University of Cambridge Local Examinations Syndicate

Assessment of Literacy at Key Stage Two
Text Level Strand of the National Literacy Framework

Sylvia Green and Clare Green

A paper presented at BERA, University of Cardiff, September 2000

Contact Details
Sylvia Green, Primary Assessment Unit, RED, UCLES, 1 Hills Road, Cambridge, CB1 2EU.
Green.s@ucles.org.uk

Autobiographical details
Sylvia Green is Head of the Primary Assessment Unit at the University of Cambridge Local Examinations Syndicate.
Clare Green is a Research Officer in the Primary Assessment Unit at the University of Cambridge Local Examinations Syndicate.

ABSTRACT
This paper will report on the second phase of the CamPAS English Project (Cambridge Primary Assessment Scheme). The aim of the project is to develop materials to support teacher assessment at Key Stage 2, focusing on the text level strand of the National Literacy Framework. Assessment units are being developed which are linked to criteria selected from the framework. Units were selected from the first trials and were trialled for a second time during autumn 1999. The paper will report on the results of the second trials and will evaluate the factors which impact on performance of Key Stage 2 pupils in reading and writing in the context of the National Literacy text level criteria. The analysis of the data will focus on the way pupils interact with different text types and question types and will consider these issues in relation to gender and levels of ability.

Introduction

The National Literacy Strategy was implemented in primary schools in September 1998, following a pilot in 16 LEAs during the 1997/98 academic year. The aim is to raise standards, so that, by the year 2002, 80% of 11 year olds will have reached Level 4 or above in the National Tests for English at Key Stage Two. The Text Level Strand is one of the three strands within the framework and is the strand which presents many challenges and changes to the work at Key Stage Two. It introduces teaching objectives in a range of text types, introducing a variety of non-fiction texts throughout the key stage. QCA has stated that a range of text types will be included in Key Stage Two tests therefore the way in which pupils access texts, and the factors which affect their performance, are important issues.

In 1997, The Exeter Extending Literacy (EXEL) Project found that:

> *'Teachers are unsure of exactly how to teach non-fiction reading and writing……teachers get a much reduced opportunity to work out for themselves strategies for supporting interactions with non-fiction texts or making judgements about criteria for and levels of success in these interactions (with children).'*
> *D Wray and M Lewis, Extending Literacy, p.17. Routledge, 1997.*

The QCA report on standards at Key Stage 2 for 1997 pointed out that:

> *The range of writing prompts included in tests from year to year will continue to vary, so it is important to encourage children to notice and use the characteristics of a variety of writing types…. Children need more experience of writing a variety of texts requiring forms and organisational strategies different from those required for story writing…. Children need more opportunity to explore a range of non-narrative texts and discuss their distinctive features and how they are organised.' p7.*

In a study of 'Reading Performance at Nine', Brooks et al (1996) found that progress in reading slowed between years 3 and 4 and they suggested that fundamental research was necessary on 'what works' in raising literacy standards, especially in low performers. They identified the need for the dissemination of effective strategies for keeping up momentum in Key Stage 2, with the early identification of those at risk. At this stage two years after implementation of the National Literacy Framework it is important to investigate the rate of learning assumed by its prescribed content as well as the effects on children according to their age, ability and gender.

The extension of the literacy curriculum has necessitated a broadening of the range of text types to incorporate a variety of non-fiction including: instructional, explanatory, recount, reports, exposition, persuasion and discussion. As Alison Littlefair (1991) commented 'there is a need for the development of flexible reading ability beyond the 'story genre'. In order for teachers to develop a valid programme of work it is important to gather information about the ability of children of different ages to access texts in terms of different text types and other factors. Littlefair's research suggests that fluent readers of narrative texts are not necessarily able to transfer their skills to other genres. Readers employ a range of strategies according to the nature of texts and the purposes of the reading and associated writing activities.

Neate (1993) commented that children lack flexibility in their reading and require different reading strategies. Wray and Lewis (1992) carried out research in UK Primary schools where they found that there was a greater emphasis on fiction rather than non-fiction texts with an emphasis on,

> *'a linear way of approaching books – in dealing with a book you begin at the beginning and work your way through it.'*
> Wray and Lewis, Reading Beyond the Story, p.57. 1992

This can be compared with models of reading for information effectively e.g. using contents and index pages to locate information. They also concluded that relatively little is known about how primary children access information texts and the way they use different strategies in changing circumstances.

Much attention is currently being paid to the differences in performance between boys and girls and to developing strategies to improve literacy levels, particularly for boys. The task of driving up standards has led to a number of initiatives to target boys' reading and writing skills and to motivate boys more successfully than may have previously been the case. Hall and Coles (1999) investigated reading choices of 8000 children aged 10, 12 and 14. Choices were gender specific with girls tending towards narrative texts and boys towards more factual reading. Boys were found to prefer a more structured kind of writing task whereas girls showed more interest in writing poetry. Boys tend to prefer learning facts, and as reported by Ofsted in 1996, their learning tends to improve when they value what they are doing in a purposeful context. Boys' reading involves a greater degree of technological information with numbers, instructions and factual reports. Such preferences may well reflect their exposure to different text types and the different ways they are socialised, based on expectations throughout the early years and primary education.

Methodology

In order to investigate the factors affecting pupil performance against selected criteria from the NLF (National Literacy Framework) assessment instruments were developed which were trialled with Key Stage 2 pupils. A model was designed which incorporated reading comprehension and writing composition activities. There was a stimulus text for each unit and the reading activities informed the writing activities which followed. By linking the activities within the trial instruments it was possible to explore the relationship between reading and writing abilities and also to question whether reading skills are transferred into the development of writing ability. The research was designed so that a Rasch Analysis could be carried out to determine the comparative difficulty of items within units, as well as units overall, and placed them on a common scale of difficulty. The Rasch model assigns a single number, the 'difficulty', to each item and a single number, the 'ability', to each pupil.

> *The Rasch Model is one of a number of item response models which are*
> *widely used nowadays for a variety of purposes including item banking.....the model links sets of items or tests through common examinees, or sets of examinees through common test items.*
> Williams and Ryan, 1999.

As reported by Pollitt et al.1998,

> *The particular advantage of using the Rasch model is that wherever there are common items (or common people), different tests can be brought onto the same scale – that is, the persons and the items across the tests can be measured in the same frame of reference.*

The research was designed so that each pupil completed two units and each unit was linked by common pupils. It was necessary to have common units between year groups. Each unit was written for a specific year group according to the criteria selected from the targets of the NLF for that year group. Units were completed not only by the year group for which they were written, but also by a group of pupils from the year group above. These 'overlap' groups provided the data necessary for the Rasch analysis. The Rasch model attempts to summarise test data with the assumption that the outcome of an interaction between a pupil and a question depends only on ability, where ability is defined in terms of the trait being investigated, e.g. reading, and the difficulty of the question.

Since many of these questions were marked out of 2 or more marks the 'partial credit' Rasch model (Wright and Masters ,1982) was used. In this model a three mark question actually has three difficulties, but this has no serious consequence for this analysis.

It is a probabilistic model, the more the pupil's ability is greater than the question's difficulty, the more likely they are to answer the question correctly. Difficulty and ability are measured using the same arbitrary units, called logits, and are related to the probability of success on a question. This is a unidimensional model with questions and people located on a single trait e.g. reading. The raw score determines the person's ability estimate and the item difficulty is estimated by its facility. Misfitting items can be identified and investigated and the success of the model can be judged according to the degree of misfit in the data. Relatively few misfitting items or pupils would suggest that the data conformed well to the model.

The trial 2 design is set out in Appendix 1, p.19. Pupils from years 4 to 7 completed the units in October, 1999, ensuring that the relevant teaching programmes had been completed. The inclusion of Year 7 middle schools pupils minimised the transfer effects associated with the change from the primary to the secondary phase. As well as Rasch analysis the data were analysed to investigate performance at item level and at unit level.

Framework of Assessment Criteria

The framework for the development of the instruments began with the selection of a range of text types. The aim was to select a range of text types across the key stage which reflected the targets prescribed within the NLF. This enabled investigation of the linear pattern of learning assumed by the framework, targeting specific skills within the prescribed timetable. By designing the trials to target the instruments at particular year groups in a cross-sectional study, it was possible to investigate the rate of learning and to explore effects in terms of ability and gender. The framework covers a wide range of text types and is cyclical, with certain criteria and text types re-visited during the Key Stage. The aim was to ensure that the instruments were based on texts which were representative of the framework overall, avoiding repetition and including fiction, poetry and non-fiction. The criteria were selected after the texts were chosen so that the tasks could be developed to ensure meaningful units of activity rather than discrete items. Key objectives were sampled from the NLF and these were grouped to provide the focus for assessment within each unit. Having identified the focus for each unit and selected the appropriate stimulus texts, the units were developed. Details of the 12 trial units are set out in Appendix 2, pp.20-21.

The range of question types included multiple choice, short answer, matching, completion and extended writing. Each question addressed a specific objective from the NLF. Mark schemes were written for each unit and, following trial 1, they were informed by pupil responses and feedback from markers.

Trials

The first trials took place in April 1999 and involved 14 schools across England and Wales. A range of school types and geographical areas were represented. The aim was to select a sample according to geographical area, number of pupils on roll and urban / rural location. Each unit was completed by 100 – 120 pupils. Overall 1250 pupils were involved with each child completing two units. The second trials took place in October 1999 and involved 8 schools and 720 pupils with each pupil completing two units. Teachers were asked to allow all pupils to attempt the units, apart from those for whom they were wholly inappropriate e.g. pupils with Special Needs Statements at Stage 4/5 of the SEN Register. Where help was given teachers annotated the pupil's work. Teachers were briefed by researchers about the administration of the trials and were asked to complete a feedback sheet. They commented on stimuli, layout, accessibility, reading level, management, timing and any other issues. Schools provided the following pupil information:
- teacher assessment levels for reading and writing
- stage on special educational needs register
- date of birth

Gender data were gathered from the pupils' activity sheets.

Marking was carried out by a team including researchers and experienced Key Stage 2 markers. Each marker was allocated a set of 60 scripts, 5 from each allocated unit, representing a wide range of performance. The teams marked from the mark schemes and then met to co-ordinate their marking and to discuss issues related to the mark schemes and the sample scripts. During the marking senior researchers carried out a sample check to monitor the reliability of marking.

A similar marking co-ordination procedure was carried out for trial 2. Scripts were allocated so that no one marker was responsible for an entire school or for an entire unit. This was to ensure that marker effects did not contaminate the data. Mark schemes were designed to capture as much data as possible. Where there were a number of correct response options available for a question, they were coded and entered into the database to enable a thorough analysis of pupils' choices.

Trial 2   Findings and Discussion

Figure 1   % Mean Scores
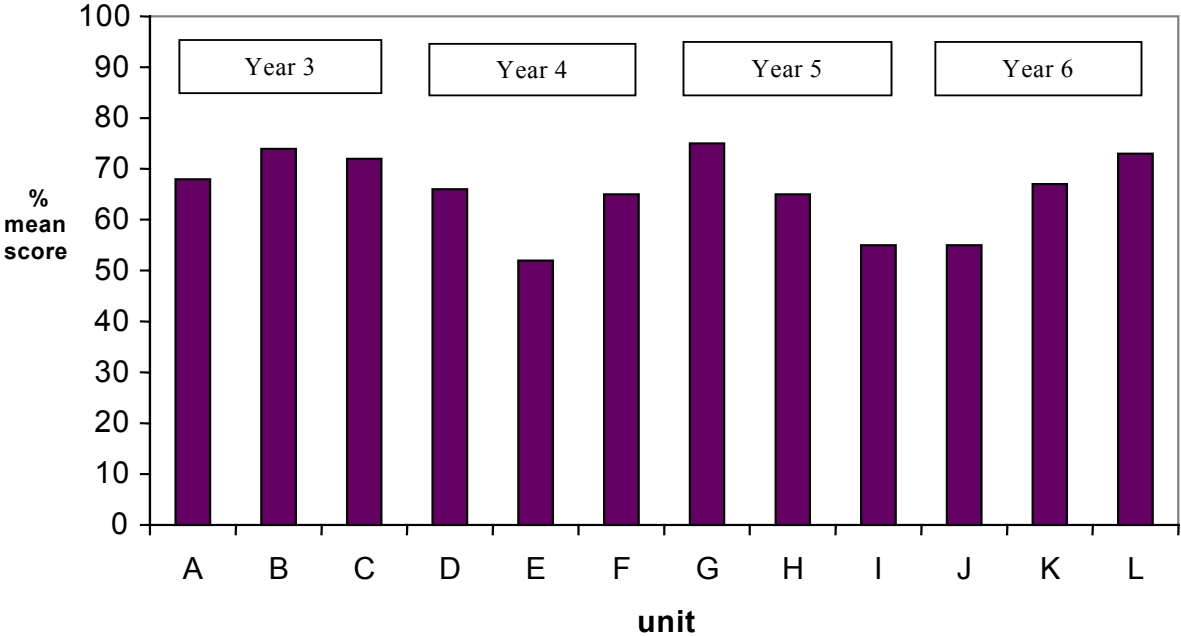
*For details of text types and assessment criteria for units A – L see Appendix 2, pp. 20-21

In Year 3 and Year 4 mean scores were higher in fiction units where the text type and assessment criteria were arguably more familiar. Narrative texts have traditionally formed a major part of teaching schemes for this age group. In Year 4 scores were lowest in unit E which focused on organising non-fiction text. It has been reported in QCA Standards Reports (1999) that this is an area where children are still having difficulty at the end of the Key Stage. Paragraphing is introduced in the National Curriculum at level 5 and it is not surprising that Year 4 pupils find this a difficult skill to master. Similarly the art of writing in a journalistic style proved problematic at Year 4 and it could be argued that this is also a high level skill to introduce at such an early stage. In the upper Key Stage, at Year 5, the highest mean score was again in the fiction unit (G). Although this was based on a play script, rather than a narrative text, some of the criteria focused on familiar areas such as character and personal response and therefore the familiarity with narrative text types may have been an advantage in this unit. The lowest mean score was in the unit which addressed persuasive techniques, bias, fact, opinion and clarity of expression (I). This evidence suggests that these areas are more challenging to both teachers and pupils and that it will take time to extend the range of text types beyond the more traditional fictional texts. In Year 6 the fiction unit (J) had the lowest mean score and this seemed to contradict the pattern in the rest of the Key Stage. The criteria in this unit focused on genre, narrative perspective, summary, voice and style. There was also an element of critical evaluation, with pupils required to comment on the success of the text in fulfilling the aims of the writer in terms of the impact on the reader. Both the nature of the text, which was classic fiction, and the difficulty of the criteria clearly created problems for Year 6 pupils. Across the Key Stage it could be argued that the extension into less familiar text types and the introduction of higher order skills are challenging pupils at this early stage in the implementation of the National Literacy Strategy.

Figure 2   % Mean Scores Years 3 and 4   Reading and Writing
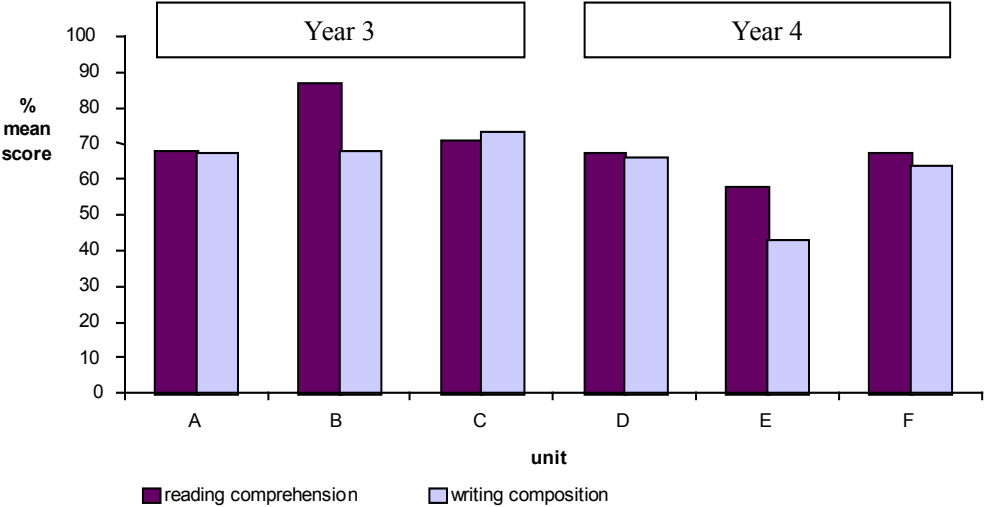
Figure 3   % Mean Scores Years 5 and 6   Reading and Writing
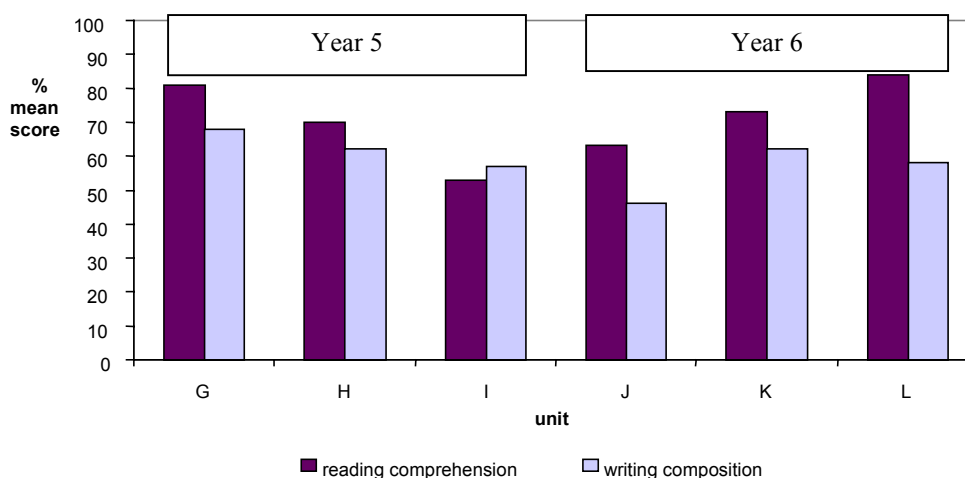
Reading and Writing

Overall in the lower key stage scores were higher for reading than for writing except for the Year 3 instructional text unit (C). This supports the view that reading skills develop in advance of writing skills. In the instructional text unit the reading criteria focused on recognising and understanding organisational features. This may have proved more challenging given the tendency to concentrate on narrative text forms at this age. The writing activity involved writing instructions modelled on the reading stimulus. The scaffolding provided by the model text may well explain why writing scores were higher in this unit.

 In Year 3 the greatest difference between reading and writing scores was in the narrative text unit (B) where the focus was on personal response, theme and character. Familiar question types which ask for descriptions, feelings and opinions about stories may have helped to increase reading scores.

In Year 4 the greatest difference between reading and writing was in the journalistic text unit (E) where the activity involved writing in the style of a reporter, paragraphing and editing writing to fit a given space.
As has been reported in QCA Standards Reports 1999, organising writing in paragraphs is still a problematic area at the end of key stage 2 and the additional challenge of writing in a particular style may well have led to the differential performance in the reading and writing activities.

Writing scores were higher than reading scores in the persuasive text (I) unit in Year 5, where the writing activity involved writing a persuasive letter. Higher writing scores may have resulted from the fact that letter writing is a familiar activity, the information for the content of the letter was provided in the stimulus text and a planning framework was provided to scaffold the activity.

The greatest difference between reading and writing scores in Year 5 was in the play script text unit (G) where the reading involved personal response to characters and situations and the writing involved writing a play script in the correct form using scripting conventions demonstrated in the reading stimulus. During the primary phase play scripts are often read and many are included in graded reading schemes. The familiarity with the form in terms of reading may have led to higher reading scores together with the fact that the types of skills required are similar to those needed when reading a narrative text however, writing play scripts may well have been a less familiar activity.

In Year 6 reading scores were higher than writing scores in all units with the greatest difference in the poetry unit and the least difference in the discursive text unit (K). The writing activity in the poetry unit involved critical evaluation and comparison of poems which require higher order skills.

Figure 4   % Total Mean Scores by Gender
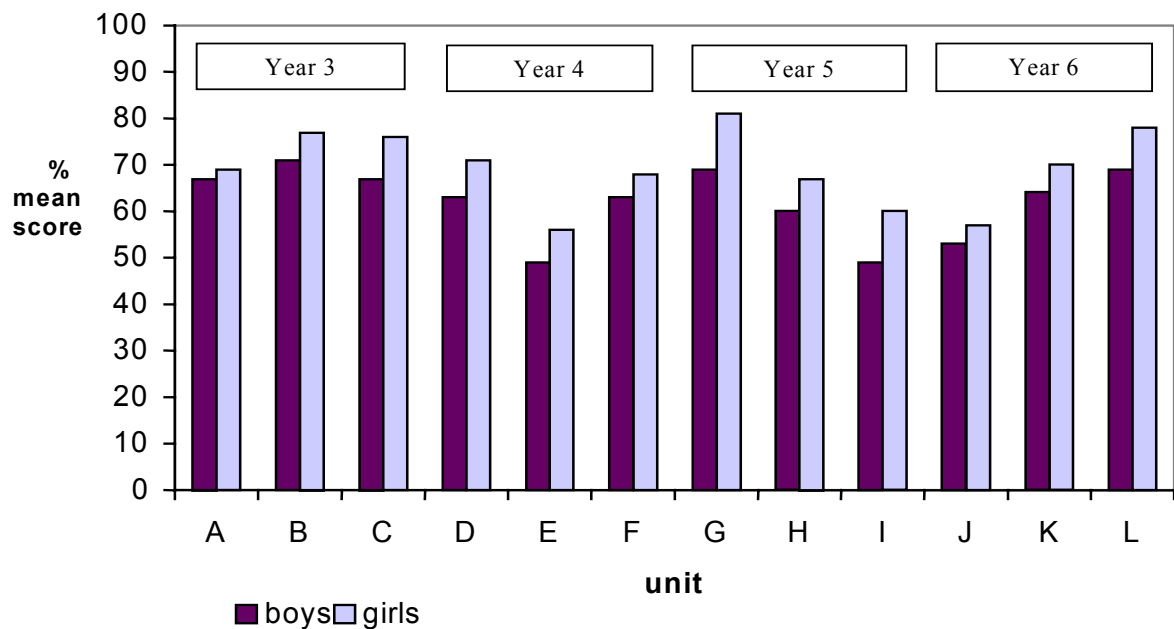
Gender differences were significant at the 5% level for units D and K, and at the 1% level for units G, I and L. Across the key stage girls outperformed boys in all units. In Year 4 differences were statistically significant at the 5% level in the narrative unit (D). In Year 5 the differences were greatest in the play script text unit (G) and the persuasive text unit (I), at the 1% level. Differences in Year 6

were statistically significant at the 1% level in the poetry unit (L) and at the 5% level in the discursive text unit (K).

Figure 5   % Mean Scores for Reading Comprehension by Gender
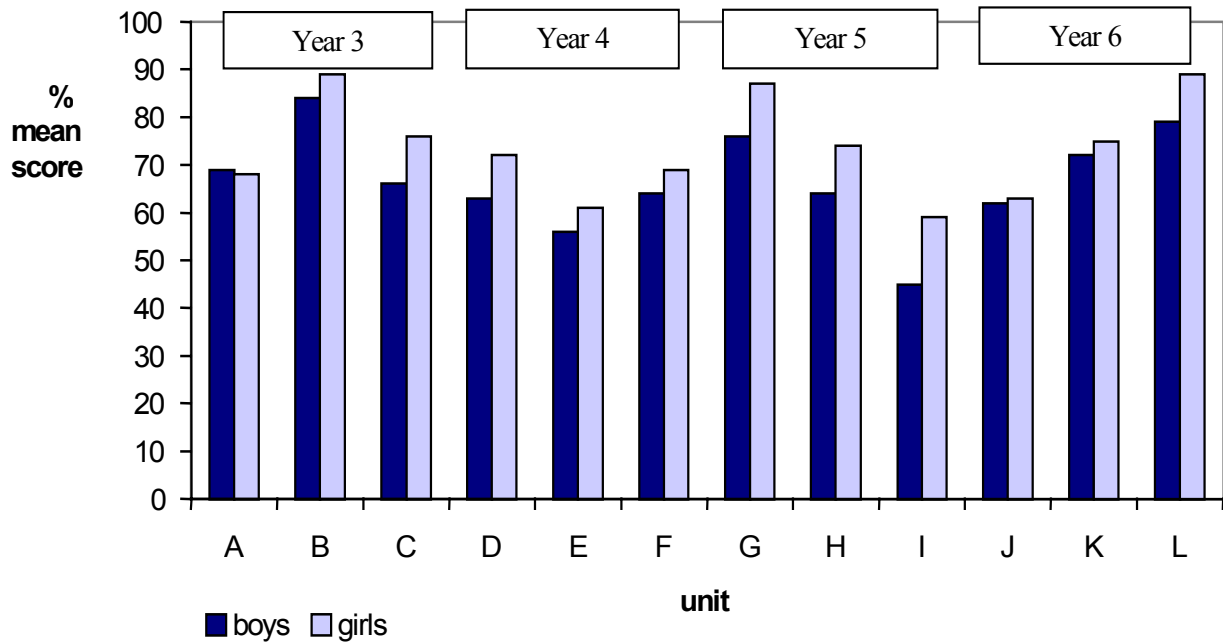
Mean scores on reading split by gender

In Year 3 boys outperformed girls in reading only in the information text unit (A), which focused on textual organisation and retrieval of information. Year 3 differences were only significant at the 5% level in the instructional text unit (C). In Years 4, 5 and 6 girls outperformed boys in all units. In the Year 5 play script unit (G) and the persuasive text unit (I) the differences were significant at the 1% level. The greatest difference in Year 6 was in the poetry unit which focused on theme, impact of language contrasts and connections, statistically significant at the 1% level. The fact that the only unit where boys outperformed girls in reading was the information text unit (A) may support the view that boys are more able when dealing with non-fiction texts especially reference materials and this reflects their choice of reading material. Boys' scores may also have increased as a result of the question types and response styles including single word and short answers related to the retrieval of facts from the text. Girls' higher scores in the poetry unit (L) support the view that girls perform better in questions where evaluation and personal response are required and may also be affected by girls' reading preferences.

Figure 6   % Mean scores for Writing Composition by Gender
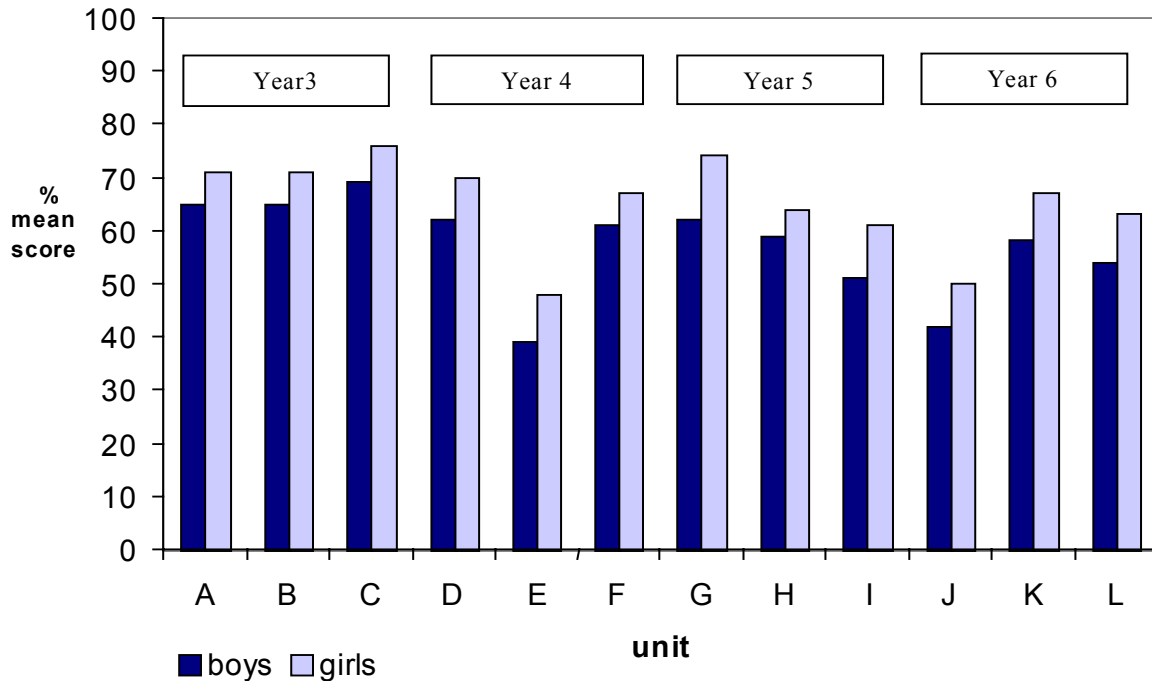
Mean scores on writing split by gender

Girls outperformed boys in writing in all units across the key stage with differences statistically significant at the 1% level in the play script unit (G) and at the 5% level in the persuasive text unit (I), discursive text unit (K) and poetry unit (L). The play script unit (G) which focused on annotating and writing a play script in the correct form using conventions from reading, proved problematic as pupils included appropriate content in an incorrect form.
(Trial 2 mean scores can be found in Appendix 3, pp.22 and 23)

Gender Differences

-   not statistically significant
*   statistically significant at 5% level
** statistically significant at 1% level

	Text Type	Unit	Reading	Writing
Y3	A Informative	-	-	-
	B Traditional Narrative	-	-	-
	C Instructional	-	*	-
Y4	D Historical Narrative	*	-	-
	E Journalistic	-	-	-
	F Explanatory	-	-	-
Y5	G Play Script	**	**	**
	H Recount	-	-	-
	I Persuasive	**	**	*
Y6	J Classic Fiction	-	-	-
	K Discursive	*	-	*
	L Poetry	**	**	*

Girls outperformed boys apart from in the reading comprehension in the informative text unit (A). The statistical significance of gender differences increases in the upper key stage as shown in the table above.

Rasch Analysis

The Rasch analysis enabled us to place 12 units from trial 2 on a common scale of difficulty (see Figure 7 below). The difficulty is measured using an arbitrary unit called a logit. Items with a negative measure are easier and items with a positive score are more difficult. The relative difficulty of units in the lower Key Stage is as might have been expected, with Year 3 units (A, B and C) lower on the scale than the Year 4 units (D,E and F). The journalistic text unit (E) was more demanding than the other Year 4 units. The pattern in the upper Key Stage is more complex. The most difficult unit overall (I) was a Year 5 unit based on a persuasive text. Unit H, another Year 5 unit, was more difficult than unit K which was a Year 6 unit based on a discursive text. Unit H was also more difficult than unit L, the Year 6 poetry unit. The Year 5 and 6 units were not split in terms of difficulty in the way that might have been expected according to the structure of the National Literacy Framework teaching objectives.

Figure 7 Trial 2 Units on common scale of difficulty

```
|Measr|+unit           |
logits
+  2 +                 +Text Type          Focus
|    |
|    | unit I          persuasive          fact, opinion, bias, persuasive technique
|    | unit J          classic fiction     personal response, viewpoint,
|    |                                     summary, voice, style
|    |
|    |
|    |
|    | unit H          recount             sequence, formality, connectives,
|    | unit K          discursive          note making, reporting
+  1 +
|    | unit G          play script         character, personal response,
|    |                                     dramatic conventions, scripting
|    | unit E          journalistic        key information, style, editing, paragraphs
|    | unit L          poetry              theme, impact, language, reviewing
|    |
|    |
|    |
|    |
|    |
*  0 *
|    |
|    |
|    |
|    |
|    |
|    |
|    |
|    | unit D          historical fiction  settings, character, narrative order
|    |                                     narrative sequence, paragraphing
+ -1 + unit F          explanatory         key words and phrases, purpose
|    |                                     connectives, write explanation
|    |
|    |
|    |
|    |
|    |
|    |
|    |
|    | unit A   unit C  A information       headings, sub-headings, key words,
|    |                                     phrases, index, page numbers, chart,
|    |                                     non-chronological report
+ -2 +                 C instructional     purpose, organisational devices,
|    |                                     sequencing
|    |
|    | unit B          traditional fiction theme, character, letter, story plan
|    |
|    |
|    |
|    |
|    |
+ -3 +
```

14

Pupil Misfit

Figure 8 below shows the results of the Rasch analysis which highlights unexpected behaviour by pupils outside the range of their ability level. Pupils with a fit statistic above +2 or below –2 are considered to be misfitting. The number of misfitting pupils is relatively small, 23 out of 720, and most of these are of high ability.

Figure 8

```
            -4      -3      -2      -1       0       1       2       3       4       5
           ++------+------+------+------+------+------+------+------+------+------++
        6 +                                 |                                      +
          ¦                                 |                                      ¦
          ¦                                 |                                      ¦
        5 +                                 |                                      +
          ¦                                 |                                      ¦
          ¦                                 |                                      ¦
        4 +                                 |                    1                 +
          ¦                                 1           1                          ¦
          ¦                                 |                                      ¦
  P     3 +                      1          | 1          1          1              +
  E       ¦                  1          |          1    1                          ¦
  R       ¦              1              1|1    1   1  11    111       1             ¦
  S     2 +--------------------1--12---|-11------11--1---1----------------+
  O       ¦                1         1 1211|   14    114    1 11   1                ¦
  N       ¦            1  1     2 1 11 |1 3313 13111 1 1 1      1                   ¦
        1 +              112221 121221 21212 41112222  111 2 1                  1 +
  O       ¦        2       1      213 12 11 214325333733426 2123     1 1   1        ¦
  U       ¦    1   1    1       1 12  1152 4732521352535411     1                   ¦
  T     0 +---------1------1--21-1-131252646444X2826493216522112-11--1-----+
  F       ¦                 121       4117423337X539XX35X75 31311 4                 ¦
  I       ¦            11      1 1   12 21435535834657521511 1                      ¦
  T    -1 +     11          1 1    52| 463332524232    1                           +
          ¦             2      11 2 222| 43333313 212                              ¦
          ¦               1 2    1 1|  1                                           ¦
       -2 +--------------------1------1|---1-2---------------------------+
          ¦                     1      | 1                                         ¦
          ¦                            |     1                                     ¦
       -3 +                            |                                           +
          ¦                            |                                           ¦
          ¦                            |                                           ¦
       -4 +                            |                                           +
           ++------+------+------+------+------+------+------+------+------+------++
            -4      -3      -2      -1       0       1       2       3       4       5
                                  PERSON MEASURE
                                      1 1
  ITEMS   1     11 1123555 7 73475758646167976868964 3612111      1
                   Q         S         M         S         Q
```

Item Misfit

Figure 9 below shows the results of the Rasch analysis in terms of item misfit highlighting unexpected behaviour affecting responses to questions near the pupils' ability level. An item with a fit statistic of above +2 or below –2 is considered to misfit. There are 9 such items out of a total of 207. These results together with those shown in Figure 8 suggest that the data conformed very well to the Rasch model, thus supporting the validity of the results. The misfitting items were identified and analysed in further detail.

Figure 9

```
           -4      -3      -2      -1       0       1       2       3       4       5
          ++------+------+------+------+------+------+------+------+------+------++
      6 +                                     |                                   +
        ¦                                     |                                   ¦
        ¦                                     |                                   ¦
      5 +                                     |                                   +
        ¦                                     |                                   ¦
        ¦                                     |                                   ¦
      4 +                                     |                                   +
        ¦                                     |                                   ¦
        ¦                     1           1   | 1                                 ¦
      3 +                        1            | 1                                 +
   I    ¦                                  1  |                                   ¦
   T    ¦                                  1  |                                   ¦
   E  2 +---------------------11----1----1------1--------------------+
   M    ¦                          111 2 1 |1131    1                             ¦
        ¦              1                 32 | 1                                   ¦
      1 +                      11   1    11  |1                                   +
   I    ¦                1         1     11 11 1    1111                          ¦
   N    ¦        1  11 13    2   2 2 2    32211 |211                1             ¦
   F  0 +1----1----11211-4-1121-1312--1-12--2---1----1--1---------------+
   I    ¦            1      12    21 1111 111   2122|22   211  1                  ¦
   T    ¦              1 1 1    1 1       1 2 3 222   31  2     1                 ¦
     -1 +              11   2        22 32 11     |  1   1                        +
        ¦                                2  11  1 1 1                             ¦
        ¦                    1   1  1 112     2| 1   1 1                          ¦
     -2 +------------------2---------1----|---1---------------------+
        ¦                                     |   1                               ¦
        ¦                                     |                                   ¦
     -3 +                                     1                                   +
        ¦                                     |                                   ¦
        ¦                                     |                                   ¦
     -4 +                                     |                                   +
          ++------+------+------+------+------+------+------+------+------+------++
           -4      -3      -2      -1       0       1       2       3       4       5
                                     ITEM   MEASURE
                       1    11121 333332334332421 11
PERSON      1 311 1 11115435728780474784652878701381148636551 9  21 11    16
                     Q       S        M         S        Q
```
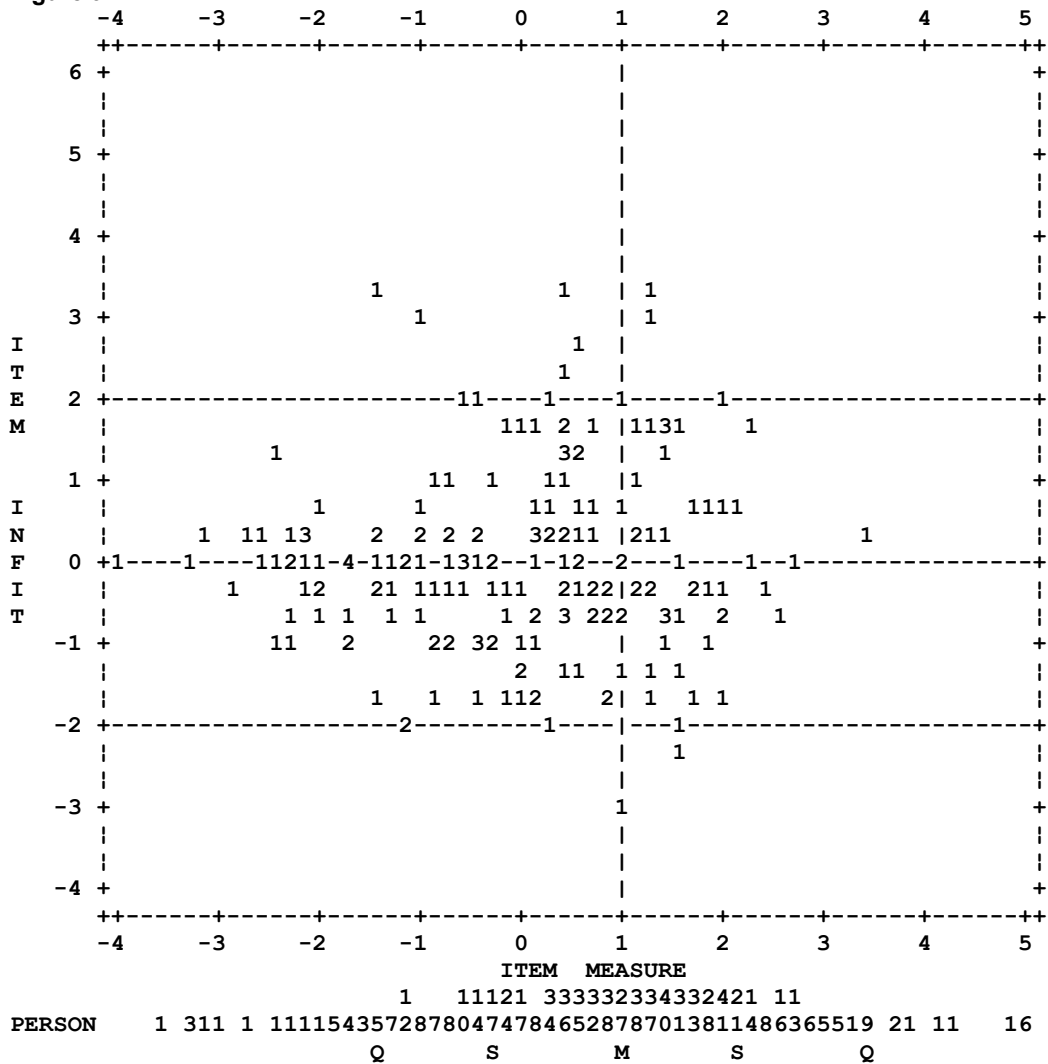
Distribution of pupils and items in Trial 2

Figure 10 shows the distribution of pupils and items with the most able persons and most difficult items at the top. The column on the left represents pupils. Each question appears twice, once in the middle column which shows the difficulty of getting one mark, and again in the column on the right which shows the difficulty of getting full marks. Sometimes these will be the same. The results indicate that the difficulty of the items matches the ability of the pupils, to a greater extent when multiple marks are taken into account, as shown in the column to the right. As the trials were completed in October, early in the academic year, each unit was completed by pupils in the year group above the year for which they were written, to ensure coverage of the teaching objectives being assessed. The materials were designed to assess mastery in the assessed areas and not as a means of grading. These two factors may explain why a small minority of the items appear to be too easy for pupils in trial 2. This explanation is supported by the results from trial 1 (see Figure 11, p.15), where the units were completed by the year group for whom they were written. The difficulty of the items matches more closely the ability of the pupils.
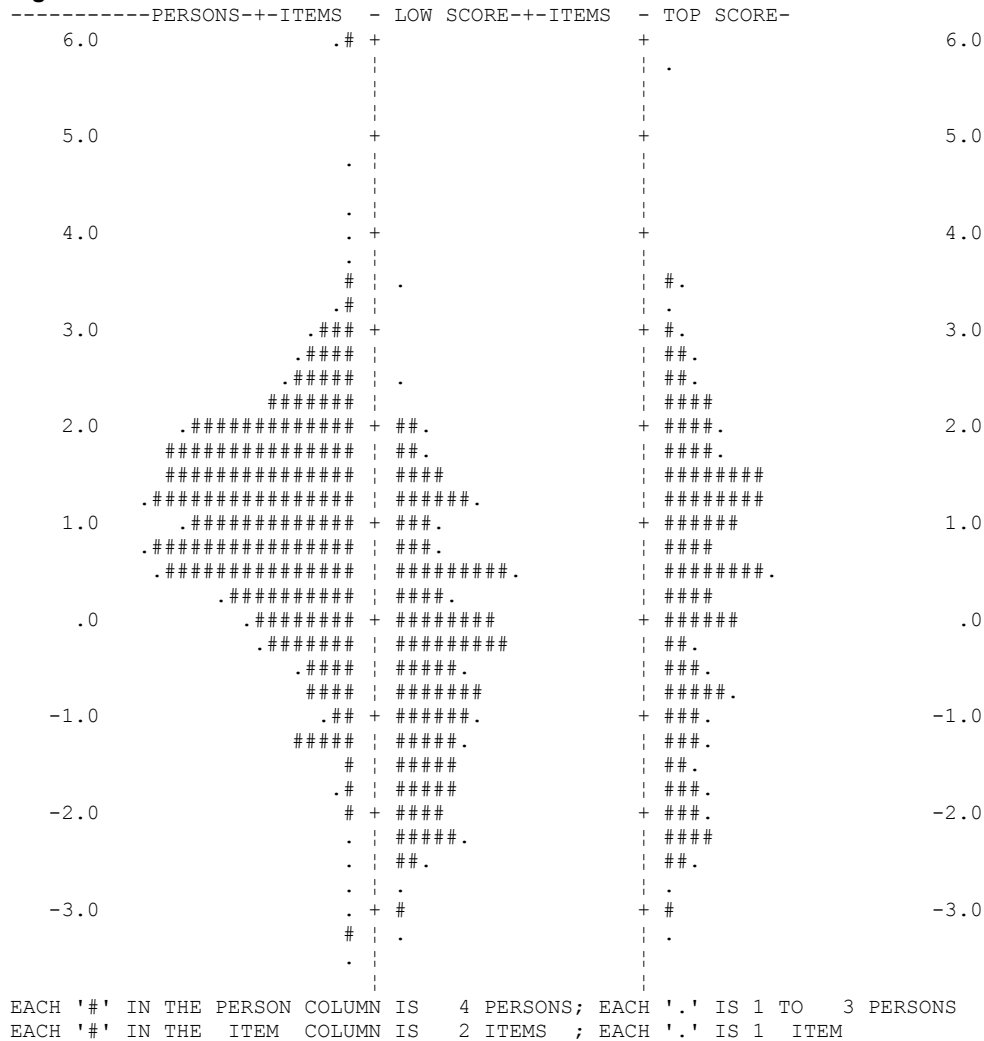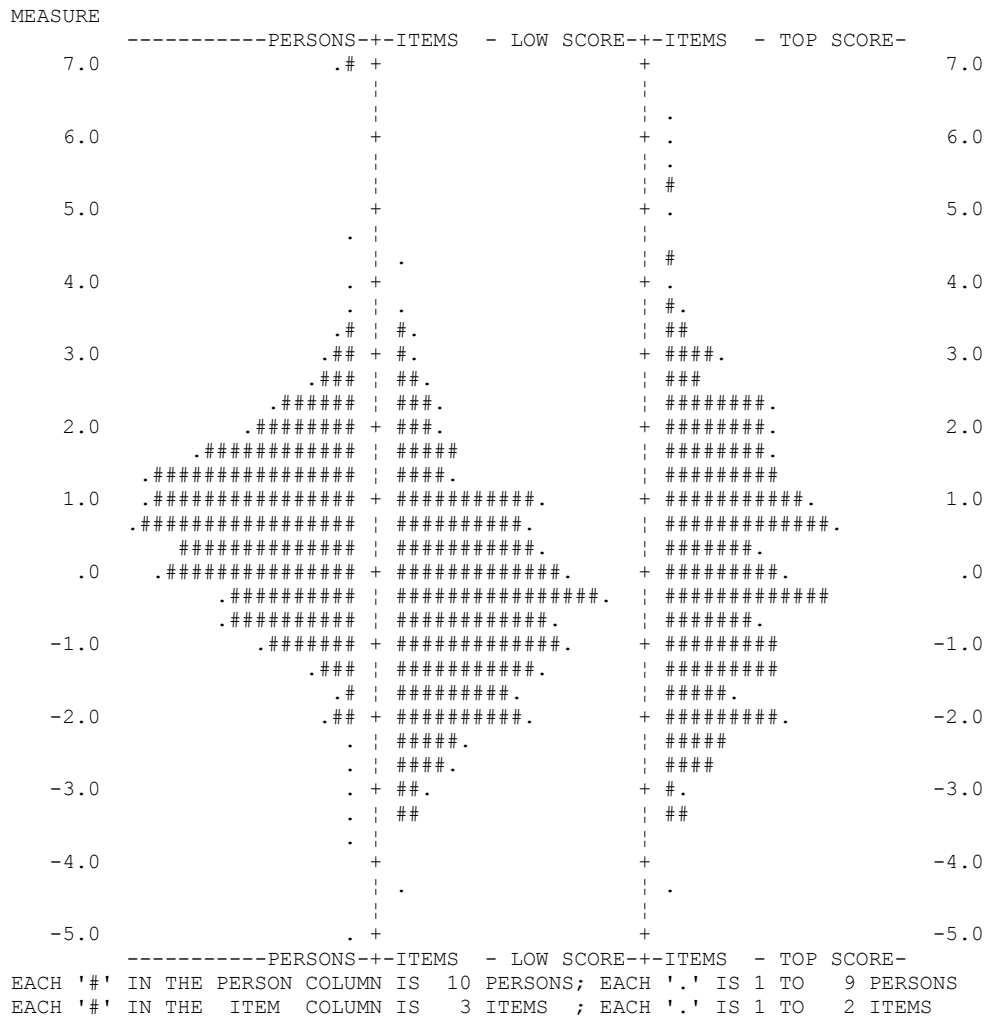
Figure 10

```
----------PERSONS-+-ITEMS  - LOW SCORE-+-ITEMS  - TOP SCORE-
   6.0                .# +              +                          6.0
                          ¦              ¦  .
                          ¦              ¦
                          ¦              ¦
                          ¦              ¦
   5.0                    +              +                          5.0
                      .  ¦              ¦
                          ¦              ¦
                      .  ¦              ¦
   4.0                .  +              +                          4.0
                      .  ¦              ¦
                      #  ¦  .           ¦  #.
                     .#  ¦              ¦  .
   3.0             .### +              +  #.                      3.0
                  .#### ¦              ¦  ##.
                 .##### ¦  .           ¦  ##.
                 ####### ¦              ¦  ####
   2.0       .############# + ##.          + ####.                  2.0
             ############### ¦ ##.          ¦ ####.
             ############### ¦ ####         ¦ ########
            .############### ¦ ######.      ¦ ########
   1.0       .############# + ###.         + ######                 1.0
            .############### ¦ ###.         ¦ ####
            .############### ¦ #########.    ¦ ########.
              .########## ¦ ####.        ¦ ####
    .0          .######## + ########    + ######                   .0
                .####### ¦ #########    ¦ ##.
               .#### ¦ #####.      ¦ ###.
               #### ¦ #######     ¦ #####.
  -1.0           .## + ######.      + ###.                        -1.0
               ##### ¦ #####.      ¦ ###.
                  # ¦ #####        ¦ ##.
                 .# ¦ #####        ¦ ###.
  -2.0             # + ####         + ###.                         -2.0
                   . ¦ #####.       ¦ ####
                   . ¦ ##.          ¦ ##.
                   . ¦ .            ¦ .
  -3.0             . + #            + #                            -3.0
                   # ¦ .            ¦ .
                   . ¦              ¦
                      ¦              ¦
EACH '#' IN THE PERSON COLUMN IS   4 PERSONS; EACH '.' IS 1 TO   3 PERSONS
EACH '#' IN THE  ITEM  COLUMN IS   2 ITEMS  ; EACH '.' IS 1  ITEM
```

Figure 11   **Distribution of pupils and items in Trial 1**

```
MEASURE
        -----------PERSONS-+-ITEMS  - LOW SCORE-+-ITEMS  - TOP SCORE-
   7.0                  .# +                    +                      7.0
                           ¦                    ¦
                           ¦                    ¦ .
   6.0                     +                    + .                    6.0
                           ¦                    ¦ .
                           ¦                    ¦ #
   5.0                     +                    + .                    5.0
                         . ¦                    ¦
                           ¦ .                  ¦ #
   4.0                   . +                    + .                    4.0
                         . ¦ .                  ¦ #.
                        .# ¦ #.                 ¦ ##
   3.0                 .## + #.                 + ####.                3.0
                      .### ¦ ##.                ¦ ###
                    .##### ¦ ###.               ¦ ########.
   2.0            .####### + ###.               + ########.            2.0
               .########### ¦ #####             ¦ ########.
            .############### ¦ ####.            ¦ #########
   1.0     .############### + ###########.     + ###########.          1.0
          .############### ¦ ##########.       ¦ #############.
           ############### ¦ ###########.       ¦ #######.
    .0     .############## + #############.     + #########.           .0
           .########## ¦ ###############.      ¦ #############
           .########## ¦ ############.          ¦ #######.
  -1.0       .####### + #############.          + #########            -1.0
             .### ¦ ###########.               ¦ #########
              .# ¦ #########.                  ¦ #####.
  -2.0        .## + ##########.                 + #########.            -2.0
              . ¦ #####.                       ¦ #####
              . ¦ ####.                        ¦ ####
  -3.0         . + ##.                          + #.                    -3.0
              . ¦ ##                           ¦ ##
              . ¦                              ¦
  -4.0            +                             +                       -4.0
                ¦ .                            ¦ .
                ¦                              ¦
  -5.0         . +                             +                        -5.0
        -----------PERSONS-+-ITEMS  - LOW SCORE-+-ITEMS  - TOP SCORE-
EACH '#' IN THE PERSON COLUMN IS  10 PERSONS; EACH '.' IS 1 TO   9 PERSONS
EACH '#' IN THE  ITEM  COLUMN IS   3 ITEMS  ; EACH '.' IS 1 TO   2 ITEMS
```

Issues arising from item level analysis

Items were investigated for difficulty, bias and misfit.

| Difficulty | facility value for the item in the context of the unit |
| | difficulty value from Rasch analysis in the context of the 12 units overall |

Bias   gender
                national curriculum teacher assessment level

Misfit   unexpected behaviour affecting responses to questions near the pupils' ability level

The analyses highlighted areas for investigation and trends in performance on certain items. The main issues relate to:

- organisational features of fiction and non-fiction texts
- task scaffolding
- National Curriculum and National Literacy Framework Expectations

Organisational Features of Text

The identification of organisational features in non-fiction texts, including bullet points, headings, paragraphs, captions, was demanding in the lower Key Stage units. This area proved more challenging in the context of reading than in the use of such conventions in the related writing activities. Littlefair (1991) commented that there is a need to develop reading skills beyond the story genre and Neate (1993) found that children lack flexibility in their reading. These findings support the views of the QCA (1997) which reported that children need more experience in writing a variety of texts requiring forms and organisational strategies different from those required for story writing.

Task Scaffolding

In items addressing organisation and presentation of text the response type affected performance where, for instance, one device was given as an example and pupils were required to complete a sentence giving another example. This may have provided a cue for the correct type of response. Where a more open-ended question was asked the facility was lower. The support provided by planning tasks, writing frames and model texts led to higher facility values. Where support was provided in the form of notes of content for writing tasks, pupils were more successful in using organisational devices in presenting information effectively. This may have been because they were able to concentrate their efforts on the organisation rather than on the content of the writing, therefore focusing on the criteria which were being assessed. Further research on the impact of writing task support is needed to investigate these issues further.

National Curriculum and National Literacy Framework Expectations

Discrepancies between the expectations of the National Curriculum and the National Literacy Framework may partially explain why some items proved to be so challenging during the trials. Items which assessed knowledge of organisational features of texts proved challenging and raised the issue of contradictory expectations in the National Curriculum and the National Literacy Framework. One example of this was in the Year 4 explanatory text unit where pupils were required to identify key features of the text and then to use the conventions identified through reading in their own writing. The relevant criteria do not appear in the National Curriculum Level Descriptions for reading until Level 5, whereas in writing they are introduced, in general terms, at Level 3. The data confirmed that Year 4 pupils found the criteria challenging in reading and less so in writing. In his report on the National Curriculum, Dearing (1993), quoted OFSTED in saying that the assessment requirements of the National Curriculum play a vital role in raising expectations. If this is the case it is important that there should be clarity and consistency in content at each stage of the taught curriculum. If there are contradictory messages about the expectations of pupils and teachers then the validity of the assessments could be questioned.

Conclusions

The empirical data suggest that, especially in the lower key stage, the introduction of texts outside the story genre created challenging contexts for assessment. The tradition of focusing on narrative texts may account for the better performance on the fiction units, while genres which may still be relatively new to the key stage, such as persuasive and journalistic texts, gave rise to lower scores and this supports the view of the QCA (1998) that 'children need to be asked questions which help them to develop skills when reading a range of texts'.
Overall writing scores were lower than those for reading with some exceptions. A range of factors may have led to the exceptions to the trend and these include issues of content, task support, familiarity, item type and response style.
As may have been expected girls outperformed boys in all units with the gap between boys and girls increasing in the upper key stage. It could be argued that the recent emphasis on boys' reading performance may be at the expense of their writing skills and that their writing is equally deserving of extra attention. This could be balanced by the fact that any improvement in reading skills will lead to increased writing competence. Poulson (1993) commented on the relationship between the domains of reading and writing and discussed the ability to transfer reading skills into the development of writing in different genres.

Overall the criteria of the National Literacy Framework were well-matched to the ability of the pupils in the trials (see Figures 10 and 11, pp.14 - 15) and the data conformed well to the Rasch model adding validity to the findings (see Figures 8 and 9, pp.12-13). For some criteria there were differences between performance in reading and writing, possibly because of mismatching between the National Curriculum and the Framework. The empirical data support research evidence that there is a greater emphasis on fiction than non-fiction in UK primary schools and that there is a need to develop skills in reading different text types for different purposes. Although the criteria matched the pupils' ability overall, there were some areas where there was a lack of consistency between the Framework and the National Curriculum expectations.

Further Research: Effective Scaffolding in Reading and Writing Assessment at Key Stage 2

The item analysis provided evidence which suggested that the scaffolding of items assessing certain criteria significantly affected pupil performance. One example of this was in the use of paragraphing to organise and sequence text which was difficult for Year 4 pupils for whom it is a teaching objective in the National Literacy Framework. The aim of further study will be to explore the rich source of data to investigate the extent to which scaffolding of tasks facilitates pupil performance in a range of contexts at Key Stage 2. Criteria highlighted in the data will be selected and parallel tasks will be designed, with different degrees and types of scaffolding. Controlled experiments will be carried out to investigate the impact of scaffolding on the performance of pupils of differing abilities and ages and in a variety of contexts.

Comments

The implementation of the National Literacy Strategy has led to greater emphasis on the development of text level skills in the primary phase. It has posed challenges for teachers in terms of curriculum content, resources and pedagogy in extending the range of text types addressed during Key Stage 2. It will take time for such a major initiative to be effectively implemented and, in the interim, the issues raised by its introduction will need to be investigated. The assumed progression which underpins the National Literacy Framework is prescribed within its content and in the time scale set out for teaching . Studies which gather empirical data about pupil performance in the context of the National Curriculum may illustrate potential problems and help to explain pupil performance which does not fit the 'expected template' of progression presented by the Framework's teaching objectives. The content of the National Literacy Framework will inevitably impact on the development of national test materials and it could be argued that this was the case in the 1999 Key Stage 2 English tests. Other assessment issues relate to the way in which tasks are designed and presented to pupils, since it is important to facilitate, rather than impede, pupil performance. Factors such as item type, response style and task scaffolding need to be taken into account to ensure valid and reliable assessment. As Desforges (1992) suggested, "the validity of assessment relates to the extent to which it promotes quality learning based on the taught curriculum in the classroom context."

References

Brooks, G. (1996)   *Reading Performance at Nine, NFER.*

Dearing, R. (1993)   *Report on the National Curriculum,* 'Theoretical Issues, Professional Solutions, ed. Beech and Singleton in 'The Psychological Assessment of Reading', Routledge.

Desforges, C. (1992)   'Assessment and Learning', *Forum, Volume 34, No.4.*

DFEE (1998)   *National Literacy Strategy: Framework For Teaching,* London.

Hall, C.
& Coles, M. (1999)   'The Reading Choices Project', *NFER Research, Spring Issue 21.*

Littlefair, A. (1991)   *Reading All Types Of Writing,* Open University Press.

Neate, B. (1993)   *The Diversity of Registers Found in Primary Children's Information Books,* in 'Reading Beyond the Story', UKRA.

Pollitt, A. et al (1998) *The Common Reading Test – Final Report,* QCA.

Poulson, L. (1992)   'Teacher Assessment at Key Stage 2', *Reading, Volume 26, No.3, Nov. pp.6-12.*

QCA (1997)   *Standards at Key Stage 2,* Great Britain: QCA Publications.

Williams, J.
& Ryan, J. (1999)   Rasch Modelling in Test Development, Evaluation, Equating and Item Banking, *paper presented at BERA Annual Conference, University of Sussex, 1999.*

Wray, D. (1992)   *Reading Beyond The Story,* UKRA.

Wray, D.
& Lewis, M. (1992)   'Primary Children and Information Books', *Reading Beyond The Story, UKRA.*

Wray, D.
& Lewis, M. (1997)   *Extending Literacy, Children Reading and Writing Non-Fiction,* London, Routledge.

Wright, B.D.
& Masters, G.N. (1982) Rating Scale Analysis, *MESA Press, Chicago.*