Using assessment data
to support the learning
of young pupils in
four Kent primary schools

Sally Patricia Hyne
A dissertation submitted in partial fulfilment of the Doctorate in Education
of the University of Greenwich
March 2007

ACKNOWLEDGEMENTS

I should like to thank the following people, without whom this thesis would not have been written with such energy and enthusiasm:

- ❖ Dr. Neil Hall and Maggie Gravelle, my supervisors on the EdD programme at the School of Education, University of Greenwich, who have supported and encouraged me throughout the course of my study

- ❖ The lecturing team on the EdD programme, who helped me move towards new knowledge and understanding

- ❖ Mary Stiasny, who inspired me with confidence

- ❖ The Governing Body of my school, for allowing me to have study time

- ❖ My school colleagues, for putting up with my frequent absences for study leave

and most of all

- ❖ My dear family: Tony, Victoria and Justin, who have always believed in me and offered their love and support. Thank you, guys!

ABSTRACT

This thesis discusses how assessment data are used to support the learning of pupils aged four to seven years in four Kent primary schools. The sample was 451 pupils in Reception, Year 1 and Year 2; the researcher collected and analysed quantitative data from pupil attainment on school entry - either as Baseline Assessment or the Foundation Stage Profile - and from results in reading, writing and mathematics at the end of KS1, both as SATs and Teacher Assessment. These data were triangulated with qualitative data collected from a semi-structured questionnaire, classroom observations and interviews with the Reception class practitioners. The author - a head teacher with many years' experience of primary schools and the Early Years – outlines recent and current government policy and links these to assessment theory and existing practice in the four schools studied. She identifies some possible influences on attainment and looks at how value-added data are currently used as measures of pupil performance. The three research questions look at whether benchmark data can be used to predict future achievement, the educational implications of using value-added data as measures of pupil performance, and whether benchmark data can be used to support learning in the primary classroom. The findings led the researcher to conclude that accurate prediction from prior attainment is not possible at the present time and that contextual value-added data are only useful when other variables are taken into account. However, the findings showed that benchmark data - when used formatively - can be useful in supporting pupils' learning. This study will help headteacher colleagues to look at data in a fresh way, and to identify and target the needs of individual pupils to optimise their performance from the beginning of the Foundation Stage to the end of Key Stage One.

Contents

Page

List of tables

CHAPTER 1: INTRODUCTION

1.1:   CONTEXT

The Education Reform Act (ERA) of 1988 brought about a number of major changes to the education agenda in England and Wales. Any one of these could form the basis of this thesis: the changing role of governors; local management of schools; the Code of Practice for special educational needs (SEN); standards for qualified teacher status (QTS); teacher appraisal and the inauguration of the National Professional Qualification for Headteachers (NPQH). However, arguably perhaps the most far-reaching changes were related to the new National Curriculum (NC) and the testing of pupils at age 7, 11, 14 and 16, that is, at the end of Key Stages 1, 2, 3 and 4 (KS1, KS2, KS3, KS4).

From the many key aspects of the Education Reform Act of 1988 I have chosen assessment as the focus for my thesis. I have examined the effectiveness of benchmark data in predicting pupil achievement, using triangulated analysis of quantitative and qualitative data from four primary schools in West Kent. This small-scale research is underpinned by assessment theory linked to recent and current government policy and existing practice in the four schools studied.

1.2:   OVERVIEW

With the new, statutory National Curriculum of 1988 came greater accountability, leading to the publication of 'league tables' and the concept of 'naming and shaming' schools who performed less well than their peers. The testing at the end of the first three Key Stages was part of a whole raft of new assessment tasks and tests, with schools - and parents - being informed of the NC levels that an 'average' pupil should attain at age 7, 11 and 14. The implication was that any pupil who achieved levels higher than the 'norm' (given as level 2 at KS1; level 4 at KS2; and levels 5/6 at KS3) was of above average ability and that those pupils who could not attain the norm had some element of special educational needs (SEN) and would need appropriate

1

intervention and support strategies. These NC levels were linked to *statements of attainment* - hundreds of them - leading to a 'tick sheet' mentality amongst some schools and local education authorities (LEAs). The Dearing review of 1999 (QCA, 1999a) led to a reduction in the number of statements of attainment, a 'slimming down' of the National Curriculum and a move towards a 'best fit' model, using NC level descriptors, all of which eased the burden on class teachers. The main changes related to:

> A more explicit rationale for the School Curriculum and the national curriculum: value, aims and purposes... A stronger emphasis on inclusion: providing effective learning opportunities for all pupils... Clearer, more flexible national curriculum subjects... Greater consistency and coherence in personal, social and health education (PSHE) and citizenship [non-statutory for KS1 and KS2; and]... New flexibility at KS4. (QCA, 1999a)

The new National Curriculum document produced in 1999 (DfEE/QCA 1999b) - the so-called 'Curriculum 2000' - was welcomed by teachers and was user-friendly: it had one volume instead of the previous ten ring binders, and was colour-coded, with the level descriptors in a fold-out end section. The concept of inclusion - a key government ideal - was outlined in an eight-page section (pp. 30-37) that detailed how to provide 'effective learning opportunities for all pupils'. Assessment was embedded in the document and cited as one of four main purposes of the National Curriculum: 'The National Curriculum... establishes national standards for the performance of all pupils in the subjects it includes.

These standards can be used to set targets for improvement, measure progress towards those targets, and monitor and compare performance between individuals, groups and schools.' (p. 12). This new document kept the 'best fit' model of the post-Dearing curriculum and made target-setting a key element of the teaching and learning process. Guidance was given to teachers on how to assess pupils at the end of Key Stage 1 using differentiated sub-levels, from 'W' ('working towards' level 1 of the National Curriculum), rising through 1c, 1b, 1a (a below-average pupil) to 2c, 2b and 2a (an average pupil), and 3c, 3b and 3a, representing the achievement of an above-average pupil (p. 18).

Baseline Assessment (BA) - a raft of tests for four-year old pupils during the first six weeks of schooling - was piloted for the 1997/98 academic year (DfEE 1998a) and became statutory for the following cohort. This BA would be used as a standard against which future levels of achievement - measured in NC sub-levels and on an 'Average Points Score' (APS) basis - would be gauged as a 'value-added' (VA) or 'residual' score.

If pupil performance was to be judged against national norms in standard assessment tasks and tests (SATs) at the end of the three Key Stages, and against prior attainment in the form of value-added data, then it became necessary to set interim targets for individual pupils, groups and classes. The government also wished to set targets for improvement in standards of achievement, particularly in the subjects of 'literacy' (encompassing reading, writing, spelling and handwriting) and 'numeracy' (being number, measures, shape & space and handling data), linked to the new government initiatives, the National Literacy Strategy (NLS) and the National Numeracy Strategy (NNS) which came into being in 1998 and 1999 respectively (DfEE 1998b, 1999c). LEAs had to set their own targets in literacy and numeracy, which were filtered down to individual school level and led to the concept of school league tables and their publication in the public domain. The government's initial targets for 2002 were that '80% of all 11-year olds achieve at least level 4 in the National Curriculum tests for English, and 75% achieve at least level 4 in the tests for mathematics' (DfEE, 1999c, p. 2.)

The concept of value-added assessment had to be made against a pupil's prior attainment, and for that to happen a clear strategy for tracking the progress of individuals and groups needed be in place as part of the school's assessment policy. If schools were expected to set targets for English and mathematics at the end of KS2, each cohort needed to be assessed from its entry into the Reception class of an infant or primary school. Commonly most primary schools had a one- or two-form entry: since the advent of the infant class size

legislation in September 2001 (DfEE, undated leaflet), many schools opted for Reception class intakes in multiples of 30 where numbers permitted. Assuming a single class group of 30, this meant that the cohort of 30 pupils in a primary school would be tracked as a group - and as individuals - using data from the Reception class to the end of KS2. The school would be aware of the SEN and 'English as an additional language' (EAL) profiles of that cohort - although these profiles might change if pupils migrated to or from other schools - and would thus be in a position to formulate targets for the group's projected results at the end of KS2. (All of these SATs results relate to English and mathematics, not science, as these are the only two subject areas for which the government requests targets.)

Pupils could thus be tracked from the Reception year (aged 4 years) to the end of Year 6, (aged 11 years) and their progress monitored against value-added criteria. This was made easier in Kent by the Advisory Service producing a matrix called *'Pathways to Progress'* (KAS, 2001). This document indicated that, with a BA score of between 34 and 37 points, a pupil could be expected to achieve a level 2c (a low average) in KS1 SATs and a level 4a (a high average) in the KS2 SATs. Similarly, a pupil who scored the same 34 to 37 points at the end of KS1 could be expected to achieve only a level 2a in the KS2 SATs (well below average), whilst a pupil who scored 46 or more points at baseline (the maximum possible is 56) should achieve a level 5a (a high 'above-average') in the KS2 SATs. There were also interim projected scores on the matrix which meant a teacher could plot the progress of an individual pupil, using projected and actual scores for the tests at the end of each year.

Whilst the Baseline Assessment scheme allowed schools to predict future performance against National Curriculum criteria, the new Foundation Stage Profile (QCA, 2003a) offered no such facility. At a meeting in May 2003 with the Team Leader of the Management Information Services department of Kent County Council (KCC), I enquired how the LEA would be using the summative data from the new Foundation Stage Profile (FSP) and how these would

compare with the data collected in previous years from the BA data. I was informed that the progress charts produced by KCC were based on actual achievement, using grouped data, and that they predicted minimum expectations for future performance. The total BA scores from the Reception year for one cohort (about thirteen thousand pupils) had been plotted against scores the same pupils achieved in KS1 SATs at the end of Year 2, using regression analysis or a 'line of best fit'. Further analysis by the Team Leader and her two colleagues - using the *Statistical Program for Social Sciences* (SPSS) - had produced a progress chart (the afore-mentioned *Pathways to Progress*, KAS, 2001) projecting future performance using correlated data from the KS1 and KS2 pupil data, based on the APS. When I enquired whether the new data from the FSPs would be analysed in a similar manner, it was explained that this new document - to collect summative data from pupils at the end of the Reception year - was, as its name suggested, a *profile*, and as such gave an overview of pupil achievement against criteria; there was no way in which such data could be used to predict future performance.

As a headteacher in a primary school I am responsible for working with colleagues and governors to set targets for achievement at the end of KS2 in English and mathematics. As assessment coordinator, I track the progress of all pupils from their benchmark data at the end of the Reception year, through non-statutory tests at the end of Year 1 to teacher assessment at the end of Year 2 (KS1), assessing progress against value-added data based on predicted outcomes produced by Kent County Council (KCC) who are the local education authority. I continue to track progress, using value-added criteria, through optional (non-statutory) tests at the end of Years 3, 4 and 5 and the statutory tests at the end of Year 6 (KS2). In our school we also use tests in verbal and non-verbal reasoning (VR and NVR) to assess a pupil's intelligence quotient (IQ) at the beginning of Years 3 and 5; these give an indication of how well the pupil is performing in tests according to their academic potential as defined by the test design.

My school is part of a small consortium of four primary schools in West Kent; we work together to share good practice across all year groups and curriculum areas, arranging frequent meetings between subject leader colleagues where issues such as planning and assessment are discussed. My interest in assessment (across the entire primary phase but within the Foundation Stage in particular) suggested the title for my doctoral research:

Using assessment data to support the learning of young pupils in four Kent primary schools.

This is examined under three research questions:
1) Can benchmark data be used to predict future achievement?
2) What are the educational implications of using value-added data as measures of pupil performance?
3) Can benchmark data be used to support pupils' learning?

The research is small-scale, involving the four schools in my consortium; data were collected from the Foundation Stage and KS1, then analysed and represented in a variety of ways to address the research questions. Quantitative data - in the form of BA, FSP and KS1 scores were supplemented by qualitative data collected from a survey of Foundation Stage practitioners and triangulated by observations I made in the Reception classes of the four schools. I used SPSS to organise and present the data and to help with my analyses. I hope that this study will generate much discussion within our consortium and lead to a wider dissemination of the issues related to assessment in the primary phase across other schools in my West Kent Cluster.

CHAPTER 2: ASSESSMENT IN SCHOOLS IN ENGLAND
AND WALES

2.1: ASSESSMENT - THE BACKGROUND

In 1974, the Assessment of Performance Unit (APU) was set up by the then Department of Education and Science (DES). The objectives of the APU, as given in a White Paper of that year, were 'To promote the development of methods of assessing and monitoring the achievements of children at school and to seek to identify the incidence of under-achievement.' (see Gipps and Murphy, 1994, p. 114). It was made clear from the beginning that performances might differ: one group might be advantaged or disadvantaged when compared to another because of the pupils' differential learning experiences. These experiences might be apparent in the context of the task, that is, its setting; the content of the task (what it was about); the task demand, for example 'write a letter to a friend'; or the mode and style of response favoured in that particular subject (Gipps and Murphy, 1994, p. 146). However, the authors noted that, where these effects occurred, differences in performance might represent differences in pupils' achievement but not necessarily their potential. (See also Gipps, 1994)

The Report of the National Curriculum Task Group on Assessment and Testing (the TGAT Report) of 1987 proposed that standard assessment tasks and tests should be used to assess pupil achievement; the Education Reform Act (ERA) the following year decreed that all children should be formally assessed at the ages of 7, 11 and 14, as well as at 16 years, through the General Certificate of Secondary Education (GCSE) system.

Pilot tests were constructed in the light of a specification of what needed to be tested. These tests were administered to groups of children in order to assess the quality of the formal instructions and to ascertain the likely duration of the tests. A process of item analysis was conducted, calculating the difficulty and

7

the discrimination indices for each item. The difficulty index was judged by calculating the percentage of children getting the correct answer. The discrimination index was rather more complex: the total test scores of the pupils in the group were arranged in order of merit, and the top and bottom 25% of the scores for each paper were then compared item by item. If 95% of children in the top 25% of scores, but only 10% of children in the bottom 25%, answered an item correctly, then such an item was making a clear distinction between high- and low-achievers. Conversely, if 80% of children in the top 25% of scores and 75% in the bottom quarter answered the item correctly, then little distinction had been made between the high- and low-achievers and that particular item would therefore be discarded (see Desforges, 1989, p. 40).

The first pilot testing of pupils at the end of KS1 (Year 2) took place in May 1990, with the first national testing for all Year 2 pupils scheduled for 1991. The pilot test items were presented in a variety of contexts, with a range of response modes and styles: practical, oral and written activities; this was to ensure that the tasks and tests had content validity, that is, that they matched normal class-based tasks with which the children were familiar. Analysis of the pilot tests (Gipps and Murphy, 1994) showed that, 'while there were many statements of attainment...which had proved easier for one sex than the other, there was no clear evidence that the SAT as a whole contained overall gender biases', and that, in a sample group of 353 Bengali-speaking pupils, the children performed better on the SATs than their teacher assessment (TA) had indicated, especially in the 'Speaking and Listening' strand. (Gipps and Murphy, 1994, pp. 188-9). A consortium from the National Foundation for Educational Research (NfER) recommended that, for bilingual pupils, the national SATs for 1991 should be administered in a mixture of English and the pupil's first language (though not for the English tests, which should be administered in English only); where no bilingual support was available, teachers were recommended 'to use mime, sign language, repetition or rewards, drawings or other methods - including children's mutual help - to explain the tasks...' (Gipps and Murphy, 1994, p. 188).

Research quoted in Gipps *et al* (1995) found that the first year of national KS1 SATs in 1991 was a traumatic experience for pupils and teachers, with problems in classroom management and time management - an average of forty to forty-five hours per class due to the long practical tasks required in mathematics and science. By 1992, the tests had been made shorter, were more standardised, and more 'pencil-and-paper'; they took an average of twenty-four hours per class. By 1993, spelling and reading comprehension tests had to be completed for all pupils except those working at level 1, mathematics now included algebra, and science included physical processes; tests this year took an average of thirty hours to complete, and teachers admitted to 'teaching to the test' in order to prepare for SATs. In 1994, KS1 SATs in science were dropped in favour of teacher assessment; teachers still had to allocate NC levels to indicate pupil achievement, and this still entailed some form of assessment against NC level descriptors. The results of the 1991 tests were reported to the LEA but not to parents; many teachers objected to the need for reporting results, and teaching unions threatened to boycott the process. The following year, results also had to be reported to parents, giving the child's own scores and a summary of the school's results; headteachers were also encouraged to include in reports comparative national information based on the 1991 NC assessment results.

This was the start of the concept of 'league tables', which were published for the first time in March 1997, based on results for KS2 pupils for tests taken in 1996. The 'Primary School Performance Tables' - as they were called – would: help parents when choosing schools for their children and allow clear comparisons of achievement at Key Stage 2. Such comparisons, set alongside local and national averages, will support the drive to raise levels of achievement in all primary schools (DfEE, 1996a, p. 3). Some of the practical mathematics and science SATs tasks, though time-consuming, had proved so successful with the children that they had influenced the teaching style and approaches found within classrooms from then on. Two headteachers said that

their school was now paying more attention to the structure, mechanics and 'basics' of writing; class teachers said that their assessment practice had become more observation-focused, leading to better formative assessment (see Gipps *et al*, 1995).

A paper published by the Organisation for Economic Cooperation and Development suggested that:

> the findings of international studies may be used to support an optimistic view of education, underpinning a belief that systems can be improved significantly...' and that 'there is...considerable agreement that national assessment is appropriate at the primary level...There is, however, considerable variation in the grade at which an assessment is first carried out. It can begin at kindergarten (Canada), first (France), second (Ireland and Sweden), third (Finland) or fourth grade (Scotland)...but similarity of the title of a subject (e.g. mathematics and science) does not ensure similarity of curriculum content or achievement. (Kellaghan and Grisay, 1995, p. 145).

The authors continued:

> A further problem in the development of assessment procedures is that curriculum representation and balance may be affected by scaling after tests have been administered... It would seem to be extremely difficult, perhaps impossible...to devise a single test that is an equally valid measure of different countries' definition of "achievement". (pp. 48-49)

They drew attention to the fact that, whilst:

> the most comprehensive approach to the use of performance testing in national assessment is to be found in the British national assessment system...experience with the first major assessment of 7-year old children in 1991 brought to light serious inconsistencies in the administration and scoring of the SATs... The lack of standardisation that was a feature of administration and scoring must call into question the use of the data obtained, not only for international comparisons but even for comparing individual pupil scores or aggregated school scores. (p. 51).

In December 1995 a report was made to the Secretary of State for Education and Employment (SCAA, 1996); among the findings were the following alternative proposals to the existing national assessment procedures:

> Revert to a situation where the test result is combined with the teacher assessments to produce a single level for the subject...; Make teacher assessment the main means of reporting attainment to parents at the end of the key stage, using the current tests as moderating instrument...; [and]...Provide a bank of tests which could be used on either an optional or a compulsory basis by teachers to support and confirm their teacher assessments during the course of the key stage. This is the so-called 'Scottish model' of assessment... (paragraph 4.2.1)

However, the report continued: 'All of the different alternatives presented have disadvantages which make them less suitable than the present arrangements...The present approach should therefore be retained, and any alternative models considered as part of the longer term review...' (paragraph 4.2.2.)

An evaluation report commissioned by SCAA (Sizmur and Burley, 1997) looked at the findings from the 1996 KS1 assessments in the core subjects (English, mathematics and science):

> Revised national curriculum Orders were introduced for all subjects in 1995. National curriculum assessment was not, however, fully aligned with the revised Orders until 1996. This year, therefore, featured some major changes to some aspects of the assessment approach... A further significant change... was the incorporation of writing, spelling and handwriting into a single English attainment target. (paragraph 1.1).

For the purposes of the report, and for two parallel reports on *Analysis of Pupil Performance* in English and mathematics produced by SCAA in 1996, questionnaire surveys were carried out - on headteachers and one year 2 teacher - in a sample of 200 schools drawn from various types (independent, first, infant, junior and infant, special). Respondents were asked to rate the tasks (mathematics, reading, writing and spelling) in terms of structure, ease of administration, and how well they thought the National Curriculum level

obtained reflected their own teacher assessment of the pupil's performance. The report concluded that:

> the assessment system overall [was] functioning reasonably satisfactorily... The overall balance between written tests and more flexible tasks appears to have been appropriate... It is recommended that the present balance of tests and tasks be maintained; [and that the] use of Key Stage 2 tests for Level 4 be reviewed (paragraph 5.1)

The preliminary report of the Numeracy Task Force (DfEE, 1998d) identified why numeracy is a key skill that pupils need to acquire:

> Numeracy and literacy are both important in enabling children to access the full curriculum, and later, to play a full part in adult working and social life... Early work in mathematics must begin to lay the foundation for the skills and insights children will use in later life... The Task Force believes strongly that being fully numerate is an entitlement for all children. (paragraphs 5-8)

The Numeracy Task Force had been set up in May 1997 by the Secretary of State for Education and Employment; its recommendations led to the establishment of the National Numeracy Strategy (NNS). The national target set for 75% of 11-year olds to achieve Level 4 in the National Curriculum tests in English, mathematics and science at the end of KS2 in 2002 was said to be ambitious but achievable (DfEE, 1998d, paragraph 12); it would be achieved by 'high expectations of staff and pupils, systematic monitoring...of teachers' planning, teaching and assessment of numeracy, and clear targets for numeracy, with a realistic plan for achieving them...' (paragraph 23)

The following year, guidance was made available to schools on 'efficient ways of recording pupil achievement to help to raise standards.' (QCA, 1999d) It outlined three types of assessment that should be kept: daily/weekly, termly and annual; these records would enable teachers to 'track the progress of individual pupils or groups, identify patterns over time...set individual and group targets for improvement, and confirm end-of-year statutory assessments.' (p. 2) Wintle and Harrison (1999) suggested that schools should compile portfolios of evidence in the form of individual pupil Records of Achievement; these were

used in primary schools for a time, but have now generally been superseded by exemplar portfolios of NC levels in the core subjects of English, mathematics, science and ICT.

In March 1998, a group of aspiring headteachers on the National Professional Qualifications for Headteachers (NPQH) course in West Kent was presented with a confidential draft copy of the Primary Pilot Document for Performance and Assessment Data (PANDA) reports that would be sent from the Office for Standards in Education to all primary schools in England on an annual basis from October 1998, containing analyses of data from the previous academic year. The introduction to this document stated:

> This PANDA Report has been sent to your school to help you in the development of your plans to raise standards. Its role is to help you see how good your school is, in comparison with other schools. Knowing where you stand is the first step in the cycle of evaluation, planning and school improvement, and should help you set targets for what you want pupils in your school to achieve...Your school's performance data are shown in comparison with national averages and in comparison with other schools with pupils from similar backgrounds. (Ofsted, 1997a.)

Data included figures for authorised and unauthorised absence, and grades given by previous Ofsted inspections of the school. The key data were in the sections: 'Basic characteristics of your school' (the number of pupils on roll, the percentages eligible for free school meals (FSM), with English as a second language (EAL), and with special education needs (SEN), and 'Your school's context' (the ward the school was in, the percentages of adults with higher education, children in high social class households, minority ethnic children and children in over-crowded households). These categories became known as 'proxy-indicators', and were used to allocate schools into groups for purposes of comparison of pupil achievement in KS1 and KS2 SATs. Page 13 of the draft document stated:

Whilst the link between individual pupils' backgrounds and their academic attainment is weak, and it is important that pupils' backgrounds are not taken as an excuse for low expectations about their achievement, there is, as already mentioned, across the country, a correlation between the background of pupils attending a school and pupils' attainment in relation to national standards. It can be useful, therefore, to look at pupils' attainment in a school in comparison with the performance of pupils in similar schools, that is, schools with pupils from similar backgrounds, as well as with the national average.

Three booklets from central government on 'Guidance for effective target-setting for pupils with special educational needs' (DfEE, 1998c, 1999a and DfES, 2001a) introduced the concept of using 'P' levels instead of the former 'W' grading when reporting to parents. The P scale moves from P1(i) to P8 (11 sub-levels) and then into the NC level 1c (the lowest NC grade). As these P levels are summative - that is, they are reported to parents at the end of each year - a pupil may be working within the P scale during the year but manage to reach NC level 1 by the end of the summer term. It was proposed that, from December 2001, maintained special schools would no longer be permitted to set zero-rated school performance targets for those pupils with moderate, severe and profound learning difficulties, but would instead be required to use the new P levels. From the end of the 2004/2005 academic year, the DfES would collect P scale data from all schools (mainstream as well as special) and publish a national summary of results. The data would inform the future extension and refinement of value-added calculations. (see QCA, 2004f).

Since July 1998, when I had first become involved in the target-setting process, data had been required by Kent County Council (KCC) that projected pupils' NC levels in English and mathematics at the end of KS1 and KS2. That first year, a letter to headteachers from the County Schools Effectiveness Officer stated:

> At this early stage of target setting, with limited commonly held data available to make reliable predictions about future pupil performance...schools are best placed to make judgements about the potential progress of individual pupils and therefore the potential to raise overall levels of achievement in a particular year

group... in cases where there are particular issues, including for example especially high SEN levels, anomalies in the pattern of targets, a high turnover of pupils or significant differences in levels of targets between schools with apparently similar intakes... (KCC, 1998)

That first round of target-setting in March 1998 asked headteachers to complete a form showing the projected grades for pupils in all classes, from Reception to Year 6, with no request for gender split or level of SEN within any cohort. Further targets were requested in December 1998, again for all seven year groups; maybe schools' previous comments on SEN had been recognised, because these new forms (one each for English and mathematics targets) asked headteachers to include levels of SEN for each cohort. One of the categories (non-statutory, but included on the forms in the hope that schools would complete it) was for the pre-Year R cohort: the form stated that these data were 'for planning purposes only: pupils not yet admitted.' The return for December 1999 was now just one form requesting targets for Year 1 through to Year 6 (note that the Year R data were no longer requested); there was still space on the form for data relating to SEN. The December 2000 return added sections for Years 3 to 6 for projected percentages for those pupils expected to achieve NC level 2+ at the end of KS2 in English and mathematics (the earlier forms had only asked for 3+). The return for 2001 was the same as for 2000, whilst the 2002 return saw the sub-division of 'English' into 'reading' and 'writing' at KS1 and the introduction of targets for level 3+ at KS1 and for level 5+ at KS2.

By December 2003, headteachers were being asked to complete a more complex set of forms: on the main form, in KS1, only data for Year 2 were required (not for Year 1 or Reception) within the three subjects of reading, writing and mathematics. The data had to be entered by gender split for every cohort and each of the two subjects of English and mathematics at KS2. In addition, further forms had to be completed; data were required for pupils who were a Looked After Child (LAC) or a Traveller, and those who had EAL: projections were required for the percentage of pupils who were likely to achieve NC level 4+ and level 5+ in English and mathematics at the end of

KS2, set against the total number of pupils and by gender split, under the Pupil Level Annual Schools' Census (PLASC) EAL categories (twelve in all). In December 2004, all these data were required again (still with no space to record levels of SEN), together with another form detailing those pupils who had been identified as the lowest-attaining 20% in the baseline to KS1 value-added data, that is, those pupils whose achievement was at or below an APS of 12.3 at the end of KS1. The 2002 guidance (KCC, 2002a) had given headteachers a new matrix from which to make their projections: both the *Pupil achievement levels: baseline to Key Stage One* table (KAS, 1999) and the *Pathways to Progress* table (KAS, 2001) had been superseded by a new chart, *Minimum expectations of progress   baseline to Key Stage 1* (KCC, 2002a). Each of these three matrices offered different baseline totals on which to project expected grades at the end of KS1; headteachers who had projected certain NC levels using one of the two earlier charts now had to undertake the whole process again, using the new data.

A letter sent to all Kent schools in October 2003 (KCC, 2003b) outlined new primary target-setting guidance, based on the recently published Primary Strategy *Excellence and Enjoyment* (DfES, 2003) which required schools to set targets for individual pupils as before but no longer required LEAs to formally agree statutory KS2 targets with schools. However, KCC still wanted some measure of tracking pupils' progress in all county schools, and the October 2003 letter suggested that LEA advisers would still 'work with schools both to support and challenge them in the targets they set and in their use of performance data'; this led to the continuing process of Local Education Officers visiting schools during the autumn term every year to 'agree' statutory targets with the headteacher. The Kent School Organisation Plan (KCC, 2004d, p. 37) announced that the most recent data for the percentage of KS2 pupils achieving NC level 4 or above were 72.8% for English and 70.6% for mathematics; the DfES target was 78% and 75% respectively.

In May 2004, some 5,000 schools in 34 volunteer LEAs across England took part in a trial of the assessment process at Key Stage 1, aimed at eliminating the need for formal tests and tasks; this had been announced in the Primary Strategy document of 2003 (DfES, 2003). All four schools in my study undertook the trial, along with the majority of infant and primary schools in Kent. This KS1 trial followed a ballot on a boycott of National Curriculum tests and tasks at KS1 and the NC tests at KS2 by the National Union of Teachers (NUT) in late autumn 2003.

Guidance issued by KCC (2004i) explained the changes in KS1 assessment as a result of the trial under eighteen bullet points, including the need to report a teacher assessment level at W, 1, 2, 3, 4+, A (absent) and D (disapplied) in science and in speaking and listening, and additionally as sub-levels of NC level 2 in reading, writing and mathematics. A combination of 2003 tests and 2004 tests and tasks could be used with the proviso that, in writing, the longer and shorter tasks from the same year had to be used, in conjunction with the spelling test from either year. It was suggested that:

> The new arrangements are intended to offer teachers more flexibility and to place greater emphasis on their judgements about children's progress throughout the year and should not lead to an increase in the amount of testing at this key stage. (KCC, 2004h, p. 5)

Despite this assurance, practitioners in all four schools of my study, as well as colleagues in other Kent schools, all felt that, far from being an 'easy option', taking part in the trial was exceedingly onerous and time-consuming. Government research indicated that the trial was successful, however, and the academic year 2004/2005 saw the full national implementation of teacher assessment at KS1, replacing the former SATs. Tasks and tests are now used by Year 2 teachers in all schools in England to underpin their assessment of pupil performance against NC level descriptors, using P levels where appropriate, continuing up through NC level 1, sub-levels 2c, 2b and 2a, to level 3 and, for some exceptional pupils in certain schools, to level 4 and above.

2.2: BASELINE ASSESSMENT

The concept of assessing children as young as five years of age against national norms has moved rapidly in England from its statutory inception in 1996 to the more complex model currently in use. In July 1995, the Government announced plans for expanding pre-school education and asked SCAA to draw up Desirable Outcomes for Children's Learning on Entering Compulsory Education and to consult on the need for guidance for providers on educational activities appropriate to the outcomes (DfEE, 1997b). In January 1996, the Department for Education and Employment (DfEE), in conjunction with SCAA, had produced a thin booklet entitled *Nursery education: desirable outcomes for children's learning on entering compulsory education* (DfEE, 1996a). This listed, under six key headings, the areas of learning that were considered essential for a child to experience before moving on to the KS1 programme of study.

These areas were categorised as:
- Personal and Social Development;
- Language and Literacy;
- Mathematics;
- Knowledge and Understanding of the World;
- Physical Development; and
- Creative Development,

and were initially called 'Desirable Outcomes' , being 'goals for learning for children by the time they enter compulsory education...[which] begins the term after the child's fifth birthday' (p. 1).

In March 1997, SCAA announced that it was to be merged with The National Council for Vocational Qualifications to form the Qualifications and National Curriculum Authority (QNCA) (SCAA, 1997a). It also announced that, following 'overwhelming support for SCAA's proposals on the baseline assessment of children entering primary school...' in the previous term's consultation process, September 1998 would mark the introduction of baseline

assessment for all children starting primary school. It also announced that work was continuing on 'materials exemplifying what children should have learnt by the time they start school'; they called these *Desirable Learning Outcomes* (note the change of name). The timescale for the introduction of BA was: April 1997   publication of national framework together with optional material available to support the development of schemes; September 1997 - pilot year for BA; November 1997 - first published list of accredited schemes; September 1998 - statutory requirement for schools to use an accredited scheme.

June 1997 brought two further publications from SCAA: *Baseline Assessment - information for schools* (SCAA, 1997b) and *The National Framework for Baseline Assessment: criteria and procedures for the accreditation of baseline assessment schemes* (SCAA, 1997c). These booklets were almost identical in text and format: both outlined the two purposes of baseline assessment, which were:

- to provide information to help teachers plan effectively to meet children's individual learning needs; and
- to measure children's attainment, using one or more numerical outcomes which can be used in later value-added analyses of children's progress.

Each publication, in slightly different ways, outlined that, from September 1998, all maintained primary schools in England would be required to use an accredited scheme for baseline assessment. These schemes would not be national but would be offered by providers who could apply to be accredited by SCAA/QCA against the criteria laid down on pages 6 and 7 of the earlier booklet (SCAA, 1997a); it was suggested that baseline assessment should normally be completed within the first seven weeks of a child starting primary school. The second booklet (SCAA, 1997b) gave a list of key dates when the new baseline assessment scheme chosen by each school would need to be implemented: a national pilot of newly-accredited schemes would take place from September 1997, with evaluation taking place between September and December of that year and published in March/April 1998. January 1998

would see the first publication of national data on baseline assessment, to be followed by further publications in May and September 1998; these would thus cover the three possible intake periods for children in the Reception year. By the end of the summer term 1998 at the latest, all LEAs and schools would need to have chosen which accredited scheme to use from September 1998 when baseline assessment would become statutory for that intake. (See DfEE, 1998a). Teaching unions objected to the publication of national data and many schools boycotted the pilot scheme - or certainly the notification of results to their LEA - and the proposed publications did not take place.

In 1999, the Qualifications and Curriculum Authority (QCA), changed the title of the 'Desirable Learning Outcomes' to 'Early Learning Goals' (ELGs). The publication in October of *Early Learning Goals* (QCA, 1999c) introduced the concept of a 'Foundation Stage' (p. 6) and its relation to the National Curriculum at KS1. The ELGs kept the same six key areas of learning, but 'personal and social development' was renamed 'personal, social and emotional development, whilst 'mathematics' became 'mathematical development' (by the following year - see QCA, 2000 'language and literacy' had become 'communication, language and literacy'). The 1999 document emphasised that these ELGs would 'establish expectations for most children to reach by the end of the Foundation Stage' but that they were **'not a curriculum in themselves'** (p. 18). The document then set out a rationale for learning in each of the six headings, with key objectives and six case studies showing a child's expected progression towards, and beyond, the Early Learning Goals from age three to the end of the Foundation Stage.

This resulted in the publication in May 2000 of a ring-bound document entitled *Curriculum Guidance for the Foundation Stage* (QCA, 2000) which had been developed by QCA, working closely with early years practitioners and experts, together with the National Literacy and Numeracy Strategies and Ofsted. In her introduction, Margaret Hodge made the first mention of the term 'Stepping Stones'. These were set out in colour-coded format through the document, with

attainments shown as: the yellow band being the most likely achievement for three-year olds; the blue band for four-year olds and the green band for five-year olds; the final grey band showed achievement of a child working within level 1 of the National Curriculum. It was emphasised that:

> although these Stepping Stones are presented in a hierarchical order where possible, not all children conform so neatly to this sequence of learning... as children move from one stepping stone to another, they take with them what they have already learned and continue to practise, refine and use their previous learning, so that learning becomes consolidated. (Introduction)

Practitioners welcomed the document, and used the idea of the Stepping Stones to help them plan their pupils' learning progress. The guidance suggested that, in order for all pupils to show achievement against the Stepping Stones, learning opportunities should be offered 'using alternative and augmentative communication, including signs and symbols' (p. 18) and 'providing bilingual support' (p. 19). By October 2001, the guidance had been supplemented by the publication of the document *Planning for Learning in the Foundation Stage* (QCA, 2001b). The document purported to 'help those [Foundation Stage practitioners] who are less experienced to plan successfully and in a way that is manageable' (p. 2); to this end, the document did give some examples of curriculum plans, with five case histories taken from different early years settings.

A pilot BA in September of 1997 and January 1998, undertaken by most infant and primary schools in Kent, entailed assessment against thirty-two statements relating to achievement in aspects of English, mathematics and 'personal and social development'. (This cohort of pupils took their KS2 SATs in May 2004 and transferred to secondary school in September 2004; their achievements in the Year 6 SATs compared to their BA scores, together with a profile of their individual learning needs, appears in Chapter 6.2.) The first formal BA of pupils during the first seven weeks of the term of entry (whether this was in September or January) commenced with the autumn 1998 intake; Kent LEA had chosen the scheme entitled *Starting Together*, which was administered

using a double-sided A4 card for each pupil. Personal data to be entered comprised the child's first forename and surname; gender; whether English was the child's first language; whether they were entitled to a free school meal; whether they had a statement of SEN; the type of pre-school attended; whether the child's attendance was full- or part-time; and the date of testing. Teachers then had to tick a 'Yes/No' box against seven statements for each of the seven aspects of English and mathematics: 'Reading for meaning and enjoyment', 'Recognising print and symbols', 'Phonological awareness", 'Writing', 'Speaking and listening', 'Sorting and counting' and 'Using mathematical language'. In addition, there were four boxes to be ticked in each of the aspects of the remaining area of learning, 'Personal and social development': 'Attention skills', 'Expressing needs and opinions', 'From dependence to independence' and 'Interaction and collaboration with others'. This made a total of sixty-five separate statements on which the teacher had to make a judgement, based on evidence from observations and from the child's oral contribution and written work. This form of BA was used in Kent from the 1998/99 academic year until it was overtaken by the Foundation Stage Profile, which started its formal life in Reception classes at the end of the summer term 2003.

These baseline assessments were reported to KCC which, in October 1998 and again in January 1999, sent an analysis to each school of the pupils' BA scores. These were sent as a summary sheet for each child, showing the individual scores achieved for each of the eight aspects: Reading A, Reading B, Reading C, Writing, Speaking and Listening, Mathematics A, Mathematics B, and Personal and Social Development (the maximum for each was seven points, making a maximum score of 56); further data gave the school average for each aspect and the pupil's age-standardised score for English and for mathematics, together with his or her total BA score. There was also a graphical representation of the scores - by gender split - for each of the assessed areas (English, mathematics and personal/social development) together with a class summary sheet giving the individual scores achieved, age-standardised scores and an average score for the school.

2.3: THE FOUNDATION STAGE PROFILE

Running parallel with the development of the Foundation Stage curriculum and baseline assessment was the evolution of a new assessment tool, to be called the Foundation Stage Profile. In May 2001, QCA wrote to fifteen randomly selected early years practitioners inviting them to join a panel to review the development of the national scheme for baseline assessment. These practitioners were drawn from infant, primary and first schools throughout England, and they joined a special educational needs panel of twelve members, together with a Foundation Stage working group of fourteen members, national partners from the DfES, the Teacher Training Agency, NLS, NNS and Ofsted, and nineteen other early years practitioners. This large group was part of the consortium set up by the NfER and Birmingham Education Service to develop what became the Foundation Stage Profile of January 2003 (QCA 2003a). The purpose of the panel was:

> to increase the number of practitioners contributing their views to the new national scheme for baseline assessment, its classroom validity and suitability for assessment of children at the end of the Foundation Stage...The input of practitioners in the development process of the new national baseline assessment scheme is crucial in ensuring that it is suitable for children at the end of the Foundation Stage. ... the material is currently confidential and practitioners' panel members will be expected to respect this fact... in signing your agreement to join this panel, you will be bound by QCA's confidentiality procedures and to the remit of the panel, which is also attached. (Letter from QCA, 2001a)

The first meeting of the practitioner panel took place on 19 June 2001; members were told the rationale for setting up the panel and it was explained how, at the end of the Foundation Stage, practitioners needed to assess each child's performance in relation to the ELGs that formed the *Curriculum Guidance for the Foundation Stage* (QCA, 2000). It was emphasised that practitioners would be expected to '...keep effective records of each child's progress throughout the Foundation Stage...to inform judgements...' but also '...to sum up each child's achievements in a way that is manageable and useful'.

However, even at this early stage, there was an indication that assessment at entry to school would be used for future value-added purposes: the scheme allowed for 'the calculation of a score over some of the areas of the Foundation Stage curriculum', using *assessment summaries* for Knowledge and Understanding of the World (KUW), Physical Development (PD) and Creative Development (CD) and *assessment scales* for Personal, Social and Emotional Development (PSE), Communication, Language and Literacy (CLL) and Mathematical Development (MD). The assessment summaries asked practitioners to grade pupils as 'progressing towards', 'achieved some', 'achieved all' or 'working beyond' against eight criterion-referenced statements for PD and five for CD. KUW was more complex, being grouped into 'A - Exploration and Investigation', 'B - Sense of Place, Time, Cultures and Beliefs' and 'C - Information and Communication Technology, Designing and Making Skills'; in all, a total of eleven statements to be assessed.

The assessment scales were even more detailed than the summaries: PSE was broken down into the three strands of 'Dispositions and Attitudes', 'Social Development' and 'Emotional Development'; CLL into 'Language for Communication and Thinking', 'Linking Sounds and Letters', 'Reading' and 'Writing'; and MD into 'Numbers as Labels and for Counting', 'Calculating' and 'Shape, Space and Measures'. Each strand comprised eight criterion-referenced statements, making a total of one hundred and four separate statements of attainment.

The handbook recommended that the judgement should represent the practitioner's assessment of the child's typical achievement based on their existing knowledge and records of the child, by informal observations of the child's performance in different contexts or by means of an assessment activity: a nineteen-page *Assessment Activities booklet* accompanied the handbook. Assessments were to be recorded in a *Pupil Record Booklet* for each child: judgements were to be made against the statements of attainment by one tick (no date was required, or evidence that the pupil had retained the concept over

24

time). The completed front cover would show the pupil's first name, surname, date of birth, gender, pre-school experience (LEA nursery, nursery centre, playgroup, accredited childminder, or none), the number of full terms spent in a Reception class, and whether the pupil was learning English as an additional language (yes/no, then additional data grading the child's use of English from 1 - 'new to English' to 4 - 'a very fluent user of English in most social and learning contexts'). The final box to be completed indicated whether or not the pupil had been identified as having special educational needs according to the Code of Practice for SEN: zero representing no needs, 1 to 5 representing the stages of the Code, and S being a pupil with a statement of SEN.

The second meeting on 12 November 2001 reiterated the rationale for using the FSP 'to summarise each child's development where this has not already been done'. Each strand within the six ELGs now comprised nine assessment scales - the summaries had been dispensed with - making a total of one hundred and seventeen criterion-referenced statements against which the teacher had to make judgement. Practitioners still had to complete the booklet cover with data relating to the pupil's level of English and of SEN and his/her pre-school experience. Two teacher's handbooks offered support in the form of *Optional Activities* and *Case Studies* to aid assessment against each strand of the ELGs.

Delegates at the third meeting on 14 March 2002 were given a handbook in A4 portrait format containing the case studies introduced at the second meeting, together with detailed sections relating to, among others, 'Access to the Foundation Stage profile for children with a range of individual needs', 'Children with English as an additional language' and 'Contributors to the assessment process'. The former offered links between the ELGs and the P levels that had been introduced in 1998; practitioners were reminded of the difference between these two forms of assessment: 'The ELGs specify expectations for children's progression by the end of the Foundation Stage, while the P scales have been written for use with children of any age who are working towards level 1 or who are working within levels 1 or 2 of the National

Curriculum for extended periods of time.' The section on 'contributors to the assessment process' highlighted three key stakeholders: the practitioners, the parents/carers and the children themselves, together with 'other adults in the setting' (being learning and SEN support assistants, lunchtime supervisors and outside agencies). More recording was being asked of the early years practitioner: the handbook now contained a pro-forma, to be completed jointly by the parent/carer and the practitioner before the child started school. This was in the form of questions about the child's knowledge, skills and understanding within the six ELGs, and complemented by a further pro-forma called a *Child Conference*, which was to be completed with the child once he/she had started school. A further pro-forma, *Ongoing Assessment* was to be completed jointly with parent/carer and practitioner, and would show 'Progress since last meeting'.

The final meeting of the practitioner panel on 26 September 2002, led by the newly-appointed Project Director for the Foundation Stage Profile, gave information on developments since the last meeting and introduced the '...final draft materials for the profile...as it will be going out in January [2003]...are there any comments you wish to raise?' Attendees watched a video and were shown a CD, both of which would be part of the final package when it was sent into schools the following term. There was also talk of an eProfile which would be made available to practitioners as an alternative to completing the booklet manually. Members of the practitioner panel were warmly thanked by the Senior Research Officer from NfER in a letter sent to them in July: '...it is necessary that these pre-test materials are thoroughly trialled with children and teachers, and we very much appreciate the extra time that you and the children at your school have given in helping us with this work.' The profile booklet was sent to schools as a single copy - with instructions on how to order a set for the Reception class - together with the file 'Foundation Stage Profile: Handbook' in January 2003 (QCA, 2003a). The cover had space for the child's photograph, their name, date of birth and gender but no mention of level of SEN or of English as an additional language, and no space to record the UPN.

Alongside the development of the FSP, in Kent the LEA was working on a Foundation Stage 'Record of Transfer'. Kent's Early Years Development and Childcare Partnership had worked with a group of school and pre-school practitioners to develop and trial the Record, which would be made available to all Reception class teachers from the summer term 2003. The Record was said to be compatible with the FSP, and its aim was to support teachers in matching their provision to the needs of the new school intake more effectively.

The Record (double-sided A4) was to be completed by pre-school providers (nursery schools and classes, private nurseries, pre-school playgroups and childminders) who formed the early part of the Foundation Stage before children went into Reception classes in infant, primary or first schools. The main section of the Record was for details of the child's movement towards - or completion of - the colour-coded 'Stepping Stones' bands of the ELGs. Teachers had to enter the child's name, address, gender and date of birth; date of entry, number of terms and number of sessions per week spent in pre-school; whether the attendance was regular or irregular; and additional personal information on the child's medical or social needs. If the child had been diagnosed with a special educational need, this had to be recorded on the Record, and a supplementary sheet completed detailing the particular need and what steps had been taken as intervention strategies. Children with SEN were designated as being at either *Early Years Action* or *Early Years Action Plus*, as described in the SEN Code of Practice (DfES, 2001b). Pre-school providers had to complete sections describing the child's areas of strength and of concern; parental involvement and the child's view; other professionals involved, and areas to be addressed in future planning. They also had to append a Play Plan or Individual Education Plan and a Care Plan if used. Upon completion of the Record, pre-school providers had to arrange a mutually convenient time to share the form with the parents and obtain their written consent to the Record of Transfer being passed on to the child's new Reception class teacher.

The Foundation Stage Profile had to be completed on a termly basis: teachers had to circle a coloured dot against either 'autumn term', 'spring term' or 'summer term' to show when the child had met the criteria for the ELG. The green section of each aspect of the ELG represented a hierarchical sequence of criteria at the basic level of achievement: these three statements had to be completed in order. The middle section of each aspect coloured grey - was not hierarchical and the five statements could be completed in any order; the final pink box containing one criterion indicating that the child was working within level 1 of the National Curriculum. As the FSP booklets had to be ordered from QCA, they were not received in schools until the end of the spring term 2003, meaning that, for this inaugural year, teachers could only record achievement up to and including the summer term 2003, that is, for the whole year, rather than on a termly basis. In all, there were now 117 statements across thirteen strands of the six ELGs; these comprised nine statements against each of the following strands:

- *Dispositions and Attitudes, Social Development* and *Emotional Development* within the ELG 'Personal, Social and Emotional Development' (PSED);

- *Language for Communication and Thinking, Linking Sounds and Letters, Reading* and *Writing* within the ELG 'Communication, Language and Literacy (CLL);

- *Numbers as Labels for Counting, Calculating* and *Shape, Space and Measures* within the ELG 'Mathematical Development' (MD); and against the final three ELGs:

- 'Knowledge and Understanding of the World' (KUW);

- 'Physical Development' (PD); and

- 'Creative Development' (CD).

In Kent in July 2004, the first year following the pilot of 2003, practitioners had to attend a full moderation meeting at local teachers' centres and bring their profile booklets and three or four lever-arch files of evidence to prove that their judgements were sound. In November 2004, an email from the KCC

'Clusterweb' (most communications from the LEA and the DfES are now sent electronically) informed headteachers that:

> It is a statutory requirement that all schools participate in LEA moderation annually... This year the focus will be on Communication, Language and Literacy (all four scales) and Creative Development (one scale). Moderators will spend time in classrooms talking to children and looking at the evidence that demonstrates their individual achievement. The following will be of particular importance: teachers' files of dated observations; annotated photographs which demonstrate achievement; recent examples of children's own mark making; paintings, drawings and models etc. The teacher's own knowledge will also form an important part of the evidence base and we would like to have approximately 30 minutes with the teacher for professional discussion. (KCC, 2004b).

Surrey County Council, in an email in October 2003, congratulated their schools on the 100% return of the FSP data, and stated that the moderation process would consist of a visit to three randomly-selected schools in each cluster group, together with moderation meetings during the spring term 2004. Data were also promised showing links between FSP profile points and NC levels, which they suggested would be useful for target-setting. Surrey schools were recommended to use the eProfile - developed by Suffolk LEA and available as a download - to record their FSP data for their own and LEA purposes; these data could also be stored on a computer program, *Assessment Manager*. In 2005, Kent primary schools were asked to return their FSP data to the LEA in one of three formats: on a spreadsheet from the DfES, as an eProfile (now recognised by central government as a valid form of recording), or in *Assessment Manager*; the data were also required to be uploaded to the DfES on the *School to School* website as a Common Transfer File.

The first pilot testing of Year 2 pupils took place in 1990, with revised National Curriculum Orders being introduced five years later; by 1998 primary schools were being asked to set targets for performance at the end of KS2 based on benchmark data from KS1. Alongside these changes, 1996 brought in Baseline Assessment for pupils in the Reception year, to be replaced in 2003 by the Foundation Stage Profile. From KS1 SATs and tasks to KS1 teacher

assessment, from BA data sheets to the FSP and the eProfile, class teachers were being presented with a plethora of instruments with which to measure pupil performance; it appeared that government was putting unwarranted faith in the complex issues related to assessment and – perhaps more worryingly – in the ability of classroom practitioners to understand and use assessment in its diverse forms to improve pupil achievement. Pressure on teachers and headteachers was increasing, and it seemed unlikely to abate in the immediate future.

CHAPTER 3: ASSESSMENT AS MEASUREMENT AND EVALUATION

3.1: MEASUREMENT AND EVALUATION

According to Reynolds, Livingston and Willson (2006):

> Teaching is often conceptualized as a straightforward process whereby teachers provide instruction and students learn. With this perspective, teaching is seen as a simple instruction-learning process. In actual practice, it is more realistic to view assessment as an integral component of the teaching process. In fact, it has been estimated that teachers devote at least one-third of their professional time to assessment-related activities. (p. 2)

The authors describe a *test* as being 'a device or procedure in which a sample of an individual's behaviour is obtained, evaluated, and scored using standardized procedures'; *measurement* as 'a set of rules for assigning numbers to represent objects, traits, attributes, or behaviors'; and *assessment* as 'any systematic procedure for collecting information that can be used to make inferences about the characteristics of people or objects'. (p. 3)

In England at the present time, primary school teachers – in common with their US counterparts - currently spend a great deal of their time in testing and assessing pupils, and in measuring their progress against various norms and criteria. Local authorities collect data from primary schools in the form of standardised tasks and test results at the ends of KS2, teacher assessments at the end of KS1 and KS2, and the pupils' scores against the various components of the Foundation Stage Profile (FSP) together with the total scores. Some data are sent direct from schools to the Department for Education and Skills (DfES) via the School-to-School (S2S) website; these data are uploaded by feeder primary schools and then downloaded by the receiving secondary schools (this process is also used when children move schools other than at the usual transfer time). Testing at the end of KS1 has effectively been abolished (although Year 2 teachers have to make summative teacher assessments against National

Curriculum (NC) levels and sub-levels, with the aid of 'old' NC tests and others that are provided each year for the purpose), but KS2 tests and teacher assessments still take place every May, and the FSP has to be completed by the end of the Foundation Stage, that is, before a pupil enters Year 1, which signals the start of measurement of progress against the National Curriculum learning objectives.

Cohen, Manion and Morrison (2001) examined the differences between a norm-referenced test that 'compares students' achievements relative to other students' achievements... (e.g. a national test... which has been standardized on a large and representative sample...)', a criterion-referenced test which 'requires the students to fulfil a given set of criteria, a predefined and absolute standard or outcome...' and the more recent domain-referenced tests, where 'considerable significance is accorded to the careful and detailed specification of the content or the domain which will be assessed.' (pp. 318-319). In recent years, an effort has been made to distinguish between assessment **of** learning and assessment **for** learning, following the work of Professors Black and Wiliam (see next section of this chapter). However, many teachers of my acquaintance administer tests and make their assessment without a full understanding of assessment theory, or indeed, of the role of mathematics in measurement.

Measurement can be made of *stability* -where the same test is administered twice to the same sample, with the scores being correlated; of *equivalence* - where two forms of a test are administered to the same sample on the same day; and of *internal consistency*, where reliability estimates are obtained from one set of test data, either by using statistical procedures such as Hoyt's Analysis of Variance or Cronenbach's Coeffient Alpha, or by applying a split-half procedure, where one test is administered in two halves, with subscores obtained for each of the two halves, which are then correlated and subjected to a correction factor such as the Spearman-Brown prophecy formula.

When we are studying the *reliability* or *validity* of a test, we need to measure the relationship between individual scores or sets of scores. When an increase in one variable is accompanied by an increase in the other variable (for example, aptitude and achievement) then the correlation is positive; when an increase in either one is accompanied by a decrease in the other then the correlation is negative. The significance of a relationship depends on the number of cases or scores being studied, so the sample needs to be large enough to answer the research question.

Reliability can be measured as *scorer reliability*, when a sample of papers scored independently by two different readers can be analysed using Pearson's r to estimate the reliability of a single reader's score; *group homogeneity*, where the more heterogeneous the group, the higher the reliability; and *difference scores*. The latter are considerably less reliable than single scores, because the errors of measurement on each test contribute to error variance in the different scores. It is generally accepted that standardised tests should have reliability coefficients of at least .85 for an individual and .65 for a group. For achievement tests, equivalence reliability estimates are essential, with internal consistency estimates to indicate the homogeneity of the test content. It is also essential to know the sample size, its representativeness, the mean and standard deviation of the sample scores, and the standard errors of measurement and how they were obtained.

Validity evaluates the degree to which the test is capable of achieving the required aims, which might be to describe the current state of achievement of the individual or group being tested, or for making predictions about future performance. Cohen, Manion and Morrison (2001) listed examples of validity, including content validity, construct validity, and predictive validity, together with both internal and external validity. *Internal validity* 'seeks to demonstrate that the explanation of a particular event, issue or set of data which a piece of research provides can actually be sustained by the data.'; *external validity* ' refers to the degree to which the results can be generalized to the wider

population, cases or situations.' (pp. 107-108)   *Construct validity* is the degree to which test scores reflect a previously-defined construct such as 'creativity'; *content validity* is related to how adequately the content of the test is able to sample the *domain* - the subject matter or area of competence - which is being tested.   Predictive use is dependent on criterion-related validity, where 'future success' is the criterion.

In general, educators and psychologists have interpreted most test data as being *interval measurement*. Nominal scales (the assignment of different numerals to categories) and ordinal scales (where scores are put into rank order) are also used; ratio scales are less common in educational and psychological measurement.   Measurement of the *central tendency*, that is, the average or typical score in the distribution, can be represented as the mean (the arithmetic average of a set of scores), the median (the point below which 50% of scores lie) or the mode (the most frequently occurring score). To know a person's raw score in a test is of little value: it is more valuable to know how far the score deviates from the mean, so measures such as the standard deviation (SD) are often made in order to indicate the variability of the distribution of scores.   The data collected can be presented in graphical form such as histograms and frequency polygons: the distribution of scores or grouped data may show normal distribution (the familiar bell-shaped curve) or be skewed either positively or negatively.   Scattergrams or box-and-whisker plots are often used to show relationships between sets of data.   These and other statistical tests can be made on the data, either with manual calculation or using a computer program such as SPSS in order that the findings can be analysed.

An understanding of basic statistical analysis is essential if teachers are to make the best use of their time when assessing the progress of their pupils against normative and criterion-referenced objectives.   Some of my Kent colleagues do not have this understanding, and leave the data analysis to their senior management team, or to the assessment co-ordinator; thankfully this is currently being addressed by KCC, with more courses and seminars being

provided for the beginning statistician. Also, in 2004 KCC sent all primary schools copies of a handy booklet, *Guide to managing data effectively in the primary* school, which suggested that:

> Schools should use data analysis as a means of identifying areas for improvement... To contribute to raising standards, data management, analysis and evaluation have to be viewed as an integral part of the school improvement process. (KCC, 2004k, p. 4)

In the wider domain, Professor Tymms (2002) gave a clear introduction to data analysis in primary schools, including formative and summative assessment, norm-referenced and criterion-referenced tests, and the value or otherwise of baseline assessment as applied to pupils in the Reception class. He discussed so-called 'value-added' data, and suggested that the term is confusing, as it has two quite different meanings (pp. 58–72). In another of his handbooks (Tymms – undated) gave a useful introduction to using analysis of pupil data to set appropriate targets in primary schools; the answers to the exercises would help practitioners to understand how basic statistical analysis could benefit their school.

Multi-level modelling is used by statisticians to enable complex analyses to be made on national data sets from schools (Goldstein, Huiqi, Rath and Hill, 2000; Schagen and Schagen, 2005); thankfully for the lay person, Professor Tymms suggested that 'the results from the multi-level analyses are invariably identical to the results from the much simpler analyses'. (Tymms, 2002. p. 65) For the purpose of this thesis, regression analysis was not used, as there was no KS2 data to compare with the baseline and FSP data sets that I had collected.

3.2: ASSESSMENT FOR LEARNING

The TGAT Report (DES, 1987) described assessment as a process that 'should be the servant, not the master of the curriculum' (Section I.4). The Task Group, under Professor Paul Black as Chair, gave priority to four key aspects of assessment, namely, that it should be:

- criterion-referenced, that is, achievement should be related to objectives;

- formative, that is, it should provide a basis for decisions about pupils' further learning needs;

- calibrated or moderated; and

- related to progression.

(Clarification of 'moderation' and 'calibration' was given thus: *Moderation* when the results are expressed as a grouped percentage distribution of pupils, adjustments can involve deciding which borderline pupils should change their group status. *Calibration* - when procedures for equating results of the assessments are statistical.) It was suggested that assessment was more than mere tests - that it could be better expressed as 'standard assessment tasks' used to assess a pupil's understanding. The art of constructing good assessment tasks was to 'exploit a wide range...of modes of [teacher] presentation, operation [mental, written, practical, oral] and [pupil] response.' (VII.48); recommendation IX was that 'a mixture of standard assessment instruments...be used...in order to minimise curriculum distortion.' Group moderation, according to recommendation X, should be an integral part of the national assessment system (although the report's authors admitted that this would be a time-consuming and costly process).

Recommendation IX.54 cast some doubt about 'the reliability of the observations made and the validity of the interpretations placed on them': Reliability refers to the scores assigned to pupils and is measured in terms of their stability over time and over different sets of equivalent tests or over other variable assessment conditions. Computation of a reliability coefficient allows the estimate of an error of measurement...Validity relates to an interpretation or use of test scores. It is obviously reduced by errors in the estimate of the scores themselves but is further reduced by unwanted influences on the relationship of the measures to the constant which they are interpreted to measure.
(appendix G).

Recommendation VIII related to possible gender and ethnic bias in assessment tasks and made reference to pupils 'whose first language is not English': 'We recommend that assessment tasks be reviewed regularly for evidence of bias, particularly in respect of gender and race.' Recommendation XII.133 stated: 'It is well-known that a variety of socio-economic factors so affect pupils' capacity to respond to school work that they can lead to very wide differences between the performances of different schools...' (citing Rutter's 1979 research *15000 Hours*).

Recommendation XIII. 147; this section related specifically to the primary stage of education and concluded that:

> national assessment at the first stage should be aimed principally at identifying children in need of help which is exceptional for a child of that age, whether because he or she is making so little or so much progress in a profile component. [a particular aspect of learning within a curriculum area] ... Care should be taken at all reporting ages, but especially at this age [seven years] to avoid giving the impression that the assessment is a prediction of future performance. (XIII. 148)

Professor Black developed his theme of assessment *for* learning rather than *of* learning when he published, with Professor Dylan Wiliam, a research report entitled *Inside the black box: raising standards through classroom assessment* (Black and Wiliam, 1998). A synopsis of the research (General Teaching Council for England, 2003) highlighted key areas of the study that had asked questions such as 'In what ways do current assessment policies inhibit effective learning?' and 'What can we do to improve formative assessment?' In their report, Black and Wiliam represented the classroom as a 'black box' with inputs (teachers, resources, curriculum, among others) and outputs (good test results, pupils with skills and knowledge, and so on); the authors' brief was for readers to identify those particular aspects that went on 'inside the black box' that helped to raise standards. The key message of the report was that learning is driven by what teachers and pupils actually **do** in classrooms, and that formative assessment has a vital role to play in this. Black and Wiliam argued that, as teaching and learning are interactive, teachers need to know about their

pupils' progress and difficulties with learning; that is, to use formative assessment, in order to adapt future lesson planning to meet the needs of the individual and the group.

The study suggested that pupils:

> look for the ways to obtain the best marks rather than at the needs of their learning which these needs ought to reflect... Pupils who encounter difficulties and poor results are led to believe that they lack ability... Whilst the high-achievers can do well in such a [rewards or grades] culture, the overall result is to enhance the frequency and the extent of under-achievement... What is needed is a culture of success, backed by the belief that all can achieve. Formative assessment can be a powerful weapon here if it is communicated in the right way. (p. 9)

Black and Wiliam argued that ways to raise pupils' self-esteem through formative assessment involves self- and peer-assessment, teacher-pupil discussions, open questions by teachers, and quality feedback to pupils after tasks or homework have been completed. They recommended a reduction in curriculum content - which they hoped would be met by the 2000 review - and in the amount of external testing at all Key Stages; they felt that the status of teachers' professional judgements needed to be raised and given more recognition by central government. (See also Black, 1998.) Shirley Clarke, at the Institute of Education, University of London, developed her own practical strategies for using formative assessment in the primary classroom, building on the work of Professors Black and Wiliam (Clarke, 1998). The research was considered apposite by Kent LEA, which commissioned Clarke for a series of conference dates for primary headteachers at which she was the keynote speaker. Later publications developed her theme of assessment for learning through constructive feedback and pupil self- and peer-assessment. (See Clarke, 2000, 2001a, 2001b, 2003, 2005.) The Assessment Reform Group, funded by the Nuffield Foundation, published a pamphlet entitled *Assessment for learning: beyond the black box* (Assessment Reform Group, 1999). This developed Professors Black and Wiliam's premise of assessment *for* rather than *of* learning, which (according to the preface of the later publication) had 'proved

without a shadow of doubt that, when carried out effectively, informal classroom assessment with constructive feedback to the student will raise levels of attainment.' The authors of this later report suggested that change was urgently needed if schools were to successfully use formative assessment to effect higher levels of learning in the classroom, citing Her Majesty's Inspectorate (HMI) reports in England that suggested much classroom practice fell short of providing adequate formative assessment practices.

Following the launch of the NLS, QCA sent guidance to all schools on setting targets and assessing children's progress; the pack contained a booklet of children's work exemplifying the learning targets, with criteria and suggested texts for independent reading (QCA, 1999b) This guidance was grounded in the concept of assessment for learning, and showed teachers how to set short- and medium-term targets, how to use assessment and link it to recording, and how to use it in summative reports at the end of a school year; the exemplars of pupils' work gave teachers benchmarks of where to pitch their own assessments against learning targets relating to the teaching objectives of the NLS.

In July 2002 King's College, London published a follow-up text from Black, Wiliam and colleagues entitled *Working inside the black box: assessment for learning in the classroom* (Black, Harrison, Lee, Marshall and Wiliam, 2002). This described research carried out since 1998 with over thirty teachers in six secondary schools in Medway and Oxfordshire, involving almost 1000 pupils. Overall, the research suggested four key components of change: asking questions in class; marking homework; peer assessment; and self assessment. The main findings were that there were practical ways to raise standards which did not involve so-called 'teaching to the test'; that emphasis on competitive testing invariably damaged learning through some pupils' loss of self-esteem; and that teachers, when released from external pressures, were eminently able to determine appropriate learning opportunities based on good assessment practice. Many aspects of this good practice were highlighted, notably that pupils' self- and peer assessment were invaluable in supporting both the self

esteem and the learning of the pupils concerned; that the marking of homework was essential (but that time needed to be set aside in lessons so such feedback could be accessed, and re-writing undertaken where necessary); and that not all class work needed to be marked, but rather that certain, specified sections should be marked in more depth. The use of appropriate questioning in class was key: teachers were recommended to allow longer for pupils to answer; it was suggested that wrong answers could be helpful, and that hands did not always need to be held up (this led, in many Kent primary schools, to the introduction of 'number fans' and 'letter fans' to signify understanding of the answer to the question posed). In line with previous research, 'open' questions were preferred to 'closed', and teachers were encouraged to let questions lead into small-group discussion in order to fully explore issues.

Assessment for learning was highlighted by the Assessment Reform Group who, in 2002, produced a research paper entitled *Testing, Motivation and Learning* and a pamphlet, *Assessment for Learning: 10 research-based principles to guide classroom practice.* Research by Black, Wiliam and colleagues found 'strong evidence of the negative impact of testing on pupils' motivation, though this varied in degrees with the pupils' characteristics and with the conditions of their learning.' (p. 3) and that 'for the less successful pupils, repeated tests lower self-esteem and the effort they put into learning. This has the effect of increasing the gap between high- and low-achieving pupils.' (p. 1) They found that 'how assessment of their learning is reported back to the pupil (feedback) affects motivation to learn. It has a central role since the feeling of self-efficacy is derived from performance in previous tasks of the same kind.' (p. 5) The team suggested that teachers should concentrate on encouraging pupils, discuss their work with them, and give feedback related to target-setting rather than to a comparison of test grades (p. 9).

The same year, the Office for Standards in Education published two pamphlets (Ofsted 2002a, 2002b) which reported an overview of the first three years of the National Numeracy and National Literacy strategies respectively. The NNS

report (2002a) noted that 'Teachers' assessment of pupils' progress remains a weakness... [and that,] During the first two years of the strategy, assessment procedures in many schools were inconsistent.' (paragraph 31). The NLS report (2002b) suggested that 'More often than not, the best-performing schools show strengths in assessment. They use information from assessment to improve teaching.' (paragraph 46) and that, 'In many ways, assessment is the key to improvement and higher standards. In schools where standards have remained static or have fallen, assessment is often poorly understood.' (paragraph 47).

Kirklees Metropolitan Council published a report by Ros Wilson (2002) that aimed to improve standards in writing across the primary curriculum. She introduced an action plan with eight specific points, including a new criterion scale in which each NC level in writing had been broken down into specific criteria against which work could be assessed (Wilson, 2002, pp. 8-14). Many primary schools in Kent recognised the value of this method of assessing progress in writing in a formative way, and schools A and C in my research have adopted the criterion scale in their assessment of pupils in the Reception class and throughout KS1.

During 2003 and 2004, the National Foundation for Educational Research (NfER) commissioned a series of research projects that examined assessment for learning and its impact on pupil achievement. A study by Kirkup (2004) and her team looked at schools' use of target-setting, peer and self assessment and feedback, linked to the increasing use of optional SATs from QCA (for Years 3, 4 and 5 in the primary phase). Kirkup and Twist (2004) suggested that assessment in the classroom can serve many purposes, and that 'many teachers appear to object to the uses to which the test data is put, rather than the tests or testing per se. They acknowledge the need for valid and reliable instruments that allow them to make comparative and summative judgements of their pupils.' In the survey and follow-up interviews of KS2 teachers, many reported that 'there is more scope for using the QCA optional tests, administered in a

formative way (as compared to the statutory key stage 2 tests...)'. The research concludes that 'The challenge for schools is to achieve an integrated approach to assessment, with formative use of assessment evidence, including that obtained from formative summative tests.' (pp. 8-9)

On 10 November 2003, QCA headlined a new 'Assessment for Learning' website and published a booklet aimed at 'using assessment to raise achievement in mathematics' (QCA, 2003b). It suggested that ways in which pupils could improve their skills and understanding might be through self-assessment, peer assessment, and setting learning goals related to individual targets agreed with the teacher:

> If learning is to be efficient, pupils must be active agents in judging the gap between what they know and what they need to know...The teacher can then help them to close the gap by sharing learning goals with them. However, only the pupil can do the actual learning. (p. 2)

In February 2004, Kent LEA produced a booklet entitled *Integrating test analysis into whole-school planning* which encouraged teachers to use 'gap analysis' to identify pupil misunderstandings (KCC, 2004a, paragraph 6). Further support on gap analysis came in the form of five colourful leaflets from the National Assessment Agency (QCA, 2004a, 2004b, 2004c, 2004d, 2004e); these provided information to teachers on the implications for teaching and learning drawn from an analysis of pupils' performance in the 2004 National Curriculum tests at Key Stages 1 and 2.

Previously, 2003 had seen the introduction of the Primary Strategy document (DfES, 2003). Describing assessment at KS1, the report suggested that:

> Good-quality assessment is essential for planning children's learning, based on their performance, so that they are stretched to achieve to their full potential. The standards that children achieve at the age of seven, particularly in reading, are strongly associated with future progress. (paragraph 2.26).

The report summarised the imminent changes to assessment at the end of KS1:

> Assessment must be as good as it can possibly be at helping teachers assess children's learning accurately and rigorously so as to plan the best possible learning for them... We propose to trial, in 2004, an approach that uses testing to underpin teacher assessment, rather than having the two things alongside each other... (paragraphs 2.29 to 2.32)

The reporting of assessment for pupils with special educational needs was discussed:

> Many children with special educational needs...are capable of average or even very high achievement...we propose to extend and refine the coverage of value-added calculations, including the use of P scales, which measure the progress of children working below National Curriculum levels, so that the value of a school's work with pupils with special educational needs is properly recognised. (paragraphs 2.33 to 2.36)

Referring to the Foundation Stage, the document stated that:

> Reception teachers can teach the elements of the literacy hour and daily mathematics lesson flexibly over the day, and are advised to have both lessons in place by the end of the year. The Foundation Stage Profile will give Year 1 teachers important information about children coming into their class, and help them plan effectively to meet their needs. (paragraph 4.16)

The Annex exhorted teachers to think about what they were going to assess:

> Be selective, focus on the key aspects of learning that you wish to assess, and highlight these on your plan. Then use a simple system for recording children's progress. Link curricular targets to your plans for groups of pupils and some individuals. (p. 78)

Following the 2003 document came the ubiquitous 'suitcase' from the DfES: the latest in line after the National Numeracy and Literacy Strategies. This time, the large plastic binder contained a 'school self-evaluation grid for assessment for learning', with two booklets and one video for each of the following aspects of *'Excellence and Enjoyment: learning and teaching in the primary years'* :

- Understanding how learning develops;
- Creating a learning culture;
- Planning and assessment for learning.

These were to be used as part of a whole-school, integrated approach to learning and teaching: headteachers were urged to use them as the basis for staff training sessions. The booklet entitled *'Assessment for learning'* recommended self- and peer-assessment, marking feedback for homework and class work (with the pupil where possible), and appropriate questioning to assess understanding: in fact, all the key proposals from the 2002 research study by Black, Wiliam and his team of researchers.

The Ofsted Handbook (Ofsted, 2003b) placed assessment for learning at the heart of its judgements. In Chapter 4 of the Evaluation Schedule under the heading 'How effective are teaching and learning? it cited 'Very good' assessment as one in which '... Pupils' work receives well-focused, diagnostic comments that help them to see how to improve...'; a grade of 'Excellent' would be given to a school where 'Assessment...enables pupils to play a very strong part in making and recognising improvement in their work...' (p. 73)

A research project undertaken by NfER (Ashby and Sainsbury, 2001) looked at 415 schools of various types containing primary-aged pupils that were surveyed in 1998 and 2000, with 229 of these schools forming the sub-sample for the comparative analyses. The authors of the report had asked such questions as 'Does your school use National Curriculum test results to inform curriculum management?...curriculum planning?'

The responses showed that:

> The most universal purposes for the use of test results in school management contexts were checking National Curriculum attainment and monitoring performance at the end of the key stage... comparisons with other schools... screening for special educational needs...planning continuity and progression...The Government's overriding agenda for primary schools since its

election in 1997 has been a rise in standards, especially in literacy and numeracy. The national tests play a central role in these initiatives, by providing a measure of these standards. Teachers, headteachers and LEAs have had responsibilities placed upon them to contribute to this improvement in standards, and their use of test results is a direct response to this policy context. (section 3)

A paper presented by Dylan Wiliam at the British Educational Research Association (BERA) 20th annual conference in Oxford in 1994 drew on the work of Samuel Messick to argue for the role that values have to play in educational assessment. Wiliam suggested that:

> Validation... involves evaluating the fitness of the model for its purpose... [and] the examination of the values and ethical considerations underpinning those assumptions and the consequences of using the models. (Wiliam, 1994, pp. 10-11).

3.3: TARGET-SETTING

My first research question, *Can benchmark data be used to predict future achievement?* will be answered by an examination of the practice of setting targets for pupils' achievement in reading, writing and mathematics based on teacher assessment of a pupil's current and expected progress.

In 1996, the DfEE published a survey of good practice (DfEE, 1996b) that suggested 'Inspection shows that internal review - and target-setting - is often the weakest parts of the school's planning cycle.' (paragraph 10) Schools were starting to use indicators - such as comparisons between a pupil's reading scores at age six years and eight years - to monitor performance and to improve achievement, both for individual pupils and for the cohort as a whole. However, target-setting at this time was not fully embedded in practice at primary level, perhaps due to the fairly recent advent of the National Curriculum. By 1998, schools were beginning to recognise the importance of using national benchmark data to improve performance; QCA published a booklet (QCA, 1998b) that explained to headteachers basic statistical terms such as upper and lower quartiles, standard distributions and correlations, and which gave, for the first time, benchmark tables for comparison between

schools' data for results at the end of KS2. Value-added information was cited for the first time as being an important aspect of data analysis at primary school level; these data could be represented, it was suggested, as scatter plots, trend lines or chances graphs, all comparing the proportions of pupils who attained certain NC levels at the end of KS2 with their performance at the end of KS1 (at this time, results were given as NC levels rather than as APS as now).

Of greater importance to this thesis is the comment on page 21 that 'Researchers and teachers are becoming increasingly aware of a range of factors that impact on pupil performance...'. These were: prior attainment; socio-economic factors; free school meals (FSM); EAL; gender (girls were outperforming boys in English by 10% at KS1 and by 15% at KS2, although there was no statistically significant difference in mathematics and science); SEN; and summer-born pupils: age-related differences were apparent from the Reception class through to the end of KS2 and on to GCSE level. (Possible influences on achievement are examined more fully in Chapter 6, section 6.1.)

The White Paper *Excellence in Schools* (DfEE, 1997a) said that from September 1998 schools would have to set challenging targets for improvement. A document produced by the newly-commissioned Standards and Effectiveness Unit of the DfEE (DfEE, 1997c) used case studies to show how performance indicators could help headteachers to reach the national targets set by the Government: that, by 2002, 80% of 11 year-old pupils would achieve level 4 in English and 75% in mathematics. The document also said that, from 1998, the DfEE would fulfil the requirement of the 1997 Education Act:

> ...to set and publish pupil performance targets annually in the core subjects of the National Curriculum... [and] to continue to publish national performance data annually, including information about similar schools which can be used for 'benchmarking'... LEAs should have sufficiently ambitious targets to enable national targets to be reached. (p. 4)

The Curriculum in Successful Primary School (Ofsted, 2002d) found that the schools in the survey (of 31 primary schools throughout England):

> achieve what many others claim is not possible. They have high standards in English, mathematics and science, while also giving a strong emphasis to the humanities, physical education and the arts.
> (p. 7)

They did this in various ways, from involving all the staff in discussions and creating a strong sense of teamwork to enriching the curriculum through day and residential visits. The key feature was that the headteachers monitored the curriculum closely through the planning and evaluation of lessons - and in particular, through using monitoring and assessment of targets set for improvement, across all subjects and with regard to the differing performances of boys and girls and pupils from ethnic minority groups.

In all the schools in the survey:

> Pupils understand the nature and purpose of their learning, in some cases contributing to the planning and evaluation of the curriculum, and they know what they have to do in order to make progress.
> (paragraph 91)

Standards in literacy and numeracy had 'risen substantially in primary schools in recent years', according to the DfES (2002d), although results at KS2 showed that only 75% of pupils had achieved NC level 4 in English compared to the target of 80%, with 73% achieving the same level in mathematics (the target was 75%). However, the document continued: 'The results of 7-year olds, which have also risen significantly in recent years, are promising signs of higher standards in the future.' (p. 1) It suggested that these higher standards could be achieved by the whole school community working together in 'shared action and support', by the introduction of 'catch-up' programmes (including Early Literacy Support, Additional Literacy Support and 'booster classes'), and by schools making a greater attempt to meet the needs of the 'more able' (target-setting for pupils expected to achieve NC level 3 at the end of KS1 and level 5 at KS2 had been introduced in 2002).

The Annual Report of Her Majesty's Chief Inspector of Schools (HMCI) for 2000/01 (Ofsted, 2002c) had found that there had been:

> a pause in the upward trend of pupils achieving Level 4 in English and mathematics [75% and 71% respectively], but a continuing rise in science. The gap between the attainment of girls and boys in English at the age of 11 has increased.
> (p. 22)

Two years later, the Annual Report from HMCI found that:

> The results in national tests [for nursery and primary schools] have not improved for several years. The gap between the standards attained by boys and girls persists, particularly in writing. (Ofsted, 2004, p. 13)

Analysis of gender differences for results in Kent reflect this national picture, and during 2004 KCC undertook 'detailed gender trend analysis to highlight where the gender gap in terms of performance exceeds 10% and examining the possible factors behind this.' (KCC, 2004i, p. 46)

Local Education Authorities (LEAs, re-named Local Authorities or LAs from the start of 2006) use data on pupil performance as part of their strategic management. A report commissioned by the Local Government Association looked at the role of the LEA in collecting, analysing, dissemination and using pupil performance data for target-setting and raising standards of attainment in schools. Respondents were keen to emphasise that test scores were not the only way in which they assessed the achievement of their pupils, and that children's wider educational development and experiences were equally important. (Rudd and Davies, 2002). In 2004, Kent Advisory Service produced a handbook of good practice in reviewing pupil achievement and setting targets (KAS, 2004).

This advised headteachers that:

> The national target of 85% of 11 year olds achieving Level 4+ in both English and mathematics remains... Floor targets of 65% Level 4 in English and mathematics remain in place, to be achieved by all schools in 2006... Changes for this year are: the removal of the statutory obligation to set Level 5 targets (although DfES and KCC strongly encourage this...); the setting of a school absence

target; increased significance being placed upon the setting of appropriate targets for pupils in potentially vulnerable groups. (p. 2)

The use of the Foundation Stage Profile is explained: it is suggested that this is 'an on-going summative assessment tool' (p. 4) – surely this is a contradiction

– and that teachers should observe pupils learning in a range of contexts and plan the appropriate next step for development. It also enables the Year 1 teacher to set targets for learning, using data from the Profile as well as knowledge gained about the pupil; 'pupils' individual progression towards these targets should be monitored throughout the year and planning adjusted to ensure that the challenge and pace of learning is appropriate.' (p. 4) – that is, that assessment data are used formatively to maximise achievement.

3.4:   VALUE-ADDED MEASURES

My second research question, *What are the educational implications of using value-added data as measures of pupil performance?* will examine how value-added data are used by central government, local authorities (LAs) and schools to improve performance in reading, writing and mathematics in England.

Lindsay and Desforges (1998) suggested that:

> Value-added analyses require both the baseline assessment and the outcome measures to be of acceptable reliability and validity. Consequently we must ask questions also of the quality of the end of Key Stage 1 assessments...early assessments were found to be far from satisfactory.
> (p. 58)

On the question of predictive validity, the authors said:

> Suppose all the children identified as being 'at risk' [of low attainment] were then provided with top quality and highly effective intervention. At the end of Key Stage 1 these children would score well on the assessment. The result is to **reduce** the apparent predictive power of the baseline assessment. (Lindsay and Desforges, 1998, p. 82)

Speaking about league tables and value-added data, Black (1998) suggested that:

> The justification that [league tables] can guide parental choice is a weak one, insofar as comparisons made without allowance for the many features which affect the results can be misleading... a school's circumstances can affect its pupils' capacity to make progress as well as their starting points. (p. 31)

Saunders (1998, 1999) looked at the debate on school effectiveness as measured by LEAs through value-added data sets, and examined a whole range of performance indicators used by schools to set challenging but realistic targets for pupil attainment. (See also Samuels, 1998)

In June 2001, KCC sent all its primary schools - for the first time - sets of data that related to their VA analysis from Baseline (1997-98) to KS1 (2000). Tables gave the names of pupils on roll when Baseline Assessment (BA) was completed at the start of the 1997 academic year, their BA total and predicted KS1 APS. The actual KS1 APS for each pupil was given, together with the individual pupil residual (VA score) and the overall school residual, all broken down into gender. This process was repeated in August 2001 in respect of BA (1998-9) to KS1 (2001) VA data, and the following year for BA (1999-2000) to KS1 (2002).

Primary school performance tables – known colloquially as 'league tables' were changed for their September 2001 publication, based on data from tasks and tests undertaken in May of that year. The changes at primary level were threefold: to include special schools for the first time; to publish a single Average Points Score (APS) in addition to other results at KS2; and to show the percentage of total sessions lost through both authorised and unauthorised absence (a session being a morning or an afternoon). Once again, special dispensation was given in respect of pupils in KS2 who had recently arrived from overseas ('recently' being after the start of the 1999/2000 academic year) and whose first language was not English: their results did not need to be reported. The 2002 league tables added data in respect of the percentage of

pupils eligible for KS2 assessment with SEN, both with and without statements; by 2003 the tables included the percentage of pupils achieving NC level 5 in English, mathematics and science, together with an APS (using equivalences adopted in the *Autumn Package*) for each pupil who was not absent for the test or disapplied. More significantly, for the first time, a value-added measure (VA) was published, showing the progress made by pupils between KS1 and KS2; the percentage of eligible pupils included in the VA measure; and the percentage of pupils included in the VA measure who were on roll for both the KS1 and the KS2 tests. This last was most important, considering the high mobility of pupils in some schools, especially Schools A and B in my study.

These 2003 league tables were late in being published because of a delay in the external marking process; they were finally published in December, three months later than planned. At their Annual Conference in October 2003, the normally liberal National Union of Head Teachers (NAHT) passed three motions that showed how the strong feelings frequently made public by the National Union of Teachers (NUT) and the National Association of Schoolmasters and Women Teachers (NASWT) about the politics of education had received sanction by the leading headteachers' union. These were: to pursue the abolition of league tables in England, due to the government's continual refusal to apply fair or reasonable criteria to their publication; to investigate all legal means of preventing their publication (including a boycott); and to abandon SATs at KS1. The Conference noted that, since the publication of the Primary Strategy document earlier that year, the strength of feelings of members had increased; this had been exacerbated by the recent adoption by the DfES of 'a seriously flawed and unacceptable version of value-added' (NAHT, 2003b).

Schagen and Schagen published a paper in the British Educational Research Journal in 2003 that looked at multi-level analysis of national value-added data sets to assess the impact of selection on pupil performance. The same year, Ian

Schagen presented a paper at the Value Added Measures Advisory Group (VAMAG) that explained the concept of a school 'residual', which Schagen described as 'the amount by which the school's actual results differ from what would be expected, taking account of various features of the school and its pupils.' (Schagen, 2003, p. 2) The research used multi-level analyses to present the data in various graphical formats, with explanations for the non-specialist teacher in primary and secondary schools in England.

By May 2004, Kent Management Information had sent primary schools a Governor Information Summary Sheet, detailing the basic school profile for KS1 and KS2, with results as percentages and as PANDA grades, with VA data for each Key Stage, with predicted and actual targets for English and mathematics at levels 4+ and 5+, with percentage of authorised and unauthorised absence, and with the number of permanent and fixed-term exclusions. All this, in addition to the annual PANDA report from Ofsted, the *Autumn Package* and the *Pupil Achievement Tracker* (PAT) from the DfES, and *'Making Figures Speak for Themselves'*, the KCC annual CD containing graphs and tables: schools were being sent the same data but in different formats and from different sources, which some colleagues found rather confusing.

The 2004 league tables were made available on the internet: these data did not include results for special schools, pupil referral units, hospital schools or independent schools. They did, however, give value-added scores 'showing how much a school had improved its pupils' achievements since they took their first set of tests in 2000'; the data were shown for each school in England as rankings on the VA score and on the aggregate score achieved in the three KS2 tests of English, mathematics and science, based on the APS. The accompanying notes on the website explained to the novice how a VA score had been calculated for each school: first, the VA score was found for each pupil by comparing their APS at KS2 with the median score in APS at KS1; then the mean for the pupils' individual scores was calculated to give the VA

score for that school (for the 2004 tables, the VA measures ranged from 94.7 to 105.6.) Again, the mobility of pupils was considered: where data were missing for 50% or more of those pupils eligible to take the test, the VA score was not published.

Data for every school in England were made available in tabular format and also as line graphs: the first graph showed a single APS for the school (from the combined totals of English, mathematics and science) as a VA score for pupils' improvement between KS1 and KS2; further line graphs showed the percentage of pupils achieving NC level 4 or above in each of the three subjects for the school, the LEA and nationally. Other data related to authorised and unauthorised absence; the percentage of pupils absent or disapplied for each test; the total number of pupils in the cohort who were eligible and the percentage of those pupils with SEN; and a trend graph for the school, LEA and national data over the past three years. All of the schools in England were ranked and 'banded' as shown in Table 1.

Table 1: Bands for national value-added measures

	Value-added score
Top 5% of schools nationally	102.1 and above
Next 20% of schools nationally	100.9 to 102.0
Next 15% of schools nationally	100.3 to 100.8
Middle 20% of schools nationally	99.8 to 100.2
Next 15% of schools nationally	99.3 to 99.7
Next 20% of schools nationally	98.0 to 99.2
Bottom 5% of schools nationally	97.9 and below

These data sets were also made available by the DfES to all schools in England in hard copy; the documentation gave tables for conversion of NC tasks and test levels 1, 2c, 2b, 2a, and 3 to equivalent APS: respectively, 9, 13, 15, 17 and 21. These APS were used in my analysis of data for the four schools in my study; other APS equivalents were given for 'W' ('Working towards level 1') and for level 4+, but these grades did not occur in my data sets. The detailed seven-page guidance to schools from the DfES gave formulae that had been used to calculate the 'input measure' (KS1 test result APS) and 'output measure' (KS2 test result APS); for calculating the VA measure (KS1 to KS2) for pupils and for schools; the 'coverage indicator' (showing the percentage of eligible pupils that were actually included in the VA calculation); and for the 'mobility indicator' (those pupils with a PLASC record who were on roll at the school for both the KS1 and the KS2 assessments).

3.5:   PERFORMANCE INDICATORS

The HMI report on the teaching of reading in forty-five Inner London primary schools (Ofsted, 1996) had found that:

> The key factors that emerge as significant predictors of reading accuracy when the background data is analysed in combination (using multiple regression techniques) are, at Year 2: the SEN rating, stage of English reached, free school meals status, ethnic group, the number of terms of infant education, chronological age and sex (in that order). For Year 6 the key factors that stand out are again the SEN rating, free school meals and stage of English with ethnic group the only other variable reaching significance. (paragraph 146)

A report the following year on the teaching of number in three inner-urban LEAs (Ofsted, 1997b) showed that:

> Girls generally performed as well as, or better than boys in Year 2, but in Year 6 girls' mean scores were below those of boys; the difference was greater in the mental test... The relationship between social factors, such as the incidence of free school meals, and pupils' performance was stronger in Year 6 than in Year 2. However, some schools serving disadvantaged catchment areas performed well... Pupils' level of fluency in English was a significant factor in their performance in number. The positive effect of fluency in English was more apparent in Year 6 than in Year 2, in both the written and the mental tests... The performance

of pupils from different ethnic groups showed considerable variation. Black African, Bangladeshi and Pakistani pupils achieved low scores in both Year 2 and Year 6. The small group of Chinese pupils performed well in comparison with all other ethnic groups. (pp. 5-6)

In 1997, the Effective Provision of Pre-School Education (EPPE) Project was set up by the DfEE as a longitudinal study to investigate over 3000 three- and four- year old children from their first entry into pre-school education to the end of KS1 and beyond; the study used both qualitative and quantitative methods (including multilevel modelling) to explore the effects of different types of pre-school provision on pupils' social, behavioural and cognitive development. The study also looked at the influences on the child of home environment and family characteristics such as gender, ethnicity, language, parental education and the educational environment of the home. (EPPE, 2003) In a symposium at the annual BERA conference in Edinburgh in 2003, Brenda Taggart described how pre-school children with SEN had been identified early by using the B Ability Scales of the General Cognitive Ability Test. The researchers had found a strong relationship between SEN and 'multiple risk' backgrounds; these they suggested were gender (boys had a higher percentage of SEN), EAL, large families, lower socio-economic status, lower parental education, and 'home' children (those with no or minimal pre-school experience).

At the same conference, Schagen and Benton presented a joint paper (2003) that looked at the relationship between school and background factors and progress from KS1 to KS2; the study had used the newly-available National Pupil Database (NPD), which combines value-added information about pupils' progression from the end of one key stage to the next with pupil background information derived from the Pupil Level Annual Schools' Census (PLASC). Two school variables were included in the multilevel model used for the research: school size and the percentage of pupils eligible for free school meals (FSM). Pupil-level variables were taken from the PLASC data sets, and related to gender, ethnicity, FSM, EAL and whether the pupil data were matched for

the end of KS1 and KS2. The study found that pupil stability (having been in the same school throughout KS2) had a positive relationship with all KS2 outcomes, whereas age had a negative relationship; that girls did better than boys (in value-added terms) for KS2 English only, but boys seemed to make better progress in mathematics and science; that pupils eligible for FSM made less progress than their peers; pupils with SEN - especially at the higher stages of School Action Plus and those with a Statement of SEN - made clearly reduced progress when compared to pupils with no SEN; and that pupils with English as an additional language (EAL) made more progress between the end of KS1 and KS2 than expected. Progress made by different ethnic groups appeared to be more variable, with White Other, Bangladeshi, Chinese and Other ethnic groups making more progress than expected; and Black Caribbean, Pakistani and Unknown ethnic groups making less progress than expected in value-added terms, based on the performance indicators included in the model. Looking at school-level background factors, it appeared that the percentage of pupils eligible for FSM had a negative relationship to KS2 outcomes, whilst the school size had a variable relationship, with no statistical significance.

At a conference in London in June 2004, Professor Tymms from the CEM Centre, University of Durham, spoke of his team's research and the problems that schools in England have with assessing pupils in the Foundation Stage. With BA and FSP data, he said, the critical influences are: the child's date of birth; whether the child has EAL or SEN (and what form the SEN takes); and what type of home background a child comes from. Typically, affluence (Tymms says this leads to good home discourse and the availability of books) affects early reading more than early mathematics development (which he says is dealt with more at school than at home). When assessing children in the Foundation Stage, it is important to find a 'window of opportunity': if this is missed, a pupil may be achieving at a certain level but this fact has not been recorded. Professor Tymms said that the total BA score - rather than scores for individual aspects - is the key indicator for future success, as it shows the

child's general capacity to do well in reading, writing and mathematics. (KCC used the score for Personal and Social Development as their key indicator when projecting pupil performance at the end of KS1, and the total BA score for projecting Year 6 performance.) He suggested that residual data as generated by the PIPS assessment package, giving value-added measures, was the most meaningful and valid form of data apart from randomised controlled testing,; other forms of data had less validity, with teacher assessment - being subjective - the least, and raw data from SATs not much higher.

In discussing value-added measures, Professor Tymms said that, until 2000, QCA had set the threshold levels for the NC subjects of English, mathematics and science at KS2 as equivalent to the attainment of pupils nationally in the immediately preceding year; he described this as a flawed system. Also, the TGAT Report of 1987 that led to the National Curriculum had suggested that the average pupil would achieve a NC level two at Year 2 and level four at Year 6: Professor Tymms said that the regression line nationally from KS1 is slightly below level 4 at KS2. He spoke of the 'three-year moving average' and gave the example of a class of 30 pupils: with a 27% confidence interval for the cohort, a success rate of 77% pupils achieving level 4 and above at the end of KS2 in one year would be comparable with the next cohort achieving 52% in the following year. Professor Tymms said that a trend graph - using the current data and those from the two preceding years - should be used to find the three-year moving average; for each year's data, the number of pupils in the cohort should be multiplied by the percentage achieved, these three sums should be added together and divided by the total number of pupils in the three cohorts.

Between 2003 and 2005, each primary school in Kent was sent a set of spreadsheets containing the names, date of birth and gender of each Year 3, Year 4 and Year 5 cohort, with their individual results at KS1 as NC levels and APS achieved; headteachers were instructed to have their class teachers complete the forecast sheets, to predict target NC levels for their pupils at the end of KS2, based on prior attainment and allowing a margin for 'challenge'.

These data were accompanied by a summary sheet showing, for the preceding three years, contextual information related to the school roll; the percentages of pupils with SEN and FSM; and the target set by the school and actual percentages achieved at or above the 'average' NC levels in reading, writing and mathematics (at KS1) and in English, mathematics and science (at KS2); and the scores as APS and the PANDA grades allocated to each Key Stage in each of the subject areas.

By September, 2005 data produced for schools by the Fischer Family Trust (FFT) were being given a high profile by Kent Management Information. The FFT was asked by central government to offer data analyses to 162 LEAs in England and Wales; the data sets used are based upon matched pupil data from KS1, KS2 and KS3; from GCSE/GNVQ data; and from PLASC data for all pupils in England and Wales. The key principles of the analyses that are presented to schools through their LEAs is that the data provide questions, not answers, and that the estimates are not targets but the vehicle for promoting discussion **about** targets. Using national pupil-based datasets, FFT produce 'predictive' models of likely achievement at pupil, school and LEA levels. Current models enable the calculation of four types of estimates: Type A are estimates of likely achievement if the progress made by a particular cohort was the same as the progress made by all pupils nationally; Type B are estimates based on progress by pupils in schools in 'similar schools'; Type C are estimates of the achievement that would be expected for a particular cohort if their progress were consistent with the overall improvement necessary to achieve national targets; and Type D are estimates of likely achievement if the progress made by a particular cohort was similar to that made by the pupils in the top 25% of schools nationally the previous year. (It is important to note that it is the progress, not the achievement, which is the key factor in all these estimates.) The concept of a 'similar school' is one which has a similar socio-economic background in terms of its pupils and their home circumstances, given as percentages - for the ward in which the majority of pupils live – of parents who are in employment and parents who have had a higher education.

CHAPTER 4: METHODOLOGY

4.1: RESEARCH DESIGN

Educational research is about the development of new knowledge and improved practices; there is, however, the problem of generalisability and the possibility of limited impact on classroom practice, especially where research findings contradict current educational policy. Theoretical - or rational research is concerned with thinking about the world rather than directly examining it; empirical research - using a quantitative or qualitative approach or a mixture of both - is based on the observation and recording of data and events and an analysis of trends and relationships which may then be used to justify a hypothesis or to predict likely outcomes.

Whilst a quantitative model is based on a linear style and assumptions that there are detailed theories to describe the world accurately, the qualitative form of empirical research relies less on statistical analysis - using data gathered from a range of sources - and takes a critical, interpretive approach. There may still be a linear style but this follows a logical, narrative argument, as in historical research. It may use the symbolic interactionist model to study the way in which people react in different ways when they are in different settings; it focuses on context and meaning, with individual actors at the centre. There is an interactive link between the researcher and the subjects of the study using, for example, an ethnographic or phenomenological design. Creswell (1998) describes qualitative research methods as a 'rigorous, systematic approach' (p. 5) used to 'explore a social or human problem' (p. 15).

There are four criteria to disciplined inquiry: internal validity (do the findings map the phenomenon?); external validity (can the findings be generalised?); reliability (can the findings be replicated or reproduced?); and objectivity (are the findings free from unacknowledged bias?).

Educational researchers also need to consider three key elements:

1. What is the nature of reality of the particular field of inquiry? (The ontological aspect.)

2. What is the nature of the researcher's own personal knowledge in the field, and what is the relationship between this and the knowledge that it is hoped to acquire from an examination of the literature and the research findings? (The epistemological aspect.)

3. How can the researcher obtain the desired knowledge and understanding; how can the data be collected and analysed? (The methodological aspect.)

My research was empirical, and involved the collection and analysis of both quantitative and qualitative data. The research model was 'testing and assessment' (see Cohen, Manion and Morrison, 2001, p. 80) and measured achievement and potential through comparing the summative scores of individuals and groups. From this model I developed a realistic time scale and broke down the investigation into specific purposes, each with a research question.

Ethics

All research necessitates ethical considerations based on the right to privacy and non-participation, to confidentiality and to anonymity. In addition, all participants - whether an individual, a group or a culture have the right to a full knowledge of the research and high expectations of the researcher, who must ensure no physical, emotional or psychological intrusion and maintain support for human dignity and welfare (see Cohen, Manion and Morrison, 2001, Chapter 2). When adults are invited to take part in a research project they are generally able to understand what is required of them and how it may impact on their privacy, their self concept and their self-esteem: when the respondents are young children, there is less understanding, Indeed, some may consider that to even study such young children is an invasion of their privacy: at what age, for example, can a child's consent be established? and is he or she

only agreeing to please the adult (especially if that adult is the class teacher or the headteacher). Of course, parents must also give their informed consent, and they may be concerned over many issues, not least that their child's name - and possibly photograph - may be published. Ethical considerations were not problematic in this study: informed consent - from the pupils or their parents - was not required, and privacy and anonymity will be preserved, as the data are non-sensitive and non-confidential, being entered into the software program as anonymous variables and cases so schools and individual pupils cannot be identified.

Choosing the sample

One of the earliest considerations within the chosen research design was related to gatekeepers and access, and to the size and type of sample that would be required. I had already identified three schools where the headteacher was a friend as well as a colleague, and with whom I share professional discussions on a weekly basis. We four schools form a small cross-district consortium which has been in operation since 1999; we share expertise through regular meetings of our subject leaders, senior management teams and headteachers and through sporting fixtures. The headteachers of the three schools were keen to participate in my research; they, too, were interested in whether benchmark data could be used to predict future achievement, and how - both individually and collectively - we could use the data to support pupils' learning through the Foundation Stage and KS1.

The sample would be convenient - I would be using schools with geographical proximity and it would probably illustrate a typical case, that is, I would expect these four schools to form a representative sample in terms of diversity of ethnicity, home background and language, and special educational needs. The size of the sample was directly related to the number of pupils in the three cohorts chosen, namely 451 across the four schools. Cohen, Manion and Morrison (2001) suggest that 'researchers must obtain the minimum sample size that will accurately represent the population being targeted.' (p. 93). These 451

cases represent a sampling error of 5% with a confidence level of 95% in a population of 1,000,000; where 'the larger the number of cases there are in the wider, whole population, the smaller the proportion of that population can be which appears in the sample.' (p. 92). There is, of course, the problem of sample mortality: some pupils for whom I had collected benchmark data in the Foundation Stage might no longer be at the school when I collected their scores at the end of Year 2 while, conversely, some pupils might have joined the school after the initial data collection.

4.2: REPRESENTATION OF DATA

As a Head Teacher, I am a practitioner who implements government policy. I am not a statistician, and while I might have some numerical fluency and statistical insights, I have declined to undertake this research as a highly statistical exercise. I approached this thesis having in mind my professional role within my school and within the consortium of four schools that work together, where we support one another in implementing government policy; I have tried to adopt, in this presentation, a genre which will be meaningful to them.

The quantitative and qualitative data I collected have therefore been represented as narrative and tables   than graphs and figures, and as projected and actual Average Points Scores (APS) and National Curriculum (NC) levels, as these are data that my fellow headteachers understand and use on a daily basis. In my analysis of the data I have included some basic statistical concepts such as measurement of the central tendency   as mean, median and mode – together with ratio to show pupil-to-teacher and boy-to-girl relationships, but have deliberately avoided representing the data using terminology that might be unfamiliar to my colleagues.

4.3: TRIANGULATING THE DATA

In order for me to understand the context for my thesis, I decided to triangulate the data. This method - according to Cohen, Manion and Morrison (2001) helps the researcher to 'map out, or explain more fully, the richness and complexity of human behaviour by studying it from more than one standpoint... by making use of both quantitative and qualitative data.' (p. 112). I would be collecting, collating and analysing quantitative data sets for individual pupils, class groups and schools, and felt that the addition of qualitative data would help me to answer my third research question: *Can benchmark data be used to support pupils' learning?* This qualitative data took three forms: a semi-structured questionnaire to survey the teachers and teaching assistants in the four schools; lesson observations in each class, where I could see for myself what assessment practices were being used; and discussions with the practitioners over any issues that had arisen as a result of their responses to the questionnaire. I arranged to visit each of the five classes for a morning; all of the Reception class practitioners are well-known to me through our close consortium links and they were happy for me to scrutinise their assessment procedures and discuss the implications with them afterwards.

4.4: DATA COLLECTION

This investigation looks at data from four primary schools in West Kent; the pupils who formed the sample for the study were at that time in Years R, 1 and 2 (cohorts 2003-2010, 2002-2009 and 2001-2008 respectively). Two of the primary schools have similar profiles; the other two form a contrast in terms of the pupils' ethnic, linguistic and home background. The focus for collection of the quantitative data was confined to results in English (reading and writing) and mathematics, as these are the subjects for which targets are required to be set for the end of KS 2 (Year 6); they are also used by the DfES as value-added measures from school entry in Year R to the end of Year 2, and from Year 2 to the end of Year 6.

Collecting qualitative data

The qualitative data were collected in three forms: one was a survey of the class teachers in the five Reception classes (School B had two parallel classes in their Foundation Stage). The instrument used was a semi-structured questionnaire (see Appendix III) which was sent by post to each of the five Foundation Stage teachers, following a verbal request I had made through my headteacher colleagues. For the survey I used my experience as a teacher and researcher over many years to decide on the questions: I wanted to know a) the extent to which each of the class teachers used benchmark data - either the former BA or the new FSP - to inform their planning for individual pupils, groups or the whole class, b) what form their assessments took, c) whether they grouped the children by ability or otherwise, and d) their opinion of the whole process of teacher assessment and evidence-gathering. School B had two Reception classes, which resulted in five pieces of evidence in the form of completed survey sheets. I typed the transcript from each completed survey sheet and made an exploratory analysis of the categories; these data were then tabulated (see Table 12). The other two sets of qualitative data were collected through observations made in each of the Reception classes of the four schools and discussions with the class teachers and teaching assistants on issues that they had identified on their questionnaires.

Collecting quantitative data

For the Year 2 cohort, I collected quantitative data related to the pupils' scores as measured by Baseline Assessment (BA), together with their scores as measured against National Curriculum (NC) levels at the end of KS1. For the Year R and Year 1 cohorts, I collected quantitative data from the pupils' Foundation Stage Profiles (FSPs) that had been completed by the class teacher at the end of the respective academic years. (The Year 1 cohort was the first to be assessed using only teacher assessment - Kent was one of the LEAs to pilot this in the summer of 2004 - whilst the Year R cohort became the first to be similarly assessed once KS1 SATs were abolished nationally for 2005.) BA

data were collected for each cohort as separate scores for the strands of the Desirable Learning Outcomes, the total BA score and standardised scores (SS); data from the FSP were collected for each of the six areas of learning of the Early Learning Goals (ELGs) and the total FSP score. I needed to build up a profile for each cohort in each school and decided to collect data relating to the pupils and their background: their name; date of birth; gender; level and category of special education need (SEN); and whether they had English as an additional language (EAL). I asked each headteacher about other aspects of the school's profile: the number of terms each pupil had spent in the Reception class, their pre-school experience, and whether they were designated Travellers, asylum seekers or a Looked After Child (LAC - see Chapter 6.1); in addition I collected data on the mobility of the pupil population for each school for Years R, 1 and 2.

I designed a matrix detailing the quantitative data that I wished to collect, with space to record the date of collection and any comments. Some data were immediately available from my initial interviews with the headteachers: those relating to pupil name, date of birth and gender from the school registers; others relating to the pupils' SEN stages and categories and to their ethnicity and home background took a little longer to acquire, as they required access to the school's computer database for the applicable codes, based on the annual PLASC return to the DfES.

Numeric values needed to be assigned to the KS1 National Curriculum levels achieved; these were, in ascending order, from W (meaning 'working towards level 1'), through level 1 (no sub-levels), 2c, 2b, 2a, to level 3 (again, no sub-levels), and I decided to use the Average Points Score (APS) assigned by the DfES. I would be analysing the data against value-added measures, according to the guidance from KCC:

> Value-added analysis relates to progress based on prior attainment.
> It is therefore important to use a prior attainment indicator that
> bears a strong relationship to the outcomes which are being
> considered. Research has shown that the best predictors of future
> attainment are **combined** prior attainment measures. The combined

score used throughout the Baseline to KS1 analysis is Baseline total, which takes into account all elements (5 English, 2 mathematics and 1 PSD) and ranges from 0 to 56 inclusive. For the KS1 to KS2 analysis, average KS1 points score was used, with values taken from an average of Reading, Writing and mathematics...The KS1 Points Scores are given below. These are identical to those used in the QCA *Autumn Package*. (KCC, 2003a).

According to the guidance:

> The [data] may be used for projecting KS1 results from total Baseline score (for pupils who have not yet taken the KS1 tests)...This approach can contribute to the target-setting process but the following points need to be considered:
> - the achievement profile of the pupils from whose results the regression equation is derived (reference group) should be mirrored by that of the pupils for whom projections are being made (the test group);
> - the tests taken by the reference and test groups must be identical.
>
> All points scores calculated in this way should be regarded as 'projections' rather than 'predictions'; they more accurately reflect minimum expectations for individual pupils since there is no inherent element of challenge...The difference between an individual's projected and actual scores is the pupil residual. (KCC, 2003a)

The stage of SEN for each pupil was coded as either 0, meaning no special educational need, 1 (Early Years Action) or 2 (Early Years Action Plus). The Code of Practice (DfES, 2001b) identified pupils with special educational needs in early years settings:

> The key test for action is evidence that the child's current rate of progress is inadequate... Where progress is not adequate, it will be necessary to take some <u>additional</u> or <u>different</u> action to enable the child to learn more effectively... (Chapter 4:13).

Special Educational Needs Coordinators (SENCos) were requested to:

> devise interventions that are <u>additional to</u> or <u>different from</u> those provided as part of the setting's usual curriculum offer and strategies *(Early Years Action)*. (chapter 4:20). *'Early Years Action Plus* is characterised by the involvement of external support services who can help early education settings with advice on new IEPs and targets... (Chapter 4:29).

The various categories of SEN needed to be assigned values; from a pilot study undertaken with a group of eight schools in my Local Learning Group (LLG) - which included the four schools in my investigation (Hyne, 2003) - I was able to identify the most frequently-occurring categories, and coded them as shown in Table 2 (the abbreviations are those used in the schools' annual PLASC return).

Table 2: Coding for categories of special educational needs

SEN category	Abbreviation	Code assigned	SEN category	Abbreviation	Code assigned
Moderate learning difficulty	MLD	1	Severe learning difficulty	SLD	2
Speech, language and communication needs	SLCN	3	Behavioural, emotional and social difficulties	BESD	4
Autistic Spectrum Disorder	ASD	5	Visual impairment	VI	6
Hearing impairment	HI	7	Physical disability	PD	8

Another key factor in my data analysis was the ethnicity and home language of the pupils in the data sets; I decided not to collect data on the pupils' ethnicity as home language appeared to have a greater bearing on achievement, according to my interviews with the headteachers. From my earlier research with the LLG I was able to identify the most common languages spoken in the home: one pupil with a PLASC code of AIND (for 'Asia: Indian') had a PLASC language code of 'Punjabi' whilst another Asian Indian spoke English at home; an ethnic code of APKN ('Asia: Pakistan') gave either 'English, 'Punjabi' or 'Urdu' as the home language. Some pupils in the survey were from mixed race: one pupil with an ethnic code of MWBC ('White/Black Caribbean') had English as the home language; another pupil, with an ethnic code of BAFR ('Black African') had 'Other' as the home language. This code of 'Other' was used for pupils with home languages in Kosovan, Creole, Thai, Lithuanian, Romany and Greek. (The DfES has not yet included these ethnic minority home languages

on its database, although there are fifty-six different ethnicity codes.) The following table shows how I coded the various home languages.

Table 3: Coding for home languages

Home language	Code assigned	Home language	Code assigned	Home language	Code assigned	Home language	Code assigned
Punjabi	1	Bengali	2	Hindi	3	Urdu	4
English	5	Cantonese	6	Portuguese	7	Turkish	8

(To the home language of 'Other', representing as it did so many different languages, I assigned the value 0, which also represented 'Refused' or 'Not Known'.)

I knew that pupils who had a high mobility - either as designated Travellers or asylum seekers, or those with particular home backgrounds - would need to be identified within the data sets; such pupils would affect the value-added aspects of my data analysis, as data for them would almost certainly be incomplete. The only indication that any pupils in the survey might be Travellers or asylum seekers was from reports run from the schools' databases which showed 'Ethnicity information held (other than British)'. The variables on the school databases, 'Ethnic description' and 'Home language', gave some clues: ethnic code 'WKOS' was a white Kosovan pupil with 'Other' as the home language - was this child an asylum seeker? Two pupils with the ethnic group shown as 'WROM - Gypsy/Roma' (one with 'English' and the other as 'Other' for the home language) were presumably from Traveller families (did the second child speak Romany?)

I had identified three other variables for my data collection: the number of terms the pupils had spent in the Reception class, their pre-school experience, and the mobility of pupils through Years R, 1 and 2. When I interviewed the headteachers, the first two variables were too difficult to include as there was no easy way to access this information. The Kent Record of Transfer - which gave these data - was only available for the whole cohort in schools A and B

who each had an attached nursery. Although the record was meant to be sent on to the new Reception class teacher from the pre-school setting, this had not happened for about a third of the pupils in Schools C and D, so these were discounted from my analysis. The third variable, that of pupil mobility, was ascertained through my interviews with the headteachers. In the case of School A, there was a great deal of missing data from pupils who had either left the school between the Reception class and Year 2, or who had joined the school between these periods. For the former, the pupil had already been entered as one of my cases, with their BA or FSP data; for the latter, I had to decide whether or not to include them on the SPSS database. In both cases, the data would be incomplete, so I decided not to include the newcomers, whilst leaving the existing data for the leavers on the spreadsheet as incomplete data. A fourth variable that I had considered earlier, and which related to whether or not any pupil had been identified as 'gifted and talented', was also discounted as there were only three pupils so identified out of 451 cases (see Chapter 6.1). As I collected the data, other aspects became apparent that I had not previously considered; for example, I noticed that School B, Class 2 had one set of twins in Year 1 and another set of twins in Year 2 but in different classes; School D had one set of twins in Year R. The data for these three sets of twins in my sample will be examined in Chapter 6.1.

As I collected the various data sets, they were entered into the SPSS variables as individual cases; some pupils had the same first initial, so I entered their first name in abbreviated form as a string variable; the school name was also entered as a string variable. Other data were entered in numerical form: dates of birth; SEN stage and category; home language coded as shown in Tables 1 and 2; and gender, coded as '1' for male and '2' for female (the accepted coding when using SPSS). School B had two classes in each cohort, and I entered the data as discrete groups, making five groups for comparative analysis. The BA data were entered as raw scores for each of the elements mentioned above, from 0 to 7 inclusive, and the BA total, from 0 to 56 inclusive; the FSP scores as raw scores within each of the six areas of learning of the Foundation Stage, from 0

to 117 inclusive. The KS1 NC levels were entered as Average Points Scores (APS) for the three aspects of reading, writing and mathematics, and again as a total score, for the Year 2 cohort.

Summative data for the end of KS1 for the Years R and 1 cohorts would not be available for my research, due to the short time-frame of the study, but the data from the Year 2 cohort was a key area of my investigation, as it would help to answer my first and second research questions: *Can benchmark data be used to predict future performance?* and *What are the implications of using value-added data as measures of pupil performance?*

I used the SPSS program as a database to help me analyse the data from each of the three year groups and four schools; I used the program to split the data into variables which I then analysed manually. I used the *'Pupil Achievement Levels: Baseline to Key Stage One'* table produced by KAS, 1999 to project likely outcomes for the Year 2 cohort in reading, writing and mathematics NC levels at the end of KS1, and compared these with the actual scores achieved by the pupils. Using the Average Points Scores (APS) used by the DfES, I made a set of tables showing the number of pupils who were projected to achieve NC levels 1 through to 3. The DfES allocates Average Points Scores thus: 9 for Level 1; 13 for Level 2c; 15 for Level 2b; 17 for Level 2a; and 21 for Level 3. The data - for each of the three strands of reading, writing and mathematics will be found in Tables 8, 9 and 10.

For each of the classes in Year 2, I examined the number of pupils who had either exceeded or not achieved the projected NC levels: these data will be found in Table 11. I analysed these data further for each of the strands of reading, writing and mathematics, and for each group, by looking at the individual case profiles - under gender, SEN stage and category, EAL, and term of birth - to try and find a pattern or correlation. For each group, I examined the ratio of boys to girls, and the number of pupils who were summer-born, had EAL or SEN. For each individual pupil (case) I looked at the projected

70

Average Points Score and the actual APS achieved at the end of Year 2; I also examined the total score for BA personal, which the DfES says gives a better indication of potential achievement than do the scores for reading, writing, mathematics, or the other three categories that made up the Baseline Assessment package. I used the SPSS program and Excel spreadsheets to help me find the mean, median and mode for the BA personal scores, and for the projected and actual APS for reading, writing and mathematics; these data are examined in Chapter 5.2.

CHAPTER 5: FINDINGS

5.1: ANALYSIS OF QUANTITATIVE DATA

This first set of findings relates to the basic group and pupil profiles; as mentioned throughout this research, the 'proxy indicators' used by government (of gender, home language, special educational needs, and summer-born pupils among others) may or may not have a causative effect on a pupil's subsequent achievement. These group and pupil profiles, therefore, might give an indication of likely outcomes: they **might** be predictors of KS1 Average Points Scores and thus of value-added data for the pupils concerned.

Data for Year 2

Data for the Year 2 cohort were analysed in detail, as this cohort - for whom I had projected and actual scores - was key to answering my first research question: *Can benchmark data be used to predict future achievement?* (see Table 4). School A had one class of 30 pupils, although only 21 of these were included in the data analysis due to a high level of pupil mobility; the male: female gender balance was 14: 7. School B had two classes: the first comprised 29 pupils with a gender balance of 18: 11, the other had 28 pupils with a gender balance of 11: 17; School C had one class of 34 pupils, with a gender balance of 16: 18, whilst School D had one class of 30 with a gender balance of 12: 18. According to HMI (Ofsted, 1996) these differing gender balances might be expected to create differing classroom climates, and thus might affect pupil progress and subsequent achievement.

The percentage of pupils with special educational needs (SEN) was 48% in School A; 24% in School B, class 1 and 14% in class 2; 15% in School C; and only 3% in School D. These pupils presented with SEN in categories moderate learning difficulties (MLD), severe learning difficulties (SLD), speech, language and communication needs (SLCN), and behavioural, emotional and

social difficulties (BESD) at both 'School Action' and 'School Action Plus'. It should be noted that, in an average class size of 30 pupils, 3% represents just one pupil. This becomes important when considering value-added scores: for example, a residual score of minus 6% may look significant, but would only indicate that two pupils did not do as well as had been expected, and by just one National Curriculum sub-level; or, in fact, that one or both pupils had been absent for the test, thus giving a 'nil' score and affecting the total percentage.

The differences between the pupils with EAL across the five class groups were equally marked: School A had 43%; School B had 38% in class 1 and 36% in class 2; School C had 6%; and School D 0%; again, this might be a significant factor in the relationship between their projected and actual scores. The home languages spoken by the pupils in School A were Punjabi, Bengali, and 'Other', although the highest percentage (63%) spoke English in the home; in School B, the data showed home languages of Punjabi, Bengali, Hindi, Urdu, and 'Other', with 63% speaking English at home; School C had just two Punjabi-speaking pupils, whilst School D had pupils whose home language was 100% English. (Data for classes 1 and 2 were not substantially different.)

The percentage of pupils born during the three terms of autumn, spring and summer across all five classes was 29%, 14%, 36%, 21% and 37% (for autumn-born pupils); 14%, 31%, 25%, 29% and 27% (for spring-born); and 57%, 55%, 39%, 50% and 37% for the summer-born pupils. Being summer-born, and therefore having fewer terms in a Reception class, is considered by some to be detrimental to high achievement, and as such is used as a proxy-indicator by central government statisticians (see Chapter 6.1).

The range and mean of BA total scores for this Year 2 cohort were similar across all the groups: scores in School A ranged from 24 to 51, with a mean of 39; School B, class 1 from 25 to 50, with a mean of 38, and class 2 from 28 to 51 with a mean of 39; School C from of 31 to 48, with a mean of 38; whilst

School D ranged from 31 to 50, with a mean of 40. All these data sets are shown in Table 4.

Table 4: Basic group and pupil profiles for the Year 2 cohort

	School A	School B class 1	School B class 2	School C	School D
Number of pupils in the cohort	21	29	28	34	30
Gender balance male: female	14: 7	18: 11	11: 17	16: 18	12: 18
% of pupils with SEN	48	24	14	15	3
% of pupils with EAL	43	38	36	6	0
% of pupils born during the autumn	29	14	36	21	37
% of pupils born during the spring	14	31	25	29	27
% of pupils born during the summer	57	55	39	50	37
BA total scores ranged from .. to ..	24 to 51	25 to 50	28 to 51	31 to 48	31 to 50
Mean of BA total scores	39	38	39	38	40

The range and mean of BA total scores for this Year 2 cohort were similar across all the groups: scores in School A ranged from 24 to 51, with a mean of 39; School B, class 1 from 25 to 50, with a mean of 38, and class 2 from 28 to 51 with a mean of 39; School C from of 31 to 48, with a mean of 38; whilst School D ranged from 31 to 50, with a mean of 40. All these data sets are shown in Table 4.

Data for Years 1 and R

Although most of my research has been centred around the Year 2 cohort, the data sets for the Year 1 and Year R cohorts have also been collated, to identify any patterns and correlations either between cohorts or between group and pupil profiles; these data are shown in Tables 5 and 6.

Table 5: Basic group and pupil profiles for the Year 1 cohort

	School A	School B class 1	School B class 2	School C	School D
Number of pupils in the cohort	28	26	25	32	30
Gender balance male: female	13: 15	16: 10	16: 9	20: 12	15: 15
% of pupils with SEN	36	19	28	9	7
% of pupils with EAL	32	50	40	9	10
% of pupils born during the autumn	54	31	32	28	20
% of pupils born during the spring	21	23	44	41	17
% of pupils born during the summer	25	50	32	31	63
FSP total scores ranged from .. to ..	46 to 109	55 to 106	58 to 111	62 to 117	60 to 109
Mean of FSP totals	86	87	84	96	89

Table 6: Basic group and pupil profiles for the Year R cohort

	School A	School B class 1	School B class 2	School C	School D
Number of pupils in the cohort	28	25	25	36	30
Gender balance male: female	16: 12	17: 8	15: 10	18: 18	18: 12
% of pupils with SEN	39	8	0	6	17
% of pupils with EAL	29	52	36	11	0
% of pupils born during the autumn	42	24	36	22	33
% of pupils born during the spring	33	36	20	36	20
% of pupils born during the summer	25	40	44	42	47
FSP total scores ranged from .. to ..	82 to 114	53 to 105	51 to 102	70 to 117	52 to 112
Mean of FSP total scores	100	84	72	97	81

Comparing the profiles of all four schools

When the three cohorts are compared across the four schools, there are some obvious similarities and differences which might influence the KS1 APS outcomes. Whilst a male: female gender balance of 16: 12 or even 18: 12 is manageable (with 11: 17 and 12: 18 preferable for a number of reasons), a class containing 14 boys and only 7 girls, or 16 boys and only 9 girls, or 18 boys and 11 girls is hard work (again, for a variety of reasons, including the 'male dominance' phenomenon, according to colleagues). When these latter data are added to those for SEN (48%, 28% and 24% respectively), and to EAL (43%, 40% and 38% respectively), together with the percentage of summer-born pupils (who are frequently socially immature compared to their peers), maybe teachers can be forgiven for saying - as do my colleagues every time they are asked to justify our TA or SATs results - that it is the **cohort** that matters: that whatever strategies are put into place, they can be nullified by the basic profile of that particular group of pupils.

No pupil in the Year 2 cohort of any of the four schools achieved the 'top' score of 56 in BA, but one pupil (female) in the Year R and one pupil (male) in the Year 1 cohorts of School C achieved the highest FSP score of 117. These pupils were each in a cohort with a low (but not the lowest) percentage of SEN; the Year 1 pupil was in the cohort with the lowest EAL percentage; neither pupil had SEN; both spoke English at home; one was autumn- and one spring-born - all favourable indications of high achievement.

The Year R pupil was in the second-largest size class, which would have decreased the adult: pupil ratio, and might have been expected to have affected pupil progress (the impact of class size at KS1 will be examined in the next chapter). I did not collect data on other adults in the classroom; some schools have full-time support in the form of one or more qualified nursery nurses, others have full- or part-time teaching assistant support, whilst others have parent or governing body volunteers or no support at all.

What could account for the fact that each of these pupils was able to achieve the highest score when the Foundation Stage Profile (FSP) was completed? The teacher was confident that the child demonstrated clear evidence of having achieved the criterion in every one of the 117 statements of achievement across thirteen strands of the six Early Learning Goals (ELGs). The FSP is completed at the end of the Reception year, unlike Baseline Assessment (BA), which used to be completed during the first seven weeks of the child starting school: there is no way of knowing whether or not the child came into school with high achievement in all areas of the ELGs, or accessed his or her knowledge, skills and understanding over the course of the school year. In order to ascertain which of these facts is the more likely, I should need to make a further study - and assessment - of the child at his or her pre-school setting or within the home. It would be useful to know whether these two pupils went to a nursery school, or a pre-school playgroup, or learned at home with Mum (and/or Dad, siblings, extended family) but, as explained earlier, there was no way that I could collect these data.

5.2:   KEY FINDINGS - YEAR 2 COHORT

My key findings relate to the Year 2 cohort, for whom I was able to analyse and compare data from the pupils' BA scores, their projected National Curriculum (NC) levels and actual results - as Average Points Scores (APS) - achieved at the end of KS1. For this analysis, I used the data I had collected on pupils' total BA scores and related these to the *'Pupil Achievement levels: Baseline to Key Stage One'* table (KAS, 1999) which projected a pupil's NC levels in reading, writing and mathematics at the end of KS1 and is reproduced - exactly as sent to schools by Kent Advisory Service - in Table 7. The table was sent 'To Reception teachers of all primary schools' by Kent Advisory Service with a letter dated 9 July 1999. Practitioners were told that:

It will shortly be possible to establish the relationship between attainment in baseline assessment and children's subsequent achievement at different levels in Key Stage 1 national assessments in English and Mathematics based upon actual patterns of performance data. This will enable the development of chances graphs for target-setting for individual pupils at the end of Key Stage 1...This year, however, guidance upon expected patterns of progression from baseline to Key Stage 1 is provided. This has only been produced by matching the percentages of pupils achieving different total baseline scores against the distribution of levels achieved in recent Key Stage 1 test. We emphasise that this is for guidance only, for use within the broader context of teachers' professional judgements. (KAS, 1999)

Table 7: Pupil achievement levels: baseline to KS1

BASELINE TOTAL SCORE	ACHIEVEMENT IN KS1 READING
< 33	L1 or Below
>= 33	L2C or Above
>= 36	L2B or Above
>= 39	L2A or Above
>= 42	L3

BASELINE TOTAL SCORE	ACHIEVEMENT IN KS1 WRITING
< 32	L1 or Below
>= 32	L2C or Above
>= 39	L2B or Above
>= 44	L2A or Above
>= 48	L3

BASELINE TOTAL SCORE	ACHIEVEMENT IN KS1 MATHEMATICS
< 30	L1 or Below
>= 31	L2C or Above
>= 36	L2B or Above
>= 41	L2A or Above
>= 44	L3

(KAS, 1999)

It will be noted that the total Baseline Assessment (BA) scores project differing achievement in reading, writing and mathematics: to obtain Level 3 (the highest possible at KS1) in reading, a pupil would need to score a total of 42 points; in writing, 48 points; and in mathematics, 44 points. The reason for these differences was explained in the letter (see preceding paragraph). It would appear from this that the scores for reading were the easiest to achieve, with mathematics next and writing the most difficult. It would also appear from the chart that Level 1 can be achieved by a lowest score of 30 for mathematics, 32 for writing and 33 for reading: an apparent anomaly when compared to the scores for Level 3, but based on actual KCC data. However, 'L1 or Below' on the chart indicates that a child may have scored between zero and 33 (or 32, or 30), which could mean that he or she has been assessed as obtaining Level 1 or is 'working towards' Level 1 (and given, at that time, a level 'W'). A pupil is expected to achieve a minimum of two complete NC levels (six sub-levels) during the two years spent in KS1: with a nebulous score of 'W' at the end of the Key Stage, how, then, can value-added criteria be applied between then and the end of KS2? With the 'P' levels that have recently been introduced (see Chapter 2.1), this problem should be eliminated.

The analysis of the data for the Year 2 cohort - NC levels projected from BA total compared to APS outcomes is key to answering my first research question: *Can benchmark data be used to predict future achievement?* Tables 8 to 10 present the data for each school, with the number of pupils who were projected from their total BA scores to achieve the six NC levels shown - using the tables given to schools (KAS, 1999) - compared to the actual NC levels obtained. (The bold figures in Tables 8, 9 and 10 indicate the important category of NC level 2 and above; these data are used by Ofsted to evaluate three-year trends and form a key aspect of each school's annual PANDA report.) Of the data presented, it will be noted that some pupils appear to have achieved the projected results, but the data are deceptive: an analysis of the case numbers showed that the data related to different pupils. Surely this suggests that BA data (and, possibly also FSP data) cannot be used to predict an

individual pupil's future achievement, merely a cohort's overall percentage scores at the various NC levels.

Year 2 pupils achieving the projected NC levels in reading (see Table 8.)

Four out of the five class groups projected more pupils to obtain Level 3 in reading than actually did so; this level is not divided into sub-levels. Within the Level 2 category (Levels 2a, 2b and 2c), of the five class groups, four showed a higher number of pupils achieving the grades than had been projected: 11 instead of 7 for School A; 19 instead of 13 for School B Class 1; 16 instead of 11 for School B Class 2; and 31 instead of 23 for School C. These data, combined with the data for Level 3 given earlier, are reflected in the numbers of pupils achieving Level 2 and above (Level 2+), whereby each of the first four groups show the same number or more pupils achieving the grades than had been projected. School D had one more pupil achieving Level 3 than projected, but six more achieving Level 1 than projected: these data, combined with the lower number of pupils achieving Level 2 than projected (12 instead of 19) meant that School D had six fewer pupils at Level 2+. A closer examination of the data contained in the cases (the individual pupil data) revealed that, whilst four pupils were projected to achieve Level 2b (APS 15) in School A and four did so, **these were not the same pupils.** Two had indeed been projected to achieve Level 2b, but the third had been projected a Level 2a and the fourth a Level 3; the other two pupils projected a Level 2b only actually managed a Level 1. These anomalies are to be found throughout the data sets for all the class groups and all three of the strands of reading, writing and mathematics; what looks at first sight as 'correct' projections is true for the overall school percentages - as shown in DfES statistics and league tables - but not for the individual pupils, so value-added measures can be applied on a cohort but not an individual basis.

Table 8:   Projected and actual number of Year 2 pupils achieving NC levels for READING

NC levels	School A		School B Class 1		School B Class 2		School C		School D	
(Average points score)	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual
L 3 (21)	7	4	10	4	10	5	8	2	10	11
L 2a (17)	3	1	2	10	4	5	8	13	7	6
L 2b (15)	4	4	5	6	5	10	10	9	8	6
L 2c (13)	0	6	6	3	2	1	5	9	4	0
L 1 (9)	7	6	2	2	4	4	3	1	1	7
L W	0	0	0	0	0	0	0	0	0	0
Level 2+	**14**	**15**	**23**	**23**	**21**	**21**	**31**	**33**	**29**	**23**

N.B. The bold figures indicate the category of NC level 2 and above; these data are used by Ofsted to evaluate the three-year trends which form a key aspect of each school's annual PANDA report.

Year 2 pupils achieving the projected NC levels in writing (see Table 9).

In School A, seven more pupils achieved Levels 2a, 2b and 2c than had been projected (19 instead of 12); two pupils had managed a Level 3 (although only one was the same pupil who had been projected to do so) and of the seven pupils projected to achieve a Level 1, four did so whilst the other pupils got a Level 2c. The data for School A show that 100% of pupils in the class group achieved the important category of Level 2+ compared to the 67% that had been projected to do so; this is pleasing when government concerns are centred at present on the low levels achieved in writing at all key stages in the UK. Across the five class groups, twelve instead of the projected nine pupils achieved Level 3, and in schools A and B (both classes) more pupils than projected got a Level 2. In schools C and D, more pupils got Level 1 than had been projected (6 instead of 3, and 10 instead of 1 respectively), resulting in these two schools having a lower Level 2+ percentage than had been projected: 82% rather than 91% for School C and - even more worryingly - 67% instead of the projected 97% for School D. I use the word 'worryingly', because Level 2+ is the important category for DfES purposes, so the value-added data from Baseline to KS1 for School D would be seriously depressed.

As in reading, an examination of the case numbers showed that the data was not matched for those pupils projected to achieve certain NC levels; the projected percentages for each level may have been met for the cohort, but not for the individual pupils.

Table 9: Projected and actual number of Year 2 pupils achieving NC levels for WRITING

NC levels (Average points score)	School A		School B Class 1		School B Class 2		School C		School D	
	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual
L 3 (21)	2	2	1	3	4	3	1	1	1	3
L 2a (17)	3	3	4	9	5	5	4	8	3	5
L 2b (15)	5	5	7	4	5	8	11	10	13	8
L 2c (13)	4	11	11	7	7	5	15	9	12	4
L 1 (9)	7	0	2	2	4	4	3	6	1	10
L W	0	0	0	0	0	0	0	0	0	0
Level 2+	**14**	**21**	**23**	**23**	**21**	**21**	**31**	**28**	**29**	**20**

N.B. The bold figures indicate the category of NC level 2 and above; these data are used by Ofsted to evaluate the three-year trends which form a key aspect of each school's annual PANDA report.

Yr 2 pupils achieving the projected NC levels in mathematics (see Table 10).

The data for mathematics show that, of the 28 pupils projected to achieve Level 3, only 23 did so. The number of pupils who had been projected to achieve Level 2 across all five class groups was exceeded in schools A and B Class 2, matched in schools B Class 1 and C, and not achieved in School D. However, School D had two more pupils achieving Level 3 than projected and School B Class 1 had one more pupil, whereas schools A and C each had two fewer pupils at Level 3, and School B Class 2 had four fewer pupils than projected at Level 3. When the percentages for Level 2+ are compared, it will be seen that School B Class 1 achieved 100% in mathematics (from a projected 96%) and School A achieved 81% (from 76% projected. This is in contrast to the other three class groups, where the percentages were less than projected: 84% instead of 92% for School B Class 2; 94% rather than 100% for School C; and 97% instead of 100% for School D.

A closer examination of the case numbers showed clearly that, once again, there was a mismatch between the actual pupils who were projected to gain a certain NC level and those who did so: in fact, only nine of the 23 pupils projected to achieve a Level 3 actually did so.

Table 10: Projected and actual number of Year 2 pupils achieving NC levels for MATHEMATICS

NC levels	School A		School B Class 1		School B Class 2		School C		School D	
(Average points score)	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual	Projected	Actual
L 3 (21)	5	3	5	6	9	5	5	3	4	6
L 2a (17)	3	2	5	8	3	8	6	6	9	4
L 2b (15)	6	0	7	4	7	4	15	13	12	7
L 2c (13)	2	12	7	7	4	4	8	10	5	12
L 1 (9)	5	4	1	0	2	4	0	2	0	1
L W	0	0	0	0	0	0	0	0	0	0
Level 2+	**16**	**17**	**24**	**25**	**23**	**21**	**34**	**32**	**30**	**29**

N.B. The bold figures indicate the category of NC level 2 and above; these data are used by Ofsted to evaluate the three-year trends which form a key aspect of each school's annual PANDA report.

Pupils exceeding or not achieving projected NC levels

Why does it appear from Tables 8, 9 and 10 that projections at Level 2 and above (2+) for School B (both class groups) are totally accurate for reading and writing? Why are School A's results at Level 2+ higher than projected in reading and mathematics - and considerably higher in writing - than the other class groups? Why do the data for School D show much higher projections than actual in reading and writing? Does this mean that the school had used intervention or 'booster' strategies to extend the writing skills of this particular cohort? If so, why had these apparently had no effect on pupils working at the lower levels of achievement? An examination of the interventions implemented in all four schools might reveal further information, but these data were not part of my study. It appeared that National Curriculum levels and Average Points Scores achieved at KS1 – and used by government as 'prior attainment' for projecting KS1 to KS2 value-added data - were only a partial indicator of a pupil's potential; I needed to analyse the data in more depth, so I looked at the number of pupils who had either exceeded or not achieved the projected levels, in all four schools and across all three subjects (See Table 11).

Table 11: Number of pupils who had exceeded or not achieved projected NC levels

	School A		School B Class 1		School B Class 2		School C		School D	
	+	-	+		+		+	-	+	
Reading	4	8	6	10	6	9	7	14	8	12
Writing	6	6	14	6	9	7	11	10	9	12
Mathematics	4	12	11	7	6	10	7	14	6	11

N.B.   + indicates exceeded projected NC levels; - indicates did not achieve them

These data ignored those individual pupils that had matched their projected grades, and revealed that, in reading, more pupils had not achieved than had exceeded their projected grades across all five class groups. In mathematics, the same was true for all but School B Class 1; in writing, the data showed no pattern across the class groups. As a corollary, the greatest match between projected and actual levels achieved was in School C, where 13 pupils out of 34 were matched in reading, writing and mathematics. Did this mean that the teachers in School C were more accurate in assessing their pupils in Year R - remember that projections of KS1 National Curriculum levels are based on the KAS (1999) chart - or was there some other explanation? Did my findings for reading and mathematics indicate that projecting grades in these subjects was more inaccurate than for writing? In order to examine these questions, I decided to undertake a fine-grade analysis of the data.

Fine-grade analysis of the Year 2 data

I used the data sets in SPSS to manually produce group profiles showing individual cases. I analysed the data more closely, to see if there was any relationship within the categories of gender, SEN, EAL or birth date. The data showed that, for those pupils who did not achieve the projected grades, the male: female gender balance across the five class groups respectively was, in reading, 7: 1, 6: 4, 6: 3, 6: 9 and 5: 7; in writing, 3: 2, 5: 1, 4: 3, 5: 5 and 7: 5; and in mathematics, 8: 4, 2: 5, 5: 5, 6: 8 and 6: 5 - there was no significant relationship. Similarly, data relating to gender balance for those pupils who had exceeded the projected grades showed no relationship across all five groups and all three subjects. The proportion of pupils with EAL, and with SEN, across all groups and within all subjects also showed no relationship: similar proportions of pupils with EAL and SEN had both 'exceeded' and 'not achieved' the projected NC levels.

The next analysis centred on the proportion of pupils born during the autumn, spring and summer terms: would these data show significant relationship, especially in the category of those pupils who did not achieve the projected grades? School B offered the closest relationship between summer-born pupils and possible lack of achievement: the combined data for classes 1 and 2 showed that 13 out of 18 in reading, 11 out of 13 in writing, and 12 out of 13 in mathematics had failed to meet projected NC levels. Does this indicate that summer-born pupils are less likely to achieve? These data would only be significant if data from the other schools in my study showed a similar relationship. Data for schools A and D were almost equally balanced between the three terms, the three subjects and the proportion of pupils who had exceeded or not achieved the grades, whilst School C showed that 8 out of 11 pupils in the 'exceeding' category in writing had been born during the summer term compared to only 3 out of 10 in the 'not achieving' category. There was therefore no apparent pattern in the data for summer-born pupils. Three of the class groups showed a high percentage of summer-born pupils: 52% in School A, 60% in School B Class 2, and 50% in School C; School B Class 1 and School D contained 36% and 33% respectively.

The expected 'average' National Curriculum (NC) level at the end of KS1 is 2b, which equates to an Average Points Score (APS) of 15. The data showed this to be the mean and median for the **projected** APS in School A in reading and mathematics, although the **actual** APS were a mean of 14 and median of 13 for both subjects. The mean and median for the projected APS in writing were 13; these were also the actual APS, perhaps reflecting the government's concern that NC levels in writing nationally are depressed when compared to the expected 'average' of 2b (an APS of 13 equates to a NC Level 2c). In writing, however, the projected mean APS for School B, Class 1, School C and School D was in fact 14; this was achieved by schools C and D and exceeded - with a mean of 15 - in School B, Class 1. However, the mean and the median could not be compared directly with using APS as measurement, so I looked at the mode for each projected and actual APS in each of the three strands of reading,

writing and mathematics, and across all five class groups. In School A, the mode for the projected APS in reading was split: 9 and 21. This indicated that, based on the KAS (1999) chart, pupils in this class were fairly equally grouped at the 'below average' (that is, NC Level 1) and the 'above average' (NC Level 3) grades in reading. However, the actual APS for this class show a split mode of 9 and 13, meaning that pupils most commonly achieved either a Level 1 or Level 2c, both below the expected average of 2b. In fact, in this school, of the seven pupils projected to achieve an APS of 21, only three did so. This pattern was repeated in School A: the projected mode for writing was split between 9 and 15, with 13 being the actual APS mode; in mathematics the identical data were found. School B (both classes) and School D projected an APS mode of 21 in reading: this was achieved only by School D, with School B Class 1 achieving a mode of 17, and Class 2 a mode of 15. Does this suggest that the teachers in School D were more accurate in projecting future achievement in their pupils? Or could it be to do with other factors: both classes in School B contained a higher percentage of pupils with EAL and SEN (32% and 40% EAL, compared with 0% in School D, and 16% and 24% SEN compared with 3% in School D)? Further analysis of the data sets in the group profiles for all five class groups in the Year 2 cohort showed no additional relationships between the data. It appeared from the quantitative data that my first research question: *Can benchmark data be used to predict future achievement?* had been answered in the negative (see Chapter 7.1.1.).

5.3:   VALUE-ADDED DATA FOR 1997-2004 COHORT, SCHOOL C

I decided that further fine-grade analysis was needed, to look specifically at the key aspects used by government as proxy indicators of a pupil's likely ability, and looked at the available data sets; the one I chose to analyse was the 1997-2004 cohort in School C, for whom I had Baseline Assessment (BA) data – the school had completed the pilot study in 1997/98 ‑ and results from the pupils' KS2 SATs taken in May 2004. Their Reception class scores had been assessed against thirty-two BA statements in reading, writing, mathematics, and personal

and social development, and their achievement at the end of KS2 had been assessed against National Curriculum (NC) levels in English, mathematics and science. I decided to compare the NC levels for the end of KS2 that had been projected from the BA data, using the *Pathways to Progress* table (KAS, 2001), with the actual scores achieved: Table 12 shows the projected and actual scores for KS2 as an average points score (APS) for each pupil in a cohort of 18 boys and 13 girls.

For my analysis I chose four performance indicators: gender, SEN, EAL and summer born (I did not have details of FSM for this cohort); the data would enable me to see whether my research supported the use by the government of these variables as proxy indicators for value-added measures. These four variables, with others, are examined in Chapter 6.1.

Table 12 shows that ten out of the eighteen boys scored as projected, a success rate of 56%; eight of the thirteen girls (62%) also scored as projected. However, this was in relation to the APS; when the three different aspects of the tests - English, mathematics and science - were explored separately, no boys failed to achieve the projected NC level at the end of KS2 in mathematics or science, but 50% of them failed to achieve this in English. With the girls, 15% failed to achieve the projected level in mathematics and science, with 38% in English. Pupils with EAL had similar residuals to those without EAL, and some achieved higher grades than expected: one boy with EAL and SEN at School Action Plus (SA+) for moderate learning difficulties (MLD) had a residual of 12. Overall, pupils with SEN achieved as well as their peers: one boy at SA+ with behavioural, emotional and social difficulties (BESD) had a residual of 6; one girl with MLD at SA+ had a residual of 12. Of the five pupils - two boys and three girls - who had negative residuals, two had experienced a breakdown in the parental relationship during the past year, two had other family problems, and one had suffered bereavement when two close family members died unexpectedly. The twins - a boy and a girl - had identical residuals of 6, despite the boy having SEN at SA+ for BESD.

Table 12: Value-added data for 1997-2004 cohort in School C

Pupil	Gender	Projected APS	Actual APS	Residual	SEN stage	SEN category	Summer-born?	Other key indicators
1	M	99	99	0				
2	M	81	87	6			Y	
3	M	87	99	12				
4	M	81	87	6			Y	
5	M	87	87	0	SA+	PD	Y	
6	M	87	87	0				
7	M	75	75	0	SA	BESD		
8	M	81	81	0			Y	EAL
9	M	99	93	-6				
10	M	87	81	-6	SA	BESD		
11	M	63	75	12	SA+	MLD	Y	EAL
12	M	81	87	6	SA+	BESD		Twin of F11
13	M	93	93	0				
14	M	93	93	0				EAL
15	M	93	93	0				
16	M	81	87	6	SA+	BESD	Y	
17	M	93	93	0	SA	PD		
18	M	93	93	0			Y	
1	F	93	93	0			Y	
2	F	93	93	0				
3	F	99	99	0				
4	F	99	99	0				
5	F	75	75	0	SA	PD	Y	
6	F	63	75	12	SA+	MLD	Y	
7	F	81	75	-6	SA	BESD		
8	F	69	69	0	SA	MLD	Y	
9	F	87	81	-6			Y	EAL
10	F	99	93	-6				
11	F	81	87	6				Twin of M12
12	F	93	93	0			Y	
13	F	63	63	0	SA+	MLD	Y	

Of the summer-born pupils (54% of the girls and 44% of the boys) eight had achieved the projected grade and so had zero residual: four of these pupils had SEN at School Action (SA) or SA+; one girl with EAL had a residual of -6 and two boys with no SEN or EAL had residuals of 6. One boy had a residual of 6 despite having BESD at SA+; one girl with MLD at SA+ achieved a residual of 12 whilst a boy with MLD at SA+ and EAL also achieved a positive residual of 12.

One obvious question that I needed to explore as a result of these data: why were all the residuals in multiples of six? Further analysis gave me the answer: Average Points Scores (APS) are allocated for National Curriculum levels achieved at the end of KS1 on the basis of 3 for level W, 9 for level 1, with 13, 15 and 17 for sub-levels 2c, 2b and 2a, and 21 for level 3. However, for achievement at the end of KS2, APS are allocated thus: 15 for level 2, 21 for level 3, 27 for level 4, and 33 for level 5; there are no sub-levels, so the 'average' score of levels 2b, 3b, 4b and 5b are used. This means that, when the data for KS2 are examined, there are exactly six Average Points Scores between each of the possible scores (15, 21, 27 and 33), so the residuals will always be in multiples of 6.

5.4:   ANALYSIS OF QUALITATIVE DATA

Qualitative data took the form of initial and ongoing discussions with my headteacher colleagues, a survey of assessment practices made by practitioners in the five Reception classes, and lesson observations and discussions undertaken in each of the four schools.

Analysis of survey data (see Table 13)

The survey results showed that all five teachers used benchmark data from the Kent Record of Transfer - and, earlier, from Baseline Assessment - to plan appropriate learning experiences for pupils in the Reception class, as both of these data collection tools gave an indication of a child's ability at the start of their schooling. From these data, teachers were able to sort pupils into ability groups and to make early identification of children who might have some form of SEN; one teacher said that she could also identify the more able pupils who would need enrichment activities. The FSP was viewed by all the respondents as a summative document: it could help the Year 1 teacher to sort her ability groups, but it was little use to the Foundation Stage practitioner (although one teacher did say that she used the statements of attainment from the Early Learning Goals when compiling her IEP targets for pupils with SEN).

Table 13: Qualitative data: analysis of data from completed survey sheets

Factor	School A	School B class 1	School B class 2	School C	School D
Teacher-directed tasks	1	1		1	
Child-initiated activities	2	1	1	1	
Group activities	1		1		
Grouped by intake	1			1	
Grouped by ability	1	1	3	2	1
Friendship groups	2		1	1	
Changed based on personal/social factors/esteem	1	3	1		
Changed to reflect gender balance		1			
Changed after assessment	1			3	2
Minimising change		1	1		
Pre-school experience	1*		1*	1*	
Nursery/playgroups	2	1		3	
ELGs/areas of learning	1		1	1	1
Foundation Stage Profile			1	1	
Parental involvement					1
Teacher/TA observations	2	1	1	1	1
Informal assessments	2	1		1	
Written evidence			2		
Artefacts/photographs			1		
Formal assessments	4	2	2	1	
Pupil portfolio			1		
Identifying pupils with SEN	2	2	1		1
Pre-School Assessment	1	1	1	1	
Identifying G & T pupils			1		
Learning styles e.g. kinaesthetic, creative			1		
Pupils with IEPs	2				
External agencies	1				
Pupils with EAL	2*				
Value of BA	1*				
Value of FSP	1*				
Value-added/residual data	1*			2*	
Lack of consistency in assessing BA/FSP cf. end of KS1			3*	1	3
Time-consuming	1*	1*	1*	1	2
Problem with getting evidence	1*	1*	1*	1	
Statements very broad			1*	1	
Professional judgement	1		1	2*	

N.B. The number indicates the amount of times the respondent mentioned the factor, whilst the star indicates that the teacher had strong feelings about that particular aspect.

Teacher assessments were made informally on a day-to-day basis, using observations of and discussions with the child, and more formally on a weekly or half-termly basis using published assessment packages. All four schools used child-initiated activities as the mainstay of their observations (which were shared equally between the class teacher and the teaching assistant); they said that way it was easier to 'catch the children learning', especially for the scales related to creative and physical development (CD and PD) or personal, social and emotional development (PSED), where evidence of learning was more ephemeral. All teachers said that they made their formal assessments during teacher-directed activities, either against criteria suggested by Shirley Clarke or Ros Wilson, or linked directly to the statements of attainment of the Foundation Stage Profile; these formal tasks were generally undertaken once a term.

Groupings were initially based on data from pre-school settings: children would be put into either friendship or ability groups. As the year progressed, groups were changed for social and behavioural reasons or to have pupils of similar ability working together; these groupings were fairly fluid, and generally changes were made termly or half-termly so the children did not become too unsettled.

Teachers used various methods of assessment to track pupils' progress through the Reception year and to make early assessment of potential ability or problems. Observations of and discussions with the child were recorded either on the FSP booklet (up to three times a year), on local authority-designed data collection sheets (as often as the teacher chose) or on the eProfile (up to seven times a year: on entry, then Autumn A, Autumn B, Spring A and so on). Whichever method the school used, through this ongoing process of assessing pupils against the Stepping Stones (points one to three on each of the thirteen scales) or the ELGs (points four to eight on each scale), practitioners were able to identify those pupils who were not making expected progress. Once a practitioner had identified that a pupil was not making progress – and therefore

that he or she was likely to have some form of SEN - further analysis of need would lead to different teaching strategies and increased support for these pupils, with external agencies being approached where necessary. At this stage - or later in the pupil's school life – the child would be assigned to Early Years Action, Early Years Action Plus or a Statement of SEN, according to criteria in the Code of Practice.

Question 8 on the survey sheet: *Do you believe that the baseline or foundation stage profile scores can predict attainment the pupil will make by the end of KS1?* was a rephrasing of my first research question, and I was interested in the responses of the Foundation Stage practitioners. Every teacher expressed concerns about the efficacy of using FSP data to predict a pupil's future achievement: one stated that "the assessment criteria is not consistent between foundation [stage] and the end of KS1" and that she had always understood that "you could only get true quantifiable value-added if you compare like with like". Another respondent suggested that differing backgrounds of pupils in her school meant that a child with, say, no pre-school experience and no English was being compared with a child who had attended Nursery for several years. Another teacher suggested that "using data in this way [to predict future achievement] can be unreliable" because "there are a lot of environmental factors that can affect children's learning". A third response was that "the curriculum is very different so a child could leave foundation being very creative (kinaesthetic learner) then be forced into a more formal education and not achieve as well. Other children may flourish in a more formal environment but not have achieved as well under foundation guidance." Another teacher suggested that "children develop at different times and in different ways...those who come in knowing more than other children we find slow down, while average children begin to catch up and even perhaps overtake by the end of KS1. These predictions make life very difficult for Year 2 teachers who have to get the children to reach a level they are not capable of." Each of the five teachers felt strongly about this issue, although most thought that BA data had been more useful as it had given a baseline on entry and thus a benchmark

against which progress during the Foundation Stage could be measured. The views of the practitioners from the four schools backed up, in qualitative data, my findings from the quantitative data: that benchmark data cannot be used to predict future achievement (see Chapter 7.1.1).

All of the teachers felt that completion of the FSP was an arduous task that involved the collection of an enormous amount of evidence; one teacher described the FSP as "a thorn in my side" that took away the pleasure of teaching Reception children, while another said it brought "a lot of additional work e.g. amassing evidence, which makes it unmanageable, especially if you have other curriculum duties." The survey data clearly showed how dissatisfied each of the practitioners was with the amount of evidence they were being asked to collect; every pupil – in a class that could contain 30 children - has to be assessed against 117 separate statements of attainment. Additionally, these assessments need to be undertaken on probably three or more occasions, in order that the teacher and her assistant can validate their judgements that the pupil has securely achieved each objective. To evaluate the size of the task of collecting and recording all this evidence, I arranged to spend a morning observing in each of the classes, sharing experiences with the children and discussing the assessment process with the class teacher and her teaching assistant.

Observations in the Reception classes
School A has one Reception class which pupils move into from the on-site nursery. The format of the day is similar to that found in many schools: in the morning, three teacher-directed tasks, two sessions of child-initiated play, a whole-class activity of 'Writedance' (used by many schools to develop a pupil's fine motor skills; details from http://www.luckyduck.co.uk) and another to encourage spoken language development for the high number of pupils in this class with EAL. The afternoon starts with Circle Time during which the pupils develop speaking and listening skills   again, essential with so many EAL children - and address the areas of learning in personal and social development.

On the day of my visit, there were twenty-five pupils (16 boys and 9 girls) whose ethnicity included English, Indian, Tamil, Lithuanian and Polish; there was also one Looked After Child (LAC) and several children from low socio-economic backgrounds. Pupils work on teacher-directed tasks in five groups of mixed ability; there are three learning support assistants who work alongside the class teacher. The class base consists of two interconnecting classrooms, with access to an outside play area which is shared for part of the day with the adjoining nursery. The child-initiated activities from which the children can choose include, in the home base, two computers, books and listening tapes, musical instruments, puzzles, a toy garage and cars, a writing area and wooden construction blocks. In the outside play area there are play houses, wheeled toys, a seesaw, a climbing frame and tunnel, pushchairs with dolls and soft toys, a sandpit and a water tray.

In School B there are two parallel Year R classes which take their pupils from the on-site nursery, from local playgroups and direct from the home environment. Both classes use the same shared resources and classroom organisation; the outside area is shared by both classes and the nursery. Pupils have a semi-structured day incorporating, in the morning, the daily act of collective worship followed by activity time, snack time, playtime, and - on the day of my visit - a session in the ICT suite. In the afternoon, the sessions followed a similar pattern to the morning: Circle Time on the carpet, then activity time, then a snack and a story before going home at 3.30 pm. The pupils are divided into four ability groups with seven pupils in each; these groups are set differentiated tasks, according to the area of learning. All pupils are monitored on a weekly basis for their progress in reading, writing and mathematical understanding through an inspection of their written work and through discussion between the child and the practitioner; these three key aspects of learning are also formally assessed at the end of every half-term. The day is a mixture of teacher-directed tasks and child-initiated learning: the latter is undertaken by the pupils for forty-five minutes in the morning and thirty-five minutes in the afternoon. Each pupil is invited to choose an activity

from a choice of fourteen; on the day of my visit, twelve of the twenty-six children present in one class chose the 'outside area', with ten of the twenty-seven from the other class choosing the same option. This outside area contained a sandpit, a play-house, a wooded 'wild area', climbing apparatus, large construction equipment, and a variety of tricycles, seesaws and other large items to develop a child's gross motor skills.

School C has one Reception class; the pupils come from the local state nursery school, local playgroups and direct from the home. They also use a visual timetable, and the day is again semi-structured, with teacher-directed tasks and child-initiated learning. Here, the class teacher and the nursery nurse each take a 'focus group' for an task based on one of the ELGs: one every morning and a different one every afternoon. Evidence is collected to demonstrate pupil progress: a piece of free writing, a drawing or painting, a photograph of an artefact made from construction apparatus or 'found objects' (formerly called 'junk modelling') or a written observation made by the teacher or her assistant. The role play area is changed every few weeks; at the time of my visit, it was a pirate ship, to reflect the theme of the 'Big Book' that was being shared that week; previously, it had been a shop, a bank, an estate agent, an optician and a hospital. As the theme changes, so do the children's tasks and the classroom displays, making a very colourful environment for the pupils. Assessment is made against the ELGs and recorded on the same KCC sheets as used in School B; the teacher also completes the FSP at present, but has decided that next year she will record just once as the evidence-gathering exercise is so time-consuming.

School D also has one Reception class; pupils enter the Foundation Stage from three local playgroups, three private nurseries, and one nursery class attached to a local infant school. On the day of my visit, the role play area was a traditional home, with sink, cooker, ironing board, beds and a dining table; pupils were demonstrating evidence of their understanding of one-to-one correspondence in mathematics, which the teacher was recording. This school

also has a semi-structured day; teacher-directed tasks involve the pupils working in ability groups, with two adults and one parent helper to scaffold the children's learning across the ELGs. This school has a covered outside play area which is used at certain times of the day; pupils also have child-initiated sessions during the morning and the afternoon, during which time the adults make written observations of the pupils' learning.

Discussions with practitioners

Although only one teacher had mentioned a pupil portfolio in her survey response, it was obvious from the discussions that followed my visits that each teacher kept a detailed portfolio of evidence for each child, containing written work in language, mathematics and KUW, together with photographs (mostly digital) showing whole-class and group activities such as snack time, assemblies, school trips, and play in the outside area. Two of the schools kept computer-based files for individual pupils that demonstrated the child's achievement of the learning objectives of PD, CD and PSED. These activities are more ephemeral than written work and thus more difficult to 'catch'; once a pupil had appeared to master a particular skill, the teacher would register the fact by photographic evidence and then enter details on her record sheet. All of the practitioners agreed that it was important to keep records of pupil progress and to track them over time, but emphasised how spending so much time in compiling portfolios and assessment records meant less time actually planning and delivering the learning opportunities for their pupils. Despite the onerous nature of the recording, every teacher described how their ongoing formative assessment against the ELGs was of great importance, as it enabled them to identify any pupil who was not making expected progress and to offer appropriate support. On the evidence of the qualitative data from survey responses, observations in class and discussions with the teachers and teaching assistants, my third research question, *Can benchmark data be used to support pupils' learning?* had been answered in the affirmative (see Chapter 7.1.3).

CHAPTER 6: DISCUSSION

6.1: POSSIBLE INFLUENCES ON ACHIEVEMENT

Since 2004, proxy indicators used by government to indicate the possibility of pupils in the primary phase being unable to reach the expected average level of achievement have been:

- o Special educational needs (SEN)
- o Ethnicity and English as an additional language (EAL)
- o Gender
- o Entitlement to free school meals (FSM)
- o Mobility
- o Looked After Child
- o Summer-born

I feel that there are many more variables than these that might impact on a child's ability to perform to their potential. Some are related to home circumstances, such as living conditions and financial stability, parents' educational background, and the number and age of siblings; others are related to the school context, and include:

- o Infant class size
- o Adult: pupil ratio

In this chapter I shall examine the seven performance indicators that the government now uses for its contextual value-added data, together with the two that I have identified: each of these could be considered a negative influence on a child's achievement (see also Chapter 3.5).

Special educational needs

The SEN Code of Practice (DfES, 2001b) defined children with special educational needs as those having 'a *learning difficulty* which calls for *special educational provision* to be made for them... Children must not be regarded as having a learning difficulty solely because the language or form of language of their home is different from the language in which they will be taught.' (paragraph 1.3) Provision would be available as a graduated response: in the Foundation Stage this would be through *Early Years Action* in the first instance, moving on to *Early Years Action Plus,* followed by a request for statutory assessment and a Statement of SEN if the child was not making sufficient progress. The Code explained how a child may be making less progress against the Early Learning Goals than his or her peers, but that the assumption should not be made that this is due to the child having SEN:

> Monitoring of individual children's progress throughout the Foundation Stage is essential. Where a child appears not to be making progress either generally or in a specific aspect of learning, then it may be necessary to present different opportunities or use alternative approaches to learning... There should not be an assumption that all children will progress at the same rate. A judgement has to be made in each case as to what it is reasonable to expect that particular child to achieve. (paragraphs 4.9-4.13)

A child identified with SEN in a pre-school setting will generally move on to a mainstream infant, primary or first school; in Kent, the Record of Transfer form that is passed on from the pre-school setting will have an additional sheet, 'Record of intervention provided to meet additional/special educational needs' which details areas of strength, areas of concern, parental involvement, and the child's own view, together with a summary of intervention provided at Early Years Action (with the outcomes) and also at Early Years Action Plus and Statutory Action (leading to a Statement of SEN) where relevant.

In November 2001, guidance was issued to schools on the practical operation of the new statutory framework for inclusive schooling for children with SEN (DfES, 2001c). It reiterated the principles of successful inclusion, where schools set suitable learning challenges and targets, respond to pupils' diverse learning needs, and overcome potential barriers to learning for individuals and groups of pupils. It outlined strategies that could be used to support pupils with SEN - either with or without a Statement - in mainstream schools, and to enable them .'... to be included without compromising the efficient education of other pupils.' (paragraph 40). A further document (DfES, 2001d) outlined how targets could be effectively set for pupils with SEN, using the P scale to give levels for pupils working below level 1 of the National Curriculum (that is, those previously graded as level W). The following year, new information was sent to schools about key developments in SEN, including the use of P scales as one aspect of measurable performance targets at KS2, 3 and 4 for pupils with SEN. The DfES was collecting data from schools and LEAs on the work they had been doing on P scales, and the *Autumn Package* would in future provide support to schools on their use; the University of Durham CEM Centre was again collecting national data on the P scales in order to provide schools with feedback. (DfES, 2002)

An Ofsted report in 2004 highlighted the main findings of HMI visits to mainstream and special schools - primary and secondary - during 2002 and 2003:

> Although moderation of teacher assessment for the majority of pupils is satisfactory across most schools, it is underdeveloped in relation to the assessment of pupils with SEN... Almost all special schools use P levels to assess individual pupils and set their targets... Mainstream schools are not required to set performance targets for pupils who are unlikely to achieve national expectations...(Ofsted, 2004, p. 3)

The Foundation Stage Profile handbook (QCA, 2003a) made clear that the P levels were not to be used to report pupils' progress in the Foundation Stage:

> Some practitioners have asked for clarification of the interrelationship between the P Scales and the Early Learning

Goals. The P Scales and the Early Learning Goals have been written for different purposes and have been constructed in different ways. The Early Learning Goals specify expectations for children's progression by the end of the Foundation Stage while the P Scales were written for use with children of national curriculum age who are working towards level 1... They were written primarily for supporting target-setting in the context of the national curriculum. (p. 113)

However, in February 2005, KCC issued guidance on Clusterweb to schools on 'Doing away with W'; this stated that:

> County moderation procedures (P scale, Foundation Stage and Key Stage 1) ensure that consistent judgements are made both throughout and at the end of Year R, whichever assessment framework is used... For children with profound, severe and complex needs Appendix 1 provides a mechanism for comparing Early Learning Goals with 'P' scales... The guidance above is aimed at clarifying the differing messages reflected in DfES and QCA guidance. (KCC, 2005c)

The Appendix outlined the ways in which the ELGs and the P scale should be used:

> The Early Learning Goals and the P Scales have been structured in different ways (i.e. areas of learning as opposed to specific subject areas). As a consequence of this, any given element of the P Scales may apply in more than one area of learning. In reading the tables in this chapter, you will therefore find that some items in the Profile scales have been cross-referenced to more than one element in the P Scales...

The PLASC return of January 2004 required schools to enter whether a pupil had SEN at Early Years Action or Action Plus, or at School Action or School Action Plus, or whether a Statement of SEN had been requested or issued. Kent LEA (KCC, 2004f) emailed guidance to schools on data collection by type of SEN, related to the new codes for special educational needs that had been introduced by the DfES the previous year. Data were required to be entered for pupils above the first level of intervention that is, for those at Early Years Action or School Action - using these new codes, indicating primary and secondary needs. I had already collected these data in 2003, when I researched profiles of SEN and Additional Education Need (AEN) across the

four schools from this thesis plus four other schools in the Local Learning Group (see Hyne, 2003).

In April 2004, Kent Management Information sent a letter to all headteachers, SENCos and SEN governors on the collection of P scale data, which 'becomes statutory from Summer 2005' (KCC, 2004g). Data using the P levels had to be entered onto an already-populated spreadsheet (a disk was provided) for those pupils who had not been assessed using NC levels, that is, those previously reported as being at level W. The acceptable P levels were P1(1), P1(2), P2(1), P2(2); then P3(1), and P3(2) through to P8 (no sub-levels for P4 to P8). Thus, from 2005 pupils with SEN were able to 'achieve' a level on the P scale when hitherto they had been given the generic class of 'W'; target-setting would be more specific with the new P levels and value-added data easier to calculate.

In my research, a fine grade analysis of the Year 2 data – see Chapter 5.2 - found that there was no relationship between the percentage of pupils with SEN who had either exceeded or not achieved the projected NC levels at the end of KS1. Analysis for the 1997-2004 cohort (see Chapter 5.3) showed that, out of four boys at School Action Plus (SA+), all had achieved at KS2 the APS that had been projected from their BA data but, more importantly, two had achieved a higher score by 6 APS, and one by 12 APS; for those three boys at School Action (SA), two had achieved their expected scores and one had a residual of - 6, meaning that he had achieved a lower score than projected. With the girls, a similar pattern emerges, as can be seen from Table 12. However, it is also obvious from my data that having SEN appears to make no difference to the likelihood of whether a pupil will achieve, not achieve or exceed his or her projected scores at the end of KS1.

Ethnicity and EAL

The Kent School Organisation Plan of 2003/4 to 2007/8 was required by the DfES to include, for the first time, data on ethnicity of pupils in Kent schools. The 2003 PLASC return (see Table 14) showed that 5.71% of the total pupil

population in Kent was from Minority Ethnic groups, compared with 16.08% nationally.

Table 14:   Ethnicity of pupils in Kent schools 2003/2004, as shown by PLASC

Ethnicity	Kent figures	National figures
White British	88.36%	79.90%
Irish	0.18%	0.39%
Traveller of Irish heritage	0.01%	0.06%
Gypsy/Romany	0.20%	0.09%
Other White background	1.30%	1.99%
White/Black Caribbean	0.32%	0.89%
White/Black African	0.12%	0.22%
White/Asian	0.39%	0.49%
Other mixed background	0.64%	0.88%
Indian	1.03%	2.27%
Pakistani	0.11%	2.58%
Bangladeshi	0.21%	1.04%
Chinese	0.23%	0.34%
Other Asian background	0.29%	0.61%
Black Caribbean	0.07%	1.43%
Black African	0.25%	1.60%
Other Black background	0.06%	0.40%
Other ethnic group	0.24%	0.80%
Unclassified/refused	5.93%	4.02%

Although only 5.71% of the total pupil population in Kent was from Minority Ethnic groups, my research data shows a considerably higher percentage, for example, School A had 29%, 32% and 43% pupils with EAL in the Year R, Year 1 and Year 2 cohorts respectively, with 21% Travellers in Year R and 11% in Year 1; in fact, only School D had lower than the county percentage of EAL in its Year R and Year 2 cohorts.

An email to headteachers in November 2004 reminded schools of the importance of developing inclusive practice for pupils with EAL; the same email described the role of the Minority Communities Achievement Service (MCAS) which:

> works in partnership with schools, minority communities, parents
> and carers to improve access to education and raise the achievement

of Ethnic Minorities, pupils for whom English is a second language, Refugees and Travellers... the service supports children and young people at risk of not accessing education or of under-achievement.' (KCC, 2004c)

Studies undertaken into results of the 1991 KS1 national tests showed, in the NfER survey, that black and Asian pupils did less well than their white peers in English, mathematics and science (although the sample of 460 was too small to draw reliable inferences). A similar study by Leeds University showed that pupils whose home language was English did significantly better in English, mathematics and science than their non-English-speaking peers: the sample group of Pakistani pupils had the lowest results in all three subjects, with the black Caribbean and Indian groups getting the lowest results in both English and mathematics. (see Gipps and Murphy, 1994, p. 188; and Gilborn and Gipps, 1996)

A report on the teaching of reading in forty-five Inner London primary schools (Ofsted, 1996) found that:

> Black African pupils performed better than other ethnic groups at both Year 2 and Year 6. Bangladeshi pupils achieved low scores in Year 2 but performed better in Year 6. White pupils from economically disadvantaged backgrounds consistently performed least well and constituted the largest group of underachievers at Year 6. (p. 7)

In a report the following year on the teaching of number in three inner-urban LEAs (Ofsted, 1997b), HMI showed that:

> In year 2, there is little difference between the scores of the Indian and the white groups, though the gap widens in Year 6 with Indian pupils performing marginally better, but with the small group of Chinese pupils performing particularly well. Black African, and Pakistani groups under-perform in Years 2 and 6 which almost certainly reflects a lack of fluency in English. Year 6 scores on the written tests are behind those of the mental test for every group save Chinese pupils. Bangladeshi pupils and those in the Black other group obtained written test scores well below their mental test scores indicating problems with reading and comprehending written English. (paragraph 149)

A study by Gilborn and Mirza (2000) found that:

> while all the principal minority groups now achieve higher results than ever, white pupils have benefited more than most. As a result, some minorities are even further behind the majority than they were a decade ago. The situation is especially serious for black, Pakistani and Bangladeshi pupils. (See Institute of Education Press Office, 2000).

In the USA, Porter (2005) looked at research undertaken by the Vanderbilt University on data from the National Assessment of Educational Progress (NAEP). This was a longitudinal study on trend assessments of the achievement gap among racial and ethnic sub-groups in reading, maths and science; the findings showed that NAEP long-term trend assessments had remained virtually unchanged since their first administration over three decades previously, despite many initiatives to close the achievement gap. Indeed, citing a study on Black and White pre-school children watching *Sesame Street* on television, it appeared that the gap between the two ethnic groups had actually widened. (p. 60).

In my own research, data show that having EAL appears to have no bearing on a pupil's ability to achieve his or her projected NC levels (see Table 12), despite there being a higher percentage of pupils with EAL in my study than in Kent overall (see earlier comment in this section).

Gender issues

A NfER study of the KS1 1991 national test results from 260 schools, covering 5345 pupils found that girls performed significantly better than boys in all areas of English, and in mathematics (although the differences were less marked). A further study by the DfES of 50,000 pupils taking the same 1991 tests showed that the girls out-performed the boys in English, mathematics and science, with the difference most marked in English; similar research of 2500 pupils by Leeds University found almost the same differences as the DfES study. (see Gipps and Murphy, 1994, p. 188)

107

A booklet for school governors (Equal Opportunities Commission, 1991) asked the question: 'Can the fact of being male or female affect a pupil's educational achievement?...what can governors do to eliminate sex discrimination and encourage equality of opportunity in their school?' In 1993, Ofsted published a report from HMI on the performance of boys in English secondary schools between 1988 and 1991 (Ofsted, 1993) which found that:

> Boys do not do as well as girls in English in schools (paragraph 5)
> ... in all year groups girls read more fiction books than boys and
> tended to have different tastes in reading (paragraph 12)...Girls
> wrote at greater length and generally received higher marks than
> boys for written work. (paragraph 16); [and that] There was little
> evidence of boys discussing the affective aspects of experience or
> of their writing with conviction about personal feelings.
> (paragraph 21)

Similar findings came out of research into the teaching of reading in forty-five Inner London primary schools (Ofsted, 1996), where HMI stated that:

> There are differences in the reading performance of boys and girls
> at Year 2 on both reading accuracy and comprehension. Girls, on
> average, outperform boys by the equivalent of about 3-4 months
> RA on reading accuracy. This is also the case within different
> ethnic groups. The gap between white boys and white girls at Year
> 2 is about 5 months RA on reading accuracy. The only exception is
> the Bangladeshi pupils, where boys marginally outperform girls,
> though the performance of Bangladeshi pupils overall is very low at
> Year 2. At Year 6, however, the differences by gender have largely
> disappeared. The overall distribution of reading performance for
> boys and girls is virtually identical for reading accuracy. This is in
> line with the Neale [Analysis of Reading Ability] national
> standardisation results. The only group where there is still a gender
> difference is the black pupils where girls tend to outperform boys
> on accuracy and to a lesser extent on comprehension (by the
> equivalent of about 3-4 months on reading accuracy). (paragraphs
> 139 and 140)

The 1997 report by HMI on the teaching of number in three inner-urban LEAs (Ofsted, 1997b) found that:

> There were differences in the performance of girls and boys in both
> Year 2 and Year 6. Overall girls do slightly better than boys in
> Year 2 whilst the reverse is true in Year 6 with a bigger margin of
> difference... When the mean scores of the girls and the boys in

each school are compared, the differences in scores are sometimes highly significant... In every case the graphs show that when the boys' mean score is low, the girls generally do better. However, this is not sustained; when boys do well, they outperform the girls... Overall the girls' mean scores are slightly higher than that of the boys in Year 2 but the reverse holds in Year 6... The boys maintain their performance on the mental test but on the written test their performance drops, though not to the same extent as that of the girls. (paragraphs 144-145)

In April 1997 KCC sent all their headteachers a copy of a report commissioned by the Review Team of Senior County Advisers (KCC, 1997) which highlighted recent research on the relative performance of boys and girls at all levels - in particular the comment by Chris Woodhead, HMCI in 1966 that 'Girls out-perform boys at 7, 11 and 14 in national curriculum assessments in English, mathematics and science. (p .3) and offered practical recommendations to improve boys' performance through a different and more stimulating approach to teaching all NC subjects and religious education.

QCA published a booklet in 1998 that set out to raise boys' achievement in English. The report identified that:

> Boys are less likely to concentrate on a task at length. Overall, boys are observed to be more frequently off-task than girls. From pre-school onwards, boys are more reluctant to be quietly reflective and prefer to be highly active. .. The average boy reads less than the average girl and his range of reading is narrower, with more focus on non-fiction and a reluctance to read both fiction and poetry...more boys than girls have statements of educational need based on problems with literacy... Many boys write less than girls and are less enthusiastic and committed writers. On average they become competent later than girls, and are more likely to have problems as writers.' (QCA, 1998a, pp. 12-15)

Teachers were recommended to take immediate action, which 'should focus on a particular language mode or on aspects of boys' attitude and approach to learning...' (p. 33)

Jon Pickering (1999) suggested that boys' under-performance compared to that of girls could be related to boys' attitudes to school, peer pressure, biological factors, and the lack of male role models in schools. He compared the performance of boys and girls at the end of Key Stages one, two and three and showed that, at the above-average NC grade of levels 3, 5 and 6 respectively, girls were out-performing boys in English, whilst generally boys were achieving higher percentages in mathematics and science, at the end of each key stage. (p. 71)

Research by Sukhnandan, Lee and Kelleher (2000) looked at various strategies schools had implemented to address gender differences; in primary schools, this had been Additional Literacy Support. Recent researchers in the field include Millard (1996, 1997, 2001, 2003) and Moss (1999, 2000); one study by Moss found that, of 21 KS1 pupils, the more able readers were virtually all girls (seven compared to 3 boys) whilst the pupils who had experienced the most problems were the boys (ten compared to one girl).

KCC have recently developed their guidance on addressing gender issues in the relative performances of boys and girls in Kent schools in their 'Different and Equal' project launched in November 2004 (KCC, 2004e). The issue was that:

At all key stages, both nationally and locally in Kent, girls out perform boys in English. The differences are even more noticeable if the focus is writing... In Kent, differences in writing at Key Stage 1 show 9% fewer boys achieve a level 2 and above than girls, at Key Stage 2 the difference is 17% (p. 4)

The project - involving fifteen schools from all phases of education, and led by advisers from KCC Advisory Service, Sheffield University, and three Kent schools - looked at ways to narrow this gender gap, by using strategies such as 'boy-friendly texts', talking and writing partners, structures to support thinking and planning, and real audiences, among others. Alongside this guidance came a set of data from KCC Management Information giving each school analyses

110

for each of the Key Stage results from 2002, 2003 and 2004 by gender split. These data were broken down into the elements of reading, writing and mathematics, with percentages given for KS1 - of pupils who had achieved NC levels 2b+ and 3+, with data for value-added residuals and the lowest 20% achieving pupils by gender.

My research indicates that gender does not appear to be an issue when projecting NC levels: analysis of data from the Year 2 cohort show that similar numbers of boys and girls achieved, did not achieve or exceeded the APS projected for them, and these data are backed up by those shown in Table 12 for the 1997-2004 cohort.

Free school meals

In their report on the teaching of reading in forty-five Inner London primary schools (Ofsted, 1996) HMI stated that:

> Using free school meals as an indicator of social and economic disadvantage, the [Neale Analysis of Reading Ability] results show a clear association with reading performance at both Year 2 and Year 6. At Year 2 the difference between pupils on free meals and those not, was on average six months RA on reading accuracy, and about the same on comprehension. At Year 6 the difference was close to 12 months RA on reading accuracy and slightly more on the comprehension. (paragraph 143)

Furthermore:

> White pupils on free meals do poorly at Year 2, only a little better than the Bangladeshis, almost all of whom come from non-English-speaking backgrounds. By Year 6, white pupils on free meals have a lower average score than any of the other ethnic groups (this applies to white girls as well as white boys). At Year 6, Bangladeshi pupils on free school meals have a very similar score on reading accuracy to the other non-white groups and are on average about 6 months ahead of the white group on free meals. Bangladeshi pupils still lag behind in terms of reading comprehension at Year 6. (paragraph 145)

A report on the teaching of number in three inner-urban LEAs (Ofsted, 1997b) found that:

> The proportion of pupils entitled to free school meals, 34.9% in Year 2 and 33.7% in Year 6, provides an indicator of social and economic disadvantage. The results show that the difference in performance between pupils entitled to free school meals and those that are not, to be much greater in Year 6 than Year 2. The mean scores...show the relative performance of pupils entitled to free school meals has declined in Year 6, markedly so in comparison to those pupils who are not entitled... In Year 6, on each test a quarter of pupils entitled to free school meals had scores below 84, and a quarter achieved scores greater than 104. In contrast, of those Year 6 pupils not on free school meals, a quarter had scores above 113. This was the case in both the written and mental tests. (paragraph 146)

In 2000, the Institute of Education at the University of London published research from a team of local authority officers from Hampshire LEA and researchers from the Institute, led by Professor Harvey Goldstein (Goldstein, Huiqi, Rath and Hill, 2000). This research had been funded by Ofsted to explore how Hampshire LEA used value-added analysis when reporting data for primary schools in its area, both from baseline to KS1, and from KS1 to KS2. The findings showed that, in the seventy-six school studied, and a sample size of 1400 pupils with matched data:

> The percentage of pupils eligible for free school meals... remains as a significant predictor throughout the analyses of the English score, with those schools having the lowest percentage having a predicted score up to 0.25 more than the average... Pupil eligibility...for free school meals is associated with a score reduction up to about 0.4 compared to not, as does being a boy rather than a girl. (p. 10)

The research team used a two-level model of multilevel analyses, using the *MlwiN* software package (this method is explained in greater detail in a later paper by Browne and Goldstein, 2002). The team concluded that there was little value in LEAs using so-called 'proxy indicators' (such as free school meal eligibility and average KS1 results for a younger cohort) as performance indicators; these measures are described as:

inadequate and likely to produce misleading comparisons among schools... [and that the results of their research] provide little support for the routine use of such proxy measures either for [Ofsted] inspection purposes or for publishing as 'league table' rankings of schools. They do not constitute valid measures of educational quality. (p. 2)

My own research was unable to use any data on FSM: although a pupil's entitlement to FSM is a key indicator used by government, the system whereby parents apply to the local authority for exemption from payment is a complex one and the criteria for claiming frequently change. Data that is accurate one day might change the next, if the parent found work; it was thus agreed with my headteacher colleagues that this performance indicator would be discounted from my research.

Mobility

Tymms and Henderson (see Lindsay and Desforges, 1998, p. 61) found that, in a sample of 3548 pupils surveyed as part of their value-added national project for the Curriculum, Evaluation and Management (CEM) Centre, University of Durham in 1995, only 53% of pupils did not change schools between joining the Reception class and the end of KS2. However, Strand and Demie (2006) found that:

> The current research suggests mobility has no long-term impact on pupils' educational progress during either KS1 or KS2. There is also no evidence that high levels of mobility are associated with long-term social adjustment problems.

The Kent School Organisation Plan (KCC, 2004d) commented thus on pupil mobility:

> A number of schools in Kent are affected by factors causing volatility in pupil numbers, the major ones being the impact of refugees, asylum seekers, travellers and army movements. The impact on individual schools can be quite significant and in many instances is extremely difficult to predict. (p. 25)

The 2005 PLASC return asked for the number of Travellers' children admitted on the school's roll and attendance register (though not necessarily in school on the day of count); a separate section asked for details of any other dually-

113

registered Travellers' children who were not in attendance on the day of count, together with their name, date of birth and gender. Guidance notes to accompany the PLASC return described how changes to the Education (Pupil Registration) Regulations 1995 had been amended in January 1998 to allow for the dual registration of Traveller children at more than one school, 'to protect the continuity of learning of Traveller children'. (KCC, 2005a)

In my own research, mobility was a key factor in School A which, as stated in Chapter 5.1, had missing data for nine pupils out of 30 in Year 2 because of high mobility, together with 11% of its Year 1 cohort and 21% of its Reception class who were Travellers. However, as this was the only school of the four with the issue of either high mobility or Traveller children, I did not analyse the data against these criteria.

When I interviewed the headteacher of School A about pupil mobility, I was presented with a thick ring binder full of pages with details for each cohort in the school. The Year R cohort was described as being 'remarkably stable' but the Years 1 and 2 cohorts - in common with the Years 3 to 6 - contained a high proportion of pupils who had joined the school, or left, or in many cases left and then come back. The headteacher explained that many parents would 'fall out' with another parent (often a relation of theirs or of their partner) and remove their child or children from the school, only to 'make up' with the other person and return the child to the school. On other occasions, parents had complained that their child was 'being bullied' and taken them to a local school that was perceived to have better behaviour management, only to find that this was not the case, and so brought their child back to the original school. Much of the mobility of School A was due to the high percentage of Travellers and asylum seekers: as the school had 'got itself a name' for being supportive of ethnic minorities and traumatised children, many were transferred into the school by the LEA or on recommendation from Social Services. One child joined the school in Year 1 but didn't settle and returned to his old school. Another child had joined the school when he was re-housed from a

neighbouring authority; he settled well and was beginning to make progress, only for the council to re-house him back to his original borough; the headteacher said that the child was very sad to leave. No asylum seekers had joined the school over the past three years unlike the previous five or six, when there had been an influx from war-torn European and African countries - although there was a child from Kosovo in Year R who had entered England through normal immigration procedures (this pupil scored 106 on his FSP).

The headteacher knew that many of the pupils in the school were from a Traveller background, although they frequently chose not to disclose this fact. She was certain that the Year R cohort comprised 21% of Traveller pupils and the Year 1 cohort 11%; this was ascertained from the family connections with pupils previously at the school, though only one parent had made her child's Traveller background apparent. The headteacher had decided to try to persuade the other parents to agree their background, as this would bring in funding for the school: there was no pupil support from the Kent Traveller Achievement Service, or additional funding in the annual school budget under the 'Additional Educational Needs' category, unless it could be proved that a pupil was a member of a Travelling family that had been mobile within the past two years (if the family was living on a designated site, no support was available).

Looked After Children

The term a 'Looked After Child' was introduced by the Children Act of 1989 and refers to children who are subject to care orders and those who are 'accommodated'. By 2000, sixty-two per cent of all children currently looked after nationally were subject to care orders (See DfEE/DoH, 2000a, p. 9, paragraph 3.2.) The guidance stated that:

> For too long the education of young people in public care has been characterised by fragmentation and unacceptable levels of failure. The underachievement of children in public care, and the failure of agencies to work together in meeting their needs, is well documented and has been re-visited in research and numerous reports. While no national statistics have before now been collected on the educational outcomes of young people in public care, leaving care studies reveal unacceptable levels of underachievement: 75 per

cent of care leavers leaving formal education with no qualifications, and between 12 and 19 per cent going on to further education compared with 68 per cent of the general population.' (p. 1, paragraph 1.4) ...It is estimated that young people in care are six to eight times more likely to have a Statement of Special Educational Needs than pupils in the general school population. (p. 53, paragraph 9.2.)

Further guidance (DfEE, 2000c) tackled the serious underachievement of children in residential and foster care, and aimed to bring their achievement closer into line with that of their peers. It suggested that children and young people separated from their families will have emotional and behavioural responses to that separation and the often traumatic events which led to it, and that schools have an important role to play in helping to ensure that unmet emotional needs do not lead to disaffection, difficult behaviour or exclusion.

When completing the PLASC 2004 return, headteachers in Kent were reminded to include details of any Children in Care - an alternative term for Looked After Children - together with a valid Kent ethnicity code for every pupil. Soon after, a letter was sent to headteachers; the enclosures offered clear procedures for 'action by the education department when a child moves into an adoptive home and starts a new school', and reminded headteachers that:

> Most schools will have pupils who are looked after by a local authority. The term "looked after children" (LAC) refers to children in public care, who are placed with foster carers, in residential homes or with parents or other relatives. Many of these children are at risk of underachieving at school, as they have experienced disruption to their family life and education, low expectations, poor emotional and psychological health and a lack of family support... (KCC, undated).

The Kent School Organisation Plan (KCC, 2004d, p. 25) stated that 'The number of Looked After Children although appearing a large number, remains less than 1% of the total number of pupils of school age in Kent.' Actual figures had remained fairly stable, being 1525 in 2002, 1437 in 2002, and 1351 in 2003.

In my research, data from the sample showed that, across the fifteen classes, there were only three pupils who were LAC so these data were not analysed. However, pupils who are LAC have an impact on the mobility factor discussed above: one of two such children in the Year 1 cohort of School A had joined the school at the start of the Reception year, then left, then returned in Year 1. As the child was not in the Reception class at the end of the year - when the teacher completed the Foundation Stage Profiles - he did not get a score; there was no record sent on from the interim school (wherever that was - the headteacher was not sure) and there was no guarantee that he would still be in the school at the end of Year 2 for his teacher assessments to be made.

Summer-born pupils

In England, children have to start school by the term after their fifth birthday, but most schools take children as 'rising fives', that is, at some time during their fourth year. Schools often have two intakes to their Reception class - in September and January - and this means that pupils in a Reception class can be aged between nearly five - for those born in early September - or just four, for those born in late August.

The National Institute of Economic and Social Research (NIESR) examined low school attainment by summer-born children (NIESR, 1997). They cited the practice in Switzerland and other European countries of deferring entry to the Reception class by one year 'for children who are slow-developers'. The study found that pupils placed in a class a year behind their normal age-range performed close to the average of the class in which they had been placed, and that, in Switzerland, 'the dispersion of pupils' mathematical attainments... is much reduced, to about half that in English classes'. They suggested that greater flexibility in age of school-entry than currently practised in England would enhance whole-class teaching and learning.

A study commissioned the National Foundation for Educational Research (Hutchison and Sharp, 1999) used reading test data from three cohorts of

children in one LEA to investigate the persistence of 'season of birth effects' between the ages of six and twelve. They found that autumn-born pupils performed substantially better in reading than their summer-born classmates throughout the primary phase but, in addition, that the 'season of birth effect' continued - though with decreasing influence - until 12 years of age.

Tymms (2002) found that:

> The difference in starting points for the average child aged 4 years and 0 months is about a Standard Deviation (Effect Size = 1) lower than an average pupil age 5 years 0 months... The general increment by age is a very important feature of the Reception class but as pupils get older the importance of age decreases. (p. 40)

Wilson (2003) findings showed that school achievement of the oldest (autumn-born) pupils in a cohort is superior to that of their summer-born classmates. His research looked at 'the birthdate effect' on pupils at a secondary school in Cambridge, and found that both season of birth and gender had a statistically significant influence on the likelihood of a pupil having SEN.

Across Tameside there are a number of pupils who are placed outside their chronological age-group (Tameside, 2005). Reasons for this practice include enabling gifted or talented pupils to access a suitably-extended curriculum by promoting them to a class containing older pupils. Other pupils are kept back a year at the end of the key stage, or enter school at a later date than their chronological peers. A clear rationale is given for these decisions, which are based on the best outcome for the pupils concerned. One reason for a pupil to be placed with a lower age-group is if 'a pupil has a birth date which is in July or August, particularly if they were born prematurely or are small for their age... there is a tendency for all summer-born children to show lower levels of achievement and maturity.'

A report in the *Observer* (Hinsliff, 2005) argued that summer-born children should be allowed to stay on an extra year at primary level to enable them to catch up academically with their peers. It described Lord Dearing's suggestion

that this practice should be introduced in England, and that such a move - made with parental consent - might help to solve the disadvantages, both academic and social, for summer-born pupils. The report cites research on GCSE results of 500,000 children: those born in August achieved on average a third lower grades than those of their autumn-born peers.

In my research I compared the data for two pupils in School B, Class 1: pupil A was aged 4 years 11 months on entry in September 2003; pupil B joined the Reception class in January 2004 aged 4 years 4 months. By January 2004 therefore , pupil A was 5 years 3 months old and had already received one term of primary education: in contrast, pupil B was just over 4 years old and had received no primary education. By the end of KS1 (Year 2), pupil B would be just 6 years and 11 months old, whereas pupil A would be almost eight; in addition, she would have received one more term of primary education than pupil B. When I examined the data for these two pupils they showed the FSP total score for pupil A was 97 whereas that of pupil B was 52; both were girls, neither had SEN, and pupil A had English as an additional language (EAL). Was the difference in their FSP scores due to the 'summer-born' effect? It would be wrong to make this assumption, as so many other factors have an influence on pupil achievement. Some are related to the home, but others are related to the school context and include the number of pupils in a Reception or infant class, and the adult to pupil ratio from school entry to the end of KS1. These two variables are examined below.

Infant class size

The School Standards and Framework Act 1998 provided that:

> from September 2001 at the latest, no infant class may contain more than 30 pupils... However, the government wants this target to be achieved earlier than the statutory deadline. It is therefore providing enough funding to enable schools to meet the pledge by September 2000... All admission authorities should be considering whether they need to adjust their admission arrangements in order to comply with the class size pledge. Many admission authorities will conclude that they should be regulating their admissions to the 1999/2000 year. The legislation makes it possible for them to do

so. The Government is providing additional funding for extra classrooms and teachers so that admission authorities can meet more preferences of parents of 5, 6 and 7 year olds. Legislation will require admission authorities to adjust standard and admission numbers so that they will be compatible with the class size limit...' (DfEE, undated leaflet, reference RICSGS)

School C was one where the standard admission number was in excess of 30, and they applied for additional funding for a fourth teacher in the infant department. In April 1999, the headteacher received a letter (KCC, 1999) stating that School C was one of three in Kent who had been allocated funding for the summer term 1999, to be joined by a further fifty-one primary and infant schools throughout Kent (including School A) who were all 'schools identified in the [Kent School Organisation] plan as requiring an additional class which will have an allocation [of the grant] for the academic year 1999/2000 when admissions are confirmed.' to all headteachers of primary and infant schools in Kent confirming that the then DfEE had announced the approval of Kent's plan to reduce infant class sizes.

All headteachers of qualifying schools received confirmation in April 2000 that a grant from central government for 2000/2001 would be paid into their budget to fund reduced infant class sizes, backdated to September 1999 when the legislation began if necessary, however:

> There is no requirement to match the grant to the actual salary of the new teacher appointed. Schools will be required to sign the certificate for each financial year...confirming that the grant has been spent in meeting the salary and employer on costs for a teacher of an infant class and that without the grant the school would have had infant classes with more than 30 pupils. (KCC, 2000).

An NfER research survey of June 1997, quoted in Jamison, Johnson and Dickson (1998, p. 15) had found that, out of 6,105 registration groups (classes), 91% had a teacher pupil ratio of 1: 5 – 1: 53, whilst the remaining 9% had registration groups with a ratio of 2: 14 – 2: 70. The authors point out the vicious circle of less popular schools having fewer pupils, leading to loss of teachers, leading to larger classes, making the schools less popular...; also, of

the fact that, as funding for school budgets is based on an age-weighted pupil unit, schools with tight budgets have traditionally added more and more pupils to the school roll. Since the new regulations on infant class sizes, this practice has had to stop, and primary schools in particular are finding it increasingly hard to manage their budgets. The financial implications of the infant class size legislation was described by headteachers in the NfER survey (p25) as resulting in higher salary costs which have had to be financed, effectively, at the expense of spending on resources, buildings maintenance, INSET, non-contact time for senior staff, and the use of teaching assistants; the loss of valuable classroom support from teaching assistants was the greatest fear expressed by headteachers and class teachers in the survey. (p. 40).

An international study (Ofsted, 2003a) compared the number of pupils in infant classes in England, Denmark and Finland: in England the average class size in Year 1 was 22.6 pupils, similar to Denmark, whereas Year 1 classes in Finland averaged 12.5. Teaching assistants in Finland were allocated to any Year 1 class that contained more than 12 pupils; in Denmark, teaching assistants - trained for three and a half years at a specialist institution of higher education - worked alongside class teachers to support pupils across the curriculum; two assistants were allocated to each class with more than twenty pupils.

My research data for the basic profiles of the three cohorts studied (see Tables 4, 5 and 6) show class sizes of 21 through to 30 (School C had mixed-age classes to keep the number within statutory limits). When data for the Year 2 cohort were analysed it will be seen that pupils in School A, with a class size of 21, achieved higher NC levels in reading, writing and mathematics than did pupils in School D whose class size was 30. School A also had the highest percentage across the five classes of pupils with SEN and EAL, and a gender balance of fourteen boys to seven girls: could the high NC levels achieved be the result of a smaller class size? Again, there are so many variables that might impact on the results that it is impossible to be certain which combination is the key.

Adult: pupil ratio

Where a school can afford to employ one or more adults to work alongside the class teacher, the performance of their pupils is likely to be enhanced. The infant class size legislation applies to Years 1 and 2, not Reception, but the implications of reduced funding and thus the possibility of fewer support staff has deep for the future of some schools. I know of many primary schools in Kent where tight budgets have meant the redundancy of support staff, to the detriment of the pupils. Two reports in 2002 highlighted the value of teaching assistants in primary schools; a paper from the National Foundation for Educational Research and the Local Government Association (Lee, 2002) looked at the history of using teaching assistants in schools, particularly those working with pupils with SEN. The author cited a recent Ofsted report that had found that:

> Section 10 inspections show that the presence of teaching assistants improves the quality of teaching. This improvement is most marked when the teaching assistant and teacher work in close partnership or when the teaching assistant is following a tightly prescribed intervention or catch-up programme. (p. 27)

A team of researchers from the Institute of Education, University of London (Blatchford *et al*, 2002) used multi-level modelling to look at class sizes and pupil: adult ratio differences in Reception classes and KS1 and the effect of these on pupils' learning. They summarised:

> The statistical analysis of the data indicates that there is a significant effect of class size differences on children's progress in reception for both literacy and mathematics. There was no evidence of an effect of class size upon progress in either literacy or mathematics at either Year 1 or Year 2. In addition it was found that the benefits of smaller classes in the reception year were still evident in both literacy and mathematics at the end of Year 1, but were no longer present at Year 2. There is no evidence that numbers of adults in addition to the teacher have an influence on children's educational progress. (p. 57)

They concluded:

> It was found in this study that the use and effectiveness of adult help in classes *varied between classes,* and that this is probably the

main reason why the quantitative analyses did not show clear evidence of the benefits of classroom support on children's educational progress. (p. 7)

The authors admitted that the timing of their report was before the government's drive to recruit and train teaching assistants; indeed, since their research was concluded, many schools in Kent - and throughout England - have recruited 'higher level teaching assistants' (HLTAs), not only to support pupils in class but to teach classes in the absence of the teacher who is using his or her entitlement to 'Planning, Preparation and Assessment' (PPA) time under the Workforce Reform agenda. This initiative has put school budgets under even greater strain and, in Kent, has led to further redundancies of support staff and even class teachers.

A research study (reported in NfER, 2002) reviewed literature that showed the 'increasingly significant' role that teaching assistants play in supporting teachers and pupils, particularly in English and mathematics lessons at primary level. The study concluded that, whilst the longer-term impact of the work of teaching assistants had yet to be established, the evidence thus far was positive. There was also a clear link between the number of teaching assistants in a classroom and improved performance of the pupils.

In my own research, each Reception class had at least one teacher and one teaching assistant. However, the skills and experiences of these practitioners varied: one class was taught by a graduate teacher (GTP), one was trained in Early Years and the others in primary education; one teacher had a PGCE qualification, another a BEd, others had a BA; all had differing years of teaching experience. Of the assistants, one was a trained nursery nurse, others were learning support assistants or teaching assistants; each was employed for a different number of working hours, and their job descriptions differed. However, in every Reception class there was a ratio of at least one adult to fifteen pupils (some classes had occasional help from parents or other TAs),

and this is to be commended; schools need to use their budget to optimise the adult: pupil ratio (see Chapter 8).

Twins

As I collected the data for my research, I noted that there were three sets of twins in my study: in Year 2, School B; in Year 1, School B; and in Year R, School D. I have used these data in my findings as I considered that being one of a pair of twins might influence a child's future achievement: family size and age of siblings could affect a parent's ability to offer their children a positive home experience. However, research by Professor Peter Tymms and his team at the Curriculum, Evaluation and Management (CEM) Centre, University of Durham in 2002 found no evidence of delayed language development in twins (Tymms, 2002, Chapter 3.12); the team's research had found that twins frequently have richer and earlier language development, and they attribute this to the fact that, since 2004, the greater availability of IVF treatment has increased the proportion of twins from 1:80 to 1:30 live births, and many of these are born to older mothers.

I decided to compare the BA and FSP data for these close siblings, although for the purpose of this thesis I do not have any results for their KS1 teacher assessments; Table 15 shows comparative data for these three pairs of twins.

Table 15: Comparative data for sets of twins

Context	Pupil ref.	Term of birth	Gender	EAL?	SEN?	BA/FSP total
Year 2, School B	Pupil A	Summer	M	Y	N	38 (BA)
Year 2, School B	Pupil B	Summer	M	Y	N	34 (BA)
Year 1, School B	Pupil C	Spring	M	Y	N	87 (FSP)
Year 1, School B	Pupil D	Spring	M	Y	N	88 (FSP)
Year R, School D	Pupil E	Summer	M	N	N	89 (FSP)
Year R, School D	Pupil F	Summer	M	N	Y	88 (FSP)

The two pupils in Year 2 were both boys, with English as an additional language but no special needs. Their BA total scores were 38 and 34: the class mean was 39, the median 38 and the mode 33. In Year 1, the twins' FSP total scores were 87 and 88; both were boys, both with EAL but no SEN. The mean, median and mode for this class were 84, 85 and 89 respectively: the twins' scores were within the norm for this cohort. The set of twins in Year R were also boys, neither had EAL but one had SEN at School Action Plus; their FSP total scores were 89 and 88. The mean, median and mode for this class were 94, 95 and 97 respectively; the scores for this pair of twins were below those for their classmates. Analysis showed that pupil F, with a high level of SEN, had achieved only one FSP point less than his brother; there seemed to be no obvious reason for this difference, and overall the three sets of twins performed similarly to their peers.

Gifted and talented

Each of the nine variables discussed earlier in this chapter could be considered to have a negative influence on a child's future performance; I wanted to examine one indicator that might affect pupil achievement: that of being 'gifted and talented' (G&T).

Pupils in Kent - previously called 'Able Children' (KCC, 1995) - were renamed 'Gifted and Talented Children', with the acronym 'G&T' (KCC, 2002b). Pupils in Bromley were called 'More Able Pupils' (London Borough of Bromley, 1996) and those in Bexley 'The Able Pupil' (London Borough of Bexley, 1992).

A report of conferences (DfE, 1994) had called such pupils 'Exceptionally Able Children' and Sir Ron Dearing, in his keynote address stated that:

> HMI have indicated that, generally, subject specific criteria have not been used to indicate high ability. Surely this is a missed opportunity, as high levels of achievement in any subject must inform the teacher of what to do next with a pupil. Once identified, the curriculum should provide the opportunity to implement appropriate schemes of work. (pp. 4-8)

Gifted pupils were described in the later Kent document (KCC, 2002b) as 'those with particular abilities in one or more subjects of the statutory school curriculum'; talented pupils were those 'with particular abilities in the creative or performing arts such as art and design, music, PE, dance and drama' (p. 5). Teachers were urged to:

> identify the educational, social and emotional needs of pupils with high ability...[and] recognise that there are underachieving pupils who conceal particular aptitudes with poor performance or behavioural problems...
> (p. 6).

The National Association for Gifted Children (NAGC, 2001) stated that:

> A recent estimate suggests that 5-10% of gifted children could have a learning difficulty. These pupils, with "dual exceptionality" (as they are called in the USA) have hitherto tended to slip through both nets. They don't appear high enough achievers to be labelled "gifted" but they seem too bright to be regarded as SEN. Yet practising teachers want to know how to support the creative, able child who struggles to achieve an "average" SATs result. (p. 2)

In January 2000, the DfEE introduced guidance for headteachers and teachers on how best to provide for 'able pupils' within the NLS and NNS, with key points for teaching literacy - at text, word and sentence level, and in speaking and listening - and mathematics, with ideas for extended tasks, problems or investigations within and outside the main curriculum (DfEE, 2000b). White and Ridley (reported in NfER, 2003) undertook a research study that 'presented practitioners with practical advice on how to provide for gifted and talented children both inside and outside the classroom.' They suggested that it was important that LEAs offered designated advisory staff to support schools, and that schools should be encouraged to look at a whole-school approach to G&T pupils. Their study showed that schools should offer clear differentiation within the curriculum, and that whilst mixed-ability classes were appropriate for many lessons, G&T pupils might benefit more from being put into ability groups (sets) for lessons in English, mathematics and foreign languages.

126

A team of five Reception teachers in Kent, led by two Early Years Advisory Teachers, set up a collaborative venture in 2002 with Brunel University and Leuven University, Belgium, entitled 'Identifying, challenging and motivating the gifted and talented in the Foundation Stage' (Primary Excellence Project, 2003). This project looked at the work of Professor Ferre Laevers at Leuven University, who had 'highlighted emotional well-being and involvement as crucial indicators when observing young children'. Over three meetings, the group explored the theoretical perspectives of 'gifted' and 'talented' (including Gardner's concept of Multiple Intelligences see Gardner, 1984, 1993, 1999) and the importance of a stimulating environment to challenge and motivate G&T pupils.

Before commencing my thesis I had considered that if a child had been recognised in the pre-school setting as being gifted and/or talented, then this would probably mean that they would have a higher baseline or FSP score than their peers, in either the reading, writing and mathematics aspects (a 'gifted' child) or in creative or physical development (a 'talented' child) and I therefore included this data set on my SPSS spreadsheet and my manual data collection sheet. However, in schools A and B, the headteachers that I interviewed said that no such early recognition had been made - despite both schools having a nursery - School C had identified one child who they thought might be gifted (he had an exceptional range of vocabulary); only School D was unhesitating in recognising that one child in the Reception class and two children in the Year 1 cohort were gifted and talented, having been so identified in their pre-school settings. As these were the only three pupils so identified out of 451 cases, I discounted them from my data analysis (see Chapter 4.4). However, it is important to consider that such children form an increasingly large group, both in Kent and nationally, and that government is asking schools to improve their SATs results at level 3 and above at KS1 and at level 5 and above at KS2 to reflect the ratio of above-average pupils compared to the norm. For the first time, the 2007 PLASC return asked primary schools to give a percentage for pupils identified as G&T; I would suggest that this indicates the likelihood of

G&T becoming a proxy indicator for the purposes of government measures and future performance tables (now called 'achievement and attainment tables').

6.2: TARGET-SETTING FOR THE YEAR 1 AND YEAR R COHORTS

As a Head Teacher, I am asked by KCC Management Information at 'target-setting time' (early December for primary schools in Kent LEA) to produce targets for the percentage of pupils in Year 1 and Year 2 who are likely to achieve NC levels 3 and above (3+), 2+ and 1+ at the end of KS1. These data are collected each year, together with targets for achievement at the end of KS2 (at levels 5+, 4+, 3+ and 2+) for pupils in Years 3, 4 and 5; targets are not collected for Year 6, as headteachers are asked in December to project likely grades for eighteen months in the future. The KS1 targets are also projected eighteen months in advance; for example, in December 2001 headteachers were asked to project percentages for the grades of 3+, 2+ and 1+ (separately for English and mathematics) for pupils who were then in Years 1 and 2, but who would not be taking the SATs at the end of KS1 until May 2004 and 2003 respectively. As this target-setting is repeated in similar format every year, headteachers are able to amend the percentages if the profile of the cohort changes. In School C, the Year 2 cohort of my research data had been projected, in December 2001, to achieve 23% at level 3+ and 100% at level 2+; there were at that time no pupils with identified SEN (there was at that time space to state this on the form). By the following December, four pupils out of a class of 35 (11%) had been diagnosed as having SEN: the percentages targeted that year had changed to 23% at level 3+ (the same) but 91% in reading, 92% in writing and 97% in mathematics (from this year, English had to be broken down into grades for reading and writing). This year - December 2002 - was the first time that targets needed to be set for the Year 1 cohort of my research; these were, at level 3+: 23% in reading, 3% in writing, and 14% in mathematics; at level 2+, 86% in reading, 89% in writing, and 100% in mathematics; at level 1+, 100% in all three subjects; the class size was 35, with three pupils (9%) with SEN. Guidance from KCC in December 2002 reminded headteachers that:

From September 2002 the Kent Baseline Assessment scheme "Starting Together" will no longer be a statutory requirement for pupils. A new end of Foundation Stage assessment to be known as the "Foundation Stage Profile" will be published in January 2003. This will be a nationally designed scheme and all Reception class pupils will be required to be assessed at the end of the Reception year... Management information will be collecting total baseline scores only to ensure that value-added measures can still be calculated for schools whilst the new assessment requirements embed... Guidance upon expected patterns of progress from baseline to Key Stage 1 is provided on page 17. **We emphasise that this is for guidance only, not for use without the broader context of teachers' professional judgements.** Schools may also wish to consider the value-added data provided by the LEA about baseline (1999-2000) to Key Stage 1 (2002) assessment. (KCC, 2002a).
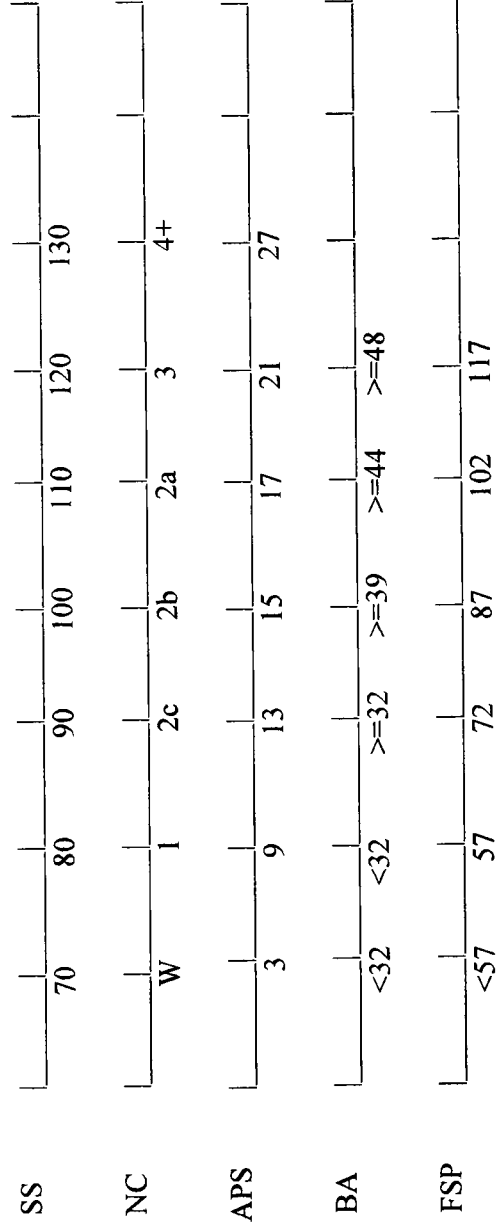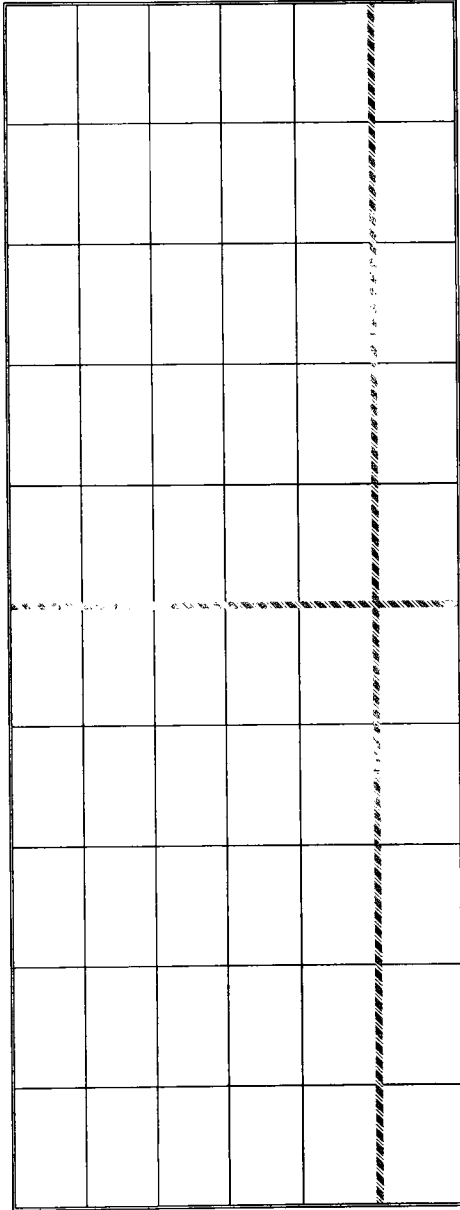
Some headteachers in my district continued to use the 'Starting Together' baseline assessment, but as it was not statutory and Reception teachers would in any case be completing the FSP a few months later, many practitioners decided not to add to their workload. This meant that, for those schools who had not continued to use the BA package, there was no easy way of setting targets for pupils in Year R: they could not just read off data from a table and enter them onto the target-setting form sent to headteachers by KCC. School C had continued with the BA scheme, and were therefore able to project grades for 2004 for their Year 1 cohort in reading, writing and mathematics, as described earlier. However, by December 2003, targets were being requested for the end of KS1 in 2004 and 2005: pupils who had not been assessed under the BA scheme, but (for many schools) only by the FSP. How, then, were headteachers to project these data? The only way was subjectively, by an evaluation of the pupils' completion of various tasks, and by teacher observation and discussion with the child, parents and colleagues. I used this method of projecting future National Curriculum levels for the 2003 and 2004 data collection exercise. However, for the purposes of my thesis the raw data for the Years R and 1 cohorts needed to be analysed by more objective means. If I had data from a number of years, they could help me to produce more accurate tables, based on a 'best-fit' or regression line; as these were not yet available from central or local government, I decided to produce a chart which might give some

indication of likely projections. I used my understanding of the links between scores for Baseline Assessment and projections for KS1 assessment, of 'average' National Curriculum levels at the end of KS1, and of standardised scores and the normal curve of distribution, to help me devise a chart that could be used to project KS1 assessment from FSP total scores. This was an important contribution to knowledge and will be examined in the next section.

6.3:   PRODUCING A PROJECTION CHART

I needed to assign projected NC levels for the end of KS1, for pupils from the 2002-2009 cohort (Year 1) and the 2003-2010 cohort (Year R); these data had been entered into the SPSS database as cohort 1 and 0 respectively. I decided to use graphical representation of three aspects of measurement, and used the normal distribution of standardised scores (from 70 to 130, with 100 as the mean) to which I added the NC sub-levels achievable at KS1 (from W to level 3, with level 2b as the mean), and the BA totals for writing from *'Pathways to Progress'* (KAS, 2001) against their projected grades. This showed that a pupil who had scored 39 points as his or her BA total in Year R would be likely to achieve a NC level 2b - the 'average' score - in writing at the end of KS1, and that this would indicate a SS of 100. I analysed the FSP total scores from the Year R and the Year 1 data and found the mean, the median and the mode for each cohort, for each school and across all eight classes. From the school profiles described earlier, I ascertained that the points ranged from a minimum of 46 to a maximum of 117 - with a mean of 87 - and added these to the projection chart (see Table 16). This gave me graphical representation showing the likely range of NC levels achievable by a pupil at the end of KS1, linked to a BA and FSP total and to normal distribution of standardised scores, and meant that I could use the continuum of FSP scores as a guide to a pupil's likely achievement at the end of KS1. I converted each piece of data from a pupil's FSP total score into a projected NC level and then into an APS at the end of KS1.

Table 16: Comparative chart for projecting possible outcomes from benchmark data in FS/ KS1

SS	70	80	90	100	110	120	130
NC	W	1	2c	2b	2a	3	4+
APS	3	9	13	15	17	21	27
BA	<32	<32	>=32	>=39	>=44	>=48	
FSP	<57	57	72	87	102	117	

Key: SS - Standardised Score; NC - KS1 National Curriculum Teacher Assessment Level; APS - Average Points Score; BA - Baseline Assessment raw total; FSP - Foundation Profile raw total.

131

I did not have enough data to enable me to separate projections for reading, writing and mathematics, so - like the *'Pathways to Progress'* (KAS, 2001) - I gave only a general target-setting grade. I then analysed and represented the data for each cohort in tabular form. When presented thus, the percentages seemed far too low for School B, Class 2 (in the Year 1 cohort) and for School B, Class 1 in the Year R cohort (see Tables 17 and 18). Whilst percentages at Level 2 and above (L2+) for the other class groups meet government requirements, those for Year 1 (32%) and for Year R (also 32%) are obviously well below an acceptable standard for any school.

Table 17: Projections from FSP to KS1 - Year 1 cohort

NC levels	School A	School B1	School B2	School C	School D
N of pupils	25	26	25	32	30
L 3	0	0	0	1	0
L 2a	5	3	2	15	8
L 2b	11	10	3	6	9
L 2c	5	10	3	7	9
L 1	5	3	17	3	4
L W	0	0	0	0	0
L 2+	**21**	**23**	**8**	**29**	**26**
L2+ as %	**84%**	**88%**	**32%**	**91%**	**87%**

N.B. The bold figures indicate the important category of NC level 2 and above.

Table 18: Projections from FSP to KS1   Year R cohort

NC levels	School A	School B1	School B2	School C	School D
N of pupils	28	25	25	36	29
L 3	0	0	0	1	0
L 2a	14	2	1	13	6
L 2b	12	3	11	14	17
L 2c	2	3	7	6	5
L 1	0	17	6	2	1
L W	0	0	0	0	0
L 2+	**28**	**8**	**19**	**34**	**28**
L2+ as %	**100%**	**32%**	**76%**	**94%**	**97%**

N.B. The bold figures indicate the important category of NC level 2 and above.

I needed to explore these data for School B in more detail, to ascertain possible reasons for the anomalies.

Exploring data for School B

Data for School B1 and School B2 in Table 17 relates to the two parallel classes in Year 1 in School B. Of the 26 pupils in Class 1, the ratio of boys to girls was 16: 10, there were 13 pupils with EAL, 5 on the SEN register and 12 were summer-born. The range of FSP scores was 55 to 106, with a mean of 87 and a mode of 84. In Class 2, the 25 pupils had a gender balance of 16: 9, with 10 pupils on the EAL register, 7 with SEN and 6 summer-born. With such similar profiles there was no apparent reason why Class 2 should have such a low projected percentage for Level 2 and above; the range of FSP scores for this class was 51 to 102, with a mean of 72 and a mode of 63.

Table 18 shows the data for the two parallel classes in Year R in the same school. Again, there seems to be no obvious reason for the low projected percentage for Class 1. These data show a range of 51 to 102, a mean of 72 and a mode of 66, compared to the data for Class 2, where the range was 53 to 105, the mean 84 and the mode 97. As with the data for Year 1, the two classes had similar profiles: Class 1 had a gender balance of 15: 10, with 9 pupils with EAL, none on the SEN register and 11 summer-born; Class 2 had a gender balance of 17: 8, 13 pupils with EAL, 2 with SEN and 9 summer-born. In fact, the data for Class 2 - with a higher level of SEN and EAL - might have suggested lower projected APS than Class 1, whereas the reverse was true.

Having analysed the data against EAL, SEN, gender and summer-born factors, I needed to look for other performance indicators that might account for the low projected APS for these two classes. A key indicator of low attainment used by central government is the percentage of pupils eligible for free school meals, but these data are based on a whole-school measure and not individual classes; in any case, I was unable to access these data for reasons described earlier.

I knew that there were two adults - a teacher and a teaching assistant - in each of the four classes, so the adult-pupil ratio was not a factor; both classes had 25 pupils, so class size was also not an issue. Also, School B has a nursery which most of the pupils attend, so pre-school experiences would have been similar for pupils in all four classes. Another indicator that might have affected the data was the actual teacher: collection of FSP data is subjective to a certain extent, and maybe the same person had taught - and assessed - the pupils in Class 1, Year R and the following year in Class 2, Year 1; when I checked this with the headteacher, this was not the case. However, the teacher in Class 1, Year R was new to the teaching profession: she was a graduate teacher who had previously worked as a teaching assistant; did this suggest that she was less confident about assessing her pupils? The data show that one pupil scored 51 on the FSP, the lowest score of the two classes, but there were also two pupils with scores of 102 and one with 100, whilst in Class 2 only one pupil, with 105, was above 99. There was no obvious reason from the data collected as to why the Level 2+ projections were so low.

The school does not moderate judgements across the two parallel classes; one recommendation I would make is that the school introduces this method of validating assessment procedures (see Chapter 8). Also, whilst most schools now have portfolios of moderated evidence against NC sub-levels in the core subjects of English, mathematics and science (and sometimes, in other subjects as well) it is rare to find school portfolios for the Foundation Stage. I know that Kent is currently compiling such an exemplar, but in the interim I would suggest that, as a consortium, we produce such a portfolio (see Chapter 8).

6.4: LINKING PRACTICE TO POLICY AND THEORY

It is evident from reading this thesis that there are three clear aspects to the study:

- ❖ the assessment practices of the four schools;
- ❖ government policy on assessment; and
- ❖ recent and current assessment theory.

Assessment as practice

Colleagues in the four schools studied all use similar methods of assessing their pupils: in the Foundation Stage, through observation of and discussion with the child, backed up by a portfolio of evidence of written achievement - in language and mathematics - and of aspects of personal, social, emotional, creative and physical development; these are compared with expected learning outcomes in the Foundation Stage curriculum and recorded - with the date of achievement - in the Foundation Stage Profile. In KS1, assessment is made against National Curriculum (NC) learning objectives and individual pupil targets through weekly tests in spellings and tables, through hearing a pupil read either in a group situation or individually, and through the response to tasks given out in lessons or as homework. Each of these types of assessment, in both key stages, is used formatively; that is, the teacher uses the data from the activities and tasks to assess the pupil against the appropriate criteria. If the pupil is making expected progress, then targets are set to move that child on to the next level of learning; if progress is slow or non-existent, then the teacher consults with colleagues, the SENCo and outside agencies for further advice. In addition to this formative assessment, teachers in KS1 undertake summative assessment against NC level descriptors, as required by statute, in order to assign levels of attainment at the end of the key stage. The levels achieved, and the Average Points Scores (APS) allocated, are then used by the local authority for value-added purposes; firstly, for Baseline to KS1 data, and secondly, for KS1 to KS2 analysis. This summative assessment at the end of KS1 is undertaken as 'teacher assessment', although in practice teachers use old NC SATs papers – and others produced by central government each year – to help them allocate NC levels in reading, writing, speaking and listening, mathematics and science for each pupil.

Assessment as policy

Government policy is for schools to improve standards over time: through data analysis and target setting for year groups, class groups and individual pupils; through Ofsted and HMI inspections; and through the publication of league tables to rank schools. Assessment is the key to this policy: the Foundation Stage Profile was brought in to assess pupils in the Early Years against the Foundation Stage curriculum guidance; KS1 SATs have been abandoned in favour of teacher assessment at the end of KS1 (although, as previously discussed, teachers still have to use NC tests and tasks to validate their judgements); and, at KS2, SATs are administered, and externally marked to give an objective dimension to the process.

Has government policy on assessment worked? Does regular testing of pupils really lead to an improvement in standards? Not according to results reported by Wherrett (2004): in October 1995, 'fewer than half of school children sitting the required English and maths tests reached the target grade'; in December 2001, 'Schools fail to achieve the government's target of 75% of all 11 year olds reaching level 4 in maths'; in August 2002, 'Schools fail to meet the expected level four pass of 80%'; and in March 2003, 'The education secretary... admits that the new pass rate targets of 85% are "optimistic"... [the] chief inspector of Ofsted argues that the dogmatic adherence to testing and standards is not only alienating teachers but preventing the very improvements they were supposed to create.' Government policy on assessment appears to be flawed: it certainly does not match with recent and current assessment theory.

Assessment as theory

Assessment theory looks closely at whether educational assessment really does measure what it purports to do. My review of the literature shows that assessment theory falls into clear categories that relate to my three research questions:

Research question 1 – *Can benchmark data be used to predict future achievement?* This encompasses the predictive capability of assessment data but more importantly, the validity and reliability of educational testing and measurement; these aspects will be discussed in Chapter 7.1.1.

Research question 2 – *What are the educational implications of using value-added data as measures of pupil performance?* The complex issue of value-added measures and the educational implications of their use by government policy-makers will be examined in Chapter 7.1.2.

Research question 3 – *Can benchmark data be used to support pupils' learning?* This is all about the dichotomy (or synthesis?) of formative and summative assessment, and about the distinction between assessment **for** learning and assessment **of** learning; it will be discussed in Chapter 7.1.3.

6.5:   A GLIMPSE AT THE FUTURE

The political agenda in education continues to change almost on a monthly basis: SATs at KS1 have now been abolished... will this happen in the future to tests taken at the end of KS2?   In September, 2001 Wales and Northern Ireland scrapped league tables (Wherrett, 2004). In Wales, the ministry of education:

abandoned KS1 testing 18 months ago, never had primary school league tables, has ditched secondary school ones, never introduced the literacy and numeracy strategy and already allows teachers to come up with their own targets. (Wendy Berliner in the Times Educational Supplement, May 30, 2003).

It is interesting to note, then, that in Wales the KS2 results in English at level 4 and above overtook those for England by 4% in 2002; in fact, whilst England scores remained static in 2000, 2001 and 2002, those for Wales steadily increased, and in 2001, the Wales scores for level 4 and above in mathematics also outpaced those for England. In an NAHT article on educational reform,

Jane Davidson - Minister for Education and Lifelong Learning, Wales - outlined her goals for her second term in office, to include:

> introducing a new curriculum for three- to seven-year olds, integrating nursery and primary school learning, focusing on child development and learning through play, so our children get the best possible start in life...reviewing testing and assessment at Key Stages 2 and 3. We will ensure that our children's achievements are measurable...One area of particular concern in the battle to raise standards is class sizes. We have already ensured that all infants are taught in classes of 30 or under, as well as reducing class sizes for other age groups... We are committed to cut all junior classes to 30 or under by the end of 2003 and reduce infant class sizes still further. (NAHT, 2003a)

A further report (NAHT, 2004) detailed proposals by a Welsh Assembly task force to review assessment for 11- and 14-year olds; they had earlier that year recommended that tests at Key Stage 3 be phased out, and one proposal was that teacher assessments would be used at the end of KS2, with additional skills tests at the end of Year 5. Initial reactions from headteachers in Wales had been positive, and their NAHT president was quoted:

> One thing that has been pretty apparent is that the abolition of league tables and Key Stage 1 tests has had no bearing on results in Wales at all. If anything results have improved...so I think there will be a groundswell of opinion in England which will look at Wales and say, they've adopted a common sense approach, why don't we do the same? (NAHT, 2004)

(SATs at KS2 and KS3 were in fact abolished in Wales in July 2004, to be replaced by the proposed skills tests at age 10; see Wherrett, 2004.)

CHAPTER 7: CONCLUSIONS

Throughout this thesis I have undertaken a detailed analysis of government policy and its implementation in four schools that I know well; I have applied a reflective and critical approach to a range of government policies. However, my audience is my professional colleagues. With this in mind my research approach, including the written version of this research report, has to take into account that the readers for whom this research is most relevant are fellow headteachers.

I have looked only at data from four schools in West Kent. In considering the various pupil assessment scores in my sample, and in trying to establish the reasons for their variation, I examined possible influences on attainment; some of these are accepted as contextual value-added data by central government, others have been suggested by my professional practice and experience. It is clear that each school in my study has some pupils classified as having special educational needs, and some with English as an additional language. These variables would certainly impact upon learning, but so might pupil mobility and family circumstance. Gender, entitlement to free school meals and being summer born might also impact on or be in some way associated with pupils' learning and hence influence their scores in tests and teacher assessments. The schools in my study use currently-accepted methods of assessing pupils' progress against national criteria, using a variety of tests. The data and findings of this study apply only to young children in Kent; broader generalisations may be useful but have to be carefully considered.

7.1.1 USING DATA TO PREDICT FUTURE ACHIEVEMENT

Research question one asked whether benchmark data could be used to predict future achievement: my conclusion is that accurate prediction is not possible at the present time, given the doubts about the validity and reliability of the assessments used.

Surely, the issue is whether the two types of assessment can be compared: whether the holistic pupil profile from the Foundation Stage, covering thirteen strands of attainment, and the summative assessments at the end of KS1, where teachers make judgements against NC level descriptors in English, mathematics and science, are assessing the same things. I think it is clear from my own experience and the findings of this study, that these two types of teacher assessment are definitely not measuring the same aspects of learning and that therefore a score achieved in the first cannot be used to predict success in the second.

Currently there are no time limits to the tests administered as part of the new teacher assessment at the end of KS1: teachers are asked to make the tests as much like normal classroom practice as possible, with support given where required and extra time allowed for less able pupils or those with emotional or behavioural needs. Also, teachers can now choose which test papers to use to validate their judgements. A study for NfER (Reed and Lewis, 2005) found that all tasks and tests from the 2005 version of the KS1 SATs were used more frequently than those from 2004. One respondent, however, stated that 'We chose the 2004 tests for maths as we felt the level 3 test for 2005 was unusually difficult' while another thought that 'the mark scheme [for 2004] is heavily weighted towards a letter.' (p. 16). The four schools in my study used a combination of NC tests from 2003, 2004 and 2005 to improve their scores. If schools are permitted to choose which tasks and tests to administer on the basis of perceived level of difficulty, this raises issues about validity and reliability, calling into question again the predictive nature of the initial test scores.

My study - of data relating to some pupils with considerable disadvantages according to government performance indicators - showed means of 84, 86, 87, 89 and 96 for the Year 1 cohort, and of 72, 81, 84, 97 and 100 for the Year R cohort; all but one class had a mean above the total FSP score suggested by central government. One pupil in School C had achieved the maximum 117 score (see Chapter 5.1); the teacher assured me that this child was indeed working well within NC Level 1, having entered the Reception class with a good command of both receptive and expressive language, and having good personal and social skills. A KCC adviser told Early Years practitioners in 2005 that the message coming from central government was that an 'average' pupil at the end of the Foundation Stage should achieve 6 points in each of the ELGs, making a total FSP score of 78. (This would imply that no pupil was working within Level 1 of the National Curriculum, evidenced by his or her achieving point 9 in any of the thirteen aspects of learning.)

If the government claim is correct, schools should be achieving more level 3s at the end of KS1 than they are at present. When this issue was put to my headteacher colleagues, all said that the two types of assessment – FSP, and TA at the end of KS1 – were measuring different things; that the first reflected a child's pre-school and social background whereas the second was more influenced by the child's learning experiences at school.

It is clear from my study of the literature and my own research (see Chapter 5) that the government's over-reliance on figures could lead to a narrow interpretation of numbers or categories and thus to errors: some pupils who have the appropriate knowledge may not be identified, and some may be identified as having the necessary knowledge when in fact they do not.

7.1.2   USING VALUE-ADDED DATA TO MEASURE PERFORMANCE

Research question two examined the educational implications of using value-added data as measures of pupil performance: my conclusion from this study is that the new contextual value-added data can be used to help schools measure pupil performance, but that other, more detailed, analysis is needed to give the full picture.

From this research (see Chapter 6.1) I would hesitate to accept the new contextual value-added data without further analysis of other variables that might influence achievement (a view, incidentally, which is supported by other researchers – see Butterfield, 1995; Lindsay and Desforges, 1998; Goldstein *et al*, 2000; Benton, 2003; Gorard, 2006; Schagen and Styles, 2006).

Benton (2003) spoke of schools that were 'particularly helpful for low ability students but less so for pupils with high attainment' and others which 'prove particularly beneficial for students from ethnic minorities' (p. 9); this was certainly true of the schools in my study. Two of the schools had a high percentage of pupils from ethnic minority backgrounds and were far more 'geared up' – in terms of expertise, resources and experience – to meet the needs of their pupils than the other two schools, who had to bring in help from outside agencies, meaning that support was delayed. One of the four schools had 'a good name' for supporting pupils with special educational needs, and parents chose to send their low-ability children there in preference to other schools in the locality; the SENCo of this school worked closely with colleagues to offer the best possible progression for the lowest achievers and those who were emotionally vulnerable.

My literature search found no information specifically for faith schools, so I cannot comment about the success or otherwise of their achieving progress in value-added terms compared to non-faith schools, although Goldstein's team found that 'School denomination was also studied but did not contribute

significantly to the predictions.' (Goldstein *et* al, 2006, p. 6). Two of the four schools in my study were faith schools, but the differences between them   as percentage of EAL, SEN, FSM and mobility - were greater than the similarity of the schools' ethos; there appeared to be no relationship between the data I collected from each of the faith schools, nor between them and the non-faith schools.

7.1.3   USING BENCHMARK DATA TO SUPPORT PUPILS' LEARNING

Research question three asked whether benchmark data can be used to support pupils' learning: the answer is yes, if the data are used to diagnose strengths and weaknesses in current practice or to inform future learning.

My research shows that benchmark data from both summative and formative assessment support pupils' learning in the primary classroom. Summative assessment – through the FSP in Year R, optional teacher assessment using old SATs papers at the end of Year 1, and the new teacher assessment in Year 2 - creates a climate of inquiry whereby teaching and support staff, senior management teams and governing bodies examine the data and decide whether class and cohort targets have been met; if they have not, then school policies and practices can be evaluated and revised to bring about the necessary improvements.

The practitioners in my study – together with most teachers in Kent - also use formative assessment to evaluate pupil progress and to set targets for improvement. Formative assessment – through observation, discussion with the child, and the marking of and feedback from tasks – is undertaken every day in primary schools, and gives an ongoing picture of each child's progress against the Early Learning Goals or National Curriculum level descriptors: it informs the teacher's medium- and short-term planning, leading to improved performance.

CHAPTER 8: RECOMMENDATIONS

Analysing data

Headteachers lead their schools in collecting and analysing data from teacher assessment at the end of the Foundation Stage and KS1, and from statutory tests at the end of KS2; from optional SATs in Years 3, 4 and 5; and from teacher assessments of various types throughout the primary phase. The results of these analyses are represented as spreadsheets, tables and graphs, and discussed at class level and by the leadership teams within the school. Data are used to set annual targets for the percentage of pupils expected to achieve certain levels in the National Curriculum at the end of KS2; these data are given to local authorities, who send them on to central government. Governing bodies and School Improvement Partners (SIPs) work with teaching and support staff to improve performance through tracking pupils' progress and analysis of data from RAISEonline reports and other data sent by central and local government. These data need to be analysed further, with regard to possible influences on attainment, and headteachers and assessment coordinators need to be familiar with the ways in which data are used by government to predict pupil achievement. Early intervention is essential if a pupil has any form of special educational needs: data from the Foundation Stage Profile should be used to inform Year 1 teachers about each child's progress and learning needs, and to access support from outside agencies where required.

Recommendation 1
School management teams should lead their staff in a fine grade analysis of all available data and an awareness of the performance indicators that might influence pupil achievement. They should use these data and analysis to examine policy and practice, leading to improved performance.

Compiling a school portfolio

Each of the schools in my study used the same form of gathering evidence on pupil progress: a portfolio for each child, with pieces of work annotated by the practitioner and showing the date, any support that had been needed or whether it was the child's own unaided work. For outcomes such as a pupil's ability to make friends or to share equipment, or for activities such as building with construction apparatus or riding a tricycle, the teacher would make written notes of her observations, or use audio- or tape-recordings, together with still photographs (digital, usually) or even videotape. This evidence base would then be used to assess pupils against the ELGs and to record the varying rates of progress throughout the year. Many of these visual data were stored as pupil files and folders on the school's central computer system or the teacher's own laptop.

However, whilst all the Reception practitioners in my study collect samples of work for individual pupil portfolios, there seems to be little evidence of agreement trialling across schools or even between parallel classes within a school. Most schools now have a portfolio of evidence, levelled to National Curriculum descriptors for each of the four core subjects of English, mathematics, science and ICT; I would suggest that schools should do the same for the six areas of learning in the Foundation Stage curriculum, matched to the Early Learning Goals, and that the evidence should be moderated at local level, either through consortia or the Cluster.

Recommendation 2

Reception class teachers should compile a school portfolio of evidence of pupil achievement which has been 'agreement trialled' within the school – particularly where there are parallel classes - and moderated with practitioners from other schools across the consortium or Cluster.

Improving practice

Statutory regulations require only that data are sent from schools to the DfES for each pupil and for each of the 117 statements of attainment; in Kent, it is up to the individual school to choose which method to use: an Excel spreadsheet, an e-Profile report or *Assessment Manager,*

collecting and collating data. Schools A, B and D use the spreadsheet (pre-populated by the local authority with pupil names, dates of birth, gender and UPNs) whilst School C uses the computer program. For data collection in 2006, all four schools are going to experiment with the e-Profile, but will probably use the computer program; this is the more familiar format, as it is the means by which data for KS1 Teacher Assessments and KS2 SATs are already sent to the local authority.

Tracking pupil progress and collating the data takes a great deal of time, and schools need to find ways to reduce the workload on teachers and management teams. Reception teachers find the expectations and requirements of central government about summative assessment in the Foundation Stage extremely arduous: evidence files that have to be presented to local authority moderators are large, unwieldy and time-consuming to compile.

Recommendation 3

Schools should examine the various ways of collecting and collating data for summative assessment in the Foundation Stage and implement procedures that help to reduce teacher workload.

REFERENCES

ASHBY, J. AND SAINSBURY, M., 2001. *How do schools use national curriculum test results? A survey of the use of national curriculum test results in the management and planning of the curriculum at Key Stages 1 and 2.* [online] National Foundation for Education Research. Available at http://www.nfer.ac.uk/research/down-pub.asp [Accessed 15.11.03]

ASSESSMENT REFORM GROUP, 1999. *Assessment for learning: beyond the black box.* Cambridge: University of Cambridge. Also available at www.assessment-reform-group.org.uk

ASSESSMENT REFORM GROUP, 2002. *Assessment for learning: 10 research-based principles to guide classroom practice.* [online] Available at http://www.aaia.org.uk [Accessed 14.4.07].

ASSESSMENT REFORM GROUP, 2002. *Testing, motivation and learning.* [online] Assessment Reform Group. Available at http://arg.educ.cam.ac.uk/CIE3.pdf [Accessed 11.1.06]

BENTON, T., 2003. *Measuring schools' value-added performance: the importance of picking the right outcome.* Paper presented at the British Educational Research Association Annual Conference, Heriot-Watt University, Edinburgh, 11-13 September 2003.

BERLINER, W., 2003. Look to the west as results rise. *The Times Educational Supplement,* May 30, 2003, p. 20.

BLACK, P. 1998. *Testing: friend or foe? Theory and practice of assessment and testing.* London: The Falmer Press.

BLACK, P. AND WILIAM, D., 1998. *Inside the black box: raising standards through classroom assessment.* London: King's College, University of London.

BLATCHFORD, P., MARTIN, C., MORIARTY, V., BASSETT, P. AND GOLDSTEIN, H., 2002. *Pupil adult ratio differences and educational progress over reception and key stage 1. Research report No. 335.* Nottingham: Department for Education and Skills.

BRITISH EDUCATIONAL RESEARCH ASSOCIATION ASSESSMENT POLICY TASK GROUP, THE, 1993. Clevedon, Avon: Multilingual Matters Ltd.

BROWNE, W., AND GOLDSTEIN, H., 2002. *An introduction to Bayesian multilevel (hierarchical) modelling using MlwiN.* [online] Available at http://www.maths.nottingham.ac.uk/personal/pmzwjb/materials/brussels.pdf [Accessed 3.2.05]

CLARKE, S., 1998. *Targeting assessment in the primary school: strategies for planning, assessment, pupil feedback and target setting.* London: Hodder and Stoughton.

CLARKE, S., 2000. *Developing practical strategies for formative assessment* (workshop handouts). [online] Available at http://www.aaia.org.uk [Accessed 14.4.07]

CLARKE, S., 2001a. *Unlocking formative assessment: practical strategies for enhancing pupils' learning in the primary classroom.* London: Hodder and Stoughton.

CLARKE, S., 2001b. *The power of focused feedback.* [online] Available at http://www.aaia.org.uk [Accessed 14.4.07].

CLARKE, S., 2003. *Enriching feedback: oral and written feedback from teachers and children.* London: Hodder and Stoughton.

CLARKE, S., 2005. *Formative assessment in action: weaving the elements together.* Oxon: Hodder Murray.

COHEN, L., MANION, L AND MORRISON, K., 2001. *Research methods in education.* 5th edition. London and New York: Routledge Falmer.

CRESWELL, J., 1998. *Qualitative inquiry and research design: choosing among five traditions.* Thousand Oaks, California; London & New Delhi: Sage Publications.

DEARING, R., 1994. *The National Curriculum and its assessment: Final report.* London: School Curriculum and Assessment Authority.

DEPARTMENT FOR EDUCATION, 1994. *Exceptionally able children: report of conferences.* London: Department for Education.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, undated. *Reducing infant class sizes — a guide for schools.* Leaflet from DfEE, undated, reference RICSGS.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1996a. *Nursery education: desirable outcomes for children's learning on entering compulsory education.* London: Department for Education and Employment/School Curriculum and Assessment Authority.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1996b. *Publication of 1996 Key Stage 2 National Curriculum assessment results in primary school performance tables.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1996b. *Setting targets to raise standards: a survey of good practice.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1997a. *Excellence in schools.* London: Her Majesty's Stationery Office.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1997b. *Nursery education voucher scheme news* (March). London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1997c. *From targets to action. Guidance to support effective target-setting in schools.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1998a *Baseline assessment of pupils starting primary school* (Circular 6/98). London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1998b. *The national literacy strategy: framework for teaching.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1998c. *The national literacy strategy: framework for teaching.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1998d. *Numeracy matters: the preliminary report of the Numeracy Task Force.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1999a. *Differentiated performance criteria for mathematics: an addendum to supporting the target setting process.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1999b. *The national curriculum: handbook for primary teachers in England.* London: Department for Education and Employment/ Qualifications and Assessment Authority.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 1999c. *The national numeracy strategy: framework for teaching mathematics from reception to year 6.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 2000a. *Guidance on the education of children and young people in public care.* London: Department for Education and Employment /Department of Health.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 2000b. *National literacy and national numeracy strategies: guidance on teaching able children.* London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND EMPLOYMENT, 2000c. *Education protects - guidance on the education of children and young people in public care* (Circular DfEE 0269/2000). London: Department for Education and Employment.

DEPARTMENT FOR EDUCATION AND SKILLS, 2001a. *Towards the National Curriculum for mathematics: examples of what pupils with special educational needs should be able to do at each P level.* London: Department for Education and Skills.

DEPARTMENT FOR EDUCATION AND SKILLS, 2001b. *Special educational needs: code of practice.* London: Department for Education and Skills.

DEPARTMENT FOR EDUCATION AND SKILLS, 2001c. *Inclusive schooling: children with special educational needs* (ref. DfES/0774/2001). London: Department for Education and Skills.

DEPARTMENT FOR EDUCATION AND SKILLS, 2001d. *Supporting the target setting process: guidance for effective target setting for pupils with special educational needs.*
(ref. DfES 0065/2001). London: Department for Education and Skills.

DEPARTMENT FOR EDUCATION AND SKILLS, 2002. *Special educational needs - Update 9, June 2002.* London: Department for Education and Skills.

DEPARTMENT FOR EDUCATION AND SKILLS, 2002d. *The national literacy and numeracy strategies. Building on improvement.* London: Department for Education and Skills.

DEPARTMENT FOR EDUCATION AND SKILLS, 2003. *Excellence and enjoyment: a strategy for primary schools.* London: Department for Education and Skills.

DEPARTMENT FOR EDUCATION AND SKILLS, 2004. *Every child matters: change for children in schools.* Department for Education and Skills. [online]
Available at http://www.teachernet.gov.uk/publications [Accessed 25.4.05]

DEPARTMENT FOR EDUCATION AND SKILLS, 2005. *KEEP: Key elements of effective practice.* London: Department for Education and Skills.

DEPARTMENT OF EDUCATION AND SCIENCE, 1987. *National Curriculum Task Group on Assessment and Testing: Report*. London: Department of Education and Science.

DEPARTMENT OF EDUCATION AND SCIENCE, 1992. *Reporting pupils' achievements to parents* (Circular 5/92, 29 April 1992). London: Department of Education and Science.

EDUCATION REFORM ACT, THE, 1988. London: Her Majesty's Stationery Office.

EFFECTIVE PROVISION OF PRE-SCHOOL EDUCATION (EPPE) PROJECT, 2003. *The EPPE symposium at the British Educational Research Association (BERA) annual conference, Heriot-Watt University, Edinburgh, Scotland, 11-13 September 2003*. Booklet containing texts of presentations, given out at the symposium.

EQUAL OPPORTUNITIES COMMISSION, 1991. *Sex discrimination in schools - a guide for governors*. Manchester: Equal Opportunities Commission.

GARDNER, H., 1984. *Frames of mind: the theory of multiple intelligences*. 2nd edition. London: Fontana Press.

GARDNER, H., 1993. *Multiple intelligences: the theory in practice*. New York: Basic Books.

GARDNER, H., 1999. *Intelligence reframed: multiple intelligences for the 21st century*. New York: Basic Books.

GENERAL TEACHING COUNCIL FOR ENGLAND, 2003. *Research of the month: Inside the Black Box*. [online] General Teaching Council for England. Available at http://www.gtce.org.uk/research/standstudy.asp [Accessed 15.11.03]

GILBORN, D. AND GIPPS, C., 1996. *Recent research on the achievement of ethnic minority pupils*. London: Office for Standards in Education/ Her Majesty's Stationery Office

GILBORN, D. AND MIRZA, H., 2000. *Educational inequality: mapping race, class and gender*. London: Institute of Education, University of London.

GIPPS, C. AND MURPHY, P., 1994. *A fair test? Assessment, achievement and equity*. Buckingham and Philadelphia: Open University Press.

GOLDSTEIN, H., HUIQI, P., RATH, T. AND HILL, N., 2000. *The use of value added information in judging school performance*. London: Institute of Education, University of London.

GORARD, S., 2006. How useful is value-added analysis of schools? *Research Intelligence,* May 2006, Issue 95, 10-12.

HINSLIFF, G., 2005. *Youngest pupils face extra year at primary school.* Available at http://education.guardian.co.uk/primaryeducation/story/0,11146,1500225,00.html [Accessed 27.1.06]

HORTON, T. ed., 1990. *Assessment debates.* London: The Open University.

HUTCHISON, D. AND SHARP, C., 1999. *A lasting legacy? The persistence of season of birth effects.* Paper presented at the British Educational Research Association Conference, University of Sussex at Brighton, 2-5 September 1999. Available at http://www.nfer.ac.uk/publications/other-publications [Accessed 27.1.06]

HYNE, S., 2003. *Directory of Additional Educational Needs for the Crossroads Local Learning Group (Dartford and Gravesham).* Paper presented at a meeting of the Crossroads Local Learning Group (Dartford and Gravesham) at Holy Trinity CEP School, Gravesend, Kent on 22 May 2003. Hard copy and disk available from St. Botolph's Church of England Primary School, Dover Road, Northfleet, Kent DA11 9PL.

JAMISON, J., JOHNSON, F. AND DICKSON, P., 1998. *Every pupil counts: the impact of class sizes at Key Stage 1.* Slough: National Foundation for Educational Research.

KELLAGHAN, T. AND GRISAY, A., 1995. *In:* Organisation for Economic Cooperation and Development, 1995. *Measuring what students learn.* Paris: Organisation for Economic Cooperation and Development, 1995, pp 45-55.

KENT ADVISORY SERVICE, 1999. *Baseline assessment 1999.* Pack sent to reception teachers of all primary schools, 9 July 1999. Maidstone: Kent County Council.

KENT ADVISORY SERVICE, 2001. *Using target setting with individual children to raise achievement.* (Includes *Pathways to Progress* chart.) Maidstone: Kent County Council.

KENT ADVISORY SERVICE, 2004. *Pupil achievement review and target setting.* Maidstone: Kent Advisory Service/Management Information.

KENT COUNTY COUNCIL, 1995. *Able children: guidance for teachers, parents and governors on supporting very able children.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 1997. *Mind the gap - the gender divide.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 1998. *Target setting.* Letter from Trevor Sandford, County Schools Effectiveness Officer, to all Kent schools with primary age pupils, 15 July 1998.

KENT COUNTY COUNCIL, 1999a. *Baseline assessment 1999.* Letter from Karen Kennedy, Management Information Officer, Kent County Council, Maidstone to all heads of primary, infant and special schools, 4 March 1999.

KENT COUNTY COUNCIL, 1999b. *Reducing infant class sizes.* Letter from Dave Shipton, Policy Officer (LM), KCC, Maidstone to the Headteacher and Chair of Governors for all primary and infant schools, 15 April 1999.

KENT COUNTY COUNCIL, 2000. *Infant teacher grant.* Letter from Dave Shipton, Policy Officer (LM), Kent County Council, Maidstone to all primary and infant schools qualifying for infant teacher grant, 20 April 2000.

KENT COUNTY COUNCIL, 2001. *Baseline (1997-8) to KS1 (2000) value-added analysis for Kent schools.* Letter from John Wiseman, Statistician/Researcher, Management Information, Kent County Council, Maidstone to primary school headteachers, 14 June 2001.

KENT COUNTY COUNCIL, 2002a. *Primary schools target setting 2002 guidance notes.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2002b. *Gifted and talented children: guidance for headteachers, teachers and governors on supporting gifted and talented children.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2003a. *Value added analysis for primary schools in Kent.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2003b. *Primary target setting guidance 2003.* Letter from Carol Parsons, Assistant Director - Schools Advisory Service (Acting), KCC, Maidstone, 7 October 2003.

KENT COUNTY COUNCIL, undated. *Child placed for adoption.* Letter from John Mitchell, Joint Commissioning Officer and Lead Officer for Looked After Children, Kent County Council, Maidstone, to all headteachers in Kent schools.

KENT COUNTY COUNCIL, 2004a. *Integrating test analysis into whole-school planning.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2004b. *Foundation stage profile moderation.* [online] Kent County Council. Available at http://www.clusterweb.org.uk/cwpages/Communication/ebulletindetail.cfm?eb -bulletin.asp [Accessed 25.11.04]

KENT COUNTY COUNCIL, 2004c. *Welcome to the minority communities achievement service.* [online] Kent County Council. Available at http://www.clusterweb.org.uk/cwpages/school/MCAS.cfm [Accessed 17.11.04]

KENT COUNTY COUNCIL, 2004d. *Kent school organisation plan 2003/4 - 2007/8.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2004e. *Boys can do better...* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2004f. *Supplementary guidance for schools in recording pupils' special educational needs in the pupil level annual schools census (PLASC).* Available at http://www.clusterweb.org.uk/G;Ed [Accessed 18.11.04]

KENT COUNTY COUNCIL, 2004g. *P-scale assessment data collection.* Letter from Liz Windon, Principal Management Information Officer, Kent County Council, Maidstone, to all headteachers, SENCo and SEN governors in Kent schools, 27 April 2004.

KENT COUNTY COUNCIL, 2004h. *Kent Area EYDC News, Autumn 2004.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2004i. *End of Key Stage 1 assessment arrangements - Trial schools 2004: Guidance to Kent schools.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2004j. *Raising standards in education. Keeping you informed.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2004k. *Guide to managing data effectively in the primary school.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2005a. *Annual Schools' Census 2005 Annexes Booklet.* Maidstone: Kent County Council.

KENT COUNTY COUNCIL, 2005c. *Doing away with 'W': LA guidance on making summative assessments of the attainment of young children which inform subsequent learning and teaching.* [online] Kent County Council. Available at http://www.clusterweb.org.uk [Accessed 12.2.05]

KIRKUP, C. AND TWIST, L., 2004. Assessment for learning. *NfER Annual Report 2003-2004,* 58, 8-9.

LEE, B., 2002. *Teaching assistants in schools: the current state of play.* Slough: National Foundation for Educational Research/Local Government Association.

LINDSAY, G., 2001. Baseline assessment: what purpose, and for whose benefit? *Education 3-13,* October 2001.

LINDSAY, G. AND DESFORGES, M. 1998. *Baseline assessment - practice, problems and possibilities.* London: David Fulton Publishers.

LONDON BOROUGH OF BEXLEY, 1992. *Primary guidelines on the able pupil.* Bexley, Kent: London Borough of Bexley.

LONDON BOROUGH OF BROMLEY, 1996. *The education of more able pupils.* Bromley, Kent: London Borough of Bromley.

MILLARD, E., 1996. *Some thoughts on why boys don't choose to read in school.* An occasional paper. London: National Literacy Trust.

MILLARD, E., 1997. *Differently literate: boys, girls and the schooling of literacy.* London: Routledge Falmer.

MILLARD, E., 2001. Boys, girls and writing. *Literacy Today,* 26, 22.

MILLARD, E., 2003. Gender and early childhood literacy. *In:* N. Hall, J. Larson and J. Marsh, eds. *Handbook of early childhood literacy.* London, New Delhi, Thousand Oaks, California.

MOSS, G., 1999. Boys and non-fiction: cause or effect. *Literacy Today,* 21, 19.

MOSS, G., 2000. Raising boys' attainment in reading: some principles for intervention. *Reading,* 34 (3), 101-106.

NATIONAL ASSOCIATION FOR GIFTED CHILDREN, 2001. *A bright challenge.* Milton Keynes: The National Association for Gifted Children.

NATIONAL ASSOCIATION OF HEAD TEACHERS, 2002. Developing good practice in reception classes. *Primary Leadership Paper 5,* March 2002, 24-37.

NATIONAL ASSOCIATION OF HEAD TEACHERS, 2003a. The battle to reduce class size will continue. *Leadership Focus,* Autumn 2003, 30-31.

NATIONAL ASSOCIATION OF HEAD TEACHERS, 2003b. Performance tables and end of Key Stage 1 assessment (England only). *Guidance paper from the National Association of Head Teachers,* First issued October 2003.

NATIONAL ASSOCIATION OF HEAD TEACHERS, 2004. Wales set to reduce testing. *Leadership Focus,* Issue 13, Spring 2004, 7.

NATIONAL CURRICULUM TASK GROUP ON ASSESSMENT AND TESTING, 1987. *Report.* London: Department of Education and Science.

NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH, 2002. The impact of teaching assistants in schools. *nferNews*, Autumn 2002, 9.

NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH, 2003. What works for gifted and talented pupils. *nferNews*, Autumn/Winter 2003, 6.

OFFICE FOR STANDARDS IN EDUCATION, 1993. *Boys and English 1988-1991.* London: Department for Education.

OFFICE FOR STANDARDS IN EDUCATION, 1996. *The teaching of reading in 45 Inner London primary schools.* London: Ofsted Publications Centre.

OFFICE FOR STANDARDS IN EDUCATION, 1997a. *PANDA Reports: 1. Primary pilot document.* London: Her Majesty's Inspectorate.

OFFICE FOR STANDARDS IN EDUCATION, 1997b. *The teaching of number in three inner-urban LEAs.* London: Ofsted Publications Centre.

OFFICE FOR STANDARDS IN EDUCATION, 2002a. *The National Numeracy Strategy: the first three years 1999-2002* (Document reference number HMI 554). London: Her Majesty's Inspectorate.

OFFICE FOR STANDARDS IN EDUCATION, 2002b. *The National Literacy Strategy: the first three years 1999-2002* (Document reference number HMI 555). London: Her Majesty's Inspectorate.

OFFICE FOR STANDARDS IN EDUCATION, 2002c. *The annual report of Her Majesty's Chief Inspector of Schools. Standards and quality in education 2000/01.* London: Her Majesty's Inspectorate.

OFFICE FOR STANDARDS IN EDUCATION, 2002d. *The curriculum in successful primary schools.* London: Office for Standards in Education.

OFFICE FOR STANDARDS IN EDUCATION, 2003a. *The education of six year olds in England, Denmark and Finland: an international comparative study.* London: Her Majesty's Inspectorate.

OFFICE FOR STANDARDS IN EDUCATION, 2003b. *Handbook for inspecting nursery and primary schools.* London: Her Majesty's Inspectorate.

OFFICE FOR STANDARDS IN EDUCATION, 2004. *The annual report of Her Majesty's Chief Inspector of Schools. Standards and quality in education 2002/03.* London: Her Majesty's Inspectorate.

PICKERING, J., 1999. *Raising boys' achievement.* Stafford: Network Educational Press Ltd.

PRIMARY EXCELLENCE PROJECT, 2003. Kent LEA, Brunel and Leuven Universities Research Project. *Peptalk*, Issue 02, Autumn 2003, 4.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 1998a. *Can do better: raising boys' achievement in English.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 1998b. *A guide to using national curriculum assessment data in primary schools.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 1999a. *The revised national curriculum for 2000: what has changed?* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 1999b. *Target setting and assessment in the National Literacy Strategy.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 1999c. *Early learning goals.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 1999d. *Keeping track: effective ways of recording pupil achievement to help raise standards.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2000. *Curriculum guidance for the Foundation Stage.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2001a. *National scheme for baseline assessment practitioners' panel.* Letter from Gail Cowmeadow, Principal Officer, Baseline Assessment, QCA to [chosen named practitioners], 14 May 2001.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2001b. *Planning for learning in the foundation stage.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2003a. *Foundation Stage Profile handbook.* London: Qualifications and Curriculum Authority /Department for Education and Skills.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2003b. *Assessment for learning: using assessment to raise achievement in mathematics.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2004a. *National curriculum tests 2004: implications for teaching and learning from the 2004 tests. Key Stage 1: English.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2004b. *National curriculum tests 2004: implications for teaching and learning from the 2004 tests. Key Stage 1: Mathematics.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2004c. *National curriculum tests 2004: implications for teaching and learning from the 2004 tests. Key Stage 2: English.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2004d. *National curriculum tests 2004: implications for teaching and learning from the 2004 tests. Key Stage 2: Mathematics.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2004e. *National curriculum tests 2004: implications for teaching and learning from the 2004 tests. Key Stage2: Science.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2004f. *Assessment and reporting arrangements: foundation stage, years 1 and 2, 2005.* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2004g. *Foundation Stage Profile: additional guidance for practitioners..* London: Qualifications and Curriculum Authority.

QUALIFICATIONS AND CURRICULUM AUTHORITY, 2005. *Continuing the learning journey.* London: Qualifications and Curriculum Authority.

REED, M. AND LEWIS, K., 2005. *Key Stage 1 evaluation of new assessment arrangements.* Slough: National Foundation for Educational Research.

REYNOLDS, C., LIVINGSTON, R. AND WILLSON, V., 2006. *Measurement and assessment in education.* Boston, New York, London: Allyn and Bacon.

SAINSBURY, M., 2001. *National curriculum assessment development and pre-testing: KS1 English 2003-2006.* Available at http://www.nfer.ac.uk/research-areas/pims-data/outlines/national-curriculum-assessment-development.html [Accessed 27.1.06]

SAMUELS, L., 1998. *'Value added' measurement of school effectiveness: an overview.* Slough: National Foundation for Educational Research.

SAUNDERS, L., 1998. *Value added measurement of school effectiveness: an overview.* Slough: National Foundation for Educational Research.

SAUNDERS, L., 1999. *Value added measurement of school effectiveness: a critical review.* Slough: National Foundation for Educational Research.

SCHAGEN, I., 2003. *Methods for presenting value-added analyses effectively for a wider audience, London, 2003.* Slough: National Foundation for Educational Research. Available at http://www.nfer.ac.uk/publications/other-publications/conference-papers/pdf_docs.VAM.pdf [Accessed 11.1.06]

SCHAGEN, I. AND BENTON, T., 2003. *The relationships between school and pupil background factors and progress from Key Stage 1 to Key Stage 2, Edinburgh, 2003.* Paper presented at the British Educational Research Association Annual Conference, Heriot-Watt University, Edinburgh, 11-13 September 2003. Available at http://www.nfer.ac.uk/publications/other-publications/conference-papers/ [Accessed 11.1.06]

SHAGEN, I., 2004. *Weighing the baby or fattening it? The use of data to inform school evaluation.* Paper presented at the NFER/ConfEd Conference, London, 8 October 2004. Available at http://www.nfer.ac.uk/pub;ications/other-publications/conference-papers/ [Accessed 11.1.06.]

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY, 1996. *Review of assessment and testing. Report from the School Curriculum and Assessment Authority to the Secretary of State for Education and Employment December 1995.* London: School Curriculum and Assessment Authority.

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY, 1997a. *Inform: news from the School Curriculum and Assessment Authority* (March). London: School Curriculum and Assessment Authority.

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY, 1997b. *Baseline assessment: information for schools* (June, ref: COM*/97/806). London: School Curriculum and Assessment Authority.

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY, 1997c. *The national framework for baseline assessment: criteria and procedures for the accreditation of baseline assessment schemes* (June, ref: COM/97/807). London: School Curriculum and Assessment Authority.

SIZMUR, S. AND BURLEY, J., 1997. *Assessment in the core subjects in 1996: Key Stage 1 evaluation report.* London: School Curriculum and Assessment Authority.

STRAND, S. AND DEMIE, F., 2006. Pupil mobility and attainment. *British Educational Research Journal,* 32 (4), 551-566.

TORRANCE, H. AND PRYOR, J., 2001. How can classroom assessment be used to support children's learning? (updated). [online] Available at http://www.standards.dfes.gov.uk [Accessed 14.4.07]

TYMMS, P., [UNDATED]. *Analysing your school's data and setting targets for primary schools* (revised edition). University of Durham: CEM Centre.

TYMMS, P., 2002. *Baseline assessment and monitoring in primary schools.* London: David Fulton Publishers.

TYMMS, P., MERRELL, C., AND JONES, P.,2004. Using baseline assessment data. *British Educational Research Journal,* 30 (5), 676-684.

WHERRETT, S., 2004. *The SATs story.* Available at http://education.guardian.co.uk/sats/story/0,13294,1289880,00.html [Accessed 29.1.06]

WHETTON, C., 2003. *The rise and rise of testing in England: why, how and what next?* Slough: National Foundation for Educational Research.

WILIAM, D., 1994. *Towards a philosophy for educational assessment, Oxford, 1994.* Paper presented at the British Educational Research Association 20th Annual Conference, Oxford, 1994. Available at http://www.kcl.ac.uk/depsta/education/pub;ications/BERA_94.pdf [Accessed 11.1.06]

WILSON, R., 2002. *Raising standards in writing.* Huddersfield: Kirklees School Effectiveness Service, Kirklees Metropolitan Council.

WILSON, G., 2003. Don't forget the pupil perspective! The birthdate effect. *Consulting Pupils Project Newsletter.* Economic and Social Research Council, No.10, July 2003. Available at http://www.consultingpupils.co.uk/Resources/Newsletter%2010.pdf [Accessed 27.1.06]

WINTLE, M AND HARRISON, M., 1999. *Coordinating assessment practice across the primary school.* London and Philadelphia: The Falmer Press.

WOOD, R., 1990. The agenda for educational measurement. *In:* T. HORTON, ed. *Assessment debates.* London: The Open University, 1990, p. 51.

Abbreviation or acronym	*Full name or title*
AEN	Additional educational needs
APS	Average points scores
APU	Assessment of Performance Unit
ASD	Autistic Spectrum Disorder (PLASC)
BA	Baseline assessment
BERA	British Educational Research Association
BESD	Behavioural, emotional and social difficulties (PLASC)
CD	Creative development (FSP)
CLL	Communication, language and literacy (FSP)
CVA	Contextual value-added
DES	Department of Education and Science
DfEE	Department for Education and Employment
DfES	Department for Education and Skills
DoH	Department of Health
EAL	English as an additional language
ELGs	Early learning goals
EPPE	Effective Provision of Pre-School Education
ERA	Education Reform Act (1988)
FFT	Fischer Family Trust
FSM	Free school meals
FSP	Foundation stage profile
GCSE	General Certificate of Secondary Education
HI	Hearing impairment (PLASC)
HMCI	Her Majesty's Chief Inspector of Schools
HMI	Her Majesty's Inspectorate
ICT	Information and communication technology
IQ	Intelligence quotient
KAS	Kent Advisory Service
KCC	Kent County Council
KS1	Key stage one
KS2	Key stage two
KS3	Key stage three
KS4	Key stage four
KUW	Knowledge and understanding of the world (FSP)
LA	Local authority
LAC	Looked After Child
LEA	Local education authority
LGA	Local Government Association
LLG	Local Learning Group
MCAS	Minority Communities Achievement Service
MD	Mathematical development (FSP)
MLD	Moderate learning difficulty (PLASC)
NAA	National Assessment Agency
NAHT	National Association of Head Teachers
NC	National Curriculum

NfER	National Foundation for Educational Research
NLS	National literacy strategy
NNS	National numeracy strategy
NPD	National Pupil Database
NPQH	National Professional Qualification for Headteachers
NUT	National Union of Teachers
NVR	Non-verbal reasoning
Ofsted	Office for Standards in Education
PANDA	Performance and Assessment Data
PAT	Pupil Achievement Tracker
PD	Physical development (FSP); Physical disability (PLASC)
PLASC	Pupil Level Annual Schools' Census
PPA	Planning, preparation and assessment
PSE/PSED	Personal, social and emotional development (FSP)
PSHE	Personal, social and health education
QCA	Qualifications and Curriculum Authority
QNCA	Qualifications and National Curriculum Authority
QTS	Qualified teacher status
SA	School Action (SEN)
SA+	School Action Plus (SEN)
SATs	Standard assessment tasks and tests
SCAA	School Curriculum and Assessment Authority
SD	Standard deviation
SEAC	School Examinations and Assessment Council
SEF	Self-evaluation form
SEN	Special educational needs
SENCo	Special educational needs co-ordinator
SIP	School Improvement Partner
SLCN	Speech, language and communication needs (PLASC)
SLD	Severe learning difficulty (PLASC)
SPSS	Statistical Program for Social Sciences
SS	Standardised score
TA	Teacher assessment; teaching assistant
TGAT	National Curriculum Task Group on Assessment and Testing
UPN	Unique pupil number
VA	Value-added
VI	Visual impairment (PLASC)
VR	Verbal reasoning
W	Working towards level 1 (NC)

CHRONOLOGY

- 1974   APU*

- 1987   TGAT Report

- 1988   ERA; NC; LM; QTS; NPQH; SEN Code of Practice

- 1989   NC introduced in state primary and secondary schools

- 1990   KS1 SATs pilot

- 1991   KS1 SATs national (reported to LEA only)

- 1992   KS1 reported to parents, with comparative national data

- 1993   KS1 spelling and reading comprehension, mathematics and science
          KS2 SATs English, mathematics and science

- 1994   Dearing Report on assessment; NC slimmed down;
          KS1 science now only teacher assessment

- 1996   Desirable Learning Outcomes

- 1997   BA pilot; March - first league tables; target-setting; Numeracy Task Force;
          EPPE Project

- 1998   BA national; NLS; PANDA reports; 'P' levels for special schools;
          Black & Wiliam - *Inside the Black Box*; Shirley Clarke

- 1999   NNS; ELGs; *Beyond the Black Box*

- 2000   Revised NC - *Curriculum 2000*; Foundation Stage; *Pathways to Progress*

- 2001   Infant class size legislation; new SEN Code of Practice;
          league tables now show APS and attendance data

- 2002   *Working Inside the Black Box*; Ros Wilson

- 2003   FSP; *Excellence and Enjoyment* (Primary Strategy); league tables
          to show % at level 5 and first value-added data; Fischer Family Trust

- 2004   League tables available on internet; target setting to include attendance

- 2005   *Continuing the Learning Journey;* PAT and PANDA amalgamated

- 2006   Launch of RAISEonline

* A list of abbreviations used in the text will be found in Appendix I.

Survey of Foundation Stage practice

As you know, my current research is on benchmark assessment data. My three research questions are:

- How can benchmark data support pupils' learning?

- Can they be used to predict future attainment?

- What are the implications of using value-added data as measures of pupil performance?

Thank you for agreeing to help me in my research. I already have the quantitative data from your school, and now I need some qualitative data.

Please answer the following questions as fully as possible. Once I have collated the responses, I should like to discuss some of the issues with you.

----

1. What type of teacher assessments do you (and colleagues) make in class? Do they take the form of observations? written or oral tasks undertaken by the pupils? discussions with colleagues? other?

2. Do you group pupils in your class by ability, friendship, age, other, or not at all? *Please discuss.*

3. *(If appropriate)* How often do you change these groupings?

4. *(If appropriate)* Are your decisions based on teacher assessment?

5. Have you used [the former] Baseline Assessment scores to help you identify any pupil who may need additional help? *If so, please give details.*

6. Have you done the same with the present Foundation Stage Profile? *If so, please give details.*

7. Do you use any other sort of benchmark data to help you decide on planning appropriate learning experiences for your pupils?

8. Do you believe that the baseline or foundation stage profile scores can predict attainment the pupil will make by the end of KS1? *Please discuss.*

9. What is your opinion of the residual scores (value-added) that Kent produces from baseline to end KS1? Do they reflect the true picture? How can they be improved?

10. How do you think the foundation stage profile compares with the earlier baseline assessment?

Any further comments:

School name ....   Class teacher ....

Many thanks for your time.