**Appendix 2**

# MEASUREMENT OF ACHIEVEMENT PROGRESS

## Design of data collection for *Reading* and *Numeracy* assessments

Achieving the aims of the project required monitoring the growth of students with learning difficulties on developmental scales of *Reading* and *Numeracy*. These scales were constructed using data from mainstream students (i.e., non-learning difficulty students) only,[1] that were then used to measure all students on the same scale. For the present 'Third Wave' project, data were collected for Year 4 to Year 6 students in *Reading* and *Numeracy* during March 2005 and again in September 2005.

To observe achievement growth and be able to make comparisons (between for example students in reference and intervention schools), two measurement scales were required. The *Reading* scale constructed is an extension of the ACER *Progressive Achievement Test* (PAT) *Reading Comprehension* scale and the *Numeracy* scale an extension of the ACER *PAT Mathematics* scale.

Traditional norming of tests does not include a calibration of all tests on a common scale, making it impossible to compare the tests according to their difficulty – independently from student achievement and the distribution of student achievement by year level. In the Rasch measurement methodology used in this project, the calibration of relative item difficulty on the scale is sample independent by first constructing a scale and then locating students' achievements on the scale.

Thus, we first constructed a scale on which we calibrated all items used to collect achievement data, and located the distribution of student achievement individually and by year level on this scale, independent of the particular test/tests that was/were administered to each student in a given year level. Note that a common scale is required to make comparisons between any two tests according to their relative difficulty, and any two cohorts according to their achievement on the scale.

## Source of items used to prepare the assessments

**Reading:** The series of *Reading Progress Tests* (RPT) and *Progressive Achievement Tests* (PAT Reading Comprehension) were used in March 2005. For the September 2005 assessments, new *Reading* tests were prepared at ACER *Reading* (labeled RC to RE in Table A2.2). The other tests used were the RPT, PAT-C, adapted from the New Zealand Council for Educational Research (NZCER) PAT-C items, together with 20 (New South Wales) NSW and 20 (Western Australia) WA items – all calibrated on scales with national benchmarks for years 3, 5 and 7.

**Numeracy:** *I can do Maths* (ICDM) and PAT Maths were used in the March 2005 assessments. For the September 2005 assessments, new tests were prepared at ACER (labeled NC to NE in Table A2.2). The other assessment consisted of adapted NZCER PAT-Maths items, 20 items from the NSW *Basic Skills Tests* (BST), and 20 items from the *Western Australian Literacy and Numeracy Assessments* (WALNA). Consistent with the Reading items, these Numeracy items have all been calibrated on scales with national benchmarks for students in Years 3, 5 and 7.

Tables A2.1 and A2.2 show the tests administered at each year level in the March and September 2005 assessments, respectively.

[1] Note that these 'mainstream' students derived from the "Effective Teaching and Learning Practices Initiative for Students with Learning Difficulties" project.

## Appendix 2

**Table A2.1 March 2005 Reading and Numeracy Tests by Year Level**

| Year level | Reading | Numeracy |
|---|---|---|
| **4** | RPT2 | PM1 |
| **5** | RPT3 | PM2 |
| **6** | RPT4 | PM6 |

RPT: Reading Progress Test
PM: Progressive Achievement Test (PAT Maths)

**Table A2.2 September 2005 Reading and Numeracy Tests by Year Level**

| Year level | Reading | Numeracy |
|---|---|---|
| **4** | RC | NC |
| **5** | RD | ND |
| **6** | RE | NE |

RC-RE: New *Reading* assessments developed by ACER for this project
NC-NE: New *Numeracy* assessments developed by ACER for this project

**Note:** For the September 2005 assessments, the *Reading* and *Numeracy* tests sent to schools were selected for each student – based on their achievements assessed during March 2005. To this end, a name label was attached on the front cover of each test booklet or answer sheet.

### Notes on the administration of the tests

Each reading assessment normally required a total time between 30 and 50 minutes and each numeracy assessment a total of 40 to 60 minutes. Students were required to attempt the tasks sequentially from the first task in a test, responding to as many questions as possible to the best of their ability. Students were given as much time as they needed to answer most questions, but sequentially. A reasonable amount of time was given to students to answer as many questions as they could. More sittings were allowed for the same test for students who 'got tired' or lost concentration, but without allowing them to review their answers from a previous sitting.

### Construction of the scales

Marking, data entry and cleaning was followed by a test-by-test analysis to eliminate any remaining noise and to ensure that each item worked satisfactorily in the test in which it was used. A single data file for *Reading* and one for *Numeracy*, with the data from each test, was then compiled. A series of joint analyses lead to the construction of the two scales. The fit indicators examined were: (a) the infit statistic of Quest, (b) the Point Biserial Correlation, (c) the mean abilities of the students who were successful and of those who were not, and (d) Differential Item Functioning and Item Characteristic Curves. The results of each analysis of fit were carefully examined, and all decisions were based on substantive grounds using the results of the fit indicators as a guide.

## Appendix 2

Figures A2.1 and A2.2 show the Reading and Numeracy scales, respectively. The distribution of the March and September 2005 achievements of all 'Third Wave' students in the project (from Year 4 to Year 6) are shown on the left of the scale, and the items from all 'Third Wave' tests are shown at their scale locations on the right.

```
LD DEST THird WAVE - March/Sep Read
-------------------------------------------------------------------------------------------------------
Item Estimates (Thresholds)                                                        23/ 3/2006 13:18
all on dest (N = 1310[2] L = 94 Probability Level=0.50)
-------------------------------------------------------------------------------------------------------
  2.0                           |     136
                                |
                                |
                                |
                                |
                                |
                                |
  1.0                           |
                        XXXXX   |
                                |      76
                                |
                           XX   |     133
                            X   |     138
                            X   |      81    128
                        XXXXX   |
  0.0                       X   |     122    131    132    137    453
                          XXX   |      94    124
                          XXX   |     139
                          XXX   |     118    125    449
                      XXXXXXX   |      82     83     93
                          XXX   |      84     92    129
                         XXXX   |      79    126
                         XXXX   |     110    116    144    627
 -1.0                 XXXXXXX   |      85     86    115    119    143
                         XXXX   |      74     80    101    117
          XXXXXXXXXXXXXXXXXXX   |      91    140    141
                      XXXXXXX   |     130
                           XX   |      97    446
                          XXX   |      88    135    450
                        XXXXX   |      77     87     89    134    444    445    454    626
 -2.0                     XXX   |      75     99    127    448    456
                          XXX   |      98    121    123
                          XXX   |      73     90    107    120
                           XX   |      69    102
                           XX   |      68     78     95    106    111
                         XXXX   |      72    108    109    112
                            X   |      70    142    455
                           XX   |      96    113
 -3.0                      XXX   |     103    452
                             X   |      71    447
                                 |     104    114
                            X    |
                           XX    |
                                 |      67    100    105
                                 |
                                 |      66    451
 -4.0                       XX   |
                                 |
                                 |
                                 |
                                 |
                                 |
                                 |
                             X   |
 -5.0                            |
-------------------------------------------------------------------------------------------------------
  Each X represents    6 students
=======================================================================================================
```

**Figure A2.1 The *Reading* scale with distributions of all students' achievements and item locations**

[2] From 819 students, 655 completed both the March and September 2005 *Reading* assessments. Hence 655*2 = 1310 cases, as each assessment (in March and in September) produced a *Reading* measure.

## Appendix 2

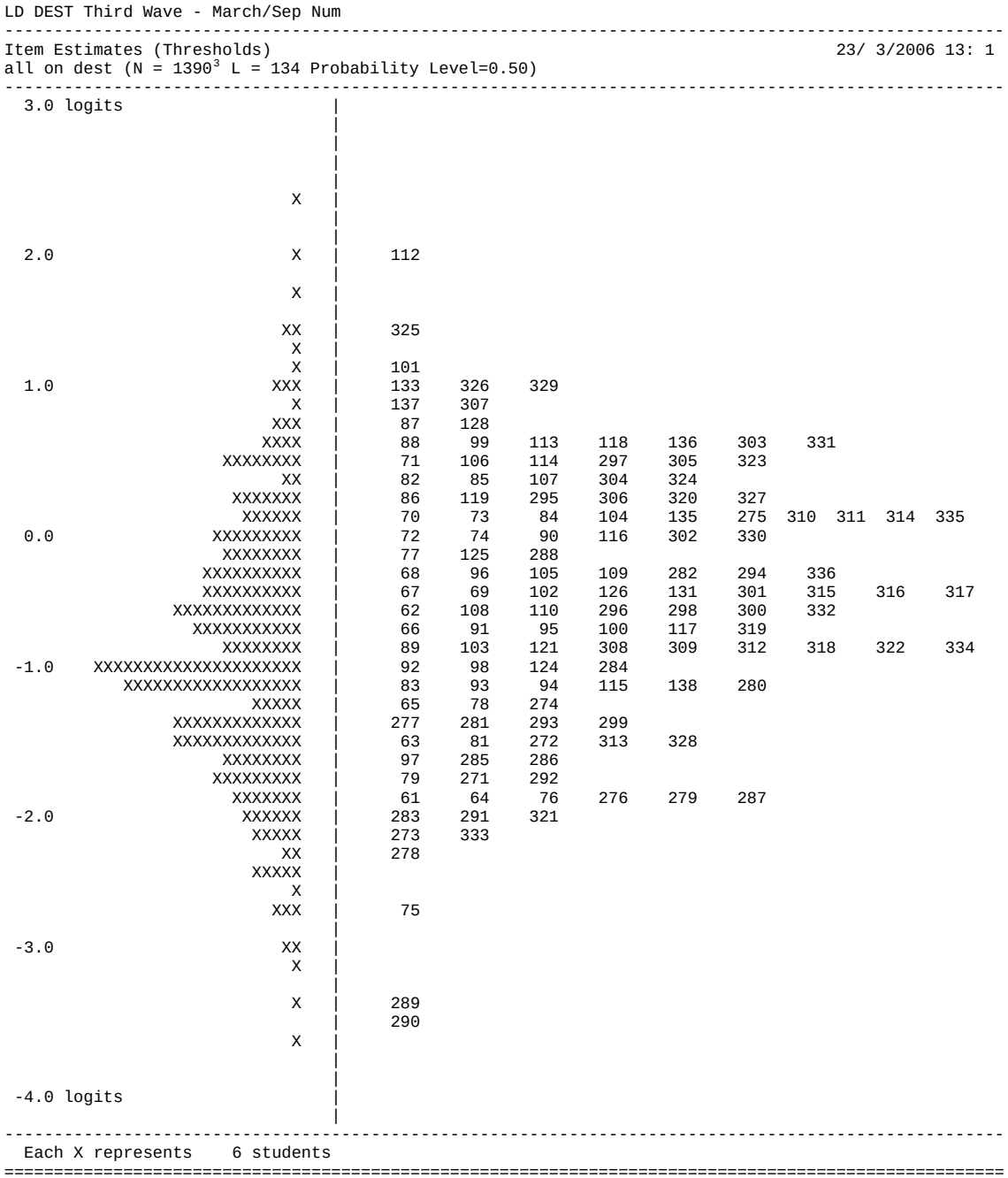



**Figure A2.2 The *Numeracy* scale with distributions of all students' achievements and item locations**

[3] Of the 819 participating students, 695 completed both the March and September *Numeracy* assessments. Hence 695*2=1390 cases, as each assessment (in March and in September) produced a *Numeracy* achievement measure.

## Description of the scales

The two scales were qualitatively described by ACER humanities and mathematics expert test developers. The scale descriptors shown on the scales are extensions of revised descriptions of the ACER *PAT Reading Comprehension* and *PAT Mathematics* scales. On the same scales the March and September students' achievements are shown in terms of percentiles. Figures A2.3 and A2.4 show the *Reading* and *Numeracy* scales, with scale descriptors on the left and achievement distributions on the right.

## Further notes on the assessments

Schools were provided with feedback from both the March and September 2005 assessments with individual reports for all participating students forwarded to schools. From the March assessments, student achievement was reported on provisional Rasch measurement scales in transformed logit units, **ldn** and **ldr**. Tests were selected to match the abilities of students from Year 4 to year 6. The scales against which the March results were reported were provisional because after the collection of the September data, all data (March and September) were analysed together to optimise the construction of the scales and to assure consistency in the estimation of scale scores.

Students' *Reading* and *Numeracy* graphical reports (see Figures A2.3 and A2.4) included both normative quantitative and qualitative information. The achievement of a student could be interpreted in terms of the tasks the student was likely to be able to do and the tasks not already mastered. Each student's identification number appeared at the bottom of the report together with an indication of how well the student has been measured. Fit indicators (Quest case infit statistic) that were close to 1 indicated that the student performed as expected; i.e., answered the 'easy' questions correctly together with some of the 'mid range' questions and made more errors in response to more 'difficult' items. Where the students' responses were 'ad hoc' (easier and harder questions answered with similar accuracy), the integer is clearly greater than 1.

The ACER research team was able to meet all the set deadlines for the September phase of assessment. There were some difficulties in sourcing a sufficient number of assessment items that were deemed suitable for use in this project. Andrew Stephanou, who had been involved in the review of the NZ PAT maths tests was able to gain support from the New Zealand Council for Educational Research (NZCER) to use them in this project. He was also given permission to use NZCER PAT reading materials. ACER Press was also given us permission to use adapted items from commercially available tests. Eighty assessment items were also sourced from the NSW and WA state-wide testing programs with permission again being granted due to the professional relationship shared with ACER. The purpose of including these 80 items in the September assessment was to locate **national benchmarks** of minimum standards on the *Reading* and *Numeracy* scales. Different tests were administered at each year level. It has been necessary to develop new tests to fill gaps we could not fill with existing assessment instruments.

By administering two tests to samples of mainstream students it was possible to calibrate different tests on a common scale and thus transform every test score to a scale score. The construction of the two scales, as well as the administration of equating tests, was based on data from mainstream students only.

The data analysis for the construction of the scales was carried out with the *ACER Ques*t software. A number of classical test theory and Item Response Modelling (IRM) fit indicators were examined to assure that each item contributed satisfactorily to the measurement the intended construct. Estimation of student scale scores with their errors and fit indicators for each response pattern were calculated using *ACER Quest*. The reports were prepared in a graphics program, *FreeHand*, and the programming for printing each of the **1638** reports was written in PostScript, making use of the mail merge facilities of Microsoft Word.

## Appendix 2

**THIRD WAVE**
**LEARNING DIFFICULTIES PROJECT**
**2005 Assessment**

Report for: *Student Name*
*School Name*
Year:

### Reading

*At the level of ability shown a student is typically able to*

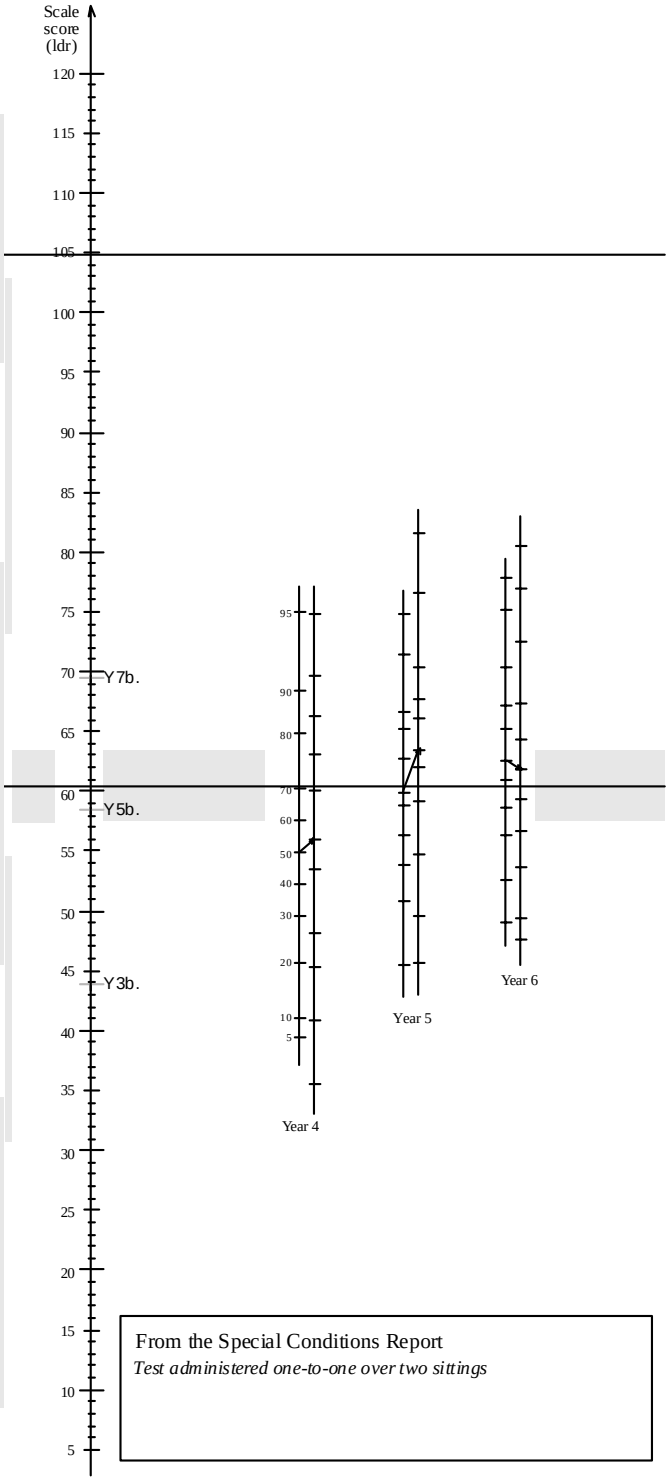
deal with a wide range of texts containing unfamiliar topics as well as sophisticated vocabulary and structures, for example, locate a detail in a complex information report; identify target information by discriminating between closely competing pieces of information; decipher meaning in a densely-written passage; interpret a colloquial expression in a narrative fiction passage; identify a main feature in a complex narrative description;

*and also*
deal with narrative and information texts containing unfamiliar topics and vocabulary, and which contain language used in relatively complex ways, for example, recognise the range of information in the details of a report; interpret challenging vocabulary in a complex sentence; recognise the author's purpose in a relatively complex passage; select information from a range of competing pieces of information; compare two parts of a report; locate information expressed in unfamiliar vocabulary; make links between ideas in a factual passage; interpret images in a complicated narrative containing competing details;

*and also*
deal with longer narrative, procedural and information texts containing some unfamiliar ideas and vocabulary, for example, identify meaning in a sentence that includes unfamiliar vocabulary; compare information from different parts of a text; locate stated information in a relatively complex sentence; synthesise information from consecutive paragraphs in a narrative passage; identify an author's attitude using clues from the text; locate information in a text with unfamiliar content;

*and also*
deal with short narrative, procedural and information texts with familiar topics, vocabulary and structure, for example, locate explicitly stated information in a short narrative, parable or accessible factual text; interpret the meaning of direct speech in a short narrative; understand the conclusion of a short narrative; recognise the motivation of a character in a short narrative;

*and also*
deal with very short narrative and information texts with familiar topics, and which have very simple vocabulary and structure, for example, they locate information in a very short, simple narrative or factual text; read and match single words and sentences to pictures; identify and differentiate different letter sounds and letter names; identify rhyming words; understand basic print concepts (such as where to start reading, full stops, capital letters, author, title); and identify simple words related to a familiar given theme.




Australian Council for Educational Research

1705
1.05

**Figure A2.3 Student report on the *Reading* scale with descriptors and achievement distributions**

## Appendix 2

**THIRD WAVE**
**LEARNING DIFFICULTIES PROJECT**
**2005 Assessment**

Report for Parker E

Report for: *Student Name*
*School Name*
Year:

### Numeracy

*At the level of ability shown, a student is typically able to*

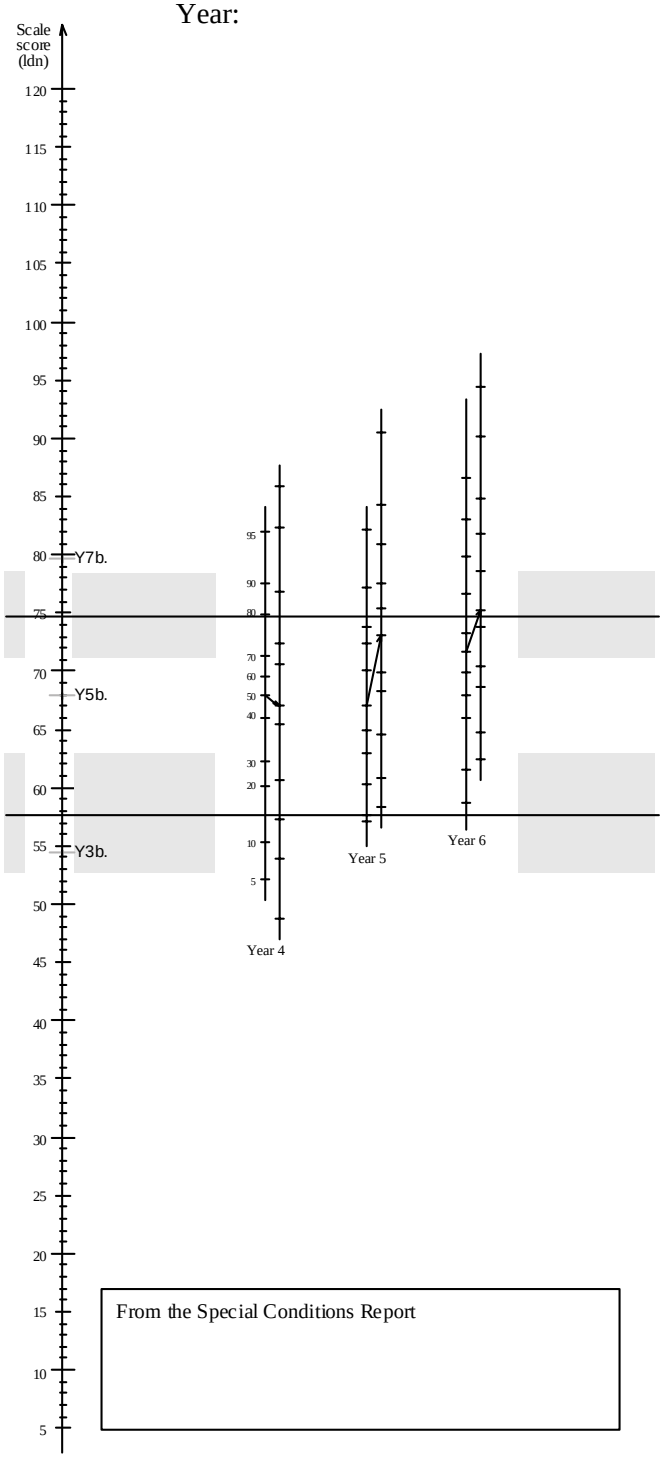
compare sizes of fractions, decimals and percentages; explain how to apply a 50% increase; choose the box with dimensions or markings that matches a given net; solve a multi-step problem involving rates; find area and perimeter of compound shapes; calculate the volume of a cuboid; recognise the relationship between circumference and diameter of a circle; calculate the side length of a right angle triangle; interpret grouped data in a column graph; interpret data shown in a box and whisker plot; recognise an event with probability closest to 1; add or remove brackets in simple algebraic expressions; apply a rule to extend a number sequence;

*and also*
write 7-digit numbers in words; put integers in order; perform single digit division with remainder; add or subtract decimal numbers with regrouping; convert fractions and decimals to percentages; solve word problems involving simple fractions or proportions, multiplication or division; recognise a prism from its description; work out a path using compass directions (NSEW); recognise the mirror image and symmetry of a picture or a 2D shape; extend a calendar to the next month; find the perimeter of a rectangle given its area; convert a time given in fractions of an hour to minutes; find volumes by counting unit cubes, visible and hidden; interpret bar charts and line graphs; work out the chance of a given random selection; find the mean of a set of numbers; read grouped data from a column graph;

*and also*
round large numbers to the nearest million; put 3-digit numbers in order; count by thirds; write 4-digit numbers in words; add 2-digit and 3-digit numbers; solve simple number problems using +, –, ×; apply 'tables' facts to division; apply simple percentages; calculate change from $1 and $2; complete a pattern using symmetry; recognise a 2D shape from a list of its properties; recognise a 3D model from different viewpoints; find a map location using coordinates; match digital time to clock face time; convert 12 hour time to 24 hour time; read a calendar; find areas by counting squares and part squares; compare information given in a pictograph or column graph; decide the most likely result of a random selection; recognise possible outcomes from a familiar event;

*and also*
put 2-digit numbers in order; recognise a half and a third of a set; use place values of 'hundreds', 'tens' and 'ones' correctly; use coins to make up a specified amount of money; solve simple word problems using + or – with totals up to 20; match a number sentence to a word problem; choose 'left' or 'right'; recognise standard 3D solids such as cylinder and cone; draw a line of symmetry on a simple 2D shape, choose two objects with the same length; recognise standard 2D shapes; read a digital clock; read a simple pictograph or column graph; complete a column graph using given data;

*and also*
recognise numbers less than 99; write numbers less than 20; add two or more small numbers; recognise coins; choose 'longest' and 'shortest' from a group; select the box with the most things in it; select the smallest or largest shape from a group.




ACER Australian Council for Educational Research

1103
0.97 1.08

**Figure A2.4 Student report on the *Numeracy* scale with descriptors and achievement distributions**

## Individual Student Reports

Student achievement was reported, not only quantitatively on interval measurement scales, but also qualitatively through scale descriptors, as illustrated in Figures A2.3 and A2.4 above. Two school reports were prepared for each student, one for *Reading* and one for *Numeracy*, and sent to schools on time to be used by teachers. Following the instructions on 'how to read the student reports' (given below), Figure A2.5 shows a sample student report with a key to reading and interpreting an individual student report.

### How to read the 2005 students' reports

These notes and the sample graphical report apply both to the *Reading* and *Numeracy* assessments.

The data collected in March 2005 have been analysed again, together with the data collected in September. The tests used were matched in difficulty to the various skill levels of the students. Test scores were transformed into scale scores expressed in "**ldr**" units for *Reading* and "**ldn**" for *Numeracy*. The measurement scale is shown at the centre of the report 1.

In the report the one of the two horizontal lines shows a student's achievement in March 3 while the other line shows the student's achievement in September.2 The probable range of achievement is shown by the shaded boxes with the most probable level shown by the horizontal line at the centre 4. If the shaded regions from March and September overlap, it is unlikely that a change in achievement has occurred between the two assessments.

Scale descriptors are shown on the left of the scale 5. The vertical shaded overlapping lines at the right of each set of descriptors show the region of the scale described by those skills 6. A student is likely to be able to do what is described at that level on the scale and below, and less likely to do what is described above.

Only students with results both from March and September have been included in norming. Normative information is shown on the right of the scale 7. The distribution of achievements of students in March and September are shown for each year level. The first line shows the percentile distributions of students in March 8 and the second line for September 9. The small divisions on each line show percentiles and can be used to find the percentage of students with a lower achievement than the achievement of the student named. In the sample graphical report, the scale score of 49 ldn units is higher than the September scale scores of 5% of all Year 1 students. The number of students in each group for *Numeracy* (N) and *Reading* (R) is given in the following table.

| Year 4 | | Year 5 | | Year 6 | |
|---|---|---|---|---|---|
| **N** | **R** | **N** | **R** | **N** | **R** |
| 192 | 189 | 225 | 228 | 278 | 238 |

The 50th percentile of the March achievements for each year has been joined with the corresponding 50th percentile of the September achievements to indicate group changes between March and September 10.

The national benchmarks for years 3, 5 and 7 have been located on the two scales and shown as Y3b., Y5b. and Y7b. 11. Individual achievement as well as group achievement can be assessed with reference to these points on the scale.

## Appendix 2

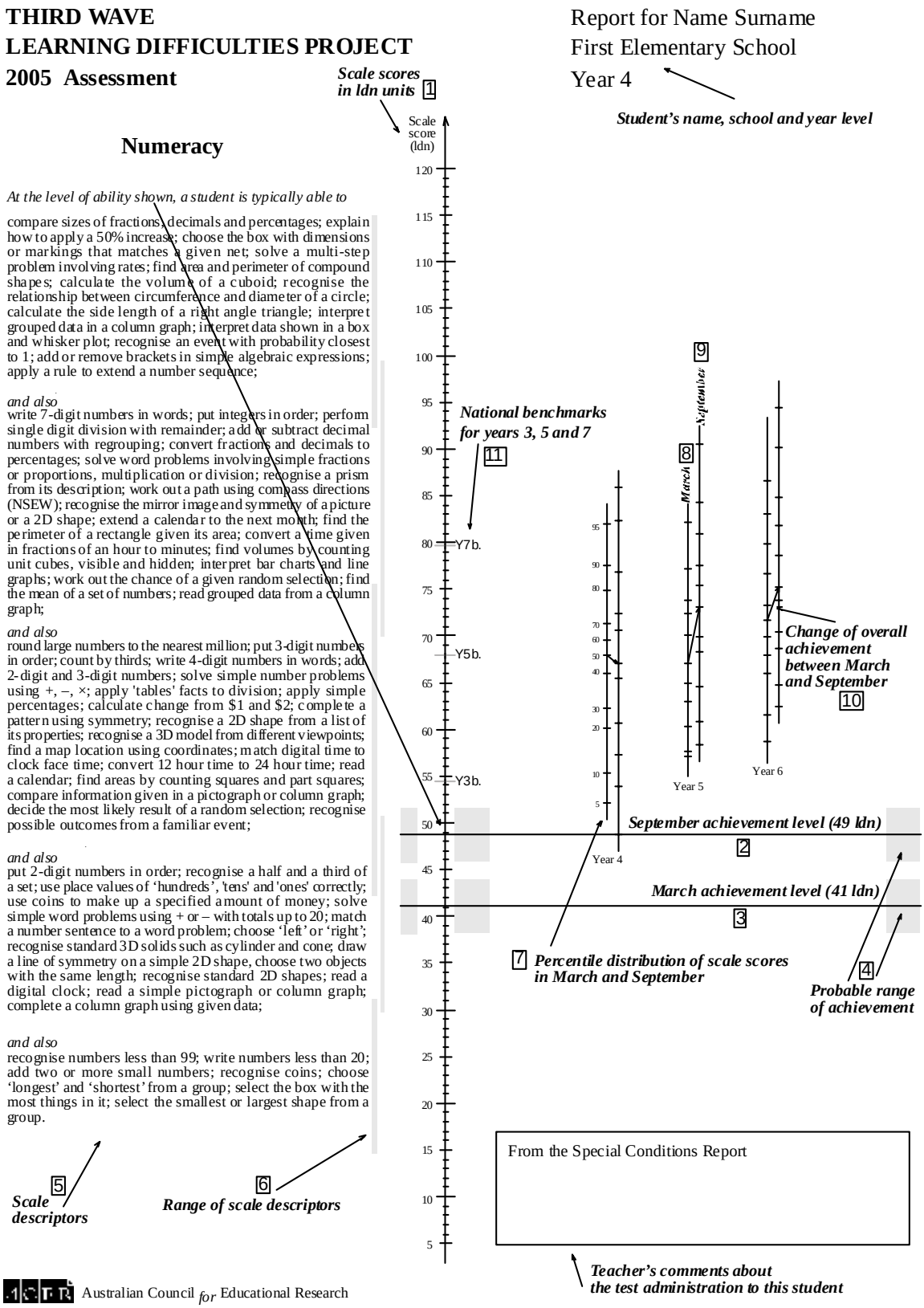



**Figure A2.5 Sample student report with key**

It was important that assessments were developed which allowed monitoring of student progress with a degree of confidence to inform both schools and systems. Such assessments

will also have future applications such as in the measurement of student progress post implementation of intervention programs. The relationships established by ACER with the above organisations, enabled this project to source items at very low cost while maintaining the highest possible psychometric standards. It is notable that whereas measures were taken to ensure quality materials at cost-effective prices, the number of students participating and the number of assessments required to construct the scale, resulted in costs exceeding expectations.

The time lapse between the inception of the project and the actual implementation resulted in materials being more expensive than previously envisaged. Whereas the target number of participating schools fell short of those initially planned, the number of students assessed currently stands was **819**, which is greater than the proposed target of **540**. In addition, the difficulties in constructing a psychometrically sound common scale sufficiently sensitive to detect progress for students with learning difficulties spanning Years 4 to 6, was greater than originally envisaged. The team at ACER did, however, indicate that the work completed was highly valuable to all involved in education in the current climate where there is a focus on valid assessment with references to nationally agreed benchmarks. Moreover, the collaboration from the various contributing parties throughout Australia and New Zealand continues to be a major asset.

Further, using *MLwiN* software, the ACER team have undertaken 'value-added' analyses of the data to identify schools in which students progressed (or regressed) further than expected, given their initial achievements (March data collection) and 'intake' characteristics. From these analyses, **5** schools were identified for qualitative site visits during March 2006, using systematic observation and interview schedules provided in **Appendix 3**. Summary syntheses of these findings are given in **Section 7** of the main report and in the second edition of the WOWW Manual (Hoad, Munro *et al*., 2007).

## Reliabilities of the Reading and Numeracy tests used in the March and September 2005 assessments

Tables A2.3 and A2.4 show test reliabilities for each of the DEST 'Third Wave' tests used in the March and September assessments. Note that *test reliability* is the proportion of the observed test score variance that is true variance (i.e., *reliability* is equal to the ratio of the true variance to the observed variance). In fact, some of the variance in test scores is due to error (the observed variance is equal to the true variance plus the error variance).

**Table A2.3 Reliabilities of March and September 2005 *Reading* Tests'**

| | March | | | September | | |
|---|---|---|---|---|---|---|
| **Test** | RPT2 | RPT3 | RPT4 | RC | RD | RE |
| **Reliability** | 0.84 | 0.82 | 0.82 | 0.86 | 0.86 | 0.86 |

**Table A2.4 Reliabilities of March and September 2005 *Numeracy* Tests**

| | March | | | September | | |
|---|---|---|---|---|---|---|
| **Test** | PM1 | PM2 | PM6 | NC | ND | NE |
| **Reliability** | 0.79 | 0.80 | 0.72 | 0.78 | 0.83 | 0.80 |

Note that these values are satisfactory and reasonably high.