The 2009 National Assessments Technical Report

Eemer Eivers, Aidan Clerkin, David Millar and Seán Close

Educational Research Centre
Foras Taighde ar Oideachas

The 2009 National Assessments
Technical Report

Eemer Eivers, Aidan Clerkin, David Millar and Seán Close

Educational Research Centre

Cataloguing-in-Publication Data

Eivers, Eemer

The 2009 National Assessments Technical Report / Eemer Eivers …[et al.].

Dublin: Educational Research Centre.

iv, 82p; 30cm

Includes bibliographical references.

ISBN: 978 0 900440 27 4

1. Reading (Primary) – Ireland – Evaluation

2. Mathematics (Primary) – Ireland – Evaluation

3. Academic achievement

4. Educational surveys – Ireland

2010

I Title. II Clerkin, Aidan. III Millar, David.

371.262

Cover design by **SilverBark Creative**

Printed in the Republic of Ireland by **Eprint**

Contents

List of Acronyms and Abbreviations

DES	Department of Education and Science / Department of Education and Skills
DEIS	Delivering Equality of Opportunity in Schools
ERC	Educational Research Centre
IRT	Item Response Theory
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PSEC	Primary School English Curriculum
PSMC	Primary School Mathematics Curriculum
SSP	School Support Programme

Chapter 1
Introduction

This chapter provides an introduction to the Technical Report for the National Assessments of English Reading and Mathematics 2009 (NA 2009). The Technical Report is intended for an academic or technical audience, and does not present the results of NA 2009. Readers who are interested in the findings of NA 2009 are referred to Eivers et al. (2010). This chapter has two main sections. The first summarises the purpose and design of NA 2009, and the second outlines the management and organisation of the study.

Purpose and Design of NA 2009

National assessments allow the objective measurement of performance at system-level. Some of the more important benefits of such national-level data include: informing policy, monitoring standards, identifying correlates of achievement, introducing realistic standards, promoting accountability, increasing public awareness, directing teachers' efforts and raising pupil achievement, and informing political debate (Greaney & Kellaghan, 1996). These are reflected in the aims of the 2009 National Assessments (NA 2009):

- to establish the current English reading and mathematics standards of Second and Sixth class pupils.
- to provide high quality and reliable data for the Department of Education and Skills (DES) to assist in policy review and formulation and in decisions regarding resource allocation.
- to examine school, teacher, home background, and pupil characteristics, and teaching methods which may be related to reading and mathematics achievement.
- to provide a basis with which to compare future assessments of English reading and mathematics at Second and Sixth classes.

The first National Assessment in Ireland was carried out in 1972, and it examined the reading skills of a sample of 10-year-olds. The first National Assessment of mathematics took place in 1977. Either mathematics or reading was assessed at regular intervals from then until 2004, when both reading and mathematics were assessed (in the same schools but at different class levels). The 2009 Assessments were the first time that the same grade levels and pupils were selected for the mathematics and reading components of the study. Since their inception, the Assessments have expanded to collect a considerable amount of contextual data (e.g., information about family background) as well as achievement data. This has allowed an examination of the relationships between achievement and characteristics of individual pupils, or of their families or school environments. Figure 1.1 provides a summary of key features of the 2009 National Assessments.

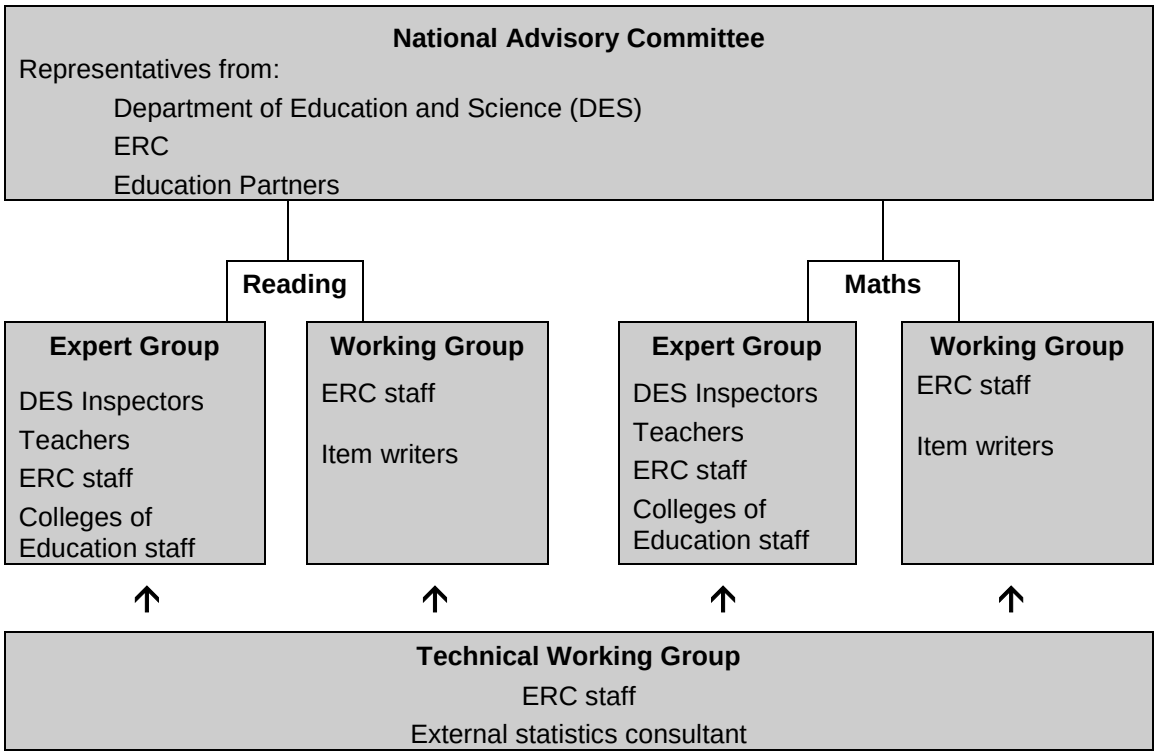
Figure 1.1: Summary characteristics of NA 2009

Sample Size	Almost 8,400 pupils (enrolled in 150 schools) were selected to participate in NA 2009. Pupils were evenly divided between Second and Sixth class. Achievement data were collected for approximately 3,900 Second class pupils and 3,800 Sixth class pupils.
Assessment Content	NA 2009 examined pupils' English reading and mathematics capabilities. The test of English reading examined the areas of vocabulary and reading comprehension. The comprehension component of the test examined the purpose (either for literary experience or to acquire and use information) and processes (Retrieval, Inference, Interpreting & Integrating, and Examining & Evaluating) involved in reading texts. The mathematics test examined mathematical content strands (Number, Algebra, Shape & Space, Measures and Data) and process skills (Applying & Problem-Solving, Integrating & Connecting, Reasoning, Implementing, and Understanding & Recalling).
Test Format	Pupils were assessed using a paper-and-pencil test. Reading: At both Second and Sixth, pupils' reading was assessed using one of four test booklets, each of which shared a common vocabulary section. Within any given class group, pupils were pre-assigned one of the four test booklets. This minimised opportunity for pupils to copy, and allowed broader coverage of content and processes. Mathematics: There were four versions of the test booklets at Second class, and six versions at Sixth, with a common block (for each grade) in each booklet. To minimise the effects of reading difficulties, the Second class tests were read aloud to pupils. Consequently, all pupils in a Second class group were assigned the same booklet. In contrast, in any Sixth class group, pupils were pre-assigned one of the six test booklets.
Test Length	Reading: The total administration time was approximately 90 minutes, of which 62 minutes (Second) and 70 minutes (Sixth) were allocated to actual testing. Mathematics: As the test was read aloud at Second class, the actual testing time varied slightly, but averaged 65 to 70 minutes, with an approximate total testing time of 90 minutes. The Sixth class administration took 130 minutes, of which 105 minutes was actual testing time.
Contextual Information	Questionnaires were completed by pupils, parents, class teachers and school principals.

Management of NA 2009

The Educational Research Centre (ERC) was responsible for the overall implementation and management of NA 2009, and was supported by a National Advisory Committee and subject-specific Expert Groups. In addition, subject-specific Working Groups (composed of external test item writers and ERC staff) sourced test materials and developed test items, while a Technical Group advised on issues such as test structure, sampling and sample size, item analysis, and test scaling. An external expert also advised the Technical Group and verified the scaling procedures used. The organisation structure for NA 2009 is outlined in Figure 1.2 (overleaf).

Figure 1.2: Organisational structure of NA 2009



The National Advisory Committee was drawn from the education partners. It advised on policy priority areas, reporting plans, and the broad assessment framework. Membership of the committee was as shown in Table 1.1.

Table 1.1: Membership and institutional affiliation of the National Advisory Committee for NA 2009

National Advisory Committee	
An Foras Patrúnachta	Carmel Nic Airt
CPSMA	Mark Candon
DES	Harold Hislop (replaced by Pádraig Mac Fhlannchadha) Margaret Kelly
ERC	Eemer Eivers Gerry Shiel Seán Close
Gaelscoileanna	Dónal Ó hAiniféin
INTO	Deirbhile Nic Craith
IPPN	John Curran
National Parents Council	Áine Lynch
NCCA	Arlene Forster
PPDS (formerly PCSP)	Catherine Shanahan (replaced by Ciara O'Donnell)

The two subject Expert Groups advised in more detail on the assessment framework, test structure and content areas, and provided guidance on questionnaire

content. Membership of the groups was composed of ERC staff, members of the DES Inspectorate, current teachers, and lecturers from Colleges of Education (Table 1.2). The subject Working Groups were smaller groups, responsible for the sourcing of test material and development of test items. They were composed of ERC staff and some primary school teachers (Table 1.3). The Technical Group was composed of ERC staff and Dr Fernando Cartwright (Statistics Canada), who provided advice on scaling the achievement data and developing proficiency levels (Table 1.4).

Table 1.2: Membership of the Mathematics and Reading Expert Groups for NA 2009

	Maths Expert Group	**Reading Expert Group**
ERC	Gerry Shiel Seán Close Rachel Cunningham (to Sept. 2009) Lorraine Gilleece (from Nov. 2009)	Eemer Eivers Gerry Shiel Aidan Clerkin Joanne Kiniry (to Sept. '09)
DES	Seán Ó Cearbhaill (to Sept. 2009) Joan Hanrahan (to Sept. 2009) John White (from Nov. 2009) Eamon Clavin (from Nov. 2009)	Diarmuid Dullaghan Pat Delea
Colleges of Education lecturers	Noreen O'Loughlin Dolores Corcoran Therese Dooley	Therese McPhillips
Primary teachers / other	Patricia Cassidy Cheryl Greene Mairéad Twohig (to Sept. 2009) Ciara O'Donnell (from Nov. 2009)	Kieran Fanning Tara Concannon-Gibney (to April 2009) Aoibheann Kelly (from Sept. 2009)

Table 1.3: Membership of the Mathematics and Reading Working Groups for NA 2009

	Maths Working Group	**Reading Working Group**
ERC	Gerry Shiel Seán Close	Tom Kellaghan Eemer Eivers Gerry Shiel Aidan Clerkin Joanne Kiniry
Teachers	Patricia Cassidy Cheryl Greene	Aoibheann Kelly Betty Behan Kate Brand Kieran Fanning Tara Concannon-Gibney

Table 1.4: Membership of the Technical Group for NA 2009

ERC	David Millar Eemer Eivers Gerry Shiel	Seán Close Thomas Kellaghan
External Consultant	Fernando Cartwright	

Chapter 2
Framework and Test Development for Mathematics

This chapter describes the framework for the mathematics component of NA 2009, and how it shaped the development of the eventual test instruments. It is divided into three main sections. The first section describes the development of the Second class mathematics test, including coverage of curriculum objectives and their underlying content and cognitive processes, and development and specification of items and test characteristics for the field trial and final test. The second section provides similar information relating to the development of the Sixth class test[1]. The final section provides sample items for Second and Sixth classes.

The assessment framework for 2009 is based on the revised PSMC which was introduced in 2000, but implemented from 2002 onwards. The framework represents an extension and modification of the frameworks used in the 1999 and 2004 National Assessments of Mathematics at Fourth class (Shiel & Kelly, 2001; Shiel, Surgenor, Close & Millar, 2006) and covers the Primary School Mathematics Curriculum (PSMC) for Second and Sixth classes (DES/NCCA 1999a). In the PSMC mathematics is described as:

> '… the science of magnitude, number, shape, space, and their relationships and also as a universal language based on symbols and diagrams. It involves the handling (arrangement, analysis, manipulation and communication) of information, the making of predictions and the solving of problems through the use of a language that is both concise and accurate.' (DES/NCCA 1999a, p. 2)

The PSMC is structured along two main dimensions – mathematical content strands and cognitive process skills – which underpin specific instructional objectives for each class level from Junior Infants to Sixth class. The mathematical content strands of the PSMC are: Number, Algebra, Shape & Space, Measures, and Data. These are further subdivided into strand units at each class level. The cognitive process skills are categorised as follow: Applying and Problem-Solving, Communicating & Expressing, Integrating & Connecting, Reasoning, Implementing, and Understanding & Recalling. These process skills are elaborated at progressive levels of complexity for each class level from Junior Infants to Sixth class. The instructional objectives associated with these two dimensions, along with exemplars, are listed in the PSMC for each class level. Unlike the Primary School English Curriculum (PSEC), the Primary School Mathematics Curriculum (PSMC) gives specific information on what is to be taught at each class level in the form of these objectives and exemplars, thus facilitating the development of an assessment framework which is directly linked to the mathematics curriculum.

[1] The mathematics framework can be found in full at: www.erc.ie/documents/nama09_framework.pdf

Development of the Second Class Mathematics Test

This section describes the framework and test development for Second Class.

Coverage of Second Class PSMC Objectives

Table 2.1 lists the 59 assessment objectives for Second class.

Table 2.1 PSMC objectives for Second Class

Number and Algebra

Numeration and Place Value
1. Count the number of objects in a set
2. Read, write and order numerals, 0 to 199
3. Estimate the number of objects in a set
4. Compare equivalent and non-equivalent sets
5. Use the language of ordinal number
6. Identify and record place value 0 to 199

Addition
7. Recognise addition as combining or partitioning sets
8. Apply commutative, associative and zero props. of addition
9. Recall addition facts
10. Construct addition number sentences and number stories
11. Solve simple word problems involving addition
12. Add numbers without and with renaming within 99
13. Use repeated addition and group counting

Subtraction
14. Recognise subtraction as deducting, as complementing and as difference
15. Recall subtraction facts
16. Construct subtraction number sentences and number stories
17. Solve simple word problems involving subtraction
18. Estimate sums and differences within 99
19. Subtract numbers without and with renaming within 99
20. Use the symbols +, - , =, <, >
21. Solve simple word problems involving addition and subtraction.

Fractions and Pattern
22. Identify half and a quarter of sets to 20
23. Recognise and extend patterns in numbers
24. Use patterns in addition facts

Shape and Space

Spatial Awareness and 2-D Shapes
25. Use the vocabulary of spatial relations
26. Give and follow simple directions
27. Describe, compare and name 2-D shapes
28. Construct 2-D shapes
29. Combine and partition 2-D shapes
30. Identify halves and quarters of 2-D shapes
31. Identify use of 2-D shapes in the environment

Symmetry and Angles
32. Identify line symmetry in shapes and in the environment
33. Recognise angles in the environment

Shape and Space (*continued*)

3-D Shapes
34. Describe, compare and name 3-D shapes
35. Identify use of 3-D shapes in the environment
36. Solve problems involving 2-D & 3-D shapes
37. Recognise the relationship between 2-D & 3-D shapes.

Measures

Length and Area
38. Estimate, compare, measure and record length using non-standard units
39. Select appropriate non-standard measuring units and instruments for length
40. Estimate, measure and record length using standard unit (metre and cm)
41. Solve simple word problems involving length
42. Estimate and measure area using non-standard units

Weight
43. Estimate, compare, measure and record weight using non-standard units
44. Select appropriate non-standard measuring units and instruments for weight
45. Estimate, measure and record weight using standard unit (the kilogram)
46. Solve simple word problems involving measures of weight

Capacity
47. Estimate, compare, measure and record capacity using non-standard units
48. Select appropriate non-standard measuring units and instruments for capacity
49. Estimate, measure and record capacity using standard unit (the litre)
50. Solve simple word problems involving measures of capacity

Time
51. Use the vocabulary of time to sequence events
52. Read and record time using simple devices
53. Read time in hours and half-hours on 12-hour analogue and digital clock
54. Read day, date month and season using calendar
55. Solve simple word problems involving measures of time

Money
56. Recognise, exchange and use coins
57. Calculate how many items may be bought with a given sum.

Data
58. Sort and classify objects by two and three criteria
59. Represent, read and interpret block graphs

In the case of six objectives (1, 4, 5, 9, 25, and 56), it was assumed that they have been mastered by the vast majority of Second class pupils as they are relatively easy and are generally covered in Senior Infants or First class at the latest, and so do not need to be tested at Second class level (e.g., Objective 9 – recall of addition facts – is an obvious prerequisite for Second class objectives where multi-digit addition is involved, and the other 5 objectives can be regarded as prerequisites for other Second class objectives). Ten objectives (3, 7, 36, 39, 43, 44, 45, 47, 49, and 52) are difficult to assess by pen and paper group test and should be assessed by other means. All other objectives were tested in the 2009 Second class assessment.

A set of items based on these objectives was prepared for the field trial, conducted in May 2008. The items were developed by individual members of the Mathematics Working Group and were reviewed by the group in plenary session. A number of resources were referred to in writing the items including: the exemplars in the PSMC and the accompanying teachers' handbook, the three main textbook series in use in the schools, and standardised tests in mathematics. The items, 120 in all, were categorised by content area and process skill.

Coverage of Content Strands and Process Skills

Table 2.2 below shows the distribution, by content strand, of items in the field trial, along with the distribution of the PSMC objectives for Second class. The distribution of items across the content strands represents a reasonable reflection of the distribution of objectives across strands in the PSMC.

Table 2.2: Distribution of the mathematics items for the NA 2009 field trial and PSMC objectives for Second class, by mathematics content strand

	Field trial		PSMC	
Content Strand	No. of items	% of items	No. of objectives	% of objectives
Number/Algebra	51	42.5	24	41.0
Shape and Space	24	20.0	13	22.0
Measure	36	30.0	20	34.0
Data	9	7.5.	2	3.0
Total	120	100.0	59	100.0

The approach to classifying process skills in 2009 was similar to the 2004 assessment. Table 2.3 shows the distribution of the field trial items across the process skill categories of the PSMC.

Table 2.3: Distribution of mathematics items for the NA 2009 field trial for Second class, by mathematics process skill

Process Skill	No. of items Field trial	% of items Field trial
Understand & Recall	20	16.7
Implement	16	13.3
Integrate/Connect	21	17.5
Reason	35	29.2
Apply & Problem Solve	28	23.3
Total	120	100.0

About 60 percent of the 120 items involved tasks embedded in a practical or environmental context of some sort (e.g. shopping, home or social activities), while the remaining 40 percent involved tasks of a purely mathematical nature. This reflects the emphasis on problem-solving in the PSMC. About one-third of the items in the field trial were multiple-choice (39 items) and about two-thirds were constructed response (81 items).

Test Specifications – Field Trial

For the field trial at Second class level, conducted in May 2008, there were 120 mathematics items in 6 blocks of 20 items each (called Blocks A to F), distributed across 5 pupil booklets so that each pupil took a core block and two other blocks, i.e. 60 items (Table 2.4).

Table 2.4: Structure of test booklets – field trial, Second class

Booklet	First Section	Second (Core) Section	Third Section
1	B	A	C
2	C	A	D
3	D	A	E
4	E	A	F
5	F	A	B

The mean percent correct scores on each of the six blocks of items field tested are given in Table 2.5 below. Although the mean percent score on all 120 items was approx. 61% , which is near enough to the intended figure of 60%, there was considerable variation in mean percent scores for each block, with Block C being the lowest at 49% and Block A being the highest at 69% – a difference of 20%. This was not the case with the five 3 x 20 item pupil booklets where mean percent scores range from 58% to 66% – a difference of 8% (Table 2.6).

Table 2.5: Mean percent correct scores by block – field trial, Second class

Block	*A*	*B*	*C*	*D*	*E*	*F*
Mean % Correct	69	56	49	53	61	67

Table 2.6: Mean percent correct scores by booklet – field trial, Second class

Booklet	1	2	3	4	5
Blocks	BAC	CAD	DAE	EAF	FAB
Mean % Correct	58	57	61	66	64

The field trial results suggested that revisions to the test in preparation for the main study should aim to increase the difficulty levels of the easier blocks and reduce the difficulty levels of the more difficult blocks, while maintaining the existing distribution of items across content strands and process skills as far as possible.

Problem Items

Analysis of the performance of the field trial pupils on each of the 120 items in the test yielded measures of the quality of the items, including: (i) difficulty level; (ii) ability to discriminate between high achievers and low achievers; and, (iii) the functioning of alternate responses in the multi-choice items. These measures were then used to identify problematic items in the test which would need to be modified or replaced. Overall, 28 items out of a total of 120 were identified as being problematic, with 12 of these considered very easy (with over 85% obtaining a correct response) and 3 considered very difficult (with less than 20% obtaining a correct response). Two items discriminated poorly (point-biserial correlation less than 0.3), two had faulty distractors, and nine others had a variety of other faults. Differential item functioning analyses did not reveal any gender bias in items.

Test Structure – Main Study

Based on the results of the field trial outlined in the above sections, the Mathematics Expert Group agreed to the following revisions to the test being made:

- Block F was deleted as it was (with Block A) one of the two easiest blocks based on mean percent correct per block.
- The "good" items from Block F (all but two poorly performing items) were used to replace the poor items deleted from Blocks A, B, C, D, and E so as to maintain the distribution of items across content strands and process skills as per the Second class mathematics framework and to narrow the spread in mean performance across the remaining 5 blocks. Twelve items across these five blocks were replaced with more suitable items (either more difficult or easier items) from Block F.

The design of the test for the main study involved 4 booklets, each containing 3 blocks of 20 items, so each pupil was presented with 60 items as part of a rotated booklet design as shown in Table 2.7.

Table 2.7: Structure of mathematics test booklets – main study, Second class

Booklet	First Section	Second (Common) Section	Third Section
1	A	C	B
2	B	C	D
3	D	C	E
4	E	C	A

Block C was the common or core block and appeared in all four booklets. Each of the other four blocks, A, B, D, and E, appeared in two booklets – once in the initial position, and once in the final position.

When these revisions were incorporated into the test the distribution of items across the curriculum content strands and process skill categories was as per Tables 2.8 and 2.9 below. The revised test maintained a satisfactory distribution of items across the content strands and process skills of the PSMC. About 30% of the items were multiple-choice and about 70% were short answer open response.

Table 2.8: Classification of mathematics items by content strand – main study, Second class

Content Area	Number	Shape & Space	Measure	Data	Total
No. of items	44	16	34	6	100
% of items*	44	16	34	6	100
% of PSMC objectives	41	22	34	3	100

*Numbers of items (N = 100) correspond with reported percentages.

Table 2.9: Classification of mathematics items by process skill – main study, Second class

Process Skill	Understand & Recall	Implement	Integrate & Connect	Reason	Apply & Problem Solve	Total
No. of items	11	17	16	28	28	100
% of items*	11	17	16	28	28	100

*Numbers of items (N = 100) are the same as the percentages.

Development of the Sixth Class Mathematics Test

This section describes the framework and test development for Sixth Class.

Coverage of Sixth Class PSMC Objectives

There are 72 objectives in the PSMC for Sixth class (shown in Table 2.10 on the next page). All of the objectives were seen as appropriate for testing by pen and paper test. Many of the more complex objectives would be tested by two or more items. A set of items based on the objectives was prepared for the field trial, conducted in May 2008. The items were developed by individual members of the Mathematics Working Group and were reviewed by the group in plenary session. As was done for Second class, a number of resources were referred to in writing the items including: the exemplars in the PSMC handbook and the accompanying teachers' handbook, the three main textbook series in use in the schools, and standardised tests in mathematics. The items, 175 in all, were categorised by content area and process skill.

Table 2.10: PSMC objectives for Sixth class

Number and Algebra

Numeration and Place Value
1. Read, write and order whole numbers and decimals
2. Identify place value in whole numbers and decimals
3. Round decimals

Operations
4. Estimate sums, differences, products, quotients of decimals
5. Add and subtract whole numbers and decimals (three places) without and with a calculator
6. Multiply a decimal by a decimal, without/with a calculator
7. Divide a four-digit no. by a two-digit no., without/with a calculator
8. Divide a decimal by a decimal, without/ with a calculator

Fractions, Decimals and Percentages
9. Compare and order fractions and identify equivalent forms
10. Express improper fractions as mixed numbers and position them on the number line
11. Add and subtract simple fractions and simple mixed nos.
12. Multiply a fraction by a fraction
13. Express tenths, hundredths and thousandths in both fractional and decimal form
14. Divide a whole number by a unit fraction
15. Use simple ratios
16. Use percentages and relate them to fractions & decimals
17. Compare and order percentages of numbers
18. Solve problems on profit and loss, discount, VAT, interest

Number Theory
19. Identify simple prime and composite numbers
20. Identify square numbers
21. Identify simple square roots
22. Identify common factors and multiples
23. Write whole numbers in exponential form

Algebra
24. Identify positive and negative numbers on the no. line
25. Add simple positive and negative numbers on the no. line
26. Know simple properties and rules about brackets and priority of operation
27. Identify relationships and record symbolic rules for number patterns
28. Use a variable in the context of simple patterns, tables and simple formulae
29. Substitute values for variables
30. Translate word problems with variable into no. sentences
31. Solve one-step number sentences and equations

Shape and Space

2-D Shapes
32. Make informal deductions about 2-D shapes and their properties
33. Use angle and line properties to classify and describe triangles and quadrilaterals
34. Construct triangles from given sides or angles
35. Identify the properties of the circle
36. Construct a circle of given radius or diameter
37. Tessellate combinations of 2-D shapes
38. Classify 2-D shapes according to their lines of symmetry
39. Plot simple co-ordinates
40. Use 2-D shapes and properties to solve problems.

Shape and Space Contd.

3-D Shapes
41. Identify 3-D shapes and analyse relationships, including octahedron
42. Draw the nets of 3-D shapes and construct the shapes.

Lines and Angles
43. Recognise, classify and describe angles and relate angles to shape
44. Recognise angles in terms of a rotation
45. Estimate, measure and construct angles in degrees
46. Explore the sum of the angles in a quadrilateral

Measures

Length
47. Select and use appropriate instruments of measurement
48. Rename measures of length
49. Estimate and measure the perimeter of regular and irregular shapes
50. Use and interpret scales on maps and plans
51. Know that the length of the perimeter of a rectangular shape does not determine it's area

Area
52. Calculate the area of regular and irregular 2-D shapes
53. Measure the surface area of specified 3-D shapes
54. Calculate area using acres and hectares
55. Identify the relationship between square metres and square centimetres
56. Find the area of a room from a scale plan

Weight and Capacity
57. Rename measures of weight
58. Rename measures of capacity
59. Find the volume of a cuboid experimentally

Time
60. Solve problem involving international time zones
61. Know the relationship between time, distance and average speed

Money
62. Compare prices to identify value for money
63. Convert foreign currencies to euros and vice versa

Data

Data
64. Collect, organise and represent data using pie charts and trend graphs
65. Read and interpret trend graphs and pie charts
66. Compile and use simple data sets
67. Calculate averages of simple data sets
68. Use data sets to solve problems.

Chance
69. Identify and list all possible outcomes of simple random processes
70. Estimate the likelihood of occurrence of events;
71. Order on a scale from 0 to 100%, 0 to 1
72. Construct and use frequency charts and tables

Coverage of Content Strands and process Skills

Table 2.11 shows the distribution of the 175 field trial items and PSMC objectives, by content strand. The distribution of items across the content strands represented a reasonable reflection of the distribution of objectives across the same strands in the PSMC.

Table 2.11: Distribution of the field trial items and PSMC objectives for Sixth class, by mathematics content strand

	Field trial		PSMC	
Content Strand	No. of items	% of items	No. of objectives	% of objectives
Number/Algebra	75	43.0	31	43.1
Shape and Space	40	23.0	15	20.8
Measure	40	23.0	17	23.4
Data	20	11.0	9	12.5
Total	175	100.0	72	100.0

Table 2.12 below shows the distribution of the field trial items across the process skills of the PSMC. This was considered reasonably well balanced with an emphasis on higher order skills such as reasoning and problem-solving, although the proportion of reasoning items could have been a little less. About half of the items were embedded in a practical context while the other half was embedded in a purely mathematical context.

Table 2.12: Distribution of the field trial items for Sixth class, by mathematics process skill

Process	No. of items	% of items
Understand and Recall	18	10.7
Implement	31	14.7
Integrate/Connect	9	10.7
Reason	55	38.7
Apply and Problem Solve	62	25.3
Total	175	100

Test Structure – Field Trial

For the field trial at Sixth class level, 175 items were distributed over 7 blocks (ABCDEFG), each consisting of 25 items. The blocks were, in turn, divided into 10 pupil booklets so that each pupil took a non-calculator block and two other blocks (for which calculators were permitted) (Table 2.13). Each booklet contained 75 items.

Sixty-six (38%) of the 175 items were multiple-choice and 109 (62%) were constructed response, which was near enough to the target ratio of two constructed response items to one multiple-choice item.

Table 2.13: Structure of maths test booklets – Sixth class, field trial

Booklet	First section (non-calculator)	Second section (calculator)	Third section (calculator)
1	A	C	E
2	B	C	F
3	A	D	G
4	B	D	C
5	A	E	D
6	B	E	F
7	A	F	G
8	B	F	D
9	A	G	C
10	B	G	E

The mean percent correct scores on each of the seven blocks of items field tested are given in Table 2.14 below. Although the mean percent score on all the blocks was 51.3% – a little below the intended figure of around 55% – there was considerable variation in mean percent scores between blocks. Blocks D and G were the lowest at 46% and Block A was the highest at 61% – a difference of 15%. The two non-calculator blocks were at about the same level of difficulty. Table 2.15 shows that, despite differences in difficulty across blocks, the booklets were evenly balanced with mean scores falling between 52 and 54%.

Table 2.14: Mean percent correct by block, field trial – Sixth class mathematics

Block	A	B	C	D	E	F	G
Mean % Correct	61	60	50	46	49	47	46

Table 2.15: Mean percent correct by booklet, field trial – Sixth class mathematics

Booklet	1	2	3	4	5	6	7	8	9	10
Blocks	ACE	BCF	ADG	BDC	AED	BEF	AFG	BFD	AGC	BGE
Mean % Correct	54	52	52	53	53	53	52	52	53	52

These results suggested that revisions to the test in preparation for the main study should aim to reduce the difficulty level of the harder blocks while to a lesser extent increasing the difficulty level of the easier blocks.

Problem Items

Analysis of the performance of the field trial pupils on each of the 175 items in the test yielded measures of the quality of the items, including: (i) difficulty level; (ii) ability to discriminate between high achievers and low achievers; and (iii) the functioning of

alternate responses in the multi-choice items. These measures were then used to identify problematic items in the test that need to be modified or replaced. No gender differences emerged from differential item functioning analyses.

Overall, 30 items out of a total of 175 were identified as being problematic (see Close, Millar and Shiel (2009) for a discussion of the considerations involved in identifying problematic items from the field trial). Four items were considered very easy (over 85% of pupils obtained the correct response), 13 were considered very difficult (with fewer than 20% obtaining the correct response), with most of these discriminating poorly (point-biserial normally less than 0.3). An additional five items had faulty distracters.

Test Specifications – Main Study

Based on the results of the field trial outlined in the above sections, the Mathematics Expert Group agreed to the following revisions to the test being made for the Main Study:

- Block G was deleted as it had the greatest number of statistically poor performing items (10) and it was also the most difficult block (with D which had 5 poor items).
- The good items from G were then used to replace the poor items deleted from Blocks C, D, E and F, so as to maintain the balance of items across content strands and process skills. Five items in these four blocks needed only minor alteration to improve them, but 12 of them needed to be replaced with more suitable items (in terms of content and difficulty).
- The two non-calculator blocks, A and B, which contained items for which calculator use was considered inappropriate, performed well with just minor adjustments to a few items needed.

The design for the main study test was then 6 booklets, each containing 3 blocks of 25 items, so each pupil was presented with 75 items as part of a rotated booklet design, as per Table 2.16.

Table 2.16: Structure of test booklets – Sixth class, main study

Booklet	First section (non-calculator)	Second section (calculator)	Third section (calculator)
1	A	C	D
2	B	C	E
3	A	C	F
4	B	C	D
5	A	C	E
6	B	C	F

A and B were the non-calculator blocks. C was the core block taken by all pupils. Each of the other 3 blocks (D, E, and F) appeared twice. When these revisions were incorporated into the test, the distribution of items across the curriculum content strands and process skill categories was as per Tables 2.17 and 2.18 below. The revised test maintained a satisfactory distribution of items across the content strands and

process skills of the PSMC, with about half of the items involving a practical or environmental context. About three-eighths of the items were multiple-choice and five-eighths were short answer open response.

Table 2.17: Classification of items by Content Strand – Sixth Class, main study

Content Area	Number	Shape & Space	Measure	Data	Total
No. of items	69	32	31	18	150
% of items	46.0	21.3	20.7	12.0	100
% of PSMC objectives	43.0	21.0	24.0	12.0	100

Table 2.18: Classification of items by Process Skill – Sixth Class, main study

Process Skill	Understand & Recall	Implement	Integrate & Connect	Reason	Apply & Problem Solve	Total
No. of items	15	30	8	47	50	150
% of items	10.0	20.0	5.3	31.3	33.3	100

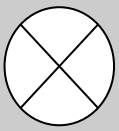
Sample Items for Second and Sixth Classes

This section contains sample items administered as part of the NA 2009 main study (in May 2009).

Sample Items for Second Class

Content Area: Shape & Space: 2-D Shapes
Process: Understand & Recall
Correct: 94%

Q. 1 Colour in half of this shape.



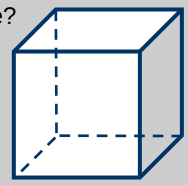
Content Area: Measures: Time
Process: Apply & Problem-Solve
Correct: 65%

Q. 2 Jane's birthday is on the 14th of March. Jack's birthday is five months later. In what month is Jack's birthday?

Content Area: Shape & Space: 3-D Shapes
Process: Understand & Recall
Correct: 66%

Q. 3 Which of these do all cubes have?

O 4 faces
O 8 corners*
O 6 edges
O 12 faces



Content Area: Number & Algebra: Operations
Process: Implement
Correct: 55%

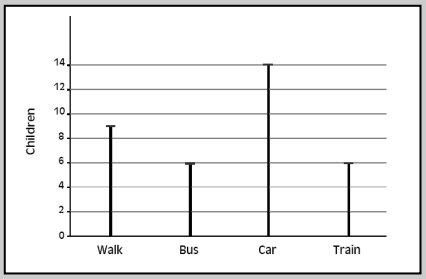
Q. 4

70
-24

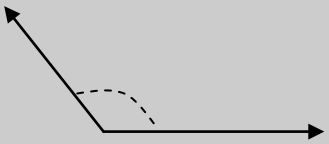
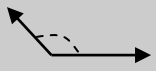
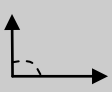
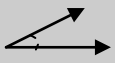
Content Area: Measures: Money
Process: Apply & Problem-Solve
Correct: 42%

Q. 5 Jim has 78c. He needs another 17c for a packet of football stickers. How much does the packet cost?

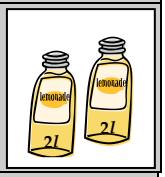
Sample Items for Second Class (continued)

Content Area: Number & Algebra: Operations **Process:** Apply & Problem-Solve **Correct:** 43%	**Q. 6** There are 30 children in Second class. Yesterday at lunchtime, 12 of them played skipping, 9 played basketball and the rest played football. How many children played football?
Content Area: Data: Represent & Interpret data **Process:** Integrate and Connect **Correct:** 39%	**Q. 7** The line graph shows the different ways pupils in 2nd Class travel to school.  How many more children travel by car than by train?
Content Area: Number & Algebra: Operations **Process:** Reason **Correct:** 25%	**Q. 8** Which of these gives the best guess of 86 – 59 ? **O** 70 – 50 **O** 90 – 60* **O** 80 – 60 **O** 80 – 70

Sample Items for Sixth Class

Content Area: Shape & Space: Lines & Angles **Process:** Understand & Recall **Correct:** 85%	**Q. 1** What type of angle is this?  Acute angle Obtuse angle* Right angle Reflex angle
Content Area: Data: Chance **Process:** Reason **Correct:** 80%	**Q. 2** The principal gave a quiz to all pupils in 6th class. It had 20 questions with one mark for each correct answer. The results are shown in the table. **Score out of 20** | **8** | **9** | **10** | **11** | **12** | **14** | **15** | **16** | **17** | **18** **No. of pupils** | **2** | **2** | **3** | **5** | **7** | **6** | **6** | **5** | **3** | **1** How many pupils got a score of 10?
Content Area: Shape & Space: Lines and Angles **Process:** Implement **Correct:** 75%	**Q. 3** Circle the letter under the angle that is about 135 degrees.  **A***  **B**  **C**  **D**

Sample Items for Sixth Class (continued)

Content Area: Number & Algebra: Operations **Process:** Reason **Correct:** 66%	**Q. 4** Which of these is the <u>best</u> estimate of 8.61 x 22? 8 x 20 10 x 22 9 x 20* 9 x 25
Content Area: Number & Algebra: Rules and Properties **Process:** Integrate & Connect **Correct:** 63%	**Q. 5** Which of these tells how to get the missing number in this sequence? 1, 2, 5, 10, 17, __ Add 7 to the last number Double the last number Add the last two numbers Add 9 to the last number*
Content Area: Data: Chance **Process:** Apply & Problem-Solve **Correct:** 51%	**Q. 6** A bag contains 4 red cubes, 6 blue cubes, and 10 green cubes. Without looking, Jenny picks a cube out of the bag. What chance has she of picking a blue cube?
Content Area: Shape & Space: 2-D Shapes **Process:** Reason **Correct:** 44%	**Q. 7** Which of these is true of all scalene triangles? They have two equal sides They have an angle greater than right-angle They have no right angles They have no sides equal*
Content Area: Measures: Capacity **Process:** Apply & Problem-Solve **Correct:** 47%	**Q. 8** 9 children at a party each drank 350 ml of lemonade. How much lemonade was left from these two 2 litre containers? 
Content Area: Number & Algebra: Decimals & Percentages **Process:** Implement **Correct:** 23%	**Q. 9** 2.25 x 0.4 =
Content Area: Measures: Money **Process:** Apply & Problem-Solve **Correct:** 23%	**Q. 10** On Thursday the Euro was worth 1.50 dollars on the currency market. A month later the Euro was worth 1.20 dollars. What was the percentage decrease in the value of the Euro over the month?

Chapter 3
Framework and Item Development for English Reading

This chapter describes the framework for the English reading component of NA 2009, and outlines how the framework shaped the development of the eventual test instruments. It is divided into three main sections. The first deals with the development of the framework for English reading, including defining the domain and the underlying content and processes. The section also relates the framework to the curriculum and textbook content experienced by pupils in order to establish the relative weighting to be attached to various framework components. The second section describes how the framework informed the development of test items, from initial text selection through to the field trial phase and to the selection of the final item pool. The final section contains examples of test units used as part of the main study.

The Assessment Framework for English Reading

An assessment framework attempts to describe what is being assessed, how it is being assessed, and why it is being assessed (Kirsch, 2001). This is achieved by describing the assessment process and the assumptions behind it, not only for those responsible for developing the assessment instruments, but also for a wider audience (e.g., policymakers, teachers, and curriculum developers). Thus, an assessment framework describes general aims, which in turn provide a basis for specifying what will be measured in terms of knowledge, skills, and other attributes.

The framework for the English reading component of NA 2009 was developed by a Reading Working Group at the Educational Research Centre, led by Dr Tom Kellaghan. As Cosgrove, Milis, Shiel, Forde and Wardle (2004) had developed an extensive framework for the 2004 National Assessment of English Reading (NAER), this was used to guide the development of the 2009 framework, bearing in mind the need for modifications due to the change in target grade levels. This section summarises the framework for English reading in NA 2009. The full framework can be accessed from www.erc.ie/NA 2009.

Definition of Reading

As in NAER 2004, reading was defined as

> the process of constructing meaning through the dynamic interaction among the reader's existing knowledge, the information suggested by the written language, and the context of the reading situation. Young readers read to learn, to participate in communities of readers, and for enjoyment (Eivers, Shiel, Perkins, & Cosgrove, 2005, p. 15).

This draws on definitions used in international studies of reading achievement, as well as research on the development of pupils' literacy skills, and the Primary School English Curriculum (PSEC) in Ireland. It emphasises reading comprehension, as this was the focus in the National Assessment, but the framework also recognises that the

foundation components of reading are important elements of an assessment. Thus, the framework explicitly refers to the need to include a set of common vocabulary items in the test instruments, to assess basic reading skills.

Dimensions of Reading Comprehension

As well as a broad definition of reading, the framework identifies different dimensions of reading comprehension. The key dimensions of any text are identified as the purpose of the text and the processes that the reader must use in order to interpret the text.

Purpose of Text

The 2009 framework differs from that used in 2004 in one very important respect – the classification of texts by purpose rather than by type. In NAER 2004, texts were described by type (narrative/expository/documents) and form (continuous/non-continuous). Narrative texts were defined as those where the main purpose is to tell a story, and expository texts as those mainly designed to inform the reader (e.g., describe something or put forward an opinion). Both narrative and expository texts were described as being presented in continuous form. In contrast, document texts were defined more by their appearance, as their distinguishing feature is that they present information in non-continuous form (e.g., using charts, tables or lists).

Although the classification is similar to that used in the PSEC, there are two main problems with it. First, documents are sometimes designed to inform or persuade the reader – thus blurring the boundary between expository and documents. Second, it can be difficult to distinguish between narrative and expository texts, as the "defining" difference is at the level of rhetorical structure. For example, the structure of narratives can be described in terms of setting-complication-resolution, or story grammar, or causal event chains. Expository texts, on the other hand, are typically described in terms of such schemata as classification, illustration, comparison and contrast, and procedural description (Weaver & Kintisch, 1991).

Given these difficulties, the 2009 framework adopted part of the structure used in PIRLS 2006 (Mullis, Kennedy, Martin & Sainsbury, 2006). This involved a change of focus from classification based on text *type* to text *purpose*. Mullis et al. note that, among young schoolchildren, the main reasons for reading are either for **literary experience** or **to acquire and use information**. These purposes were adopted as part of the framework for the 2009 assessment.

When **reading for literary experience** 'the reader engages with the text to become involved in imagined events, setting, actions, consequences, characters, atmosphere, feelings, and ideas, and to enjoy language itself. To understand and appreciate literature, the reader must bring to the text his or her own experiences, feelings, appreciation of language and knowledge of literary forms. For young readers, literature offers the opportunity to explore situations and feelings they have not yet encountered.' (Mullis et al., 2006, p. 19). Children's experience of literary text is usually via narrative fiction, but poetry and plays also fall under this category.

In contrast, when **reading to acquire and use information***:* 'the reader engages not with imagined worlds, but with aspects of the real universe. Through informational texts, one can understand how the world is and has been, and why things

work as they do. Readers can go beyond the acquisition of information and use it in learning and in action' (Mullis et al., 2006 p. 19). The structure of informational texts can vary considerably. For example, they can be continuous or non-continuous, and do not always need to be read from beginning to end. The information can be ordered chronologically (e.g., a set of instructions, in sequence) or logically (e.g., outline a problem, then propose a solution).

Processes of Comprehension

The framework identified four distinct processes which readers use to understand text (**retrieve**, **infer**, **interpret and integrate**, **examine and evaluate**). These processes also formed part of the NAER 2004 and the PIRLS 2006 frameworks, and are described below.

Table 3.1: Processes of reading comprehension, and related examples

Process	Examples
Retrieve requires the reader to read a text, and to understand how what is stated in the text relates to the information that is sought.	Look for specific information, events, ideas, definitions or phrases; identify the setting of a story; find the main theme of a text when explicitly stated.
Straightforward inference requires the reader to make inferences about something not explicitly stated in the text. The inferences are usually simple, and based on information that is explicitly stated in the text.	Deduce or infer that one event caused another; determine the main point of a series of arguments; identify generalisations in a text; describe the relationships between two characters.
Interpret and integrate requires a more holistic understanding of the text, beyond the level of sentence. Some integration of personal knowledge or experience with text content may be required.	Discern the overall message or theme of a text; consider an alternative to actions of characters; compare and contrast text information; infer the mood or tone of a story; apply text information to a real world situation.
Examine and evaluate involves evaluation of a text, either from a personal perspective or a more critical and objective viewpoint. Emphasis changes from understanding the text to critiquing it.	Evaluate the plausibility of what the text describes; identify and comment on the structure and organisation of texts; judge the completeness or clarity of information in a text; identify or comment on the writer's purposes and viewpoints.

Mapping the Reading Framework Onto Pupil Texts

The assessment framework for NAER 2004 contained in-depth content analyses of the English textbooks encountered by pupils in First and Fifth classes at that time (Cosgrove et al., 2004). The current framework does not contain a similar level of analysis, partly because content for Second and Sixth class can reasonably be extrapolated from the content at First and Fifth, and partly because (in most cases), pupils' English textbooks represent only one element of the materials they read.

Cosgrove et al.'s analyses revealed that, based on word counts, narrative short stories fell from 80% of texts at First class to 65% at Fifth class, while expository and representational texts increased from 20% to 35% at Fifth class. Mapped onto the present framework, this would suggest that 70% to 80% of Second class texts involve reading for literary experience, while 20% to 30% involve reading to acquire and use information. For Sixth class, the split between the two purposes is more even, with a little more than half of texts likely to involve reading for literary experience.

The processes in which pupils are expected to engage when reading can be inferred from the PSEC. First/Second class pupils are expected be able to recall details and events, assimilate facts, and retell stories. In addition, they should be able to respond to characters and events in a story, to imagine what it would be like to be certain characters, and to give an opinion of a text. However, these latter skills are considerably less emphasised in the PSEC than the former. Thus, in terms of reading processes, Second class pupils are expected to be able to retrieve and infer, and to a lesser extent, to interpret and integrate. They will also have encountered texts that require examination and evaluation, but their experiences of such processes will be somewhat limited.

At Fifth/Sixth class, the PSEC indicates that pupils should be able to use an array of higher order skills, including using comprehension skills to aid deduction, problem-solving, and prediction; using evidence from the text to support arguments and opinions; relating their own experiences to text content; distinguishing fact from opinion; and comparing various types of text. Relating these to reading processes, pupils at these grades are expected to be able to infer, interpret and evaluate (with a gradually increasing emphasis on interpretation and evaluation).

Although not dealt with in the 2009 framework for English reading, the types of themes and topics which pupils encounter in their books is also of relevance to test construction. Cosgrove et al. (2004) found that in First class, animals, monsters, fantasy, books and reading, playing, sleeping and transport were recurring themes. Topics in Fifth class were more varied and included nature and science, sports and hobbies, history and geography, people and culture, art, personal health and safety, and transport. These findings were used as guidelines for the development of Second and Sixth class test materials.

Specifications for Test and Item Format

As it was considered important to assess not only reading comprehension but also core reading skills, the reading framework for NA 2009 proposed that test booklets be split into two broad sections – Vocabulary and Comprehension. The aim of the Vocabulary section was to assess the ability to decode and process word and sentence meanings. As such, a multiple-choice format, composed of independent, multiple-choice items was deemed appropriate. In contrast, the Comprehension elements of the test booklets were intended to assess pupils' ability to construct meaning from a piece of text, rather than simply the ability to process word meanings. Thus, "test units", whereby a stimulus (e.g., text/diagram/table) was accompanied by a set of items related to the content of the stimulus were deemed the most appropriate test format.

In NAER 2004, First class test booklets used multiple-choice only, while Fifth class booklets contained a mixture of multiple-choice and constructed response items (i.e., where pupils are asked to write their answer, whether a word or a sentence). For 2009, constructed response was judged unsuitable for Second class, as many pupils might not have developed sufficient writing/spelling skills to demonstrate their knowledge on constructed response items. However, it was agreed that for Sixth class, the use of the constructed response format might facilitate the assessment of higher-level interpretative and evaluative skills. Thus, all Second class Comprehension units contained multiple-choice items only, while at Sixth class, units contained a mixture of multiple-choice and constructed response items. As a broad guideline, the framework suggested that approximately two-thirds of items for Sixth class test materials should

be in multiple-choice format, and one-third in a constructed response format, and that constructed response items be scored solely on the appropriateness of the answers, rather than on the quality of the writing or spelling.

Specifications for Item Type and Process

Based on the content of Second class textbooks, it was decided that the Second class tests for NA 2009 would mainly assess the retrieve and infer processes, with a lesser emphasis on interpret. Given that Second class pupils had limited exposure to class texts requiring evaluative processing, it was decided that the tests should not contain any evaluative items. In contrast, it was decided that Sixth class test materials should mainly target the interpret and evaluate processes, but also include items assessing the retrieve and infer processes (so that the full range of pupils' reading skills could be reported).

As noted, Second class English texts were mainly composed of reading for literary experience, while Sixth class texts reflected a mixture of literary and informational purposes. However, from an assessment viewpoint, both are important processes and both should be familiar to pupils, particularly as informational texts feature heavily in many other subjects. Thus, it was decided that items should reflect a relatively even split between both processes. Combining the framework, information about pupils' texts, specifications from NAER 2004, and the 2009 assessment requirements, the item specifications shown in Tables 3.2 and 3.3 were produced as a guide for item development for NAER 2009. Equivalence between text types in NAER 2004 and reading purposes in NAER 2009 was assumed, as was equivalence between process categories in the two years.

Table 3.2: Table of specifications – numbers of items by reading purpose and process – Second class

Section	Processes	Purposes		Total	Item format
		Literary	Informational		
Comprehension	Retrieve information	30	30	60	All multiple-choice
	Make inferences	25	25	50	
	Interpret & integrate	20	20	40	
	Examine & evaluate	0	0	0	
Vocabulary	Core reading skills	—	—	20	
Total		75	75	170	

Table 3.3: Table of specifications – numbers of items by reading purpose and process – Sixth class

Section	Processes	Purposes		Total	Item format	
		Literary	Informational		Multiple -choice	Constructed response
Comprehension	Retrieve information	30	50	80	113	57
	Make inferences	30	20	50		
	Interpret & integrate	20	10	30		
	Examine & evaluate	5	5	10		
Vocabulary	Core reading skills	—	—	20	20	0
Total		85	85	190	133	57

Test Development

Subsequent to the first meeting of the Working Group, a short review of pupil textbooks was conducted. As well as content, the layout and general presentation in textbooks, and passage lengths were reviewed. Findings from this exercise, in conjunction with the content of the PSEC were used to identify some key differences between the target grade levels in NAER 2004 and 2009.

Subsequent to the review, a number of changes to the design of the test materials were agreed. First, the intended structure (test units) for assessing comprehension introduces a clustering effect. Thus, pupil scores on a given unit tend to cluster together, relative to scores for other test units. This can become problematic if a test booklet is limited to a small number of large units. It is more efficient, statistically speaking, to have many small test units, as this reduces the inter-dependence of items. Some of the test booklets in NAER 2004 contained inter-related test units, while others had units in which the stimulus extended over three pages (at Fifth class). This had the effect that a) most stimuli had a large number of associated items, b) pupils were exposed to relatively few units (and topics), and c) there was a significant amount of item clustering. For 2009, efforts were made to reduce clustering by increasing the number of units and reducing their average length. However, there is a limit to how short a stimulus that assesses reading comprehension can be. Thus, upper and lower boundaries (less than one page, but with sufficient content to generate at least six related items) were set for stimuli.

Second, it was agreed that the test materials should be made more visually appealing to pupils, and the presentation should be more like that typically found in children's literature and textbooks. For example, colour booklets were used for the first time in 2009. Previous assessments had been constrained by the retention of blocks of test items from previous years, meaning that some materials had become slightly dated in appearance, but could not be modified. However, as the 2009 assessments were collecting baseline data, it was possible to design the materials to maximise pupil interest (thereby reducing the effects of boredom on test scores) and to use a style of presentation that was familiar to pupils.

Sourcing Units and Developing Items

Subsequent to agreeing broad changes to the test format, a Reading Working Group was constituted, composed of five experienced primary school teachers and four researchers at the Educational Research Centre. The group was asked to source suitable texts and to develop questions based on the texts. The following considerations were given as guidelines for text selections:

- level of interest evoked by text
- context
- familiarity with content/prior knowledge
- coherence of macrostructure/organisation (ease in computing relationships among successive words, phrases, sentences)
- coherence of microstructure
- number of bridging inferences required in a text
- word length/number of syllables per word and sentence length

- word difficulty
- grammatical complexity
- texts not to exceed 300 words at Second and 400 words at Sixth class (roughly equivalent to less than a page for each grade level, as the Sixth class text was presented in a smaller font size than the Second class text).

In writing items, the group was advised to:

- Focus on the main ideas and viewpoints
- Select items representative of the item processes described in the framework
- Generate between 6 and 12 items per text.

The group met on a regular basis to review texts and test items. Among the main criteria for selection were the links between the text and the framework, whether or not the topic was covered in other texts, the "appeal" of the text to the target age group, and the readability of the text. Text sources included narrative and information books for children, magazines and newspapers, and the Internet. A small number of texts were original pieces written by group members. Some test units used in NAER 2004 and rated as "difficult" were considered for use in 2009 (e.g., a First class unit rated as hard for that cohort was examined for its suitability as a Second class test unit). Care was taken to ensure that as wide a range of topics as possible would be covered.

Each text and accompanying items from the large pool of initially submitted texts was reviewed in three phases. First, the text and items were reviewed and (if considered acceptable) revised by two ERC staff members, neither of whom had sourced the text. If the text passed this stage, it was discussed at a meeting of the Reading Working Group. As well as general appeal, and the possibility of developing sufficient items, texts were vetted for cultural fairness. For example, texts were examined for gender stereotyping or bias, and contexts were vetted to ensure that they were reasonably familiar to urban and rural pupils. Answers and possible misunderstandings were discussed, as was the appropriateness of distractor responses. During this second stage, considerable revisions were made to test items, with smaller revisions made to the source texts, and coding guides were developed for all constructed response items. Individual texts were also reviewed in an overall context, meaning that where two texts covered very similar topics, the "weaker" of the two was identified and dropped. The third phase of text and item selection involved administering the units to a small sample of Second and Sixth class pupils, and discussing each unit with the pupils afterwards. Subsequent to this, there was some further revision of items and units.

Item Pool for the Field Trial

For Sixth class, a final pool of 26 texts was agreed upon for inclusion in the field trial, with 209 associated test items (69 constructed response and 140 multiple-choice, mirroring the one-third/two-thirds split proposed in the item specifications). The pool of selected texts included adventure stories, brochures, biographical pieces, dictionary pages, DVD jackets, information pieces, timetables, and web pages. The texts (and the related items) were distributed over 5 test booklets, so that each pupil would be expected to read 5 or 6 texts, and answer about 40 questions. Forty items were

prepared for the Vocabulary section, each consisting of an underlined target word embedded in a low-context sentence, and four response options. Pupils were instructed to select the option that was closest in meaning to the target word. All pupils were asked to attempt the set of 40 vocabulary items, which was located at the start of each test booklet.

For Second class, 22 texts and 151 items (all multiple-choice) were included in the field trial. The texts included descriptions, lists, narratives, instructions, recipes, timetables and weather maps. As with Sixth class, the texts were distributed over 5 booklets with individual pupils asked to read either 4 or 5 texts (depending on length) and answer about 35 questions. Twenty items were also prepared for the Vocabulary section for Second class, with all pupils asked to attempt the full set.

Final Item Pool

As will be described in detail in Chapter 5 of this report, a field trial was conducted in 29 randomly selected primary schools in May 2008. At Second class, 1,191 pupils (92% of the selected sample) completed the reading assessment, while at Sixth class level 1,015 pupils (89.2%) did so. The results of the study were used to inform final item and unit selection. At Sixth class, average booklet-level difficulty levels ranged from .62 to .67 (i.e., the percentage of items answered correctly ranged from 62% to 67%). Thirty-seven items were dropped, either because they were too easy or too difficult, or provided poor discrimination. Differential item functioning (DIF) analyses revealed that none of the retained items was significantly gender-biased. The reduced unit and item pool was re-distributed across four new test booklets. As well as 5-6 texts and 42-44 items, each booklet contained a common block of 20 vocabulary items selected from the 40 used in the field trial on the basis that they provided a range of difficulties, good discrimination, no gender DIF and had an average difficulty of 0.65. Of the remaining 172 comprehension items, 60 (34.9%) were constructed response and 112 (65.1%) were multiple-choice – very close to the one-third/two-thirds split proposed in the item specifications.

At Second class, average booklet-level difficulty ranged from .59 to .69. Eighteen items were dropped because of problematic difficulty or discrimination, and the remaining items did not reveal any significant DIF bias. The reduced unit and item pool was re-distributed across four new test booklets. As well as 5 texts and 33-35 questions, each booklet contained 20 vocabulary items. Tables 3.4 and 3.5 show the remaining items classified by reading purpose and process for each grade level.

Table 3.4: Final item pool for reading by purpose and process – Second class

Section	Processes	Purposes		Total	Item format
		Literary	Informational		
Comprehension	Retrieve information	26	45	71	All multiple-choice
	Make inferences	25	16	41	
	Interpret & integrate	17	4	21	
	Examine & evaluate	—	—	0	
Vocabulary	Core reading skills	—	—	20	
Total		68	65	153	

Table 3.5: Final item pool for reading by purpose and process – Sixth class

Section	Processes	Purposes			Item format	
		Literary	Informational	Total	Multiple -choice	Constructed response
Comprehension	Retrieve information	35	48	83	112	60
	Make inferences	33	19	52		
	Interpret & integrate	21	8	29		
	Examine & evaluate	5	3	8		
Vocabulary	Core reading skills	—	—	20	20	0
Total		94	78	192	132	60

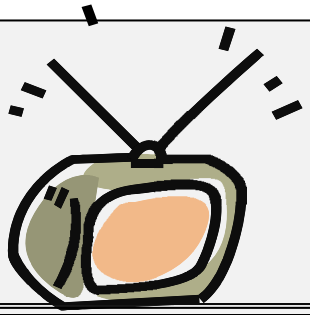
Sample Units and Items

This section contains examples of two units (one for Second class, one for Sixth) and associated items that were used as part of the main study in 2009. These units were selected because they demonstrate quite different types of stimuli, and contain items assessing a mixture of processes. For example, they include items assessing the retrieve, infer and interpret and integrate processes.

Second Class

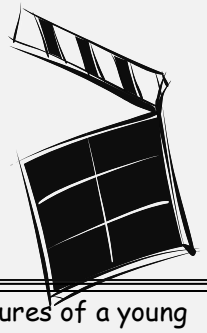
The unit "TV Timetable" – shown overleaf – is a released unit of Second class material. This unit was used in the main study in 2009. The stimulus is a non-continuous, informational piece of text with eight associated test items.

The eight items from the TV Timetable unit are shown in Figure 3.1. Five of the items assess the pupils' ability to retrieve information, while three assess their ability to infer the correct response. All items in this unit use a multiple-choice format, as none of the Second class items was presented in constructed response format. Readers should also note that the layout of the stimulus presented here is slightly different to that presented to pupils, while the layout of the test items differs considerably.



TV TIMETABLE

Saturday 20th January



08.00	**Arthur**	Animated series following the adventures of a young aadvark and his friends.
08.30	**Captain Planet and the Planeteers**	Animated series about a superhero out to save the environment with the help of the five planeteers.
09.00	**The Cobblestones**	Prehistoric cartoon fun with Terry Dactyl and Stacy Saurus.
09.15	**Yuck Yuck!**	Cartoon action with crime-fighting duo Ben and Belinda O'Brien, who take on the cases that become too yucky for adults to handle.
09.30	**Lucy McGurken**	Cartoon about a **junior** inventor who has all sorts of adventures with her best friend and sidekick Jamesie Woo.
10.00	**Cook 4 You**	Cookery series with cooks Dara and Alice. Together they run a special cafe where every day a different surprise guest calls in for a tasty treat.
11.00	**Freaky Friday**	An exciting movie where a mother and daughter wake up in each other's bodies after eating magical biscuits.

Figure 3.1: Example of items administered to Second class pupils, by process

Process		Item
Retrieve	**1**	At what time does 'The Cobblestones' begin? a) 08.30 b) 09.00 c) 08.00 d) 09.30
Retrieve	**2**	In which TV show would you find Jamesie Woo? a) Arthur. b) Yuck Yuck! c) Lucy McGurken. d) The Cobblestones.
Retrieve	**3**	'Cook 4 You' is about a) cooking magical biscuits. b) cooking for a special guest. c) a special guest cooking a meal. d) a cookery class for children.
Infer	**4**	'Cartoon about a **junior** inventor…'. The word **junior** means a) curious. b) young. c) fun. d) gross.
Infer	**5**	Which of these is a film? a) Yuck Yuck! b) Freaky Friday. c) Cook 4 You. d) Lucy McGurken.
Infer	**6**	Which show would you watch if you enjoy watching crimes being solved? a) The Cobblestones. b) Cook 4 You. c) Freaky Friday. d) Yuck Yuck!
Retrieve	**7**	Which of these programmes is the **shortest**? a) Cook 4 You. b) Yuck Yuck! c) Arthur. d) Lucy McGurken.
Retrieve	**8**	'Captain Planet and the Planeteers' lasts for a) 15 minutes. b) half an hour. c) one hour. d) over an hour.

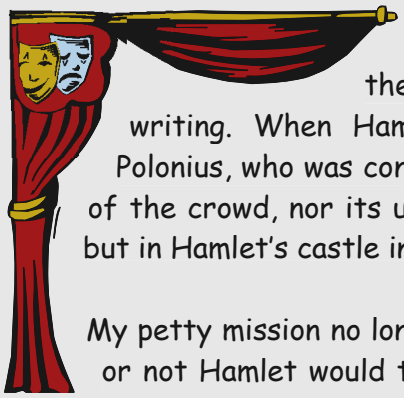
Sixth Class

The unit "Theatre Trip"[1] was included as part of the assessment of Sixth class pupils. The stimulus text is an example of a continuous piece of text, in which the main purpose is literary (in contrast to the informational purpose of "TV Timetable").

It is the 1600s. The writer is sent on a mission by his master, Falconer. He goes to the Globe Theatre in London to secretly copy Hamlet, a play by William Shakespeare, but finds some unforeseen problems...

Theatre Trip

I had been informed that, because many people considered acting to be an unsuitable occupation for women, they were forbidden by law to act upon the stage. All women's roles were played by men and boys. That fact did not occur to me now. I was totally convinced that the Queen and Ophelia were what they seemed to be. So drawn in was I by the events on the stage that it seemed less important to me to copy down the lines than to find out what these people would say or do next.



When the ghost of Hamlet's father appeared upon the balcony and called to him, I gasped but kept on writing. When Hamlet thrust his sword through the curtains, killing Polonius, who was concealed there, I was lost. I no longer noticed the press of the crowd, nor its unwashed smell for I was no longer there among them, but in Hamlet's castle in Denmark.

My petty mission no longer seemed to matter. All that mattered was whether or not Hamlet would take action to avenge his father. Every now and again, there was a passage of much talk and very little action, and I came to myself and quickly began to write. But eventually, I was drawn into the world of the play again, forgetting the world about me and the world outside, where Falconer waited.

From the start of the fencing match between Hamlet and Laertes until Hamlet's death, I believe I did not write down more than ten lines. I did get down every word of the last few speeches, but that was small comfort.

I had gone into the theatre fearful of being discovered and punished for writing down the play. I left with a dread of being punished for not having written it down. I need not have worried about being found out; no one in the audience or on the stage had paid the least attention to my writing.

[1] Thanks to The O'Brien Press Ltd for permission to adapt an extract from *The Shakespeare Stealer* by Gary Blackwood.

Three of the items relating to the Theatre Trip unit assess the pupils' ability to retrieve information, two assess their ability to infer the correct response, and two require pupils to interpret and integrate information.

Figure 3.2: Example of items administered to Sixth class pupils, by process

Process		Item
Retrieve	**1**	Why were women forbidden to act in plays? _____
Infer	**2**	The author forgot the uncomfortable conditions in the theatre because a) he was too busy writing down the words of the play. b) he was too interested in the events of the play. c) he was too afraid of being caught. d) he was too tired and hungry.
Infer	**3**	'I was totally convinced that the Queen and Ophelia were what they appeared to be.' What does the writer mean by this? _____
Retrieve	**4**	Which two characters had a fencing match on stage? _____
Retrieve	**5**	Which part of the play was the author most successful in writing out? a) The part where a ghost appears. b) The part where two men fence. c) The speeches towards the end. d) Ophelia's entrance onto the stage.
Interpret & Integrate	**6**	Why do you think members of the audience paid no attention to the writer copying down parts of the play? _____
Interpret & Integrate	**7**	Why do you think the writer's master wanted a copy of the play in writing? _____

Chapter 4
Sample Design, Weighting and Scaling

This chapter is divided into three main sections. The first section details the sample design for NA 2009. The second describes how weights were calculated. The third describes the procedures used in scaling the test items to produce estimates of item difficulty and pupil achievement.

Sample Design

The design of the sample for NA 2009 is similar to that of previous national assessments. However, whereas previous assessments have been undertaken separately for mathematics and reading, NA 2009 was designed to assess the same sample of pupils on both domains.

Survey Population

The primary sampling unit of NA 2009 was the school. The target population comprised all pupils in Second or Sixth class in ordinary classes in mainstream primary schools. This population excludes pupils in special classes in mainstream schools, as well as pupils in special schools and in private primary schools. The sampling frame for the NA 2009 main study was based on data for the 3,158 schools listed in the DES primary school database 2007-08, as this was the most up-to-date information available at the time the sample was drawn. Based on DES data for 2007-08 (DES, n.d.) pupils in ordinary classes in mainstream schools accounted for 97.0% of pupils in the class-relevant age range for Second class and 94.3% for Sixth class.

A total of 54 schools were excluded from sampling – one school had closed in June 2008, two were small schools with no pupils listed in either Second or Sixth class, and 51 were junior schools with pupils in Junior Infants through First class or Junior Infants and Senior Infants only. The remaining 3,104 schools were split into nine strata based on school size (the number of pupils in First through Sixth class) and whether or not the school had pupils at all classes, or was a junior or senior school. Stratification by school size gives more control over the sample size. Stratification by junior, senior or vertical school type was necessary as junior and senior schools could contribute data for only one class level. Table 4.1 shows the number of schools and pupils at each class level in each stratum.

Table 4.1: Numbers of schools and pupils at each class level by stratum in the sampling frame (based on DES enrolment figures for 2007/08)

Stratum		Schools		Pupils: 2nd class		Pupils: 6th class		Sample schools
		n	%	n	%	n	%	n
Excluded	No 2nd / 6th class pupils	54	1.7	0	0.0	3*	0.0	-
Junior	Small (<21)	62	2.0	356	0.6	0	0.0	1
	Medium (21-34)	10	0.3	253	0.4	0	0.0	1
	Large (35+)	56	1.8	4718	8.1	0	0.0	8
Senior	Small (<21)	49	1.6	0	0.0	200	0.4	1
	Medium (21-34)	7	0.2	0	0.0	211	0.4	1
	Large (35+)	54	1.7	0	0.0	4207	7.8	8
Vertical	Small (<21)	2025	64.1	21523	37.0	18361	33.9	48
	Medium (21-34)	527	16.7	14993	25.8	14077	26.0	36
	Large (35+)	314	9.9	16316	28.1	17056	31.5	46
Total	Total	3158		58159		54115		150

*As noted in the text above, one school (with 3 pupils in 6th class) in the 2007/08 DES listing was marked as having closed in June 2008.

Sample Size

A number of factors were taken into consideration in deciding on the number of schools (and pupils) to be selected. First there was a desire to ensure an effective sample size in excess of 400. The effective sample size refers to the size of the equivalent simple random sample. A simple random sample would mean that all pupils in the target population would have an equal chance of selection. However, such sampling is costly and impractical in most circumstances. The two-stage cluster sample methodology used in NA 2009 and in previous national assessments is more cost-effective but is less efficient in terms of the accuracy of population estimates derived from sample data. This is because clusters of pupils (in schools or classes) are selected, and pupils in these clusters tend to be more similar to each other than they are to pupils in the target population in general. The degree to which pupils in the same schools are more similar to each other than they are to the broader population of pupils is measured by a statistic called the intraclass correlation. The degree of clustering differs from variable to variable but has been found to be greater for mathematics than for reading (Cosgrove, Kellaghan, Forde & Morgan, 2000, Cosgrove et al., 2004; Shiel & Kelly, 2001; Shiel et al., 2006). Since rho (the statistic measuring intraclass correlation) for reading is lower than for mathematics, the estimates for mathematics guided the estimates for the required sample. As rho is lower for reading the effective sample for any given achieved sample will always be larger. Therefore, using the rho for mathematics sets a reasonable lower bound for the estimated effective sample size. Estimates for intraclass correlations were available from the 2004 national assessments (albeit at First, Fourth and Fifth class).

The effective sample size is important because it is directly associated with the accuracy of the survey estimates. An effective sample of 400 pupils will result in 95% confidence intervals of ± 4.9% for a percentage and ± 10% of the sample standard deviation for the reported mean. In terms of pupil performance, this means that we can be confident that sampling error on estimates of pupil achievement will be less than ± 5 scale score points, since domain scores are reported on a scale with a standard

deviation of 50 and a mean of 250. In other words, there is only a one in 20 chance that the true population mean is more than five scale score points above or below the sample estimate for the population as a whole.

A second factor determining sample size was the need to ensure that a sufficient number of pupils (800-1000) would see any particular item for the purposes of producing stable item statistics for the IRT scaling. As there was no overlap of content in the four reading test booklets at each class level (apart from the common 20-item vocabulary section) an achieved sample of approx 3200-4000 was required.

Third, two classes per grade (if there were two or more classes in the selected school) were to be selected to allow analysis of within school variation. This obviously has the effect of increasing the number of pupils selected (the cluster size) in medium and large schools. Again, this is the same procedure as was used in previous assessments. If within school differences were not a matter of interest in later surveys, and reliable estimates of item parameters exist, it may be possible to take only one class per school and reduce the actual sample size without a substantial reduction in effective sample size. The mean achieved number of completed test forms was estimated from the 2004 assessments of reading and mathematics.

Based on the three design criteria outlined above, a sample of 140 schools per class level (130 vertical schools plus 10 junior or 10 senior schools) was decided on, with 150 schools to be selected in total. The number of schools to be selected from each stratum was roughly in proportion to the number of pupils in such schools. Junior and Senior schools were a little less well represented in the sample than their pupil numbers would indicate. Table 4.2 shows the estimated achieved sample and estimated effective sample for a sample of 140 schools at one class level. The estimate of mean cluster size was derived from the mean number of completed reading test forms at Fifth class (Eivers et al., 2005). Rho, the intraclass correlation, was based on Fourth class data for mathematics as gathered in 2004 (Shiel et al., 2006).

Table 4.2: The estimated achieved and effective sample sizes for a domain at one class level based on the sample design

		n Schools	Mean Cluster n	n Pupils	Rho	Effective Sample
Senior	Small	1	12	12	0.27	3
	Medium	1	25	25	0.27	3
	Large	8	43	344	0.27	28
Vertical	Small	48	12	576	0.27	145
	Medium	36	25	900	0.27	120
	Large	46	43	1978	0.27	160
Total		140		3835		460

Sample Selection

The sample was selected using two-stage cluster sampling. At the first stage the appropriate number of schools was selected from each stratum. Within each stratum schools were sorted by four implicit stratification variables – SSP/DEIS status, area/language of instruction (Gaeltacht, Gaelscoil, Ordinary School), school gender

composition, and school measure of size (MOS). This sorting has two purposes. First, the sample selected is similar to the population in the stratum in terms of these variables. Second, replacement schools can be flagged (i.e., the following/preceding school) for use if selected schools do not participate. Any such replacement school is therefore similar to the non-participating school in terms of the implicit stratification variables. Schools were selected within strata using random start systematic sampling with probability proportional to MOS. MOS for junior schools was number of pupils in Second class, for senior and vertical schools it was number of pupils in Sixth class.

In schools with one or two classes at a class level all classes were selected. In schools with three or more classes two were selected (randomly, with equal probability) by the ERC from the class lists provided by the schools.

All pupils in selected classes were part of the target sample, except in a limited number of cases where a pupil's classroom teacher felt they should be exempted from testing. Pupils could be exempted for several reasons, e.g., if they had a learning or physical disability, or if they had less than one year's instruction in English and had limited language proficiency.

Sampling Weights

Although the sampling procedure outlined is intended to produce a representative sample of the population, bias can arise from two principal sources. The first source of bias occurs where schools (and therefore pupils) are sampled disproportionately with regard to their representation in the overall population. Although not a major feature of NA 2009, this type of sampling is common in national surveys where larger schools are selected more often than smaller schools due to the lower cost, or where certain sub-populations are over-represented because of a desire to focus on these sub-groups in later analyses. The second source of bias occurs due to non-response, i.e., where selected schools, or pupils within selected schools, do not participate. The size of the bias depends on the absolute size of the non-response, and the degree to which non-response is associated with the measure of interest (in this case the reading and mathematics test scores).

It may well be that low achieving pupils are more likely to be absent on the test day than higher achievers. Indeed, there is evidence from NAER 04 to show that this is the case (Eivers, Shiel, Perkins, & Cosgrove, 2005). Pupils who did not take the NAER 04 reading test were more likely than those who did to be rated towards the lower end of a number of ranking scales by their class teacher. This second source of bias is minimised by making efforts to ensure a high response rate, both at school and pupil level. At the pupil level the survey weights have the effect of replacing missing pupils with the mean score for pupils in the same class who participated in the testing. At the school level the effect of any non-response is addressed by replacing schools that opt out with replacement schools that are similar in terms of the sampling stratification variables. These replacement schools are selected as part of the sampling process. Otherwise the sampling weight has the effect of replacing the non-participating school with the mean score for pupils in participating schools in the same stratum. Both sources of bias are minimised by calculating sampling weights prior to analysis of the test data.

The procedure for the calculation of weights was the same at both Second and Sixth class. The weights for NA 2009 were calculated as follows:

Sample weight=n/N (sbw x scnr x cbw x pcnr)

Where:

n = number of pupils in the achieved sample

N = number of pupils in the population

sbw = school base weight (the inverse of the probability of the school being selected)

scnr = correction for non-response at the school level

cbw = class base weight (the inverse of the probability of the class being selected)

pcnr = correction for non-response at the pupil level.

Scaling

Just as representative samples of pupils were selected from the Second and Sixth class populations, so too were samples of items selected on which to assess pupil ability. These item samples were representative of the population of items that would be appropriate based on the respective mathematics and reading frameworks, and were designed and selected so as to be of appropriate difficulty. A feature of the test design was that pupils at a particular class level only saw a subset of the test items. The advantage of this approach is that a wider range of items can be used, thus improving the curriculum coverage and content validity, without overburdening pupils with very long tests. However, as pupils see generally different sets of test items, steps need to be taken to ensure that pupil scores derived from different sets of test items are comparable.

Comparability of results from pupils taking different test booklets was ensured firstly by the random assignment of booklets. Random assignment means that there should be no systematic differences between the ability levels of pupils taking any particular booklet. Second, all pupils within a grade level and domain were presented with a common set of items. In the case of reading, these were 20 vocabulary items presented at the beginning of the test. For mathematics, there were 20 common items in Second class and 25 at Sixth class, appearing as the second of three blocks. Third, the data were scored and scaled using the Item Response Theory (IRT) framework. IRT provides a difficulty estimate for each of the test items and an ability estimate for each of the pupils. Most importantly, the item difficulty and pupil ability estimates are on the same scale, and these estimates are not dependent on the ability levels of different sub-samples of pupils seeing any particular test booklet (having adjusted for any differences in the sample means and standard deviations).

IRT makes the assumption that a pupil's likelihood of getting an item correct is influenced only by their proficiency in the domain being assessed. The more able the pupil, the more likely a correct response. IRT models this relationship by using up to three parameters for an item. These are the item difficulty, the item discrimination, and the likelihood of a correct answer as a result of guessing. In the case of NA 2009 the items were scaled using difficulty and discrimination only, since the guessing

parameter is generally unstable and difficult to estimate, and the probability of doing well as a result of guessing alone is vanishingly small.

Figures 4.1 and 4.2 show how IRT models pupil proficiency and item difficulty for two items with item characteristic curves (ICCs). Pupil proficiency is shown on the horizontal access and probability of a correct answer on the vertical. The vertical columns show the proportion of pupils (split into 15 proficiency groupings) getting the item correct. Comparatively few pupils provide correct responses at the lowest proficiency levels. This proportion rises as proficiency increases. The sloping line, rising smoothly from left to right, shows the ICC – the IRT model of the probability of a correct response, given a specific proficiency. Item difficulty is defined as the point of inflection on the ICC, that is, where the slope of the ICC stops increasing. In the case of one- and two-parameter scaling this is the proficiency level where a pupil has a 50% chance of getting the answer correct. The item discrimination is the slope of the ICC, that is, the degree to which the item differentiates between pupils at different proficiency levels.

Figure 4.1 shows a moderately difficult item with a high discrimination. Pupils at the low end on the proficiency scale have a very low probability of answering the question correctly. However, as shown by the bars, some low proficiency pupils do get the item correct – perhaps by a fortunate guess in a multiple choice item, perhaps by knowing about a specific topic in a constructed response item. Pupils at the high end of the proficiency scale are not expected to get the question wrong, although some still do.

Figure 4.1: The item characteristic curve (ICC) for a moderately difficult item with good discrimination

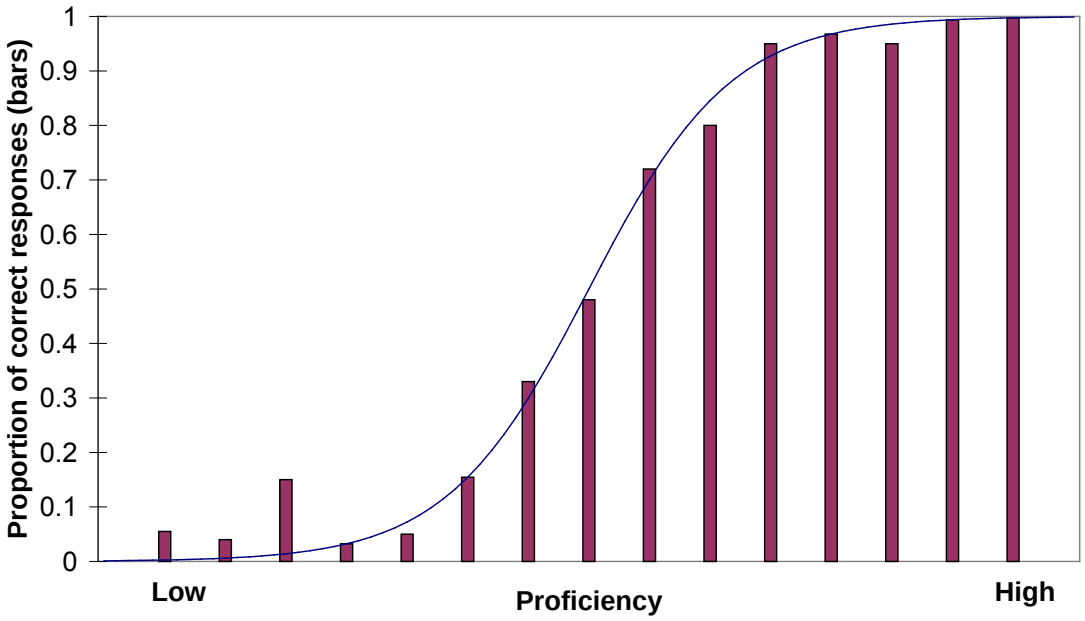
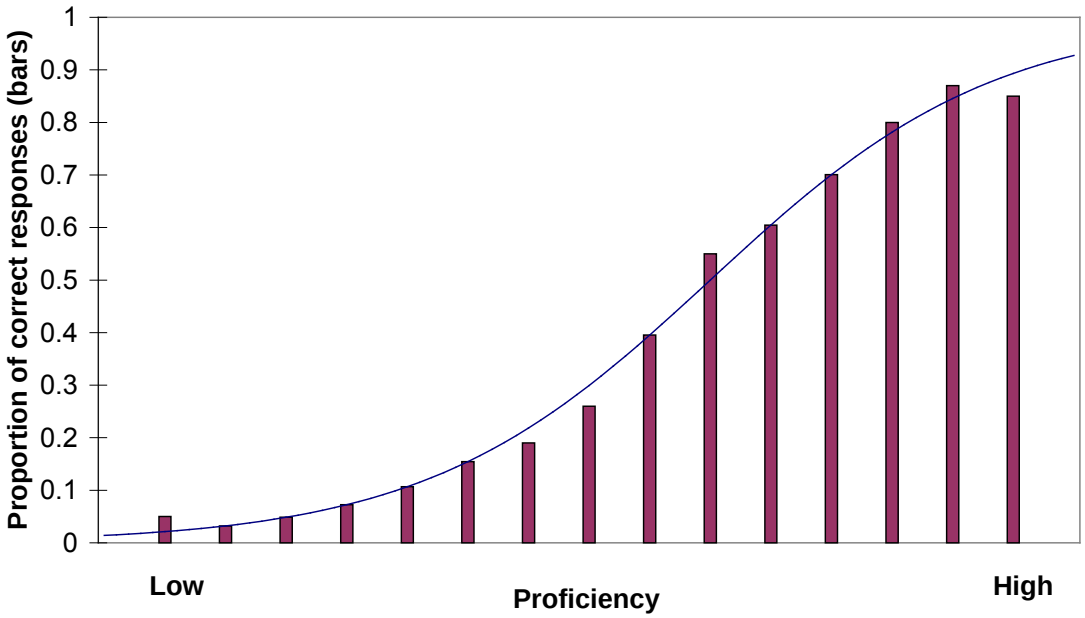


Figure 4.2 shows a more difficult item with a lower discrimination. We can see that the item is more difficult because the point at which pupils have a 50% chance of getting the item correct is further towards the "High" end of the proficiency scale. However, the item discriminates less well between pupils, as shown by the fact that the slope of the ICC is flatter. Getting the item right or wrong gives us less information about where to place the pupil on the proficiency scale.

Figure 4.2: The item characteristic curve (ICC) for a more difficult item with poor discrimination



All items, within each domain and level, were scaled on the basis of their difficulty and discrimination. All pupil scores, within each domain and level, were calculated on the basis of these item parameters. Item parameters and pupil scale scores are estimated iteratively within the IRT scaling software (BilogMG) with the final parameters and scores representing the "best fit" solution.

Percent correct scores and IRT scale scores were calculated for both domains at both class levels. As well as an overall test score, scores were created for reading purposes and processes, and mathematics content strands and process skills, as outlined in chapters 2 and 3. In line with the practice of previous national assessments, the IRT scale scores for overall test and each individual subscale were scaled to have a mean of 250 and a standard deviation of 50. The overall test scores are approximately normally distributed. The normal distribution is a symmetrical, bell-shaped, probability distribution which is often used to model data. It is particularly useful because the properties of the distribution are well-known (in part due to the fact that a great many variables measured in the social sciences are found to follow the normal curve). For example, when data are normally distributed about two thirds of cases lie within one standard deviation of the mean, and 95% of cases fall on or within two standard deviations. Thus, roughly 68% of pupils obtained test scores that fell between 200 and 300 (250 ± 50) and about 95% of pupils scored between 150 and 350 (250 ± 100).

The 2009 study was the first year in which the present National Assessment tests were administered. As baseline data, the 2009 results are the benchmark against which performance of pupils will be compared in future cycles.

Chapter 5
Field Operations

This chapter is divided into three main sections. Section one describes the field trial for the NA 2009, conducted in May 2008. The next section describes the implementation of the main study in 2009, and is followed by a section reporting reviews of the main study by DES inspectors and participating teachers.

Implementation of Field Trial

The following section describes the field trial for the National Assessments which took place in May 2008.

Overview of Field Trial

The aims of the field trial were to:

- gauge the appropriateness of the tests at both Second and Sixth classes (in terms of both difficulty and length).
- eliminate problematic items (due to gender bias, extreme difficulty/easiness, or other psychometric problems).
- gauge the appropriateness of the questionnaire measures.
- evaluate and refine administrative procedures.

The field trial was conducted during the last three weeks of May 2008. Administration of tests and questionnaires was carried out by classroom teachers. Test administration was observed by ERC staff and DES inspectors in seven of the 31 schools taking part, with comments on administration and procedures relayed to the ERC. Teachers were provided with a detailed manual, including a 'script' for administering the tests and questionnaires. Inspectors were asked to provide feedback on the assessment, and teachers completed a short form seeking their views on the timing and content of the tests and questionnaires. This feedback, together with achievement data, was used to guide revisions to the tests and questionnaires for the main study in May 2009.

Field Trial Sample

The field trial used a convenience sample of 32 schools randomly selected from all schools in Dublin city, Dún Laoghaire/Rathdown, Fingal, Co. Kildare and Co. Meath. Schools in these areas that were participating in other studies run by the ERC were exempted. All schools selected to take part were vertical (i.e., included both Second and Sixth classes).

Of the 32 originally selected schools, 29 agreed to take part in the field trial. Replacement schools were found for two of the three non-participants, giving a total school sample of 31. Of the 31 participating schools, four were all-boys, three were

all-girls, and 24 were mixed. Six schools had between 100 and 200 pupils, 10 schools had 200-300 pupils, 11 had 300-500 pupils, and four had more than 500 pupils. Six schools were designated as disadvantaged under SSP/DEIS.

Five reading booklets were piloted at both Second and Sixth class level. For mathematics, five booklets were piloted at Second class, and 10 booklets were piloted at Sixth class. As 200 responses per item were required to generate reliable item statistics, this meant that 1000 (5 X 200) participants per grade were required. Thus, 1294 pupils at Second class and 1138 pupils at Sixth were included in the full sample, to allow for non-response and ensure 1000 responses per grade.

The Second class sample of 1294 pupils comprised 639 boys (49.3%), 595 girls (46.0%), and 60 pupils for whom information on gender had not been supplied. The Sixth class sample contained 570 boys (50.1%), 513 girls (45.1%), and 55 gender-unspecified pupils. Just under one in five pupils were in SSP/DEIS schools (246 pupils (19.0%) at Second, and 198 pupils (17.4%) at Sixth class).

At Second class, a response rate of 91.7% was achieved for reading and 92.5% for mathematics. At Sixth class, the response rate was 89.2% for reading and 89.1% for mathematics. Among pupils attending SSP/DEIS schools the response rate was slightly lower, ranging from 77.8% (Sixth class mathematics) to 84.6% (Second class mathematics).

Field Trial Administration

Within each participating school, all Second and Sixth class pupils were invited to participate in testing for both the English reading and mathematics assessments. Parent Questionnaires were supplied to half of participating schools and Pupil Questionnaires to the other half. This reduced the administrative burden on schools and teachers, while allowing both sets of materials to be field trialled. Schools were advised of the test window (the period during which the administration of the tests should take place) and asked to advise the ERC of their chosen dates and the number of pupils at each grade level and in each class group.

All administration materials (test booklets, administration manuals, questionnaires and support forms) were posted to schools in advance of the agreed test dates. All booklets and questionnaires were pre-labelled with an ID code (a combination of the school and class IDs, and an individual pupil number), using the information on class size supplied by schools. Each teacher was provided with a Pupil List pre-filled with the ID codes for his or her class, and asked to assign IDs by filling in the names of participating pupils next to each ID on the list. They were asked to refer to the list to when matching the labelled IDs on all materials to pupils.

Teachers who received Pupil Questionnaires were advised that they should administer them to their pupils. Those who received Parent Questionnaires were supplied with sealed envelopes, with individually labelled ID numbers. Each envelope contained a letter to parents from the ERC, a labelled Parent Questionnaire, and a sealable return envelope. Pupils brought home the envelopes to their parents and parents were requested to complete the questionnaire, place it in the return envelope, and return it to their child's teacher. Due to the small sample size, School Questionnaires were sent to principals in all participating schools, and teachers of all classes involved received a Teacher Questionnaire. The content of the Teacher

Questionnaires were similar for Second and Sixth class, but contained some curriculum-specific differences.

Testing order (mathematics, then reading / reading, then mathematics) for each school was pre-assigned by the ERC. In 16 of the schools, the reading assessment was completed on the first morning of testing, with mathematics on the second; in the other 15, the order was reversed. Teachers were advised to encourage all pupils in ordinary classrooms to take part in the assessment, and to keep the number of pupils exempted to a minimum. Main exemption criteria included:

- pupils with less than one year's instruction in English and limited proficiency in English.
- pupils with moderate to severe learning disabilities.
- pupils with a physical disability that would prevent them from participating.

Class teachers administered tests and questionnaires, with the test administration observed (by ERC staff or DES inspectors) in seven of the 31 schools taking part. Teachers were provided with a 'script' for administering the tests and questionnaires, and advised that testing should take place over two mornings, with one full assessment completed on each morning. Participation rates were close to or in excess of 90% for all tests and questionnaires except the school and parent questionnaires (Table 5.1).

Table 5.1: Response rates for instruments used in the field trial

Grade		Tests		Questionnaires			
		Maths	Reading	Pupil*	Parent*	Teacher	School
Second	N	1204	1191	626	434	48	
	%	93.0%	92.0%	89.1%	74.9%	88.9%	26
Sixth	N	1017	1015	553	396	45	83.9%
	%	89.4%	89.2%	86.3%	71.4%	91.8%	

*Rates for the Pupil and Parent Questionnaires are based on approximately half the total sample at each grade.

After completion of the test booklets on the second day, the booklets, questionnaires, and support forms were collected and bundled together in class groups by teachers. The completed materials were returned to the ERC by the school co-ordinator.

Analysis of Field Trial Data

All test and questionnaire data were analysed, and all administrative procedures reviewed. For the test data, items were first scored, then examined using classical item analysis and Item Response Theory (IRT).

Development of Coding Guide

A provisional coding guide was used to score the open-ended items at Sixth class. Discussion and criticism of the provisional codes were encouraged, with between-coder differences used to inform a revised guide. More than 30% of open-ended responses in each booklet were double-coded, with approximately 10% of the total

response pool coded by three or more people. Differing interpretations were discussed as they arose, and a consensus agreed (with new examples added to the coding guide) before coding continued. Therefore, inter-rater reliability increased as coding progressed.

Analysis of Test Materials

Items were analysed using ITEMAN 3.5. Doing so resulted in some items being flagged as potentially unsuitable, based on one or more of the following criteria:

- low point-biserial correlation (the correlation between the correct response on a specific item and overall performance on the test).
- distractor responses displaying a positive relationship with overall test performance.
- extreme percent correct (for example, where less than 10% or greater than 90% answer correctly).

Subsequent to ITEMAN analyses, IRT was used to carry out several additional checks. These included examination of gender DIF (i.e., taking overall ability into account, does a significant gender difference on individual items remain?), and checks for group equivalence (if test booklets were distributed at random, there should be no significant difference in the performance of pupils by field trial test booklet).

The retaining or dropping of an item was not an automatic process. In each case where an item was flagged – either by statistical analysis or by teacher / inspector comments – it was carefully examined for logical and other flaws. The information from the analyses above was used to correct, where possible, any logical flaws with retained items, and to ensure that the selected items adhered to the overall framework specifications. Further details on the analyses of field trial items (by grade and by subject area) are available in Chapters 3 and 4 of this report. Similar logical checks were used to refine the questionnaire items.

Analysis of Administration Procedures

A review of administrative procedures was conducted on completion of the field trial. This review drew on observation reports from inspectors and ERC staff who had been present in some schools while tests were being administered, and on feedback from teachers and principals.

The ERC and the participating schools were generally satisfied with the administration of the field trial. However, some weaknesses were identified. For instance, teachers had to assign pupil names to a list of ID numbers, and then cross-reference the list with multiple test and questionnaire materials, in order to assign the correct material to the correct pupil. Many teachers felt that this method was unnecessarily labour intensive. ERC staff also identified this approach as open to unnecessary error. For example, there might be poor matching (by teachers) of pupils to IDs, or pupils might inadvertently mix up their parent questionnaires. A second problem which was noted was a low response rate for the teacher, school and parent questionnaires.

Implementation of the Main Study

Three main changes were made to the administration procedures for the main study:

- the use of pupil names or other identifiers in addition to an ID.
- questionnaires were sent well in advance of test dates.
- DES inspectors were assigned to all participating schools (to assist with the administration and the delivery and return of materials).

To reduce teachers' workload, schools were asked to provide a list of pupil names (or initials, or other identifier) by class group, in advance of testing. ID numbers were assigned to each pupil by the ERC, and all materials were pre-labelled with both an ID number and the pupil name (or identifier). This reduced the administrative workload for teachers, and greatly reduced the possibility that a pupil might accidentally complete another pupil's test booklet or questionnaire.

To address the issue of low response rates for some questionnaires, questionnaires for the main study were delivered to schools a number of weeks before the test dates. This allowed them to be completed at a convenient time, prior to the test days. Principals were also emailed an individualised link to enable them to complete an online version of the school questionnaire, should they so prefer. School co-ordinators were advised that their school would be assigned a DES inspector, who would be in attendance on both test days. The inspectors delivered and monitored the security of all test materials, dealt with school queries on the day, advised on administration, conducted review interviews with the assigned school co-ordinator, and collected all completed questionnaires. These changes led to a reduction in teacher workload, and increases in the response rates for questionnaires and the accuracy of supporting administrative forms.

Administration Procedures

Thirty-seven DES inspectors were assigned to the participating schools, to assist with the assessment and to act as quality monitors. All were briefed on the aims and procedures of the assessment, after which they contacted their assigned schools to confirm test dates and other arrangements. To ensure test security, test materials were not sent directly to schools, but were delivered to inspectors shortly before the start of the overall testing window (May 11th to 29th).

In each school, the ERC liaised with a designated co-ordinator. In mid-April each co-ordinator received all ancillary materials. These included an information booklet for the co-ordinator, a School Questionnaire, and a class pack for each participating class. The class pack contained a Teacher Questionnaire, sets of Pupil and Parent Questionnaires, and an Administration Manual (containing information on aspects of the survey aims, design and administration, including a 'script' for administering the tests and questionnaires). Teachers were asked to have all ancillary materials ready for collection by the inspector.

In each school, testing was conducted over two mornings. Half of participating schools completed the mathematics test first, while the other half completed English reading first (test order was pre-assigned by the ERC). At Second class, the mathematics test was read aloud by the class teacher, to minimise the effects

of pupil reading skills on mathematics performance. Thus, all pupils in a given Second class completed the same mathematics test booklet. In all other cases, pupils were randomly assigned different test booklets, with teachers and inspectors ensuring that the pupil completing a booklet was the pupil whose name was on the booklet label. Schools teaching through the medium of Irish were offered a choice of English or Irish language versions of tests (mathematics only) and questionnaires.

Each inspector made two visits per assigned school, during which they distributed the appropriate tests, oversaw the administration, assisting where appropriate, and collected all assessment and ancillary materials for return to the ERC. For each school, inspectors also conducted a short, informal review with the school co-ordinator, and completed their own review form. With the exception of the amount of work required by teachers (with which 12% of co-ordinators expressed some dissatisfaction), co-ordinators were very satisfied with the administration of the assessment. For example, only one person expressed dissatisfaction with the quality of the test materials, although almost 6% were dissatisfied with the suitability of the test materials for pupils *in their school*. Similarly, inspectors were universally positive in their reviews of teacher adherence to testing procedures and administration guidelines. However, teacher preparation was a relative weakness, as it was described as poor in most or all cases in over 4% of schools. Full details of the views of school co-ordinators and inspectors are available in the next section (*Review of the Administration*).

Response Rates for the Main Study

Generally, response rates for the assessment and ancillary materials were very high. Table 5.2 below shows the response rates for the main instruments used in the assessment.

Table 5.2: Response rates for instruments used in NA 2009 main study.

Instrument	2nd class (N=4199)		6th class (N=4189)	
	N	%	N	%
Maths Test Booklet*	3905	93.0	3832	91.5
Reading Test Booklet*	3839	91.4	3803	90.8
Pupil Questionnaire	3992	95.1	3979	94.9
Parent Questionnaire	3843	91.5	3847	91.8
	No. of classes = 202		No. of classes = 193	
Pupil Rating Form	200	99.0	191	99.0
Teacher Questionnaire	202	100.0	192	99.5
	No. of schools = 150			
School Questionnaire	N=149		% = 99.3	

* Data for tests refer to fully completed tests. Pupils who completed parts of the test booklets are not included.

The response rates for the School and Teacher Questionnaires reached almost 100%, while approximately 95% of selected pupils at each grade completed a Pupil Questionnaire. At 92%, response rates for Parent Questionnaires were slightly lower (but still very high), while test booklet completion rates ranged from approximately 91-93%. Response rates by sampled school strata for the reading and mathematics test booklets are shown in Tables 5.3 and 5.4 below.

Table 5.3: Response rates for NA 2009 main study, by school stratum, Second class

Stratum	Eligible pupils (N = 4199)	Achieved reading* (N=3839)		Achieved maths* (N=3905)	
		N	%	N	%
Junior school – Small (<21)	4	4	100.0	4	100.0
Junior school – Medium (21-34)	27	24	88.9	24	88.9
Junior school – Large (34+)	403	369	91.6	370	91.8
Vertical school – Small (<21)	614	576	93.8	574	93.5
Vertical school – Medium (21-34)	993	904	91.0	936	94.3
Vertical school – Large (34+)	2158	1962	90.9	1997	92.5

* Data for tests refer to fully completed tests. Pupils who completed parts of the test booklets are not included.

Table 5.4: Response rates for NA 2009 main study, by school stratum, Sixth class

Stratum	Eligible pupils (N = 4189)	Achieved reading* (N=3803)		Achieved maths* (N=3832)	
		N	%	N	%
Senior school – Small (<21)	3	3	100.0	3	100.0
Senior school – Medium (21-34)	41	36	87.8	38	92.7
Senior school – Large (34+)	385	354	91.9	356	92.5
Vertical school – Small (<21)	581	535	92.1	540	92.9
Vertical school – Medium (21-34)	953	861	90.3	873	91.6
Vertical school – Large (34+)	2226	2014	90.5	2022	90.8

* Data for tests refer to fully completed tests. Pupils who completed parts of the test booklets are not included.

At Second class, 733 pupils (17.5%) attended a SSP/DEIS school. Of these 673 (91.8%) supplied fully completed mathematics test booklets, and 666 (90.9%) completed reading test booklets, while 698 (95.2%) completed Pupil Questionnaires. At Sixth class, 711 pupils (17.0%) attended a SSP/DEIS school. Of these, 626 (88.0%) completed a mathematics test booklet, and 636 (89.5%) completed a reading test booklet, while 671 (94.4%) completed a Pupil Questionnaire.

Fewer than 2% of pupils overall were exempted from taking the tests (Table 5.5). Most of these pupils were exempted from both the reading and the mathematics tests. The most common reasons given for exclusion were a diagnosed moderate or severe general learning disability, or a limited proficiency in English (where the pupil had been receiving instruction through English for less than one year).

Table 5.5: Number and percentages of pupils excluded from one or both tests in NA 2009, by grade and by reason for exclusion

		N	%
Second (N=4199)	Diagnosed specific learning disability	6	0.1%
	Diagnosed moderate or severe general learning disability	20	0.5%
	Physical disability	1	<0.1%
	Limited proficiency in English	17	0.4%
	Other reason	6	0.1%
	Unknown	14	0.3%
	Total	64	1.5%
Sixth (N=4189)	Diagnosed specific learning disability	7	0.2%
	Diagnosed moderate or severe general learning disability	14	0.3%
	Physical disability	3	0.1%
	Limited proficiency in English	13	0.3%
	Other reason	10	0.2%
	Unknown	3	0.1%
	Total	50	1.2%

Review of the Administration

A two-level review of the administration of the National Assessments was conducted in almost all participating schools. Inspectors completed a review form for each of the schools to which they were assigned, outlining their views on the administration. They also conducted a short, semi-structured interview with the school co-ordinator in each school, to establish the views of the co-ordinator.

Inspector Reviews

Inspector reviews were received for 146 of 151 schools (96.7%). The first section of the review was a 4-point Likert scale, covering aspects of test administration. Reviews were overwhelmingly positive (Table 5.6). Teacher preparation for the test day administration was reported to be *good in all* or *most cases* in 95.7% of schools, while in 6 schools, most teachers were described as being poorly prepared for the assessments. The suitability of the test area, the testing atmosphere, and teacher adherence to timing guidelines were rated positively in all but two schools. For the remaining three areas rated – adherence to exemption guidelines, matching IDs to pupils, and the appropriateness of assistance given to pupils during testing – all schools received a positive rating.

In the second part of the review form, inspectors were invited to comment on the administration or to expand upon the ratings supplied in the first part of the form. Thirty-six of the 146 schools (24.6%) were explicitly praised for their co-operation, enthusiasm, and attention to detail (e.g., "very good school – all thoroughly prepared"; "teachers were very co-operative [and] conscientious"; "excellent administration"). In contrast, the level of preparedness in 14 schools (9.6%) was questioned. The most common criticism related to preparation was failure to read the instructions provided

ahead of time, followed by failure to adequately store the questionnaires and manuals between receipt of the materials and the days on which testing was conducted.

Table 5.6: Percentages of schools receiving various ratings from inspectors on the quality of administration of NA 2009

	Good in all cases	Good in most cases	Poor in most cases	Poor in all cases
	%	%	%	%
Matching name on booklet to pupil in classroom (N=145)	95.9	4.1	0.0	0.0
Appropriateness of assistance given (N=143)	91.6	8.4	0.0	0.0
Adherence to exemption guidelines (N=141)	89.4	10.6	0.0	0.0
Suitability of testing area (N=144)	83.3	15.3	0.7	0.7
Adherence to time guidelines (N=145)	82.1	16.6	1.3	0.0
Testing atmosphere (N=144)	80.6	18.1	0.7	0.7
Teacher preparation (N=142)	73.2	22.5	2.8	1.4

Minor issues on testing days were reported in 26 schools (17.8%). For example, one testing session was interrupted by noise coming from building works outside the school, and in another school two pupils were caught copying and were reprimanded. The bulk of the remaining observations (18.5% of schools) were typically general, rather than school-specific, comments on either or both of the reading and mathematics test booklets (e.g., regarding the quality of the materials, phrasing of questions, or the time allocated to testing). A small number (2.7%) commented that the pupil questionnaire was a positive addition to the assessment. Finally, some comments (7.5%) dealt with co-ordinators' views on miscellaneous aspects of NA 2009, such as the time of year of administration. These views are reported in more detail in the next section.

Views of School Staff

Inspectors conducted semi-structured interviews with a staff member in 147 schools (97.4% of participating schools). In the vast majority of cases, the interviewee was the school co-ordinator, while in a small number of cases it was the principal. As with the inspectors, staff were asked to provide ratings of specified aspects of NA 2009. They were also asked if they had any positive comments about the assessment, any negative comments, or any suggested changes that might improve future assessments.

Respondents were generally positive about their schools' participation in NA 2009 (Table 5.7). Ninety-five percent were *satisfied* or *very satisfied* with the helpfulness of ERC staff and the speed with which their enquiries were dealt, while the remainder had not dealt directly with ERC staff and therefore could not comment. The administration manuals and the quality of test materials both received very positive ratings (only one respondent was dissatisfied, in both cases). On the suitability of the tests, both in the school in question and more generally, and the suitability of the pupil questionnaires, 92-96% gave positive ratings. These ratings were more evenly split between 'satisfied' and 'very satisfied', and included 3-6% dissatisfied or very

dissatisfied responses. The highest dissatisfaction ratings were reported for the amount of work required of school co-ordinators (8.6% dissatisfied) and the amount of work required of class teachers (12.1% dissatisfied). However, even in these cases, a large majority of reviews were positive.

Table 5.7: Percentages of school co-ordinators reporting various levels of satisfaction with elements of their experience of participation in NA 2009.

	Very satisfied	Satisfied	Unsatisfied	Very unsatisfied	N/A Don't Know
	%	%	%	%	%
Helpfulness of ERC staff (N=140)	85.0	10.0	0.0	0.0	5.0
Speed with which queries were dealt (N=140)	85.0	9.3	0.0	0.0	5.7
Administration manuals (N=139)	78.4	20.9	0.7	0.0	0.0
Quality of tests (N=142)	66.9	32.4	0.0	0.7	0.0
Suitability of Pupil Questionnaires (N=139)	53.2	39.6	3.6	1.4	2.2
Suitability of tests generally (N=141)	48.2	47.5	2.8	0.0	1.4
Suitability of tests for pupils in your school (N=141)	45.4	47.5	5.7	0.0	1.4
Amount of work required of school co-ordinator (N=140)	43.6	47.9	5.7	2.9	0.0
Amount of work required of class teachers (N=141)	34.8	53.2	9.9	2.1	0.0

The next three tables summarise some of the additional comments made when asked about a) positive aspects of NA 2009, b) issues/problems with NA 2009, and c) suggestions to consider for future National Assessments. Many respondents offered no additional comments, while others had multiple comments. In Tables 5.8 to 5.10, the *n* shown represents the number of respondents who made at least one comment on a given category, while the percentage is based on all 147 respondents who were interviewed (i.e., not the subset who offered additional comments).

Table 5.8 summarises respondents' comments relating to positive features of NA 2009. The most frequent positive comment (29% of all respondents) was in relation to the high quality of the test materials – typically praising the appeal or pitch of tests. Sixteen percent felt that participation in NA 2009 was a useful experience for the school, in that it provided an external source of validation and assessment. Fourteen percent referred to the assistance provided by ERC staff, 13% commented that participating in a national study was a positive experience for the pupils, while 12% felt it was a positive experience for the school as a whole. Five percent were impressed by the use of a Pupil Questionnaire, and felt that the pupils' views on school, reading and mathematics would be interesting. Three percent commented that reading the mathematics test aloud (in Second class) was a good idea as it reduced the effect of reading ability on the mathematics assessment.

Table 5.8: Number and percentages of school respondents supplying positive comments about aspects of NA 2009.

Category	Sample comments	N	%
Tests appealing - attractive appearance / suitable pitch	*'A fair & true test generally'*		
	'Teachers commented on high quality of test materials'		
	'Scripts were modern, realistic, colour (6th class made very positive comments)'	43	29.3
	'The assessments 'looked' good with the colour illustrations, pictures and typeface'		
Opportunity for external review	*'Will be a good reference point in relation to the assessment of pupils' learning in reading & maths.'*		
	'Delighted that the school will receive feedback – this will be an additional, objective perspective re: attainment levels in numeracy/literacy in our particular school.'	24	16.3
Communication with ERC	*'ERC very supportive & quick in answering phone-calls'*	20	13.6
Positive experience for pupils	*'Pupils enjoyed it and it made them feel important.'*		
	'6th class pupils … showed extra interest in maths, particularly in the days leading up to the assessment'	19	12.9
Positive experience for the school	*'Delighted to participate in this critical area of research.'*		
	'Supports self-esteem of school to have been chosen.'	18	12.2
Pupil Questionnaire	*'Interesting comments by pupils during completion of pupil questionnaire'*		
	'Pupil q'aire was a good idea. It was good to get their views.'	8	5.4
Reading 2nd class maths tests aloud	*'Calling out the questions v. helpful (2nd maths)'*	4	2.7
Other	*'The professionalism shown by our assigned cigire'*		
	'*Liked the name "Maths survey" rather than "Maths test".*	11	7.5

Table 5.9 summarises comments about any issues or problems that arose during the assessment. The most common criticism related to the tests – either the length of time allocated or the difficulty level (31%). While a small number felt the tests were relatively straightforward, most criticisms related to them being too difficult, particularly the mathematics tests. Fourteen percent criticised the length or content of the questionnaires, some of which were reported to have been worded slightly ambiguously. Other concerns focused on the amount of work required of teachers in selected classes – either because of extra paperwork (11%), or because the assessment took place in May, which was described as a very busy time for schools (13%). Eleven percent indicated that there were no negative aspects to the assessment, while 7% indicated concerns related to excluding pupils, and 7% were unsure of the purpose of the assessment. Other comments related to difficulties accommodating exempted pupils or pupils in multi-grade classrooms (i.e., those not taking part in the assessment), and to the fact that test results would not be given to schools until September.

Table 5.9: Number and percentages of school respondents commenting on various problematic aspects of NA 2009.

Category	Sample comments	N	%
Tests – length / difficulty	*'2nd maths – sections quite difficult for middle and lower ability students.'*	45	30.6
	'The tests are quite long.'		
	'2nd reading – some questions very 'text heavy' (passages to read prior to answering questions).'		
Questionnaires - length / clarity	*'Some of questions on teacher questionnaire need clarifications.'*	20	13.6
	'"Negative consent" confusing for parents.'		
Dates for the assessment	*'Work not overly-onerous in itself, but added to very busy period.'*	19	12.9
	'The timing of the study is not good – end of year / Communion / school testing – better if done in 1st term.'		
	'Poor timing (May) – between Dec & Easter better.'		
No problems – testing went smoothly	*'Good to experience a different form of testing.'*	16	10.9
Amount of administrative tasks	*'Considerable workload for teacher in getting paperwork completed and administered.'*	16	10.9
	'Just another burden.'		
Issues related to exemptions	*'Child with SEN to be withdrawn if not participating? If withdrawn, may be a self-esteem issue.'*	10	6.8
	'The only issue I had was with absent & exempt students.'		
Lack of clarity on study rationale	*'The purpose of the testing?'*	10	6.8
Arranging multi-grade cover	*'Difficulty with accommodating pupils in dual-class situations'*	7	4.8
	'Splitting of classes (e.g. my 5th/6th class mix) – 6th class had to be in a room of their own for test and same with 2nd/3rd mix, so many class teachers and resource teachers had to relocate to accommodate testing. Maybe one class chosen in school would have been enough at one time and then the other class on a different day (practical issues that came to light during testing!).'		
Results delayed until September	*'No standardised results available for summer reports.'*	6	4.1
Other	*'It would be unfair if the same school was picked the next time.'*	14	9.5
	'Substitution cover would have been helpful, because of small scale of school.'		

Table 5.10 summarises suggestions for how future National Assessments might be improved. Some relate to issues already raised in Table 5.9. For example, the most commonly suggested change (22% of respondents) was that the National Assessments should take place earlier in the school year – although many different alternative dates were suggested. Sixteen percent made suggestions relating to the test administration and instructions, while 10% felt no changes were needed. Seven percent wanted

simplified questionnaires, while 7% wanted a better match between their school timetable and the questionnaire and testing sessions. Six percent suggested improvements to the mathematics tests, 4% suggested improvements to the reading tests, 3% wanted a reduced teacher workload, and 3% suggested more consideration for the needs of schools with multi-grade classes.

Table 5.10: Number and percentages of school respondents offering suggestions on how the implementation of National Assessments could be improved.

Category	Sample comments	N	%
Change the test window	*'Move the testing time to first term as in this school it coincides with communion & confirmation preparation'*	32	21.8
	'Not in May –end of February/early March'		
	'Timing – perhaps April/October – but aware this might yield different results.'		
Testing and administration issues	*'Make a rough-work column.'*	24	16.3
	'Guidelines should state questions will only be asked once.'		
	'More sample questions to help weaker child (average to low-average).'		
No suggestions – NA 2009 worked well	*'Very satisfied with study and involvement of parents.'*	14	9.5
Questionnaire issues	*'Readability of questionnaire to be made simpler.'*	11	7.5
	'Parent questionnaire –unclear about what was being looked for – seemed vague.'		
Length / timing of tests and questionnaires	*'Concentration levels – if length of test was broken into different chunks it could fit into school day better.'*	11	7.5
	'Very long for able children who complete quickly – need to address this.'		
Suggestions for maths test	*'Maths – 2nd & 6th – few more easy sums at the start of the test to build up confidence.'*	9	6.1
	'A lot of calculator work which students weren't used to.'		
Suggestions for reading test	*'Make Section A Reading more visual.'*	6	4.1
	'2nd class – reduce the number of comprehension or vary style of questions.'		
Reduce workload for teachers	*'Teacher profiles are extremely detailed & time consuming.'*	5	3.4
Multi-grade classes	*'In a school like ours with mixed classes, it'd be good to have one class tested at one time.'*	5	3.4
Other	*'It would be useful if there were some communication between the WSE/Inspectorate and ERC – school had both within the same term. It worked out but did initially cause anxiety.'*	8	5.4
	'As a teaching principal with two classes I had to employ a substitute to teach 5th class. It might be a good idea for the ERC to cover that cost for the days involved.'		
	'A national study in a disadvantaged school can reinforce failure – more work involved in running the study in a disadvantaged school (getting questionnaires from parents, etc.). – study will to little justice to the good work done in the school.'		

Chapter 6
Proficiency Levels

Scale Scores and Proficiency Levels

NA 2009 test data are scaled to have a mean score of 250 and standard deviation of 50, normally distributed. Both pupils and test items can be located on this continuum of achievement (the ability scale). Thus, a pupil whose ability estimate is located at a certain point on the reading or mathematics scale (e.g., a pupil with a mathematics scale score of 225) will probably be able to answer test items whose difficulty estimates place them below that point on the scale.

The further the item difficulty estimate is below that point, the greater the probability of the pupil answering correctly. Conversely, the pupil will probably not be able to answer items whose difficulty estimates place them above that point on the scale, and the further the item difficulty estimate is above that point, the lower the probability of the pupil answering correctly[1]. Because NA 2009 puts both pupils and items on the same scale, we can examine the properties of test items and use them to describe the skills or abilities of pupils at various points on the scale. Put simply, item X assesses skill Y. If a pupil is above item X on the continuum, it is likely that the pupil can demonstrate skill Y.

Proficiency levels consist of sets of items that pupils of a certain ability are likely to answer correctly, so that the continuous scales for items and pupils are grouped into discrete levels. This facilitates description of the skills likely to be demonstrated by particular clusters of pupils. They represent another way of presenting achievement data, and are useful educational and policy tools because they permit a more practical understanding of what it means to fall into a score range on a test.

Developing Proficiency Levels

Proficiency levels were developed using a PISA-style methodology (Programme for International Student Assessment) (OECD, 2009). However, the implementation differed from PISA in two regards. First, PISA scaling uses a 1-parameter model, where items differ only in terms of their difficulty. NA 2009 scaling used a 2-parameter model, where items differ in terms of both difficulty and discrimination. Second, PISA defines the bottom of a proficiency level as being where a pupil gets a score of 0.5 – i.e., where a pupil at a specific proficiency level would answer 50% of the questions at that level correctly. NA 2009 set the bar a little higher, at 62.5%. In other words, a pupil would be assigned to a proficiency level only if it were estimated that they would answer correctly at least 62.5% of the questions at that level.

[1] Assuming items with equal (or similar) discrimination values.

The development of proficiency levels begins by defining thresholds that correspond to specified proportions of pupils at or above each level. The number of levels and the proportion of pupils at each level will be based on a number of factors. These factors may include pragmatic reasons (meaningfulness and ease of interpretation), current policy, previous research, or a combination of these.

In order to allow useful comparisons, and since the same pupils were taking the two tests at a particular class level, it was decided to have the same number of proficiency levels for mathematics and reading. After preliminary analysis, taking into account item difficulties and the skills required to answer items correctly, and given the number of items available, it was determined that the number of proficiency levels it would be practical to define was five.

The simplest method of defining five proficiency levels is to divide them into quintiles, each accounting for 20% of the sample. However, the DES *Learning Support Guidelines* note that "In selecting pupils for diagnostic assessment and supplementary teaching, priority should be given to those pupils who achieve scores at or below the 10th percentile" (DES, 2000, p. 58). Pupils performing in the bottom 10%, as measured by standardised tests of mathematics or reading, are regarded as having insufficient levels of numeracy or literacy, relative to other pupils at their grade level. As the Guidelines (and day-to-day practice in schools) target pupils at or below the 10th percentile, the distribution of pupils across proficiency levels was designed to reflect this.

The distributions for mathematics and reading scores for both class levels were very close to being 'normal'. The normal distribution is a symmetrical, bell-shaped, probability distribution which is often used to model data. It is particularly useful because the properties of the distribution are well-known and because a great many variables measured in the social sciences are found to follow the normal curve. The lowest-achieving level is labelled Below Level 1, since the 10% of pupils in this category were not able to answer some of the easiest items on the reading and mathematics tests. Having determined the proportion of pupils in the lowest category, the other levels were defined symmetrically, as shown in Table 6.1.

Table 6.1: The percentage of pupils and approximate scale score range for each NA proficiency level

Level	% of pupils[1]	Scale score[1]
Level 4	10%	315+
Level 3	25%	270-314
Level 2	30%	231-269
Level 1	25%	187-230
Below level 1	10%	≤186

[1] These figures are correct for a normal distribution and are close approximations for the NA 2009 scales, given that the score distributions were not perfectly normal.

After defining the proficiency levels in terms of proportions of the normal distribution and the minimum raw score for a pupil to be eligible for a level (62.5% correct) the items were allocated to the levels using an algorithm developed by Dr Fernando Cartwright of Statistics Canada, who advised on the development of proficiency levels.

The algorithm estimates the item success probability at each of the thresholds. Beginning with the lowest proficiency level, the items are sorted, and the item probabilities are summed to produce an expected test score for a student at the bottom threshold of a proficiency level. Items will continue to be added to the current level until the minimum score specified is reached. Once that threshold is reached, the routine then goes to the next threshold, and again finds the selection of test items (not including the previously assigned items) that will allow the person at the bottom of the level to get a raw score equal to the specified minimum. This process continues until the top level is reached and all remaining items are assigned to it. Table 6.2 shows the number of test items falling into each proficiency level.

Table 6.2: The number of mathematics and reading test items at each proficiency level by class level

	Maths		Reading	
	2nd class	6th class	2nd class	6th class
Above Level 4	–	6	–	1
Level 4	22	26	22	23
Level 3	36	59	11	33
Level 2	17	39	110	79
Level 1	24	19	10	56
Total	99	149	153	192

Proficiency Level Descriptors for Reading

Subsequent to the establishment of four distinct proficiency levels (and a below Level 1) for Second and Sixth class, proficiency level descriptors were developed. The clusters of items exemplifying the skills that pupils at each Level had mastered were examined. For each Level, a short description was developed of the skills that pupils at that Level might be expected to demonstrate. This section describes how those descriptions were developed.

In NA 2009, each mathematics item was linked to a specific curriculum objective, and the collated objectives were used in developing proficiency level descriptors. For Comprehension reading items, no such specific objectives existed, and, due to the use of "test units", items could only be interpreted in conjunction with the stimulus unit text. Thus, the approach used to develop reading proficiency level descriptors was quite different to that used to develop mathematics proficiency level descriptors.

Although the scaling process created difficulty scores for each item, and these could be used to rank order items by difficulty, they provided no descriptive information about the skills that the items assessed. Further, items could be classified by the process (e.g. Retrieve) used to answer the question asked, and by the type of response required (e.g., multiple-choice or constructed response). However, items displaying these broad distinctions appear at each proficiency level. Thus, it was necessary to examine items by process type, and to distinguish levels of difficulty within each process. Each item was rated on the level of difficulty and the underlying skills assessed. Item ratings were grouped by proficiency level, and the collated ratings used to generate proficiency level descriptors. Particular prominence was given to features that distinguished a Level from adjacent Levels.

General Characteristics Related to Item Difficulty

There are a number of general characteristics that were considered when assessing item/stimulus text difficulty, summarised in Table 6.3.

Table 6.3: General item characteristics considered when rating item difficulty

Phrasing match	Items phrased in a similar manner to the relevant section of the stimulus text are easier than items where phrasing differs significantly. While this was especially relevant for Retrieve items, it also applied to items assessing other processes.
Amount of information required	Item difficulty increases in parallel with the amount of information required in order to establish the answer. If the reader must access information from multiple points in the text, this increases the complexity of the task, particularly if sequencing is important.
Conflicting information	The presence or absence of potentially conflicting information is important. For example, a reader may have to rule out potential responses in order to identify the correct answer.
External knowledge	Items requiring the reader to draw on their own experience – i.e., knowledge that is external to the stimulus text – are typically more difficult.

Rating Retrieve Items

As noted, although items assessing the Retrieve process are, overall, the easiest type of item, there was considerable variation within Retrieve items on the level of difficulty. A three-level rating of difficulty was applied. The lowest difficulty level required the item to be a literal match to the stimulus text, and to be located in a very small area of the text (Table 6.4). An item was rated as a medium difficulty Retrieve item if only one of those two criteria (literal or local) applied, while high-level difficulty ratings were assigned to Retrieve items that required the reader to search either the whole text or large parts of the text, while using different phrasing or requiring cross-checks to rule out other potential answers.

Table 6.4: Three-tier classification of item difficulty for Retrieve items

Low	Literal AND Local: The phrasing of the item is a literal or almost literal match to the phrasing used in the text **AND** the item requires the reader to search only a very localised section of the text – typically, an adjacent sentence or two.
Medium	Local: the item requires the reader to search only a very localised section of the text. However, the phrasing is not a literal match to the text. In some cases the phrasing may be reversed – e.g., *"Which is NOT…"* **OR** Literal: The phrasing of the item is an literal or almost literal match to the phrasing used in the text. However, establishing the answer may require more than a localised search of the stimulus text, or checking multiple sections.
High	Whole text / large section(s) of text: The reader may have to review a large part or all of the stimulus text. S/he may have to re-read sections to rule out certain options and to establish that the answer being searched for does not also appear elsewhere.

Rating Infer Items

Infer items were also classified using a three-tier level of difficulty (Table 6.5). Items classified as at the lowest difficulty level required the reader to make a simple inference, based on a localised section of text. Medium difficulty items required the reader to use more information, including information that was not explicitly stated in the text. The most difficult items were those that required the reader to use multiple pieces of information or to process the text at a global level, or to link local and global meanings.

Table 6.5: Three-tier classification of item difficulty for Infer items

Low	Most or all of the information required to answer the item is explicitly stated in the text. The answer can be inferred from a small section of text (e.g., a sentence), and the amount of information required is very limited.
Medium	The reader must link two discrete pieces of information in the text to establish the answer. Although at least some of the relevant information is explicitly stated in the text, the link between them is not, and must be inferred by the reader.
High	More than two discrete pieces of information or text are required to establish the answer **OR** The reader must draw on broader theme(s) in the text, and information at global text level to establish the answer

Rating Interpret & Integrate Items

Unlike Retrieve and Infer items, Interpret & Integrate items may require the reader to draw on their own experience and interpret the text from a personal perspective. While the information needed to answer Retrieve and Infer items typically resides in the text itself, Interpret & Integrate items may involve constructing meaning from a mixture of text and personal knowledge. As with Retrieve and Infer, Interpret & Integrate items were classified using a three-tier level of difficulty (Table 6.6). Low difficulty items were those where the answer was found in a small part of the text[2] while medium difficulty items required a more global analysis of the text, and some incorporation of outside knowledge. High difficulty Interpret & Integrate items were those requiring global analysis of the text, as well as considerable integration of personal knowledge.

Table 6.6 Three-tier classification of item difficulty for Interpret & Integrate items

Low	The answer is found in a small part of the text. The reader needs only to identify an idea or a theme in the text from a localised section of text.
Medium	Global meanings, require to discern main ideas and overall themes. Some outside knowledge needed.
High	Global meanings, and considerable integration of personal knowledge or experience with text content

[2] Items were also classified as low difficulty if the answer was apparent in many places or throughout the text, but where reading a small part of the text was sufficient to identify the answer.

Rating Examine & Evaluate Items

No Second class items assessed the reader's ability to Examine & Evaluate, while only a small number of Sixth class items assessed the process. Consequently, there are only two difficulty levels for this process – basic and advanced (Table 6.7). Basic level Examine & Evaluate items assessed pupils' ability to identify rationales or explanations, where clearly flagged in the text. Advanced Examine & Evaluate items required readers to engage in more complex evaluation (such as the "appeal" of a text) or to identify the rationale behind a text, where not explicitly outlined.

Table 6.7: Two-tier classification of item difficulty for Examine & Evaluate items

Basic	Identify the rationale behind a piece of text, where it is clearly flagged in the text, and where little external knowledge is required. For example, identifying a writer's main reason for writing a text, where the writer's motivation is self-evident. **OR** Deduce the most likely explanation for an outcome in the text where the explanation is clearly flagged in the text.
Advanced	Evaluate the structure and organisation of texts, or the clarity of information, the manner in which it is presented, the "appeal" etc, using external knowledge and opinions. **OR** Use external knowledge and viewpoints to evaluate a text, using much or all of the text in order to do so. **OR** Identify the rationale behind a piece of text, where extensive use of external knowledge, and global evaluation of the text required. For example, the text may cover one or more events/topics/people, but they may not necessarily be the main purpose of the text.

Process by Which Reading Items Were Rated

The process by which items were rated for difficulty was an iterative one. Initially, a small number of Second and Sixth class items were reviewed by two ERC staff working together, and each item was rated jointly, with clarifications added to the rating instructions where necessary. Next, all Sixth class reading items were rated separately by two ERC staff (only one of whom was involved in the preceding stage). In most cases, the same ratings were assigned to items. However, 36 of the 172 items (20.1%) in the Comprehension sections of the tests were either assigned different ratings or flagged as difficult to rate. In almost all cases, ratings that differed did so by one point only – for example, low versus medium difficulty, rather than low versus high difficulty. These items were reviewed, a rating agreed, and further clarifications made to the rating guide.

In the next phase, these jointly agreed ratings (and similar jointly agreed ratings for Second class) were compared against ratings provided by a third ERC staff member. Areas of differences were examined and resolved. Next, all items were rated by a primary school teacher familiar with the rating process, and these ratings were compared against the ratings supplied by the ERC. Differences were found for 11 Sixth class and 4 Second class items. Again, areas of differences were examined and resolved, and final difficulty ratings assigned to each item.

Developing Proficiency Level Descriptors for Reading

Once each item had been assigned a difficulty rating, ratings were aggregated by process and proficiency level to identify patterns of difficulty across Levels. For example, at Sixth class, Retrieve items at proficiency level 1 were overwhelmingly low difficulty items, while none of the Retrieve items at proficiency level 4 were low difficulty (Table 6.8). Similarly, half of the 10 Infer items exemplifying proficiency level 4 skills were rated as high difficulty items, while two-thirds of proficiency level 1 Infer items were rated as low difficulty items.

Table 6.8: Number of items at proficiency levels 1 and 4 (PL 1 and PL 4), by process, rated by difficulty

Retrieve			Infer			Interpret & Integrate			Examine & Evaluate		
	PL 1	PL 4		PL 1	PL 4		PL 1	PL 4		PL 1	PL 4
Low	20	0	Low	6	4	Low	4	0	Basic	3	0
Med	8	3	Med	2	1	Med	2	2	–	–	–
High	3	1	High	1	5	High	0	2	Adv.	0	1

Constructed response items were also examined for the type of response required. They were divided into 4 categories: Closed Constructed Response – straightforward retrieve; Closed Constructed Response – some deeper processing required; Open Constructed Response - straightforward retrieve; and Open Constructed Response – some deeper processing required. Items were described as Closed Constructed Response (CCR) items if only a single word or short phrase was required to answer the question. An example of this type of item would be "What was Mary's nickname?". Within CCR, items could require a straightforward retrieve (the word or phrase needed is retrieved directly from the text) or some processing. In contrast to CCR items, Open Constructed Response (OCR) items required a more detailed response. An example of this type of item would be "Why do you think that the writer described X?". Table 6.9 summarises the characteristics of constructed responses that exemplify proficiency levels 1 and 4. Proficiency level items were typically CCR, while Level 4 items were typically OCR.

Table 6.9: Number of items at proficiency levels 1 and 4 (PL 1 and PL 4), by type of constructed response

	PL 1	PL 4
CCR – straightforward retrieve	8	0
CCR – some processing required	1	0
OCR - straightforward retrieve	0	1
OCR – some processing required	2	6
Total	11	7

Information from the distribution of difficulty levels across the proficiency levels, and the types of constructed responses were used to develop descriptors of the skill set that could be demonstrated by pupils at each Level. For example, the preponderance of low difficulty Retrieve items at Level 1 suggested that pupils at this Level could answer questions where the phrasing of the item was a literal or almost literal match to the phrasing used in the text and where only a very localised section of the text needed to be searched. Conversely, the majority of medium and high difficulty

Retrieve items exemplified the skills of pupils at higher proficiency levels, indicating that Level 1 pupils could not consistently demonstrate these skills. Therefore, the skills required to answer medium or high level Retrieve items were included in the descriptors for higher proficiency levels. As well as summary difficulty ratings, individual items exemplifying proficiency levels were examined and used to clarify the distinctions between the proficiency level descriptors.

All descriptors for reading proficiency levels were developed and written by one ERC staff member. The descriptors were reviewed by two other ERC staff, with slight alterations made subsequent to their review. Next, they were examined by the external reviewer who had previously contributed to the item ratings, and finally, examined by the Reading Expert Group. The summary descriptors contained in the next section reflect the final agreed descriptions.

Final Proficiency Level Descriptors for Reading

This section outlines the final proficiency level descriptors, and the percentages of boys and girls falling into each proficiency level.

Second Class

Table 6.10 summarises the skills that Second class pupils at each proficiency level are likely to be able to demonstrate, based on a summary of the characteristics of items exemplifying the skills demonstrated by pupils at each Level.

Table 6.10: Proficiency levels on the reading scale and percentages of pupils (overall and by gender) achieving each level, Second class

Level & score range	What pupils can typically do	Total	Boys	Girls
4 **320+**	As well as lower proficiency level skills, pupils at level 4 can retrieve complex information (e.g., the information needed is located in multiple parts of the text). They can link multiple pieces of information to draw inferences. They can integrate text-wide information in order to identify the main themes in a text. As well as using discrete or explicit information, they can use the text as a whole to interpret character behaviour.	10	8	12
319 **3** **269**	As well as Level 1 and 2 skills, pupils can process texts at a whole-text level, in order to retrieve information. They can make basic-level inferences, sometimes linking one or two discrete pieces of information. They can infer word meanings if the context provides clear clues.	25	22	28
268 **2** **225**	As well as Level 1 skills, pupils can retrieve explicitly stated information where the wording of the question and the text differ. However, the information sought must be specific to a small section of text. They can make low-level inferences, including character motives, if the required information is explicitly stated in a specific section of the text.	30	29	31
224 **1** **187**	Level 1 pupils show basic reading skills. They can retrieve simple, explicitly stated, pieces of information, when there is a direct match between the wording of the question and the text. They are most successful on tasks that require comprehension of smaller units of text, such as sentences. They can perform some very basic interpretation and integration of text (e.g., identifying the theme of a text, where the theme is explicitly stated in the text).	25	28	22
<187	Pupils below proficiency level 1 have a less than 62.5% chance of correctly answering a Level 1 question. Their reading skills are very low, relative to other 2nd class pupils and are not properly assessed by the National Assessment.	10	13	7

Sixth Class

Table 6.11 summarises the skills that Sixth class pupils at each proficiency level are likely to be able to demonstrate, based on a summary of the characteristics of items exemplifying the skills demonstrated by pupils at each Level.

Proficiency Levels

Table 6.11: Proficiency levels on the reading scale and percentages of pupils (overall and by gender) achieving each level, Sixth class

Level & score range	*What pupils can typically do*	*Total*	*Boys*	*Girls*
4 **317+**	As well as skills exemplifying lower Levels, pupils at proficiency level 4 show advanced retrieval skills. They can find answers where the phrasing of the text and question differ considerably. They do not need to rely on explicitly stated information or connections, but can infer answers from multiple pieces of text, and use broad themes at whole-text level to infer an answer. They can evaluate the rationale behind a piece of text, even where the text covers multiple events/topics, and the overall rationale is not apparent unless analysed at a global level.	10	10	10
316 **3** **271**	As well as Levels 1 and 2 skills, pupils at Level 3 have complex retrieval skills. They can examine multiple elements of the text to locate the correct response and rule out incorrect responses. They can answer items where the phrasing in the text and question are not identical, and locate detail in dense texts such as advertisements or dictionaries. Pupils at level 3 have more strongly established inferencing skills (e.g., they are consistently able to link two pieces of information from a text to infer the correct response). They can interpret meanings at whole-text level, and integrate this with personal knowledge or experience, in order to identify a correct response. They can use opinion and external knowledge to evaluate arguments made, the clarity of information presented, or the structure and "appeal" of texts.	25	25	25
270 **2** **230**	Pupils at Level 2 can carry out multipart retrieval processes, such as answering questions that use a modified version of the phrasing in the text. They can also match question content with information in the stimulus text that extends beyond one or two adjacent sentences, provided that the question is an almost literal match with text content. They can combine two pieces of non-adjacent information in the text to infer a response, but their skills at this level are not consistent. They demonstrate integration skills such as identifying overall themes from texts, or drawing on outside knowledge.	30	28	32
229 **1** **183**	Pupils at Level 1 can carry out basic retrieval processes and can match words and phrases in the question with the same words and phrases in the stimulus text to answer items. They can also make low-level inferences, where at least part of the information required for the answer is *explicitly* stated in the text, or where a discrete piece of explicitly stated text coupled with very basic external knowledge is sufficient to answer the question. Pupils at this level can also engage in some interpretation and integration of information, such as identifying an idea or theme in a section of text. They can identify the rationale behind a piece of text where it is clearly flagged (for example, in the title).	25	25	25
<183	Pupils below proficiency level 1 have a less than 62.5% chance of correctly answering the easiest questions. Their reading skills are very low, relative to other 6th class pupils and are not properly assessed by this assessment.	10	12	8

Proficiency Level Descriptors for Mathematics

Mathematics achievement scales were divided into four proficiency levels based on the proportions of pupils falling into each level (Table 6.1) and items were allocated to the levels using an RP value[3] of 62.5 as the least likelihood that pupils at that level would answer items at that level correctly. Next, summary descriptions of the mathematical skills pupils at each of the four proficiency levels might be expected to demonstrate were developed.

Development of the Summary Proficiency Level Descriptions

The process started with an examination of the clusters of items associated with each proficiency level. Figures 6.1 and 6.2 are visual representations of the scales in which the items on the Second and Sixth class mathematics tests (identified by an ID code I001 to I100 for Second class and J001 to J1501 for Sixth class)[4] are located to the right of the two figures with the percent and cumulative percent of pupils at each scale point to the left. The scale, in logits, is shown on the far left of the figures. The items and proficiency levels are colour coded to show the clusters of items associated with each proficiency level. The vertical location of an item on the scale is determined by its difficulty index (the higher up the scale the more difficult the item) and the horizontal location by its discrimination index (the further right the item the lower its discriminatory power).

For example, the cluster of items belonging to Level 3 on the Second and Sixth class item maps are coloured yellow and the Level 3 section of the scale, containing pupil scores from +0.4 to +1.2 in logits or 272 to 315 in scale points[5], is also coloured yellow. As described earlier, a pupil whose score is at the bottom of this yellow band, i.e. a logit score of ~0.5 or a scale score of 272, is likely to respond correctly to at least 62.5% of the yellow group of items. A pupil whose scale score is further up the yellow band (e.g. +1.0 logits or 300 scale points) is likely to respond correctly to considerably more than 62.5% of the yellow items whereas a pupil whose scale score is in the gray Level 2 band, i.e. less than ~0.5 logits or 272 scale points, is likely to respond correctly to considerably fewer than 62.5% of the yellow Level 3 group of items.

The 99 Second class items and the 149 Sixth class items coded in Figures 6.1 and 6.2 were then listed in terms of their item descriptors (appendices A and B) as originally derived from the PSMC objectives, and subsequently ordered by scale score and by proficiency level. The following procedure was then applied to these lists of item descriptors to produce summary descriptions of the mathematical skills and knowledge pupils at each proficiency level for Second and Sixth classes are expected to demonstrate:

---

[3] The Response Probability value is the probability that a student at any proficiency level would get an item at the same level correct. The PISA studies use an RP value of 50%.

[4] Two items (one Second and one Sixth class) did not meet the requirements for scaling and were omitted from these listings.

[5] scale score = (logit score * SD) + mean. For Second class the mean was 249.945 and SD 51.404; and for Sixth class, the mean was 249.989 and SD 51.141

1. Any duplicate item descriptors *within* each of the four proficiency levels were deleted as only one was required to develop the summary description of the level. Since the items in the tests had been linked to the PSMC objectives – 43 objectives for Second class and 72 objectives for Sixth class – as per the NA 2009 framework for mathematics outlined in chapter 2, objectives were represented in the tests by one or more items, hence leading to duplicate item descriptors in the lists in Appendices A (Second class) and B (Sixth class).
2. Duplicate item descriptors *across* proficiency levels were then deleted. The item descriptor at the highest point on the proficiency scale was the one retained on the basis that any pupil who obtained a correct response to the item representing the descriptor at the higher level had a considerably better chance of a obtaining a correct response on the item representing the same descriptor at the lower level on the scale, whereas any pupil who obtained a correct response to the item representing the descriptor at the lower level would not necessarily have had an equal or better chance of obtaining a correct response to the item representing the same descriptor at the higher level.
3. Where two item descriptors within the same proficiency level were judged to describe two substantially overlapping or closely related item domains, for the sake of brevity the item descriptor higher on the scale was modified to encompass both items and the item descriptor lower on the scale was then deleted.
4. Finally, the item descriptors within each proficiency level were then grouped and ordered, in accordance with the mathematics framework, firstly by content strand (Number & Algebra; Shape & Space; Measure; and Data), and secondly by cognitive process category (Recall & Procedural Skills; Reasoning and Connecting Conceptual Knowledge; Applying & Problem-Solving). Descriptors relating to Applying & Problem-Solving were placed last in each level as this category generally included descriptors referring to content from all four strands[6]. These groupings were then integrated into short descriptive statements of the mathematical competencies pupils at each level are expected to demonstrate.

Inspection of the item descriptors and summary descriptions for each of the mathematics proficiency levels suggests the following general factors underpin the increasing levels of mathematical proficiency:

1) Kind of operation involved in a calculation (addition, subtraction, multiplication, division).
2) Size (single vs multi-digit) and type (e.g., fractions, decimals) of numbers.
3) Familiarity of word problem contexts (routine vs non-routine).
4) Number of steps/operations in a calculation or word problem.
5) Amount of reasoning and deduction involved.
6) Complexity of shapes (e.g., no. of sides/faces) in Shape & Space items.

[6] Some levels did not have item descriptors for all four content strands. At Second class, Level 4 had no Shape & Space item descriptors and Level 3 had none for Data. At Sixth class, Level 4 had no Data item descriptors. Also, for both Second and Sixth classes, items descriptors relating the cognitive process of Applying & Problem-solving are concentrated in Levels 3 and 4 with very few at Levels 1 and 2, which were generally routine one-step problems.

Figure 6.1: Map of Second class pupils and test items, ordered by difficulty (difficult to easy)

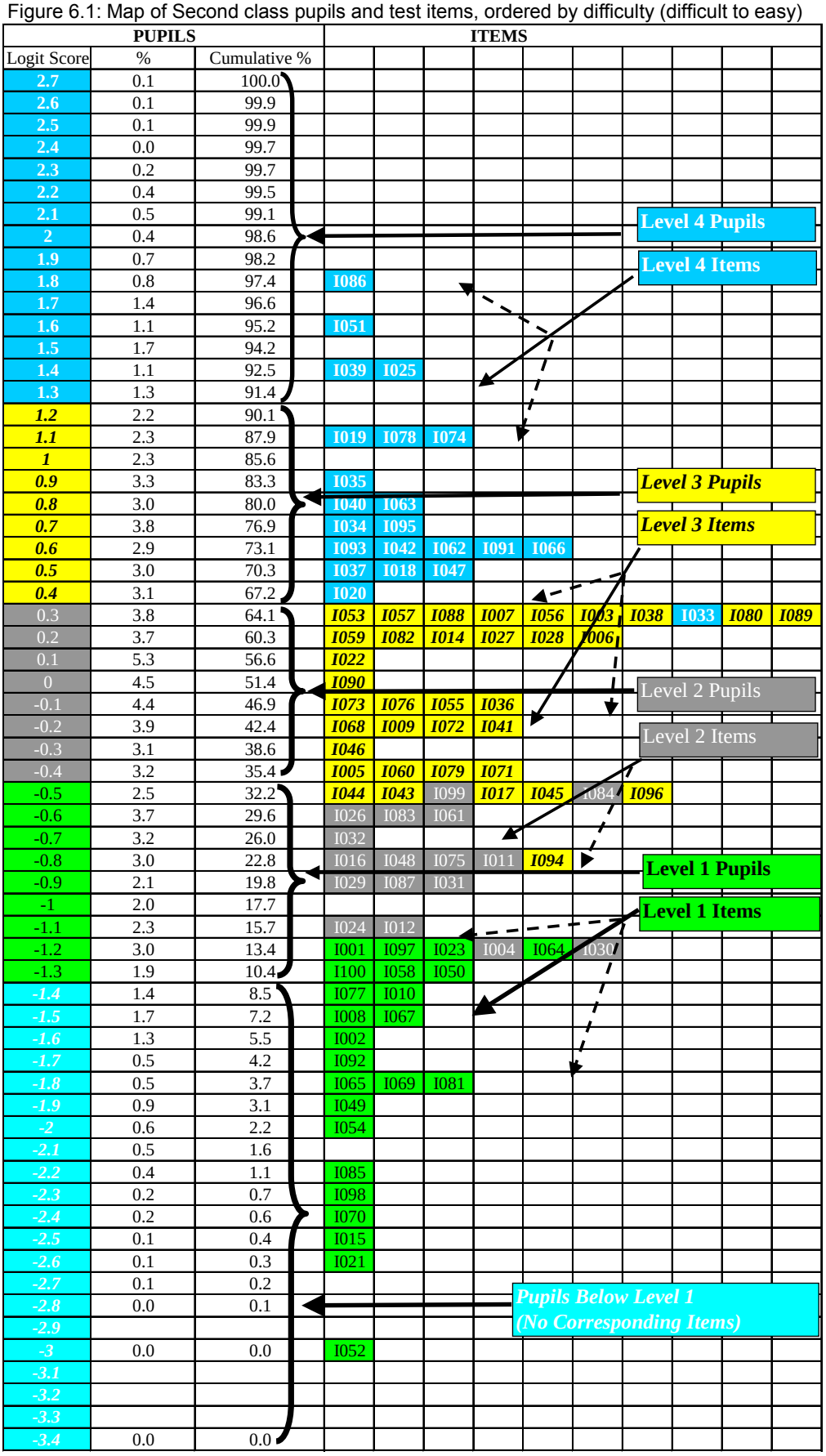
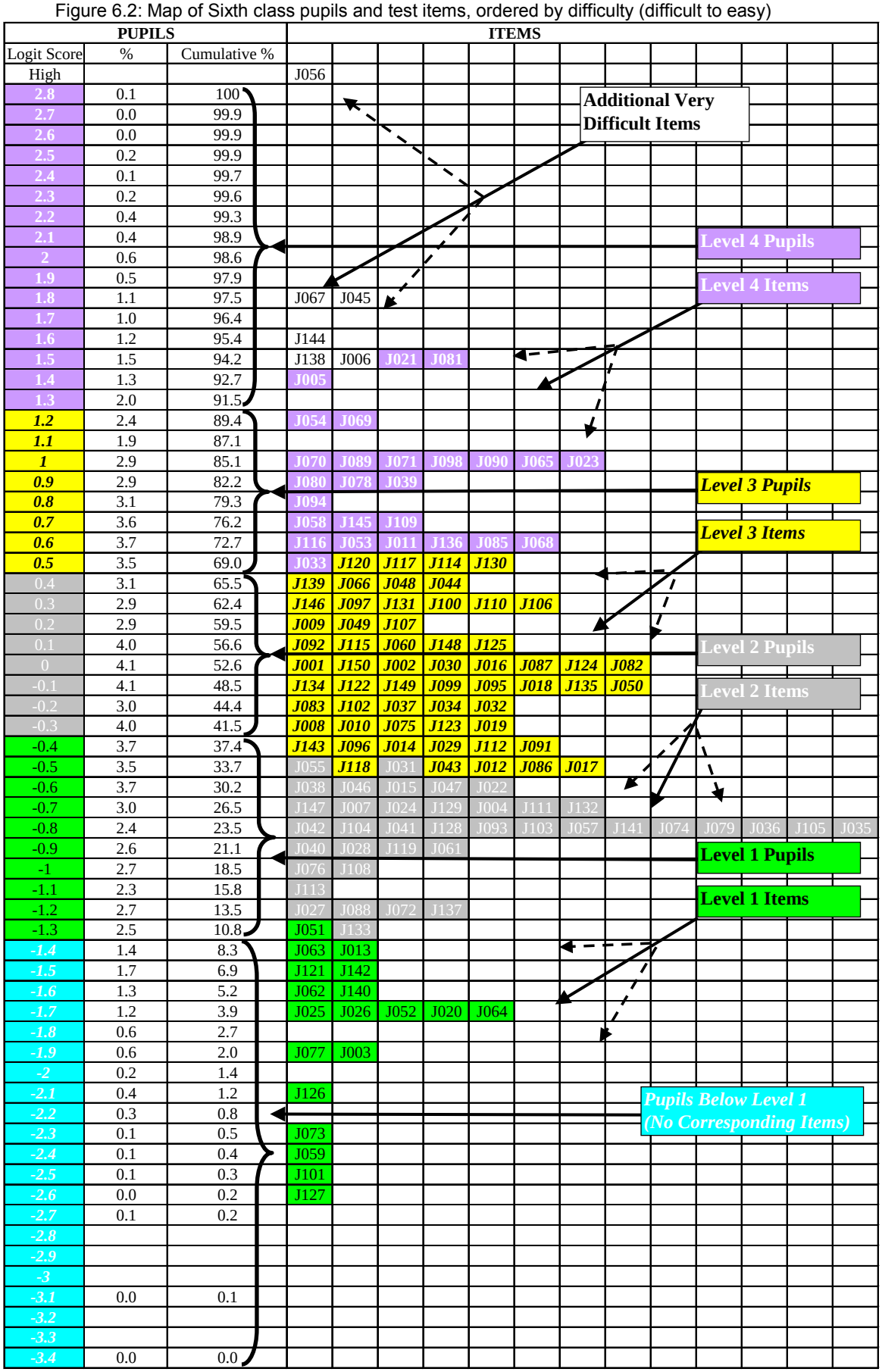


Figure 6.2: Map of Sixth class pupils and test items, ordered by difficulty (difficult to easy)



Validation of the Summary Proficiency Level Descriptions

To gauge how well the summary proficiency level descriptions reflected the mathematical knowledge demanded by the test items and the reliability of the procedure used to produce them, a validation exercise was carried out. One member of the Mathematics Expert Group (not involved in producing the proficiency level descriptions) examined the Second class proficiency level descriptions, and another member examined the Sixth class descriptions. Each was given copies of: (i) the item descriptors ordered by scale score and proficiency level (as per Appendix A and Appendix B); (ii) the items ordered by scale score and proficiency level; (iii) the summary proficiency level descriptors; and, (iv) a validation sheet to record the results of the exercise. Part of the validation sheet for Second Class is shown in Table 6.12.

Table 6.12 Section of Validation Sheet for Second Class item descriptor, item, proficiency level description links

Item Descriptor: Level 4	Code*for Matching Item	Comment on Match between Item and Descriptor	Located in Proficiency Level Description
Identify the best estimate of the difference between two two-digit nos.			
Measure area using a non-standard unit			
Connect a two-step word problem to a numerical expression			
Solve a two-step word problem on addition and subtraction of money			
Solve a one-step word problem involving repeated addition			
Read information from block graph and make a calculation with it			
Solve one-step word problem involving kg			
Solve non-routine word problem involving quarters of litres			
Calculate how many items may be bought with a given sum.			
Solve a two-step word problem involving addition and subtraction.			
Solve one-step word problem involving subtraction of kgs			
Solve one-step word problem involving addition of clock times			
Solve one-step word problem involving m and cm			
Use the associative property of addition to complete a number sentence			
Identify the best estimate of the sum of two two-digit nos.			
Solve one-step word problem on subtraction of clock times			
Use the associative property of addition to complete a number sentence			

*The code refers to the 4-character label found in the third column for each item e.g. I015.

They were asked to link the items to the item descriptors (using the item codes) and then, for each item and item descriptor, to indicate to which of the four proficiency level descriptions they could be linked. They were also invited to comment on any of the links if needed. The results of the validation exercise were as follows:

a) All 149 Sixth class items were linked to their appropriate item descriptors.

b) 96 out of the 99 Second class items were linked to their appropriate item descriptors with three not being linked due to problems with the wording of their descriptors.

c) 128 of the 149 Sixth class items were linked to the appropriate line of text in the overall proficiency level description. Step 2 of the procedure used in developing the proficiency level descriptions (i.e., lower level item descriptors being subsumed by higher level item descriptors) accounts for 11 of the items not appropriately linked, while Step 3 of the procedure (i.e., integration of same-level descriptions) accounts for the remaining 10 items not appropriately linked.

d) 76 of the 99 Second class items were linked to the appropriate proficiency level description. Step 2 of the procedure used in developing the proficiency level descriptions accounts for 18 of the items not appropriately linked, while Step 3 accounts for the remaining 6 items.

In the case of the non-linked items in findings (c) and (d) the experts indicated on their validation sheets that they had linked the items to another level and following discussion were satisfied with the outcomes of Steps 2 and 3 and the manner in which the mathematical demands of the non-linked items were incorporated into the summary proficiency level descriptions.

The high levels of accuracy in initially matching items to item descriptors to proficiency level descriptions and the fact that the experts indicated that they felt almost all of their non-matched items were an artifact of steps 2 and 3 of the procedure for developing the proficiency level descriptions provides preliminary evidence of the validity of the proficiency level descriptions and the reliability of the procedure used to produce them.

Subsequent to the review, some small modifications were made to the proficiency level descriptions. The final, agreed descriptions for Second and Sixth class are shown in Tables 6.13 and 6.14. The sample items referred to in the Tables are those shown at the end of Chapter 2.

Table 6.13: Proficiency levels on the mathematics scale, and percentages of pupils (overall and by gender) achieving each level, Second class

Level & score range	Sample items	What pupils can typically do	Total	Boys	Girls
4 **315+**	Q8 Q7	Pupils at Level 4 can calculate items which may be bought with a given sum of money; and can calculate the best estimate of the sum or difference of two two-digit numbers. They show understanding of the associative property of addition; the connection between two-step word problems and their corresponding numerical expressions; and the correct use of the symbols =, <, >. They can measure length using metres and centimetres and measure area using a non-standard unit. They can interpret information from a bar-line graph and make a calculation with it. They can solve one-step word problems involving: repeated addition; addition or subtraction of clock times; halves and quarters of metres, kg, and litres. They can solve two-step word problems involving addition and subtraction of two-digit numbers and money.	10	12	8
314 **3** **270**	Q6 Q5 Q4	Pupils at Level 3 can recall the subtraction facts, add a row of three numbers with renaming within 99, and find the difference between two two-digit numbers. They can use the vocabulary of ordinal number, and convert tens and units to numbers from 10 to 199. They can extend number patterns, identify quarters of 2-D shapes, and partition a 2-D shape into two other shapes. They can use the concept of an angle as a rotation, use a calendar to read days, dates, months and seasons, and select appropriate non-standard units for measuring capacity. They can exchange coins. They can also solve one-step word problems involving: addition or subtraction of two-digit numbers; halves and quarter of sets of up to 20 objects; addition or subtraction of money, cm and m, kg or litres; time in hr and min on 12-hour clock. They can solve one-step and two-step word problems involving minutes, hours and days.	25	25	25
269 **2** **232**	Q3 Q2	Pupils at Level 2 can be expected to add columns of three numbers with renaming within 99. They can identify odd and even numbers. They can use the symbols +, - to complete number sentences. They can identify halves of sets with up to 20 objects. Pupils at this level can combine two 2-D shapes to make other shapes. They can identify properties of 3-D shapes and compare lengths of objects in non-standard units. Pupils at this level can convert analogue to digital time (to the half-hour), and interpret information in simple block graphs. They can solve one-step word problems involving addition or subtraction of simple whole numbers.	30	28	32
231 **1** **184**	Q1	Pupils at Level 1 can be expected to count objects in groups of threes and fives; use ordinal number; locate numbers within specified intervals up to 199; connect verbal and numerical forms of numbers, up to 199; and to recall the addition facts. They can use the vocabulary of spatial relations to locate objects; identify and classify simple 2-D and 3-D shapes and list some of their properties. They can identify half of a regular 2-D shape. Pupils at this level can use the vocabulary of time to sequence events; and identify a date in a calendar. They can find the value of a group of coins. They can read a simple block graph.	25	25	25
<184		Pupils below proficiency level 1 have a less than 62.5% chance of correctly answering a Level 1 question. Their mathematical skills are very low, relative to other 2nd class pupils and are not properly assessed by the National Assessments.	10	10	10

Proficiency Levels

Table 6.14: Proficiency levels on the mathematics scale, and percentages of pupils (overall and by gender) achieving each level, Sixth class

Level & score range	Sample items	What pupils can typically do	Total	Boys	Girls
4 **316+**	Q10 Q9	Pupils at Level 4 can multiply and divide decimals by decimals, and carry out simple algebraic procedures involving evaluation of linear expressions and one-step equations. They can demonstrate a high level of understanding of signed integers and number theory concepts such as prime and composite numbers. They can deduce symbolic rules for simple functions. At this level pupils can also analyse geometric shapes in detail and deduce rules about them. They can construct circles. They can plot coordinates and use scales on maps or plans to calculate distances and areas. They can solve non-routine and multi-step practical problems involving ratios, mixed numbers, percentage gain or loss, value for money comparisons, currency conversions, speed, and time zones.	10	11	9
315 **3** **273**	Q8 Q7 Q6	Pupils at Level 3 can add and subtract mixed numbers and decimals. They can demonstrate understanding of decimal notation, factors and multiples, exponents, and square roots. They can connect verbal and symbolic representations of word problems. They can construct and measure angles and construct triangles and rectangles given selected sides and angles. Pupils at this level can classify triangles and quadrilaterals based on angle and line properties and rules. They can identify properties of 3-D shapes. They can manipulate commonly used units of area, capacity and weight. They can read, interpret, and analyse pie-charts, multiple-bar bar-charts and trend graphs. They can estimate simple probabilities. They can solve routine and non-routine word problems involving operations with fractions, decimals and percentages, length and perimeter, capacity, and time.	25	27	23
272 **2** **230**	Q5 Q4 Q3	Pupils at Level 2 can multiply fractions and decimals, estimate products, calculate common factors and multiples of whole numbers, and convert fractions and decimals to percentages. They can identify prime numbers within 30 and identify rules for number patterns. They can demonstrate understanding of a letter as a placeholder in algebraic expressions, and complete two-step number sentences involving addition and subtraction. Pupils at this level can construct lines and circles, estimate angles and use properties of shapes to calculate line and angle sizes. They can make logical deductions from simple data sets. They can solve multi-step word problems involving operations with integers, fractions and percentages.	30	29	31
229 **1** **184**	Q2 Q1	Pupils at Level 1 can add, subtract, and round whole numbers and decimals. They show understanding of whole number notation and can connect numeric and verbal representations of large numbers. Pupils at this level can classify angles and identify templates of simple 3-D shapes. They can manipulate commonly used units of length. They can read and interpret, without calculation, simple frequency tables, pie-charts, bar charts and trend graphs. They can solve routine word problems involving the four operations with whole numbers.	25	23	27
<184		Pupils below proficiency Level 1 have a less than 62.5% chance of correctly answering a Level 1 question. Their mathematical skills are very low, relative to other Sixth class pupils and are not properly assessed by the National Assessments.	10	9	11

References

Close, S., Millar, D., & Shiel, G. (2009). Characteristics and curriculum validity of test items for the national assessment of mathematics achievement in 6th class. In D. Corcoran, T. Dooley, S. Close and R. Ward (Eds.), *Proceedings of the third national conference on research in mathematics education – Mathematics for all: Extending mathematics capacity* (pp. 77-89). Dublin: St Patrick's College.

Cosgrove, J., Kellaghan, T., Forde, P., & Morgan, M. (2000). *The 1998 National Assessment of English Reading.* Dublin: Educational Research Centre.

Cosgrove, J., Milis, Shiel, G., Forde, P., & Wardle, H. (2004). *Assessment framework for the 2004 National Assessment of English Reading (NAER): First and Fifth classes.* Dublin: Educational Research Centre.

Department of Education and Science. (2000). *Learning support guidelines.* Dublin: Stationery Office.

Department of Education and Science. (n.d.). Table 2.1 — Number of Pupils in Primary Schools by age and Schooltype (2007/2008). Retrieved October 4th, 2010, from http://www.education.ie/servlet/blobservlet/stats_statistical_report_2007_2008.xls#'Table 2.1'!A1

Department of Education and Science/National Council for Curriculum and Assessment. (1999). *Primary school curriculum: Mathematics.* Dublin: Stationery Office.

Eivers, E., Close, S., Shiel, G., Millar, D., Clerkin, A., Gilleece., L. & Kiniry, J. (2010). *The 2009 National Assessments of Mathematics and English Reading.* Dublin: Stationery Office.

Eivers, E., Shiel, G., Perkins, R., & Cosgrove, J. (2005). *The 2004 National Assessment of English Reading.* Dublin: Educational Research Centre.

Greaney, V., & Kellaghan, T. (1996). *Monitoring the learning outcomes of education systems.* Washington DC: World Bank.

Kirsch, I. (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured.* Princeton NJ: Educational Testing Service.

Mullis, I.V.S., Kennedy, A.M., Martin, M.O., & Sainsbury, M. (2006). *PIRLS 2006. Assessment framework and specifications.* Chestnut Hill MA: International Study Center, Boston College.

OECD (Organisation for Economic Co-operation and Development). (2009). *PISA 2006 Technical Report.* Paris: OECD.

Shiel, G., & Kelly, D. (2001). *The 1999 National Assessment of Mathematics Achievement.* Dublin: Educational Research Centre.

Shiel, G., Surgenor, P., Close, S., & Millar, D. (2006). *The 2004 National Assessment of Mathematics Achievement.* Dublin: Educational Research Centre.

Weaver, C.A., & Kintisch, W. (1991), Expository Text. In R. Barr, M.L. Kamil, P. Mosentha, & P.D. Pearson (Eds.), *Handbook of reading research. Volume 11.* (pp. 230-245). New York: Longman.

Appendix A

Second Class item Descriptors by Content Strand, Difficulty, and Item Code

UID	Diff. Logits	content strand	Item Descriptor
Level 4			
I086	1.82	Number & Algebra	Identify the best estimate of the difference between two two-digit nos
I051	1.57	Measures	Measure area using a non-standard unit
I025	1.40	Number & Algebra	Connect a two-step word problem to a numerical expression
I039	1.37	Measures	Solve two-step word problem involving addition and subtraction of money
I078	1.12	Measures	Solve a one-step word problem involving repeated addition
I019	1.09	Data	Read information from a block graph and make a calculation with it
I074	1.08	Measures	Solve a one-step word problem involving kg
I035	0.85	Measures	Solve a non-routine word problem involving quarters of litres
I040	0.81	Measures	Calculate how many items may be bought with a given sum.
I063	0.81	Number & Algebra	Solve a two-step word problem involving addition and subtraction.
I034	0.73	Measures	Solve one-step word problem involving subtraction of kgs
I095	0.66	Measures	Solve a one-step word problem involving addition of clock times
I093	0.62	Measures	Solve a one-step word problem involving m and cm
I042	0.61	Number & Algebra	Use the associative property of addition to complete a number sentence
I066	0.60	Number & Algebra	Identify the best estimate of the sum of two two-digit numbers
I091	0.58	Measures	Solve a one-step word problem involving subtraction of clock times
I062	0.55	Number & Algebra	Use the associative property of addition to complete a number sentence
I047	0.47	Number & Algebra	Identify correct use of =, <, > in number sentences
I018	0.46	Data	Interpret a block graph
I037	0.46	Measures	Solve one-step word problems involving minutes
I020	0.44	Data	Interpret block graphs
Level 3			
I007	0.35	Number & Algebra	Use the associative property of addition to complete a number sentence
I033	0.33	Measures	Measure length using cm
I080	0.32	Measures	Solve a one-step word problem involving subtraction of money
I003	0.31	Number & Algebra	Extend number patterns
I038	0.30	Measures	Calculate how many items may be bought with a sum of money - cents
I089	0.30	Measures	Solve a one step word problem involving subtraction of clock times
I056	0.29	Measures	Use calendar to identify day of specified date
I088	0.28	Number & Algebra	Extend number patterns
I053	0.28	Measures	Solve a one-step word problem involving cm and m
I057	0.27	Measures	Solve a one-step word problem involving kg

I082	0.25	Number & Algebra	Use the associative property of addition to complete a number sentence
I006	0.23	Number & Algebra	Subtract two-digit numbers with renaming within 99
I028	0.22	Number & Algebra	Solve a two-step word problem involving addition and subtraction.
I014	0.21	Number & Algebra	Solve a two-step word problem involving addition and subtraction.
I059	0.19	Measures	Solve a one-step word problem involving litres
I027	0.16	Number & Algebra	Subtract two-digit numbers with renaming within 99
I022	0.08	Number & Algebra	Convert tens and units to numbers 0 to 199
I090	-0.02	Shape and Space	Identify quarter of a 2-D shape
I055	-0.06	Measures	Read day, date month and season using calendar
I076	-0.07	Number & Algebra	Find the difference between two two-digit numbers
I073	-0.09	Measures	Solve a one-step word problem involving subtraction of lengths - m and cm
I036	-0.11	Measures	Solve a one-step word problem involving time in hours and minutes on 12-hour analogue clock
I068	-0.15	Number & Algebra	Solve one-step word problem involving quarter of a set of up to 20 objects
I072	-0.17	Shape and Space	Partition a 2-D shape into two other shapes
I009	-0.19	Number & Algebra	Subtract a two-digit numbers with renaming
I041	-0.25	Number & Algebra	Use the vocabulary of ordinal number
I046	-0.26	Measures	Solve a one-step word problem involving litres
I079	-0.38	Measures	Solve a one-step word problem involving addition of money
I060	-0.38	Measures	Exchange coins.
I071	-0.42	Shape and Space	Solve a word problem involving angle as a rotation
I005	-0.44	Number & Algebra	Recall and order subtraction facts
I045	-0.45	Number & Algebra	Recall and order subtraction facts
I044	-0.48	Number & Algebra	Add a row of three numbers with renaming within 99
I043	-0.50	Number & Algebra	Solve a one-step word problem involving addition of two-digit numbers
I096	-0.50	Measures	Solve two-step word problems involving hours and days
I017	-0.52	Measures	Measure area using a non-standard unit
Level 2			
I084	-0.54	Number & Algebra	Find the complement of two numbers
I099	-0.55	Data	Interpret a block graph
I061		Number & Algebra	Identify the ordinal number of an object in a line of objects
I083	-0.56	Number & Algebra	Solve a one-step word problem involving addition
I026	-0.59	Number & Algebra	Solve a one-step word problem involving subtraction
I032	-0.74	Measures	Convert analogue to digital time (on the half-hour)
I075	-0.75	Measures	Solve a one-step word problem involving use of calendar
I048	-0.81	Number & Algebra	Identify half of a set with up to 20 objects
I094	-0.81	Measures	Select appropriate non-standard unit for measuring capacity
I011	-0.83	Shape and Space	Identify properties of a 3-D shape
I016	-0.84	Measures	Compare lengths of two objects in non-standard units
I029	-0.86	Number & Algebra	Identify a two-digit odd number
I031	-0.87	Shape and Space	Combine two 2-D shape to make another shape
I087	-0.93	Number & Algebra	Use one of the symbols =, +, - to complete a number sentence
I024	-1.08	Number & Algebra	Add a column of three numbers with renaming within 99
I012	-1.11	Shape and Space	Partition a 2-D shape into two other shapes

I030	-1.16	Shape and Space	Partition a 2-D shape into two other shapes
I004	-1.17	Number & Algebra	Add a row of three numbers with renaming within 99
Level 1			
I097	-1.22	Measures	Recognise and find value of a group of coins
I064	-1.22	Number & Algebra	Count objects in groups of three
I001	-1.22	Number & Algebra	Identify number in specified interval 0 -199
I023	-1.24	Number & Algebra	Identify addition combination that gives a specified sum
I058	-1.31	Shape and Space	Identify properties of 2-D shapes
I050	-1.33	Shape and Space	Partition a 2-D shape into two other shapes
I100	-1.35	Data	Interpret a block graph
I010	-1.41	Shape and Space	Classify 2-D shapes
I077	-1.44	Number & Algebra	Connect a subtraction word problem to a number sentence
I067	-1.50	Number & Algebra	Use one of the symbols =, <, > to complete a number sentence
I008	-1.52	Number & Algebra	Connect a subtraction word problem to a numerical expression
I002	-1.58	Number & Algebra	Count objects in groups of five
I092	-1.69	Shape and Space	Identify properties of 3-D shapes.
I065	-1.76	Number & Algebra	Connect a subtraction word problem to a numerical expression
I002	-1.58	Number & Algebra	Count objects in groups of five
I092	-1.69	Shape & Space	Identify properties of 3-D shapes.
I065	-1.76	Number & Algebra	Connect a subtraction word problem to a numerical expression
I081	-1.80	Number & Algebra	Use the vocabulary of ordinal number
I069	-1.83	Measures	Use the vocabulary of time to sequence events
I049	-1.95	Shape & Space	Use the vocabulary of spatial relations
I054	-2.04	Measures	Identify a date in a calendar
I085	-2.24	Number & Algebra	Connect a subtraction word problem to a number sentence
I098	-2.27	Data	Read a block graph
I070	-2.43	Shape & Space	Identify half of a 2-D shape
I015	-2.5	Shape & Space	Identify properties of 3-D shapes.

Appendix B

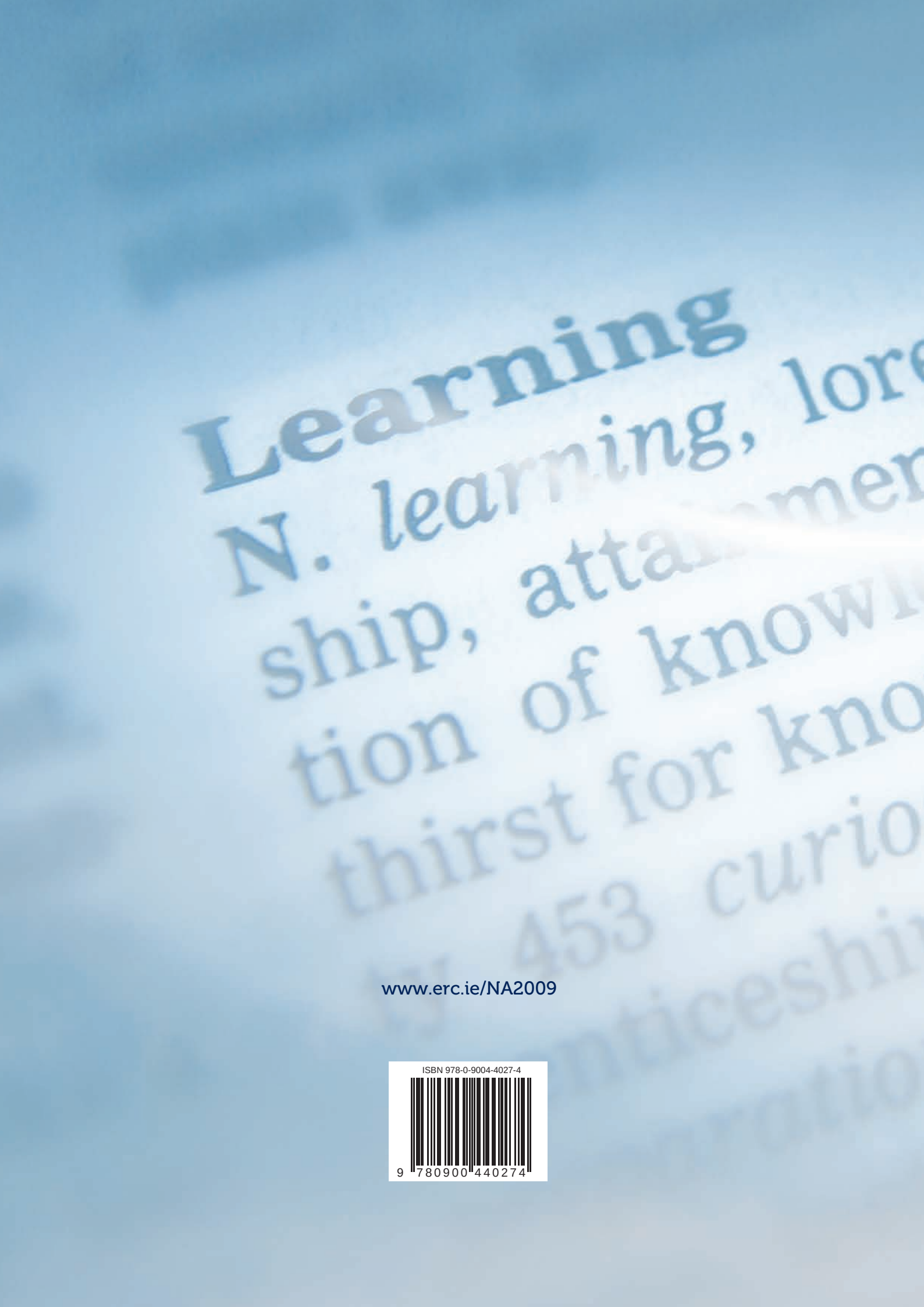
Sixth Class item Descriptors by Content Strand, Difficulty, and Item Code

UID	Diff. Logits	content strand	curriculum objective
J056	3.00	Shape and Space	Make informal deductions about 2-D shapes
J067	1.83	Measures	Calculate the surface area of a 3-D shape
J045	1.77	Number & Algebra	Identify a negative number on the number line
J144	1.61	Measures	Solve a problem involving use of international time zones
Level 4			
J006	1.51	Number & Algebra	Divide a decimal by a decimal
J081	1.48	Shape & Space	Make informal deductions about 2-D shapes
J021	1.47	Number & Algebra	Identify positive and negative numbers on the number line
J138	1.47	Measures	Solve a multi-step problem involving measures of length
J005	1.44	Number & Algebra	Multiply a decimal by a decimal
J069	1.20	Measures	Solve a multi-step problem involving measures of capacity
J054	1.19	Number & Algebra	Solve a multi-step problem on calculation of percentage profit
J071	1.04	Measures	Solve a non-routine multi-step problem on currency conversion
J089	1.03	Measures	Use the scale on a plan to calculate distance
J023	1.03	Number & Algebra	Identify symbolic rule for a simple function
J065	1.02	Measures	Use the scale on a plan to calculate perimeter
J090	0.98	Measures	Use the scale on a plan to calculate distance
J098	0.97	Data	Interpret a pie chart
J070	0.95	Measures	Calculate average speed given distance and time
J039	0.95	Number & Algebra	Identify prime numbers
J080	0.90	Number & Algebra	Solve a one step equation
J078	0.86	Number & Algebra	Solve a word problem using a simple ratio
J094	0.76	Measures	Find the volume of a cuboid
J109	0.73	Shape & Space	Draw all lines of symmetry in a 2-D shape
J145	0.73	Measures	Solve a problem on comparing prices to identify value for money
J058	0.67	Shape & Space	Find the radius of a circle
J116	0.64	Measures	Calculate the area of an irregular 2-D shape
J136	0.60	Shape & Space	Solve non-routine problem using angle of rotation of a clock hand
J068	0.60	Measures	Find the area of a rectangular space from a scale plan
J053	0.59	Number & Algebra	Solve a word problem using a simple ratio
J011	0.57	Number & Algebra	Identify prime numbers
J085	0.57	Shape & Space	Plot simple co-ordinates
J114	0.55	Measures	Solve a multi-step word problem using measures of length
J130	0.54	Shape & Space	Make informal deductions about 2-D shapes
J117	0.53	Measures	Calculate the area of an irregular 2-D shapes
Level 3			
J033	0.51	Number & Algebra	Solve a problem involving addition and subtraction of fractions

J120	0.50	Measures	Solve a problem involving comparing prices on value for money
J066	0.45	Measures	Solve problem on calculation of width given perimeter and length
J048	0.44	Number & Algebra	Identify simple square roots
J139	0.39	Measures	Construct a rectangle given its area and perimeter
J044	0.38	Number & Algebra	Write a whole number in exponential form
J131	0.35	Shape & Space	Construct a triangle given 2 sides and an angle
J110	0.34	Shape & Space	Plot simple co-ordinates
J100	0.33	Measures	Calculate the perimeter of irregular shapes
J097	0.31	Data	Read a pie chart
J146	0.29	Measures	Solve a one-step problem involving currency conversion
J106	0.28	Shape & Space	Make informal deductions about 2-D shapes
J049	0.23	Number & Algebra	Find the missing value in a table of data on a functional relationship
J107	0.21	Shape & Space	Classify triangles using angle and line properties
J009	0.17	Number & Algebra	Add simple mixed numbers
J125	0.14	Data	Solve a problem involving selected outcomes of tossing two dice
J148	0.12	Data	Estimate likelihood of drawing any number from a set of numbers
J060	0.10	Shape & Space	Construct triangles from given sides or angles
J115	0.08	Measures	Rename measures of capacity
J092	0.06	Measures	Calculate area using acres and hectares
J087	0.04	Shape & Space	Recognise angles in terms of a rotation
J016	0.03	Number & Algebra	Identify common factors and multiples
J124	-0.01	Data	Read and interpret trend graphs and pie charts
J030	-0.02	Number & Algebra	Divide a four-digit number by a two-digit number, without/with a calculator
J082	-0.02	Shape & Space	Use angle and line properties to classify /describe triangles and quadrilaterals
J002	-0.03	Number & Algebra	Identify place value in whole numbers and decimals
J001	-0.04	Number & Algebra	Read, write and order whole numbers and decimals
J150	-0.05	Data	Estimate the likelihood of occurrence of events;
J095	-0.06	Measures	Know the relationship between time, distance and average speed
J135	-0.06	Shape & Space	Draw the nets of simple 3-D shapes and construct the shapes.
J018	-0.07	Number & Algebra	Identify simple square roots
J099	-0.11	Data	Read and interpret trend graphs and pie charts
J122	-0.11	Data	Estimate the likelihood of occurrence of events;
J149	-0.11	Data	Estimate the likelihood of occurrence of events;
J050	-0.13	Number & Algebra	Translate word problems with a variable into number sentences
J134	-0.13	Shape & Space	Use 2-D shapes and properties to solve problems.
J034	-0.17	Number & Algebra	Add and subtract simple fractions and simple mixed numbers
J037	-0.23	Number & Algebra	Divide a whole number by a unit fraction
J102	-0.24	Number & Algebra	Divide a four-digit number by a two-digit number, without/with a calculator
J032	-0.24	Number & Algebra	Express improper fractions as mixed nos. and position on the number line

J083	-0.25	Shape & Space	Identify the properties of the circle
J075	-0.26	Data	Analyse a trend graph
J123	-0.27	Data	Interpret a multiple bar chart
J010	-0.28	Number & Algebra	Multiply a fraction by a fraction using area representation
J008	-0.31	Number & Algebra	Express an improper fraction as a mixed number
J019	-0.32	Number & Algebra	Solve a problem involving use of common multiple
J112	-0.36	Shape & Space	Solve problem on calculation of angle of rotation of clock hand
J143	-0.36	Measures	Solve a word problem involving renaming capacities
J014	-0.37	Number & Algebra	Solve a word problem on relationship between percentages, fractions and decimals
J029	-0.38	Number & Algebra	Solve a word problem involving estimation with lengths
J096	-0.38	Measures	Solve a word problem on conversion of euros to other currency
J091	-0.44	Measures	Rename measures of length (mm and m)
J012	-0.45	Number & Algebra	Express hundredths in decimal form
J017	-0.45	Number & Algebra	Identify square numbers
J043	-0.46	Number & Algebra	Identify square numbers
J086	-0.46	Shape & Space	Identify relationships among 3-D shapes
J118	-0.47	Measures	Solve multi-step word problem involving measures of weight
Level 2			
J031	-0.50	Number & Algebra	Compare and order fractions
J055	-0.50	Number & Algebra	Extend a number pattern involving decimals
J038	-0.56	Number & Algebra	Convert a fraction to a percentage
J015	-0.59	Number & Algebra	Identify a prime number
J046	-0.63	Number & Algebra	Solve a word problem involving addition of a positive and negative number
J022	-0.64	Number & Algebra	Solve word problem involving subtraction of positive and negative numbers
J047	-0.64	Number & Algebra	Construct a verbal rules for a number pattern
J004	-0.67	Number & Algebra	Estimate a product involving multiplication of decimals
J129	-0.67	Number & Algebra	Complete number sentence involving division
J024	-0.69	Number & Algebra	Use a letter as a placeholder in a number pattern
J132	-0.70	Shape & Space	Construct a circle of given radius
J111	-0.70	Shape & Space	Identify properties of a 3-D shape
J147	-0.74	Measures	Calculate the area of a regular 2-D shape
J007	-0.75	Number & Algebra	Solve multi-step word problem involving equivalence of fractions
J104	-0.76	Number & Algebra	Identify a common multiple of three numbers
J042	-0.76	Number & Algebra	Solve a word problem on division of a whole number by a unit fraction
J128	-0.76	Number & Algebra	Solve multi-step word problem involving ordering fractions
J035	-0.77	Number & Algebra	Multiply a fraction by a fraction
J141	-0.77	Data	Intepret a frequency table
J057	-0.80	Shape & Space	Draw a line parallel to another line
J103	-0.81	Number & Algebra	Solve multi-step word problem involving percentages
J079	-0.81	Number & Algebra	Identify common factors of two numbers
J041	-0.81	Number & Algebra	Complete a number pattern
J105	-0.82	Number & Algebra	Complete a number sentence involving subtraction

J093	-0.83	Measures	Rename measures of weight (g and kg)
J036	-0.84	Number & Algebra	Express tenths, hundredths and thousandths in decimal form
J074	-0.85	Data	Interpret a trend graph
J061	-0.86	Shape and Space	Use 2-D shapes to solve problems.
J040	-0.87	Number & Algebra	Express thousandths in decimal form
J119	-0.89	Measures	Solve a routine problem involving international time zones
J028	-0.91	Number & Algebra	Round a 4 digit whole numbers to nearest 100
J108	-0.95	Shape & Space	Construct a circle of given radius
J076	-0.97	Number & Algebra	Complete a two-step number sentence on addtion and subtraction
J113	-1.11	Shape & Space	Estimate size of angles in degrees
J088	-1.17	Shape & Space	Estimate size of angles in degrees
J027	-1.19	Number & Algebra	Identify place value to order decimals (3 places)
J137	-1.20	Shape & Space	Use the angle sum in a quadrilateral to identify unknown angle
J072	-1.22	Data	Use a data set to solve a classification problem
J133	-1.27	Shape & Space	Identify a line of symmetry in a 2-D shape
Level 1			
J051	-1.32	Number & Algebra	Solve routine word problem on addition and subtraction of five-digit numbers
J013	-1.36	Number & Algebra	Extend a number pattern
J063	-1.41	Shape & Space	Identify angle of rotation on a compass
J142	-1.51	Data	Interpret information in a frequency table
J121	-1.52	Data	Estimate the likelihood of occurrence of an event from a pie-chart
J140	-1.55	Data	Read a frequency table
J062	-1.64	Shape & Space	Classify angles
J026	-1.65	Number & Algebra	Connect verbal and numeric representations of large numbers
J020	-1.65	Number & Algebra	Write a exponential expression in multiplication form
J052	-1.66	Number & Algebra	Complete a one-step number sentence involving subtraction
J064	-1.71	Measures	Rename measures of length (cm and m)
J025	-1.75	Number & Algebra	Evaluate a simple algebraic expression
J003	-1.86	Number & Algebra	Use rounding to order decimals (3 places)
J077	-1.91	Number & Algebra	Solve routine word problem on division of a four-digit no. by a two-digit no.
J126	-2.12	Number & Algebra	Subtract a decimal (3 places) from a whole number
J073	-2.34	Data	Interpret a simple trend graph
J059	-2.43	Shape & Space	Identify the net of a 3-D shape
J101	-2.52	Number & Algebra	Add a column of whole numbers and decimals (3 places)
J127	-2.62	Number & Algebra	Divide a decimal (2 places) by a decimal (one place),



www.erc.ie/NA2009

ISBN 978-0-9004-4027-4