes it was the other way around. This

¹⁸ Comparing EC versus control classes from Year 3 to Year 7.

sample of schools was drawn from ELBS outside Belfast and was more socially advantaged than the High Deprivation schools — the distribution of %FSM ranged from 0-30. There was wide variation between the schools in terms of the mean ability of the school's intake. The schools had introduced the EC under less favourable funding and support circumstances than the High Deprivation schools. Their training was distributed across the 4 ELBs, and was probably less cohesive than that experienced by the High Deprivation schools, resulting in different interpretations. Responses to the training by teachers and the school principals varied from being generally very satisfied to being highly critical. There were unresolved tensions in some of the schools about teaching reading, particularly in the first year of implementation.

- Important findings about factors that influenced the educational development of the pupils were identified through the general patterns in the analyses which were true irrespective of the curriculum. Many of these statistically significant effect sizes were greater than those found in the EC versus control groups comparisons.
 - There was a widening gap between higher and lower ability children as they progressed through primary school. Higher ability children tended to perform above their baseline expectations and lower ability children tended to perform below their baseline expectation. The effect was more pronounced for reading than for mathematics.
 - Girls consistently outperformed boys in reading and writing; boys performed better in science (Y7 only); there were no differences between boys and girls in mathematics.
 - Children who were young for their age in class were disadvantaged compared to older children. The effect was stronger for mathematics than for reading (both PIPS measures), but strongest of all for writing.
 - Social deprivation had the strongest effect on pupils' outcomes — for mathematics, reading, compositional writing and science.

6.3 The impact of the EC on Pupils' Learning Dispositions and Attitudes over Time

- The pupils learning dispositions and attitudes were assessed through the children's self-ratings and through teachers' ratings.
- Differences between school groups did not emerge as an important factor for interpreting the outcomes for learning dispositions and attitudes. Where school group differences were detected, they tended to be in the teachers' ratings rather than in the children's ratings.

Children's Self-Ratings:

- The impact of being in an EC class in the earlier years had statistically significant positive effects on the pupils' learning dispositions and attitudes as they progressed into Key Stage 2; this was particularly true as the children got older. In Y7, they rated themselves significantly more pro-learning than the control classes on 4/7 scales, and on two other scales, their mean scores were higher, but not significantly so. Compared to control classes in Year 7, EC pupils had stronger beliefs that they could influence their future learning through their own efforts; they were more motivated through interest and the desire to improve their knowledge and skills; they were more curious; and were prepared to accept more mental challenge and take on more difficult work. The effects sizes were small but consistently in the positive direction. For Years 5 and 6 the effects were less consistent across the scales but also in a positive direction.
- The emergence of this consistent positive effect in Y7, compared to Years 5 and 6, probably means that these learning dispositions and attitudes become more salient for children as they get older and/or that Year 7 provides more open learning opportunities for children to exercise them.
- EC pupils had significantly higher scores than control pupils on one of the negative learning oriented scales — performance orientation. This scale measures extrinsic motivation and indicates that, more so than the control children, the reasons that EC children wanted to do their work was to achieve 'good' marks and to be thought 'smart' by their teachers, their friends and their family. However, high scores on this scale were also positively related to avoiding challenge, avoiding making an effort, and just 'getting the work done'. There is much debate in the research literature about the meaning and interpretation of this motivational style, and the reasons for the increase in the EC classes are difficult to interpret.
- Overall, the children's self-evaluations showed that they held relatively positive learning dispositions and attitudes to work, and this was the case whether they were in EC classes or in the control classes. For example, on a four-point scale, the pupils rated the frequency with which they engaged in positive learning activities between (2)'sometimes' and (3)'often'; and they tended to say that they avoided challenge or making an effort by guessing or skipping difficult work only 'sometimes'(2) or 'almost never' (1).

Teachers' Ratings

- The teachers' composite ratings were constructed from the annual ratings completed by the teachers for each child in EC and control classes, and provided an 'accumulating' perspective on the children's pro-learning dispositions across the years.
- The composite teachers' rating of the children's pro-learning dispositions were very high and the vast majority of the children were categorised as 'good' or 'ideal' children¹⁹ (between 84-93%). This was true both for EC and control classes, and for CA schools and high deprivation schools.
- In high deprivation schools, almost twice the numbers of children (14-15%) were categorised as 'problem' or 'difficult' compared to CA schools (7-8%). In addition, fewer children were in the 'ideal' category compared to the CA sample (~43% versus 62%), signifying that deprivation had a negative effect on pro-learning dispositions, as rated by teachers.
- There were very small differences between the percentages of children in the different categories for EC and control classes. Nevertheless, the patterns revealed a complex dynamic between teachers' perceptions of children's pro-learning orientations, participation in EC versus control classes, and school groups. In the high deprivation schools, compared to the control classes, the teachers' ratings indicated that there were more 'good' children in EC classrooms, fewer 'problem' and 'difficult' children, but also fewer 'ideal' children. In the CA schools, the teachers' ratings seemed to say that there were more 'problem' children but also more 'ideal' children.

Children's Self-Ratings, Composite Teachers' Ratings and Scholastic Attainment

- Composite teachers' ratings and the children's self-ratings were moderately correlated and in the expected directions. Teachers' ratings were positively correlated with the children's ratings of their own self-efficacy, and negatively correlated with challenge avoidance, work avoidance, and performance orientation. These relationships show a degree of congruence between the teachers' and the children's perspectives and indicate that the ratings have some validity in terms of interpersonal and intrapersonal perceptions.
- In the high deprivation schools that had high levels of funding and support and the most cohesive approach to training, the EC had several positive effects on Key Stage 2 pupils' outcomes in reading, mathematics, science and writing, illustrating that a play-based early years curriculum is consistent with enhanced attainment.

¹⁹ These labels were assigned to categories of children by the research team, not by the teachers.

- In the remaining CA schools, where levels of support were not so high, and where training was more distributed and less cohesive, the EC had a more mixed longer-term impact, particularly for reading. In the majority of cases, the pupils' outcomes were similar to the controls, showing that a play-based curriculum can at least match the attainment levels achieved by the control classes. In a minority of cases, the EC pupils' reading outcomes were poorer than the controls, particularly so for higher ability children.
- Issues related to resources, support and training, and a cohesive framework both within early years practice and further up the school, seemed important for successful implementation and outcomes. A significant lesson for schools is to recognise that the impact of introducing a play-based curriculum cannot be confined to the early years.
- This story of the longer-term reading outcomes in the context of the EC clearly illustrates the 'not-fully-resolved tensions' between adopting a play-based and informal approach in the early school years and, at the same time, using the best research evidence and the most effective methods for teaching reading to young children in the context of a language-rich curriculum. These tensions need to be resolved — and can be resolved — without compromising the general principles of a play-based curriculum which appears to have such a wide range of other positive effects on children's educational development and well-being.

IMPORTANT NOTE:

An End-of-Phase 2 Summary Report provides an overview of the findings for all four strands of the evaluation. In that report, the findings for the pupils' outcomes are considered within the context of the other aspects of the evaluation.

References

- Alexander, P. A. & Murphy, P. K. (1998). Profiling the differences in students' knowledge, interest, and strategic processing. *Journal of Educational Psychology*, 90, 3, 435 – 447.
- Cohen, J. (1969). Effect sizes. In *Statistical Power Analysis for the Behavioral Sciences*, 1st Edition, Lawrence Erlbaum Associates, Hillsdale (2nd Edition, 1988).
- Duncan, G. J. Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagini, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, 43, 6, 1428 – 1446.
- EPPE (2004). Technical Paper 11: *Report on the continuing effects of pre-school education at age 7*. ISBN: 085473 602 6. Available at <http://eppe.ioe.ac.uk/eppe/eppepubs.htm>
- Hogan, A.E., Scott, K.G. & Bauer, C.R. (1992). The Adaptive Social Behaviour Inventory (ASBI): A new assessment of social competence in high-risk three year olds. *Journal of Psychoeducational Assessments*, 10 (3), 230 – 239.
- McCombs, B. L. (1997). Background and research to date on the Learner-Centered Battery. Paper presented in the symposium, 'Self-Assessment and Reflection Models for Teacher Professional Development.', conducted at the annual meeting of the American Psychological Association, Chicago, August 1997.
- McCombs, B. L. & Lauer, P. A. (1997). Development and validation of the Learner-Centered Battery: Self-assessment tools for teacher reflection and professional development. *The Professional Educator*, 20, 1, 1 – 21.
- McGuinness, C. (2006). Building thinking skills in thinking classrooms. *Research Briefing for the Teaching and Learning Research Programme*. No. 18. September 2006.
- McGuinness, C., Sproule, L., Trew, K. & Walsh, G. (2009). *The Early Years Enriched Curriculum Evaluation Project (EYECEP). End of Phase 2, Report 1: Overview: Evaluation Strategy and Curriculum Implementation*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).
- NFER. (2005). *A review of Literacy Initiatives to inform advice on the Foundation Stage*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).
- NFER. (2009). International thematic probe: The influence of relative age on learner attainment and development. Report commissioned by QCA. NFER reference QSB.
- Pintrich, P. R. (2003). Multiple goals and multiple pathways in the development of motivation and self-regulated learning. *British Journal of Educational Psychology*, Monograph series 11, No. 2. 'Development and Motivation'.
- PIPS. (2001). PIPS technical paper. Available to download from the CEM centre, University of Durham, at <http://www.cemcentre.org/pips/pips>
- Sheehy, N., Trew, K., Rafferty, H., McShane, E., Quiery, N., & Curran, S. (2000). *The Greater Shankill Early Years Project: Evaluation Report*. Report produced for the Greater Shankill Early Years Project and CCEA.
- Sproule, L., Murray, R., Spratt, V., Rafferty, H., Trew, K., Sheehy, N., McGuinness, C. & Wlsh, G. (2001). *A Review of Recent Research Relevant to the Early Years Enriched Curriculum Project*. Privately produced review for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).
- Sproule, L., Rafferty, H., Trew, K., Dudgeon, K., Sheehy, N., McGuinness, C. & Walsh, G. (2001). *The Early-years Enriched Curriculum Evaluation Project: First Year Report*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

Sproule, L., Rafferty, H., Trew, K., Walsh, G., McGuinness, C., Sheehy, N. & O'Neill, B. (2002). *The Early-years Enriched Curriculum Evaluation Project: Second Year Report*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

Sproule, L., Rafferty, H., Trew, K., Walsh, G., McGuinness, C. & Sheehy, N. (2003). *The Early-years Enriched Curriculum Evaluation Project: Rationale for the Research Design*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

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

Sproule, L., Rafferty, H., Trew, K., Walsh, G., McGuinness, C., Sheehy, N. & O'Neill, B. (2004). *The Early-years Enriched Curriculum Evaluation Project: Interim Fourth Year Report*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

Sproule, L., Rafferty, H., Trew, K., Walsh, G., McGuinness, C., Sheehy, N. & O'Neill, B. (2005). *The Early-years Enriched Curriculum Evaluation Project: Final Report Phase 1*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

Sproule, L., Harvey-Smith, D., McGuinness, C., Trew, K., Rafferty, H., Walsh, G., Sheehy, N. & O'Neill, B. (2007). *The Early-years Enriched Curriculum Evaluation Project: Year 5 Report* (Belfast: School of Psychology, Queen's University Belfast).

Sproule, L., McGuinness, C., Trew, K. & Walsh, G. (2009). *The Early Years Enriched Curriculum Evaluation Project (EYECEP). End of Phase 2, Report 4: Outcomes for Children*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

The Education [Northern Ireland] Order (2006). Published by the UK Office of Public Sector Information. Available at http://www.deni.gov.uk/22-ppa-final_draft_education_bill_23_november_2005.pdf

Trew, K. McGuinness, C., Sproule, L. & Walsh, G. (2009). *The Early Years Enriched Curriculum Evaluation Project (EYECEP). End of Phase 2, Report 3: Parents' Views and Perceptions of the Enriched Curriculum over Time*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

Tymms, P. B., Merrell, C. & Henderson, B. (1998). *Baseline Assessment and Progress during the First Three Years at School*. Paper presented at the conference of the British Psychological Society, Brighton, March 1998.

Tymms, P. B., Jones, P., Albone, S. & Henderson, B. (2009). The first seven years at school. *Educational Assessment, Evaluation and Accountability* **21**(1): 67-80.

Walsh, G., McGuinness, C., Sproule, L. & Trew, K. (2009). *The Early Years Enriched Curriculum Evaluation Project (EYECEP). End of Phase 2, Report 2: Inside EC Classrooms and Schools, Children, Teachers and Principals*. Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).

Wechsler Objective Language Dimensions. (1996). *WOLD Test of Written Expression*. London: The Psychological Corporation.

Wechsler Objective Language Dimensions. (1996). London: Harcourt Brace.

Appendix

Figure 3 reproduced from Report 1. Timeline of data collection for the three groups of schools and external factors that might have affected outcomes

