

# Summative and Formative Assessment: Creating a Tool for Improving Learning

Brian Doig  
ACER  
<doig@acer.edu.au>

John Lindsey  
ACER  
<Lindsey@acer.edu.au>

This paper describes part of a continuing program of research and development into improved reporting of student performance in mathematics. The goal of the work reported in this paper is to derive summative and formative information from a single assessment. Developers of multiple-choice test items typically have a good knowledge of the research evidence about children's errors and misconceptions. These common errors are often employed as item distractors. Descriptions of item distractors were made for a multiple-choice test (scored correct/incorrect) to provide descriptions of errors likely to be made by students of different abilities. Through the linking of formative and summative analyses a 'map' was constructed that provides a way of alerting teachers to possible flaws in students' mathematical understanding.

## Introduction

The use of standardised achievement tests for assessing students' mathematical development is well known, and appears to be increasing its dominance in both national and international large-scale assessment programs. Many of these tests use multiple-choice as their response format, a format that critics claim, leaves students' understanding unassessed. More recently, however, attempts have been made to use summative assessment data from these tests to describe developing understanding and provide some formative information.

Mathematics educators, among others, have been provoked to call for achievement assessment to be replaced with alternatives that provide information that is of value for improving teaching and learning. Harlen and James (1997) for example, argue that

there is a need to recognise in theory and in practice the differences in function and characteristics between formative and summative assessment and to find a way of relating them together that preserves their different functions (p. 366)

and Helmke (1995), in his review of approaches to the diagnosis of student needs, concludes that

attempts to analyse students' understanding and their content-specific knowledge structures, including the analysis of errors and misconceptions, are a necessary counter-weight against the widespread use of standardised multiple-choice achievement tests (p. 289).

However, to date, rarely have the summative and formative aspects of students' learning been described simultaneously (see, for example, Doig, Fox, Ryan, & Williams, 1997; Williams, Fox, Ryan, & Doig, 1999). Williams's (2001) suggestion that "the same assessment can serve both formative and summative functions, although in general, the assessment will have been designed so as to emphasise one of the functions" (p. 176) echoes Biggs (1998) who argues that

there is a powerful interaction between FA [formative assessment] and SA [summative assessment] that could usefully be incorporated in an overall synthesis so that both ... are conceptualised within the same framework (p. 106).

Attempts to link both summative and formative information can be seen in the descriptive reporting formats offered for many current tests. An example is shown in Figure 1, where descriptions of what students can do and understand are linked to their (summative) scores.

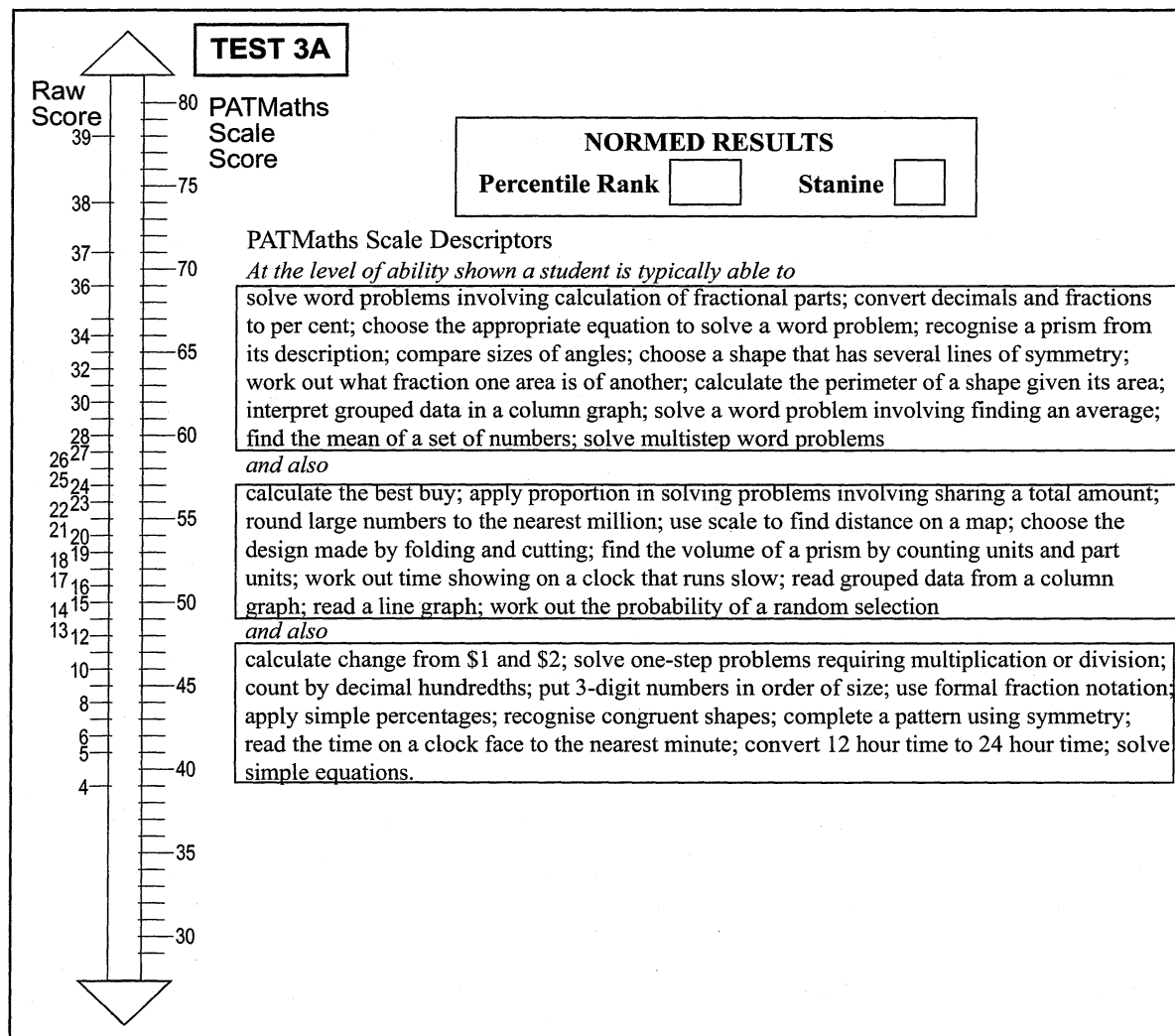


Figure 1. Example of a descriptive report from a multiple-choice test.

This style of report employs a form of scale anchoring (Kelly, 1999) that emphasises, rightly, the positive aspects of student performance. However, we believe that there is more that can be done with the student response data from a multiple-choice test.

In this paper we outline a strategy for producing a summative–formative ‘map’ that we believe points the way forward to achieving the goal of deriving both summative and formative information from a single assessment procedure and reporting the links between these two aspects of assessment.

### Methodology

The basis of this procedure is the revised ACER Progressive Achievement Tests in Mathematics (PATMaths Revised) (ACER, 1997; Lindsey, 1998). The PATMaths is a series of three pairs of two tests, linked to the same scale and spanning the mathematics curriculum of the fourth to ninth years of schooling. The tests contain between 37 and 41

multiple-choice items in number, space, measurement, chance and data, and algebra. As the name suggest, PATMaths is an achievement test, and student achievement is reported on a PATMaths scale across all six tests. The scale was constructed from data collected in a norming study in late 1997. The data were collected from students in all parts of Australia with the exception of the Northern Territory. In all, nearly six thousand primary students and nearly four thousand secondary students were assessed during the norming exercise.

### *Data*

The data for this present procedure are those from the PATMaths 2A test, which is designed for students in years 5 to 8 (in the Victorian numbering system) and covers the content of the National Profiles in Mathematics Levels 2 and 3. Data were selected from the complete PATMaths Test 2A data-set by eliminating all records that were incomplete; that is, contained items for which the student had not provided a response. This was done in order to provide the same number of response data for all items, thus eliminating the possibility of items being estimated as more difficult, due to lack of responses rather than on incorrect student understandings. After the elimination procedure, 1310 student records remained providing complete information on all thirty-nine items in the test. No elimination was carried out based on a student's sex, age or place of origin.

### *Analysis*

The analysis had two phases, quantitative, conducted first, and qualitative.

The quantitative phase of the analysis was of correct–incorrect responses only. Items answered correctly were scored as 1, and incorrectly answered items as zero. The student's total number of correct items is their raw score. All incorrect responses are taken as indicators of lack of achievement (e.g., they are wrong answers) and are scored zero to provide a summative perspective of the student's mathematical achievement.

These scores were used to establish a scale for reporting student achievement, using the Rasch software Quest (Adams & Khoo, 1993) for a simple logistic model. The raw score scale for the test is at the extreme left of Figure 2. Item order on this map is from easier items at the bottom, to more difficult items at the top, and analogously, the lower achieving students at the bottom, and the higher achieving students at the top, of the student distribution.

The Rasch model used for this analysis places a student whose ability estimate is equal (in logits, the scale's metric) to an item's difficulty estimate (also in logits), has a 0.50 probability of answering that item correctly. For example, in Figure 2, a student whose ability estimate is about 1.5 logits (a raw score of 30), has a 0.50 probability of answering items 3 and 6 correctly, and a greater likelihood of answering items below 1.5 logits in difficulty (for example, items 32, 5, and 7). Items whose difficulty is greater than 1.5 logits (for example, items 20, 25, and 39) are less likely to be answered correctly by this student.

In the qualitative phase, the correct answer was ignored and the item distractors examined for evidence of student misunderstandings. This is the formative analysis of the students' responses. The focus of the analysis was on describing the student understanding that led to the selection of that particular distractor. In some instances it was impossible to establish the underlying meaning of selecting a distractor, and so this distractor did not contribute to our knowledge of the student's understanding. An example of this is distractor C in Item 22.

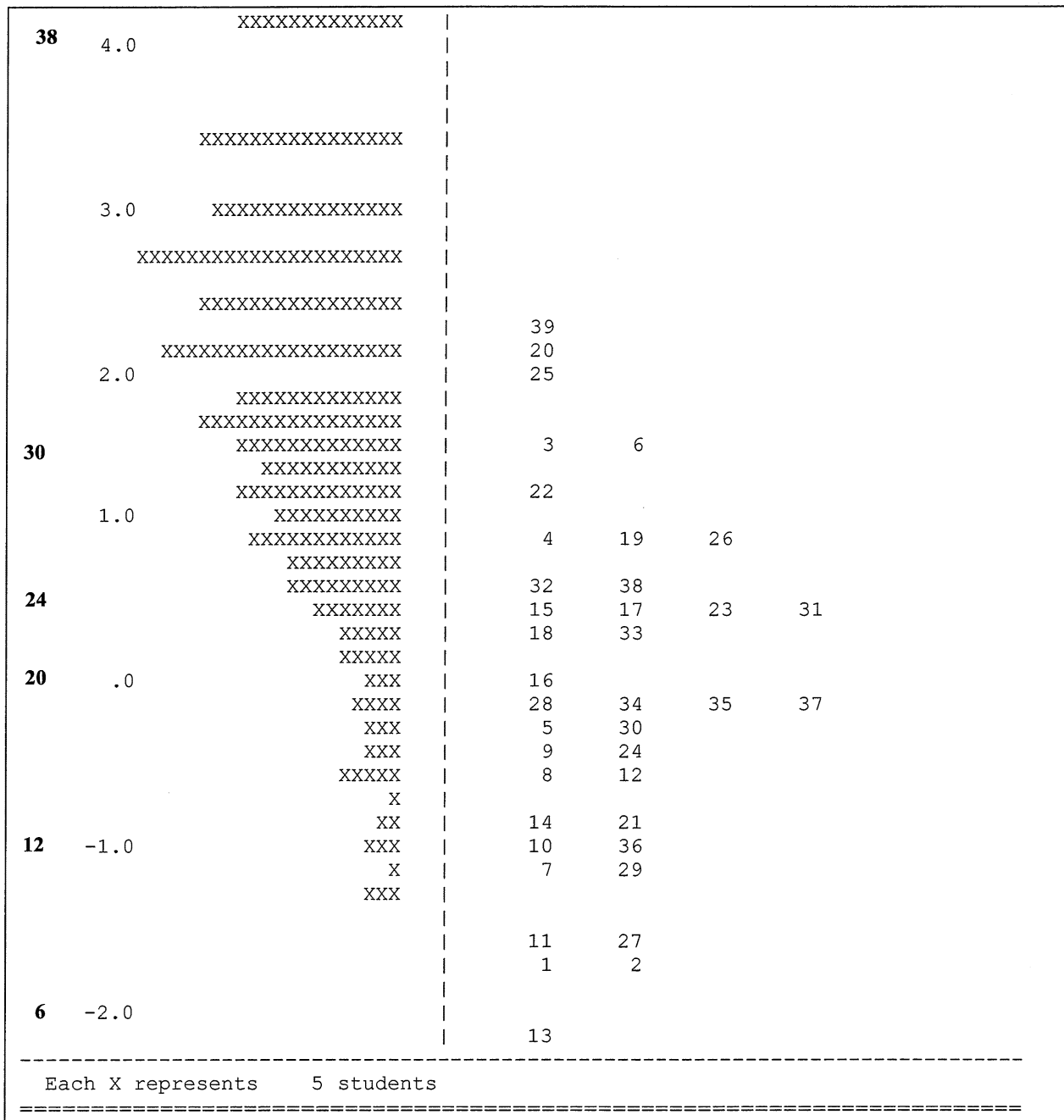


Figure 2. Item map showing student ability and item difficulty on the same (logit) scale.

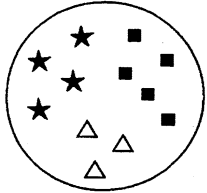
The 39 items in the test address the following mathematics content: number, geometry, measurement, and chance and data. Examples of items from each of these content areas, and the analysis of their distractors, are set out below.

*Item 6 (Number).* The student who has selected distractor A has used part of the designated subset as a fraction of the total in the set (i.e., 4 black stars are one-third of 12 objects), or, used an incorrect subset as a fraction of the designated subset as their answer (i.e., 3 triangles are one-third of 9 black shapes).

The student who has selected distractor B has given part of the designated subset as a fraction of another part of the set as their answer. That is, the 'ratio' of 5 squares to the squares and triangles.

A student who selected distractor C has given part of the designated subset as a fraction of the remaining part of the set as their answer. That is, the 'ratio' of 5 squares to the remaining 7 shapes. distractor D is the correct answer. The student who has selected distractor E has given part of the designated subset as a fraction of the remaining part of the designated subset as their answer (i.e., the 'ratio' of 4 stars to 5 squares). As can be seen from the above analysis, students either misunderstand which sub-sets to use to form the required fraction (distractor A), or have a 'ratio' or 'part-to-part' view of fractions (distractors B, C, and E).

*Item 20 (Space).* A student who has selected distractor A has given the value of the angle corresponding to the 'minutes' position of the minute hand (i.e., 25 minutes gives 25°). Students who select distractor B give a value to the angle corresponding to the time given (i.e., 5:00 gives 50°). Students who select distractor C have given the value that is half of the correct value, most likely because the size of a 'straight angle' is thought to be 90°. Distractor D is the correct answer. Our analysis suggests that students fall into two categories; those who distinguish between angle and time (distractors C and D) and those who do not (distractors A and B).



What fraction of the shapes in this circle is coloured black?

A	$\frac{1}{3}$
B	$\frac{5}{8}$
C	$\frac{5}{7}$
D	$\frac{3}{4}$
E	$\frac{4}{5}$

Figure 3. Item 6.

What is the smaller angle between the hands of a clock at 5:00 pm?

- A 25°  
 B 50°  
 C 75°  
 D 150°

Figure 4. Item 20.

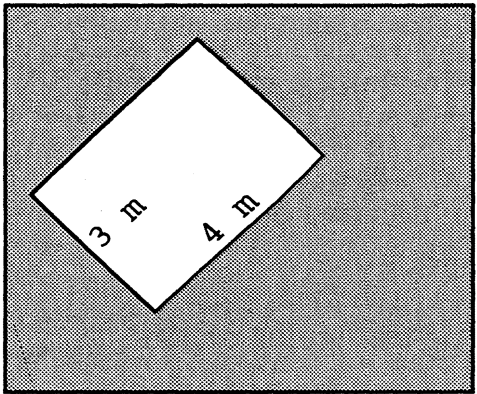
*Item 22 (Measurement).* A student who selects distractor A calculates the difference between the two perimeters (i.e., they confuse area and perimeter). Students who select distractor B use the difference between one perimeter (the larger rectangle) and the area (the smaller rectangle). Again, like distractor A, there is a confusion of perimeter and area.

We are unable to reliably interpret the understanding of students who select distractor C, thus it provides no formative information. Distractor D is the correct answer. Students who select distractor E ignore the 'cut-out' section and calculate the larger area only. It is clear that the first two distractors elicit the well-known perimeter–area confusion, while mis-reading the question would appear to lead to the choice of distractor E.

*Item 32.* Distractor A is the correct answer. Students who select distractor B work out the chance of choosing one object from a set of objects use an inappropriate attribute (i.e., the ball colour rather than number of balls). A student who selects distractor C uses only part of the total set when working out the chance of choosing one object (i.e., choose one ball out of two). Why they should do this is not clear. We found it impossible to infer the reasoning behind the selection of distractor D.

In summary, we found that the selection of any one of the incorrect distractors gave little information about the understanding of the students.

*Item 38.* Items of this type are included in all PATMaths tests and are significant in that students are not permitted to use a calculator for answering them (calculators are allowed to be used for most items).



8 m

What is the area of the shaded region above?

A  $16 \text{ m}^2$   
 B  $18 \text{ m}^2$   
 C  $35 \text{ m}^2$

Figure 5. Item 22.

Robby has four marbles in his pocket: a red one, a blue one, and two green ones. If he takes out a marble without looking, what is the chance that it is blue?

- A.  $\frac{1}{4}$   
 B.  $\frac{1}{3}$   
 C.  $\frac{1}{2}$   
 D.  $\frac{1}{1}$

Figure 6. Item 32.

A student selecting distractor A performs a simple division without using regrouping and ignores the zero in the units place (i.e., zero has no value, so omit it).

Students who select distractor B perform a division that is correct except for the units place (e.g., four cannot divide zero, so ignore this part of the question). An alternative explanation is that these students may only disregard zeroes. Distractor C is selected by students who perform a division that is partly correct, with regrouping, but only regroup once. Again, as with distractor B, an alternative explanation is that these students may simply disregard zeroes. Distractor D is the correct answer.

$4560 \div 4 =$ A 111 and 2 remainder B 114 C 1115 D 1140
---

Figure 7. Item 38

It is likely that, in addition to regrouping problems, students have problems when a zero appears in the problem. This is consistent with the problems caused by zeroes in the other three basic operations with whole numbers.

A similar analysis may be applied to other items in the test and their distractors.

### *Linking the Summative and Formative Analyses*

The critical question that arises from the summative and formative analyses of an achievement test is how do we link the results of the analyses. The use of an Item Response Theory (IRT) quantitative analysis suggests a possible strategy for doing this, as the IRT analysis provides, among other statistics, the mean ability of the students who selected each distractor. That is to say, we can find a position on the summative scale where we may place the distractor description from the qualitative analysis. This position is at the mean ability of students who selected that distractor. For convenience we shall term this the mean ability level for the distractor.

The report, shown in Figures 8 and 9 (next page) have the distractor descriptions placed on the scale at the mean ability level for each distractor. Thus, one can see the relationship between a student's score and their typical choice of distractor. As the qualitative analysis of the distractors provides evidence of a student's mathematical understanding, this is a linking of students' scores and mathematical understanding. In other words, the summative aspects of the analyses are linked to the formative information.

## Discussion

The value of being able to put student responses to individual item distractors on a scale derived from their achievement data lies in the opportunity to expose the likely mis-understandings of students linked to their achievement score.

Establishing a relationship between achievement and understanding creates a pedagogical framework from which suitable teaching and learning strategies can arise. It has been argued that teachers who are aware of their students' mis-understandings are able to teach more effectively (Ryan, Williams, & Doig, 1998) and the strategy outlined in this paper is intended to give teachers that knowledge.

As stated at the beginning of this paper, this is current work, the next step of which is to complete the distractor analysis to prepare a complete framework for PATMaths users.

## References

- Adams, R. J., & Khoo, S-T. (1996). *Quest: The interactive test analysis system*, Version 2. Melbourne: ACER.

- Australian Council for Educational Research (1997). *Progressive Achievement Tests in Mathematics—Revised: Teacher’s Manual*. Melbourne: ACER.
- Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: principles, policy and practice* 5(1) 103–110.
- Doig, B. A., Fox, P., Ryan, J. T. & Williams, J. S. (1997). *Understanding children’s mathematics to raise standards: 1997 response analysis of Key Stage 1 Mathematics test*. Research report summary and technical supplement. University of Manchester: Mechanics in Action Project.
- Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education: principles, policy and practice* 4(3) 365–379.
- Helmke, A. (1995). Diagnosing students’ needs. In Anderson, Lorin W. (ed). *International Encyclopedia of Teaching and Teacher Education 2nd edition* (pp 286-290) Cambridge: Cambridge University Press.
- Kelly, D. (1999). Interpreting the Third International Mathematics and Science Study (TIMSS) Achievement Scales Using Scale Anchoring Unpublished doctoral dissertation, Boston, MA: Boston College.
- Lindsey, J. (1998). *Progressive Achievement Tests in Mathematics – Revised: Norming Manual*. Melbourne: ACER Press.
- Ryan, J. T., Williams, J. S. & Doig, B. A. (1998). National tests: Educating teachers about their children’s mathematical thinking. In A. Olivier & K. Newstead (Eds). *Proceedings of the Twenty-second Conference of the International Group for the Psychology of Mathematics Education*. (Vol. IV, pp. 81-88) Stellenbosch, South Africa: PME.
- William, D. (2001). An overview of the relationship between assessment and the curriculum. In Scott, David (ed). *Curriculum and Assessment: International Perspectives on Curriculum Studies Volume 1* (pp. 165-181). Westport, CT: Ablex Publishing.
- Williams, J. S., Fox, P., Ryan, J. T. & Doig, B. (1999). Error analysis for 7 and 14 year-olds mathematics STAs in 1997: A way forward to integrate National testing and diagnostic teaching? Paper presented at the International Conference on School Effectiveness and Improvement, Manchester, United Kingdom: University of Manchester.



<b>SCORE</b>	<b>DESCRIPTIONS</b>
30	Confuses right- and straight-
28	Does not understand how to form a required fraction
26	Has only a 'part-to-part' view of fractions
24	
22	
20	

Figure 8. Part of a summative–formative map (fractions and angle)

<b>SCORE</b>	<b>DESCRIPTIONS</b>
30	
28	Perimeter–area confusion
26	Performs a simple division
24	
22	Mis-reads question
20	Zero has no value, so omits

Figure 9. Part of a summative–formative map (perimeter and division)