Chapter 5

RASCH ANALYSIS

The previous chapter presented the results of students' responses to the two test papers, using Mayberry's method of assessment of the van Hiele Levels of geometric understanding. In this chapter, a quantitative analysis of the results is presented. The results for the Mayberry items from both Papers I and II have been processed using the QUEST software application (Adams and Khoo 1993) of the Rasch partial credit modelling process provided by Masters (1982). The resulting analysis leads to a more precise estimate of a student's ability than the simple pass/fail scoring of the Mayberry evaluation. In addition, the analysis gives a degree of difficulty ranking for the items by comparing them with the students' success rates, and produces fit statistics that facilitate the identification of poorly discriminating items. The resultant analysis is presented in this chapter in four main sections. These are: overall performance, item analysis including question difficulty, student performance, and, finally, an analysis of van Hiele levels.

Overall Performance

The reliability of the estimates for both items and cases is the proportion of the observed estimate variance that is considered true (Adams & Khoo 1993, p.24). The reliability of the item and case estimates for Paper I were both 0.93 and for Paper II, both 0.92. That the reliability parameters are close to 1 shows that there is good separation between the items and between the cases. Internal consistency parameters (Cronbach's alpha measure) of 0.92 and 0.91 for Papers I and II indicate there is a good fit between the data and the Rasch model, i.e., the items come from the same underlying construct, geometric understanding. The fit statistics for the item estimates are given in Table 5-1, and those for the case estimates are given in Table 5-2. All of the infit mean squares are close to 1, and the transformed infit (Infit t) means are close to zero, which again indicates that the data fit the Rasch model well.

Table 5-1

Fit Statistics for Item Estimates

Paper I				
	Infit Mean Square		**Outfit Mean Square**	
	Mean	0.98	Mean	1.04
	SD	0.18	SD	0.85
	Infit t		**Outfit t**	
	Mean	0.08	Mean	0.26
	SD	0.72	SD	0.81
Paper II				
	Infit Mean Square		**Outfit Mean Square**	
	Mean	1.00	Mean	0.98
	SD	0.22	SD	0.95
	Infit t		**Outfit t**	
	Mean	0.10	Mean	0.21
	SD	0.66	SD	0.69

Table 5-2

Fit Statistics for Case Estimates

Paper I				
	Infit Mean Square		**Outfit Mean Square**	
	Mean	1.01	Mean	1.04
	SD	0.27	SD	1.12
	Infit t		**Outfit t**	
	Mean	0.00	Mean	-0.06
	SD	1.29	SD	0.95
Paper II				
	Infit Mean Square		**Outfit Mean Square**	
	Mean	1.00	Mean	0.98
	SD	0.31	SD	0.61
	Infit t		**Outfit t**	
	Mean	-0.02	Mean	-0.02
	SD	1.14	SD	0.93

Item Analysis

The Rasch method of analysis is based on a comparison between item results, and, hence, there are no comparative statistics (infit mean square and difficulty threshold) for the seven items which had either a perfect score or a zero score. All students were able to name correctly the circle (Item 5), to state a property of the sides of a square (Item 16a), and to list a property common to both a square and a rectangle (Item 25a). No student was able to give relationships of parallel lines with regard to skew lines (Items 38a, 38b and 39), nor to demonstrate understanding of a proof of the angle sum of a triangle (Item 53). As a consequence, these items were removed from the analysis.

Item fit

The infit statistics for each item are the weighted residual-based statistics described by Masters (1982), and indicate quantitatively, how appropriately each item fits the model, i.e., whether each item measures the same underlying construct as the other questions, namely, the students' understanding of geometry. The infit mean square maps produced by the QUEST package show the magnitude of the fit statistic for the items as asterisks, and indicate the acceptable limits for the fit statistics by a pair of vertical dotted lines, items outside these limits being considered aberrant.

The infit mean square maps for Papers I and II are shown in Figures 5-1 and 5-2, respectively. With one exception, Item 8 (Figure 5-2), the items show values sufficiently close to the acceptable limits indicated by the dotted lines, implying that the parameters for each question fitted sufficiently well into the model to contribute to the estimate of geometric understanding (Vine, K. 1997, pers. comm., 6 February). Items showing values greater than the acceptable limit indicate that a larger than expected number of poor performing students were correct for the question, and a larger than expected number of high performing students were incorrect. Items lying outside the right-hand dotted line, although showing values sufficiently close to the limits to be acceptable, are mostly van Hiele Level 1 items, indicating that the better students may have had difficulty in interpreting the thrust of the Level 1 questions. An item showing a lower than acceptable value indicates that a greater than expected number of low-ability students scored correctly, i.e., the item is more highly correlated with the total test score than has been modelled by other questions (Masters 1982, pp.171-172). The value for Item 8 in the infit mean square map for Paper II is very low, lying well to the left of the acceptable limits.

Figure 5-1
Infit Mean Square Map — Paper I

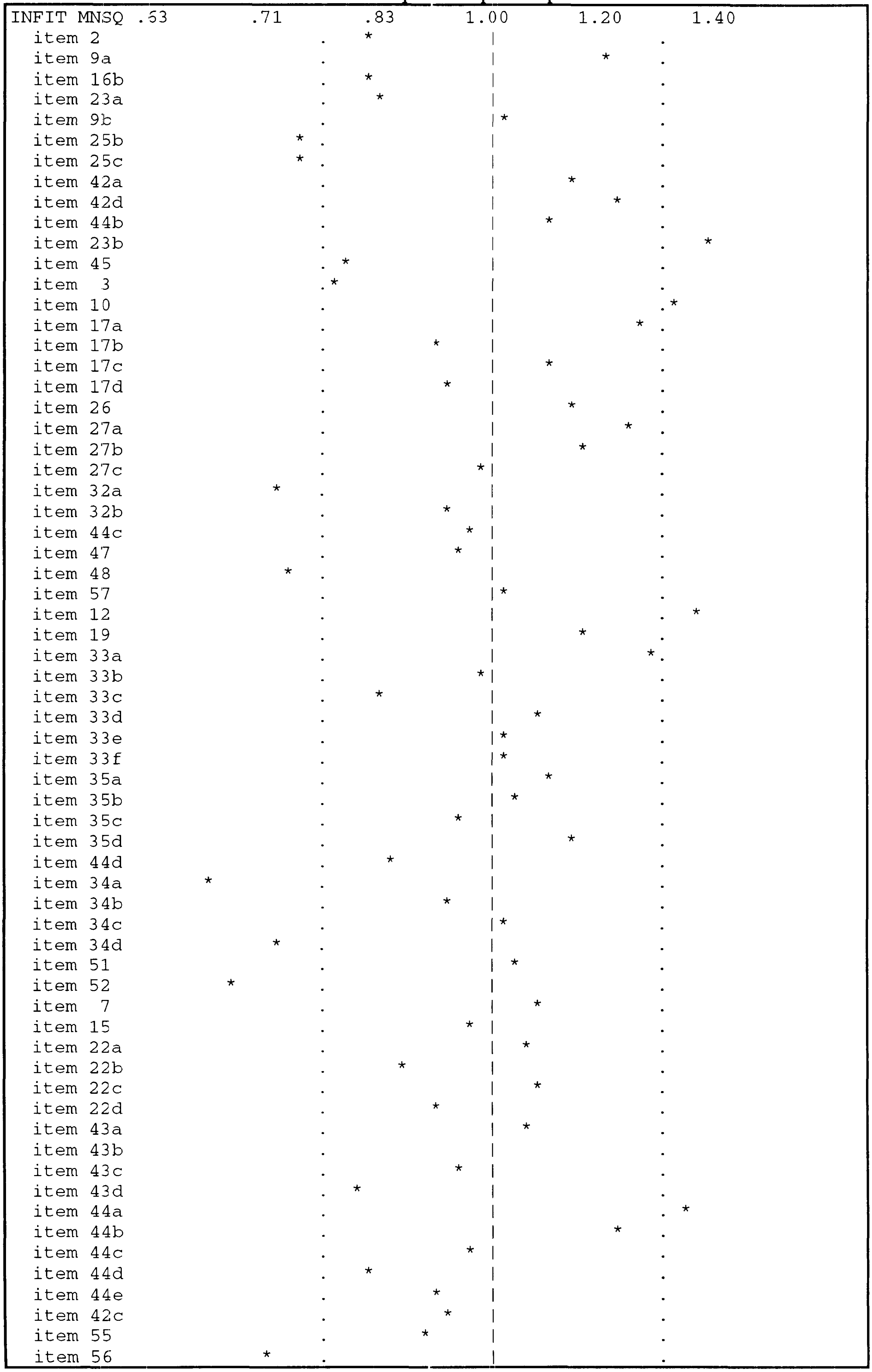
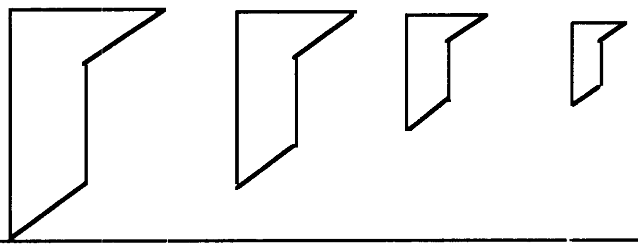


Figure 5-2
Infit Mean Square Map — Paper II

```
INFIT MNSQ   .53       .63       .77      1.00      1.30      1.60
  item 9a                         .   *     |         .
  item 16b                        .         |        *.
  item 23a                        .         |  *      .
  item 24                         .        *|         .
  item 9b                         .         |   *     .
  item 25b                        .   *     |         .
  item 25c                        .      *  |         .
  item 42a                        .         |*        .
  item 42d                        .         |      *  .
  item 44b                        . *       |         .
  item 23b                        .         | *       .
  item 45                         . *       |         .
  item 46                         .       * |         .
  item 4                          .         |         *
  item 11                         .         |         .       *
  item 18a                     *  .         |         .
  item 18b                        *         |         .
  item 28                         .         |         .   *
  item 29                         .*        |         .
  item 30a                       *.         |         .
  item 30b                        .         | *
  item 30c                        .       * |         .
  item 49a                        .         |       * .
  item 42b                        .*        |         .
  item 31a                        .         *         .
  item 31b                        .*        |         .
  item 31c                        .       * |         .
  item 32a                        .    *    |         .
  item 32b                        .        *|         .
  item 48                         *         |         .
  item 50                         . *       |         .
  item 52                         .     *   |         .
  item 6                          .         |         .  *
  item 13                         .         |         .      *
  item 20a                        .     *   |         .
  item 20b                        .        *|         .
  item 37                         .         |     *   .
  item 38c                        .         *         .
  item 38d                        .         |  *      .
  item 38e                        .*        |         .
  item 36a                        .         |   *     .
  item 36b                        .         |   *     .
  item 54                         .         |  *      .
  item 8 *                        .         |         .
  item 14                         .         |         .           *
  item 21a                        .         *         .
  item 21b                        .  *      |         .
  item 21c                        .         | *       .
  item 21d                        .        *|         .
  item 40                         .         |      *  .
  item 41a                        .     *   |         .
  item 41b                        .       * |         .
  item 41c                        .    *    |         .
  item 41d                  *     .         |         .
  item 41e                        .  *      |         .
  item 41f                        .        *|         .
  item 42a                        .         |   *     .
  item 42b                        .      *  |         .
  item 42c                        .         | *       .
  item 42d                        . *       |         .
  item 42e                        .         |       * .
  item 44e                        .       * |         .
  item 58                         .         |*        .
```

Mayberry designed Item 8 to test students' understanding of similarity through their demonstration of ability to recognise and name similar figures (van Hiele Level 1).

Item 8



Are these figures alike in any way?
YES NO

What word describes this?

Only one of the thirty students, S54, failed to give a correct response for this question, resulting in a difficulty threshold (the lowest) of -4.83. A mean threshold of 0.103 for the similarity items indicates that students found similarity relatively more difficult than the other concepts (in QUEST, the threshold mean is set at zero). The QUEST analysis ranked S54 with the lowest performance estimate (-4.34) These statistics indicate that Item 8, in being answered correctly by all except the weakest student, is more highly correlated with the total test score than has been modelled by the other items.

Question difficulty

The threshold is the representation of the item difficulty, and is the ability level that is required for an individual to have a fifty per cent chance of passing that item (Adams and Khoo 1993, p.86). The charts of item difficulty (thresholds) for Papers I and II are given in Figures 5-3 and 5-4, respectively. The figures on the extreme left of each map represent the logit scale on which both items and cases (students) are calibrated. The XXXs on the left-hand side of the map represent the distribution of case estimates (i.e., student estimates) over the logit scale. Each X represents the estimate for one student, and represents a fifty per cent likelihood that the student will answer correctly the item(s) at the same position on the logit value. The -------- line shows the mean estimate for all students. For Paper I, the mean estimate for the students is -0.08, with a range of -2.83 to +2.27, and for Paper II, the mean estimate is -0.67, ranging from -4.34 to +1.79. The items are plotted on the right-hand side of the map according to their degree of difficulty, the higher the threshold, the greater the estimated degree of difficulty for that question. Each item number has a coded subscript which indicates the concept being tested. Additionally, the items are colour coded to indicate the van Hiele level for which each has been designed; Level 1 - magenta, Level 2 - blue, Level 3 - black, and Level 4 - green. The item difficulty estimates for Paper 1 range from -4.19 to +4.02, and for Paper 2, from -4.83 to +3.34.

Figure 5-3

Item Estimates (thresholds) — Paper I

```
  4.0                              | 44crt   47rt   51ci   44cco
                                   |
                                   |
                                   |
                                   |
                                   | 9bsq  45sq  57rt  44dco  44eco
  3.0                              |
                                   | 4?rt  44dci
                                   |
                                   | 5?ci  55co
                               X   |
  2.0                              | 44bsq  43bco  42cco
                            XXXX   |
                               X   | 43dco
                                   | 43cco
                               X   | 42asq  32art  43aco
                                   | 25bsq  25csq  26rt  56co
  1.0                              |

                              XX   | 42dsq  23bsq  44bco

                               X   | 32brt
                                   |
                             XXX   | 44aco
---.0 -----------------------------|---------------------------  -0.08
                             XXX   | 27crt
                               X   | 33fci  22cco
                               X   | 27brt
                               X   | 27art  35cci
                             XXX   |
 -1.0                          X   |
                            XXX    |
                               X   | 17brt  34dci  22aco  22dco
                              XX   | 19ci 33bci 33eci 35aci 34aci 34brt
                                   | 35dci
 -2.0                              | 17drt  35bci
                                   |
                                   | 23asq  33dci      22bco
                                   |       17crt  34cci
                               X   |
                               X   | 17art
 -3.0                              |
                                   |
                                   | 33aci  33cci
                                   |                   KEY
                                   |                   sq -square
 -4.0                              |                   rt -right triangle
                                   |                   ci -circle
                                   |       16bsq       co -congruency
                                   |                   Level 4  green
                                   |                   Level 3  black
                                   |                   Level 2  blue
 -5.0                              |                   Level 1  magenta
```

Figure 5-4

Item Estimates (thresholds)— Paper II

```
                                     |  45sq   46sq   50it   54p
                                     |
  3.0                                |
                                     |
                                     |  24sq   9bsq   36ap   36bp   58si
                                     |
                                     |
                                     |  44bsq   48it   38ep   42dsi   44esi
  2.0                                |
                                X    |
                                     |  37p   41esi   42csi
                                     |
                                     |  28it   42bit   31ait   41fsi   42bsi
                                X    |
                                X    | 25csq   42dsq   23bsq
  1.0                          XX    |
                                     |  25bsq   31cit   38cp   41csi   41dsi
                                X    |
                                X    |  42esi
                                     |  30cit   41asi   42asi
                                     |  42asq   32ait   52it
                                     |  38dp   41bsi
   .0                         XXX    |
                               XX    |  40si
                               XX    |  32bit
                                     |  49ait
-------------------------------------|--     --21dsi--------------     -0.67
                               XX    |
                                     |
 -1.0                           X    |  21csi
                                     |
                            XXXXX    |  18bit
                                X    |
                                     |
                                X    |
 -2.0                          XX    |
                               XX    |  30bit
                                X    |
                                     |  23asq   21bsi
                                     |
                                     |  29it   31bit   20bp   21asi
                                     |
 -3.0                                |         30ait   20ap
                                     |
                                     |  18ait
                                     |                   KEY
                                     |                   sq -square
                                     |                   it -isosc.triangle
 -4.0                                |                   p -parallel lines
                                     |                   si -similarity
                                X    |                   Level 4  green
                                     |                   Level 3  black
                                     |                   Level 2  blue
                                     |  16bsq            Level 1  magenta
```

Each question part was designed by Mayberry to evaluate a student's understanding of geometry for a specific van Hiele level. The question parts were grouped by concept within each van Hiele level, then plotted against their difficulty thresholds. Any clusters appearing in the resultant graphs could indicate patterns worthy of analysis. The graphs are given in Figures 5-5 and 5-6. Each graph shows an increasing degree of difficulty from lower left to upper right, corresponding approximately with van Hiele Levels 1 to 4. However, the change in the degree of difficulty in both figures is much greater between Levels 2 and 3 than between the other levels. There is little discernible difference, and quite a degree of overlap between the degrees of difficulty for the questions set to test Levels 1 and 2, and, again, for Levels 3 and 4. This suggests that the students found the Mayberry questions testing for understanding at Levels 1 and 2 to be similarly easy, while the questions testing for Levels 3 and 4 were found to be similarly difficult. Such an interpretation supports the results in Chapter 4, in which it was found that the most common van Hiele level of understanding for 69% of the students, was Level 2.

Four other features have been highlighted in the figures and deserve comment. First, Figure 5-5 shows a cluster of nine items, indicated by a circle. These question parts were all designed to measure for understanding of the circle at van Hiele Level 3. However, the graph indicates that the degree of difficulty of these nine question parts is similar to the degree of difficulty of the Level 2 items. Second, Figure 5-6 shows a cluster of four items indicated by an ellipse. These question parts, which are across more than one concept, were also designed to test for Level 3 understanding. The four questions require either a simple yes/no response, or ask for recall of factual information. Their degree of difficulty is shown as being similar to that of the middle order Level 2 items, indicating that a yes/no question, or a question requiring the recall of factual information without explanatory support, does not necessarily provide insight into ability to reason at van Hiele Level 3. Third, again in Figure 5-6, Item 11 is indicated, its position and threshold of (-0.65) suggesting that it is much more difficult than other Level 1 items. This confirms the results given in Chapter 4, where is has been shown that several students were unable to identify the equilateral triangle as an isosceles triangle (class inclusion), a Level 3 skill, required by Mayberry to score a correct response for the Level 1 question. Fourth, a rectangle in Figure 5-6 identifies two items, both testing for understanding of necessary conditions in defining an isosceles triangle, yet showing noticeably different difficulty thresholds. This feature is discussed specifically in the analysis of patterns in student responses for van Hiele Level 3.

Figure 5-5
Thresholds of Items for Each van Hiele Level — Paper I

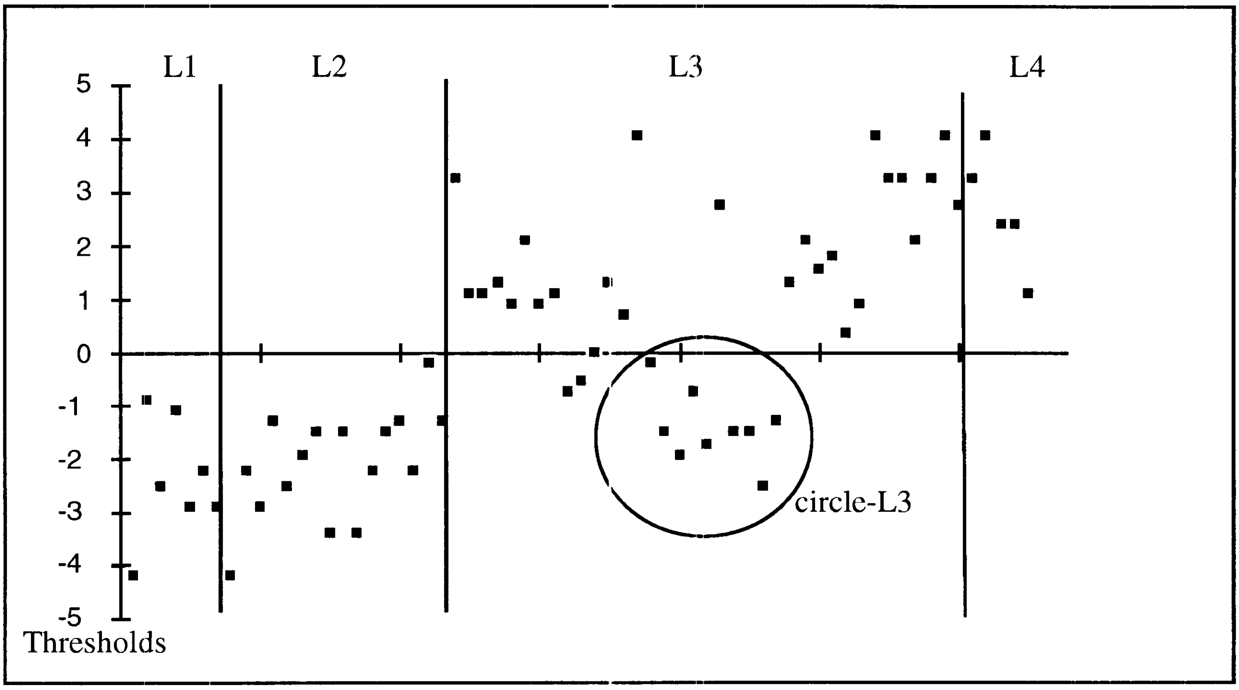
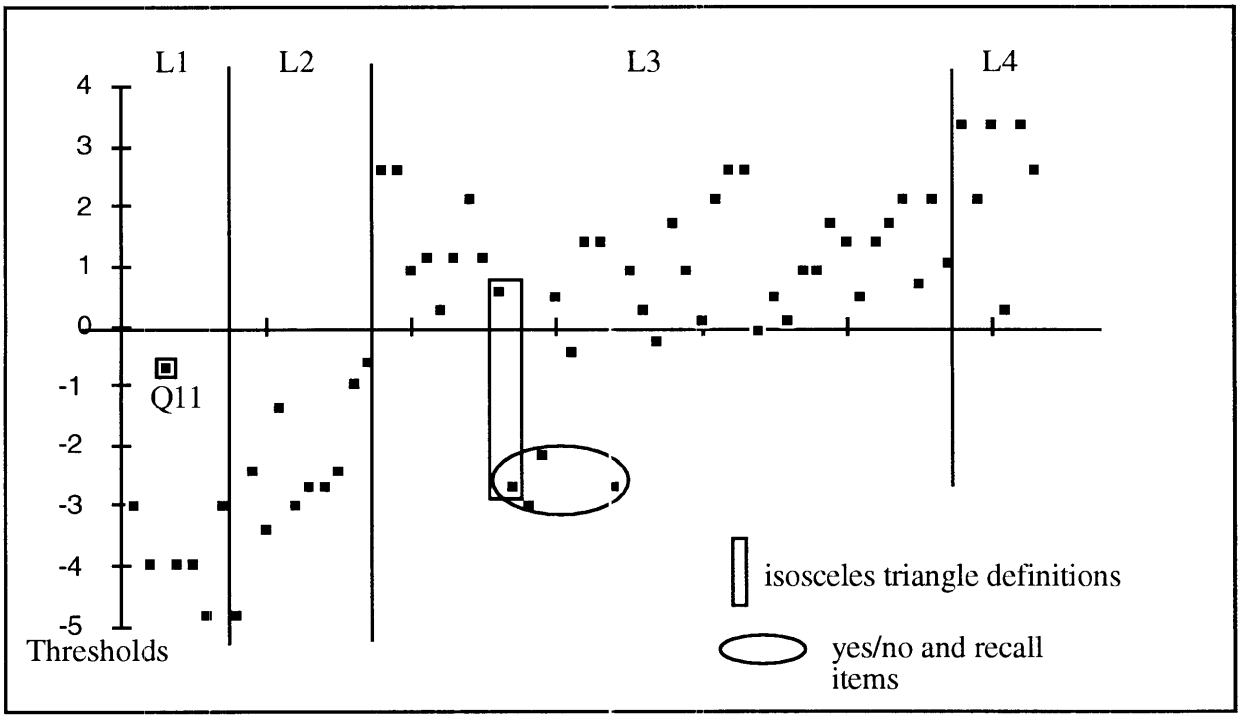


Figure 5-6
Thresholds of Items for Each van Hiele Level — Paper II



In summary, inspection of the statistics produced by the QUEST package showed all except one item to have infit mean square values sufficiently close to the acceptable limits, implying that the parameters for each question fitted sufficiently well into the model to contribute to the estimate of understanding. The plotting of the ordered items against their difficulty thresholds suggests that not all the items may be measuring the van Hiele level for which they were designed. In particular, there appear to be several items designed to test for Level 3 understanding which display a degree of difficulty similar to that of the Level 2 items.

Student Performance

Of the seventy question parts in each paper, the mean test score for students was 32.29 for Paper I and 25.17 for Paper II. With 25 question parts in Paper I and 19 in Paper II being designed to test for van Hiele Levels 1 and 2, these mean test scores indicate that the average student was answering the questions for Levels 1 and 2 correctly, as well as some of the Level 3 questions. This is consistent with the position of the line indicating the mean performance estimate for the students for Papers I and II in Figures 5-3 and 5-4. It is consistent also with the determination in Chapter 4, of each student's most common level of performance. Although 77% of the students did not achieve criteria for levels higher than van Hiele Levels 1 and 2, many of them successfully answered some of the questions designed to test for Level 3 understanding.

Standard deviations, for the mean test scores of 10.37 for Paper I and 9.26 for Paper II, indicate a broad spread of results, confirming the findings in Chapter 4, i.e., some students had difficulty achieving the Level 1 criterion in a concept, while other students demonstrated early Level 4 understanding of geometry. This corresponds with the range of the students' geometric backgrounds.

Patterns in student responses for van Hiele levels

Mayberry (1981, pp.47-49) defined a set of behavioural terms for each van Hiele level, designing her questions to test whether a student displayed understanding which matched the behavioural terms. An examination of the item estimates reveals that, in the analysis of the responses, patterns emerged showing that many students demonstrated a greater familiarity with some behavioural terms or types of questions than with others. Patterns which became apparent on examination of the item estimate maps are detailed below for each level.

The Level 1 questions were designed by Mayberry (p.47) to test whether students could 1) recognise and name figures, and 2) discriminate a given figure from others which look somewhat the same. Item estimates in Figures 5-3 and 5-4 indicate that, in general, students found it easier to recognise and name figures or concepts (e.g., Items 2 to 8) rather than to identify figures (e.g., Items 9a, and 10 to 15). This can be seen by comparing the difficulty thresholds for pairs of items for each of the concepts, namely, square, right triangle, isosceles triangle, circle, parallel lines, congruency and similarity. For example, for the square in Figure 5-3, the threshold for Item 2 is -4.19, and for Item 9a, the threshold is -0.90. The two concepts with different patterns were parallel lines and congruency. In the case of the former, the difficulties were equivalent, and for congruency, they were reversed. In the case of parallel lines, the familiarity with the word parallel (as in parallel parking, train lines) was sufficient to overcome any difficulties in question focus. In the case of congruency, it seems that Item 7 was not well phrased and students were struggling to identify the intent of the question.

Competency with Level 2 skills, i.e., the ability to recognise and name properties of geometric figures (Mayberry 1981, p.48) was tested by another set of items. The questions focused on the sides and angle properties for all concepts except the circle and parallel lines. The results indicate that the students found the notion of properties of sides easier than properties of angles. For example, all students were able to state that a square has four equal sides (Item 16a), but not all could make a statement concerning the angles of a square (Item 16b).

Correspondingly, the thresholds of questions requiring the recognition and naming of properties of sides of figures were lower than for properties of angles for all concepts except right triangle. This is apparent when the thresholds are compared for isosceles triangle Items 18a with 18b (-3.44 and -1.37), congruency Items 22a,b with 22c,d (-1.28, -2.22 and -0.19, -1.28), and similarity Items 21a,b with 21c,d (-2.72,-2.44 and -1.0, -0.65). In the questions concerning the right triangle, the identification of the longest side (threshold -2.91) was easier than the identification of the largest angle, i.e., the right angle (threshold -2.53). However, when students were asked to name the longest side and the largest angle, the naming of the hypotenuse (Item 17b, threshold -1.28) gave more difficulty than the naming of the right angle (Item 17d, threshold -1.95).

In the case of Level 3 items, patterns were less obvious. The students were required to be able to 1) give definitions, 2) recognise and name relationships, and 3) recognise class inclusions and implications (p.48). The formulating of a definition, and the notion of class inclusion tended to be the more difficult questions for Level 3, although the degree of difficulty depended on the phrasing of a question. This is illustrated in Items 28 and 29, which are both designed to test for awareness of a definition of an isosceles triangle.

Item 28

Circle the smallest combination of the following which guarantees a figure to be an isosceles triangle? a. It has two congruent angles. b. It is a triangle. c. It has two congruent sides. d. An altitude bisects the opposite side. e. The measure of the angles add up to 180°.

Item 29

Give a definition of an isosceles triangle.

Item 28 probes for understanding of necessary conditions in the defining of an isosceles triangle, whereas, Item 29 asks for recall of a definition. A rectangle in Figure 5-6 indicates the two items, showing clearly the difference between their difficulty thresholds. Item 28 has a threshold of +1.41, while the much lower threshold for Item 29 is -2.72. The difference shown in the graph between the degrees of difficulty for the two items indicates again that a question asking for a recall of factual information does not necessarily measure Level 3 reasoning.

The effect of the difference in the phrasing of questions is also shown when comparing results for the four questions, Items 41 to 44, designed to test for Level 3 understanding of several concepts, but focusing particularly on congruency and similarity. The questions have a similar structure, each requiring students to work with generalised figures, yet the thresholds shown in Figures 5-3 and 5-4 indicate that students tended to find the two questions, Items 41 and 43 somewhat less difficult than Items 42 and 44. Examination of the questions revealed that for Items 41 and 43, the phrasing is more direct, specific figures are detailed, and several short sentences are used instead of a single complex sentence as in Items 42 and 44. Items 41 and 42 are detailed below for comparison of the phrasing.

Item 41

Triangle ABC is similar to triangle DEF (in that order). Are the following a) certain b) possible, or c) impossible? Give reasons for your answers. a) $AB = DE$ d) $\angle A > \angle E$ b) $AB > DE$ e) $AB = EF$ c) $\angle A = \angle E$ f) $\angle A > \angle D$

Item 42

Will figures A and B be similar	I-always	II-sometimes or	III-never?
Give reasons for your answers.			
	A		B
a)	a square	a)	a square
b)	an isosceles triangle	b)	an isosceles triangle
c)	a Δ congruent to B	c)	a Δ congruent to A
d)	a rectangle	d)	a square
e)	a rectangle	e)	a triangle

None of the four items included figures. Van Hiele commented that questions of the nature of Items 41 to 44, are not testing Level 3 if a figure is given (Mayberry 1981, p.57). However, many students found they were unable to work without a figure. Thirty-eight of the sixty-one students drew figures on their test papers, two of them commenting, *"you have forgotten the figures."* Additionally, thirty students failed to give reasons for their attempts, indicating they were unable to perceive the relationships between properties. Table 5-3 shows a Lewis Carroll diagram relating the two sets of conditions for the students' responses, figures/no figures, and reasoning/no reasoning, and hence, representing the distribution of the nature of the sixty-one students' responses in the written test papers.

Table 5-3

Nature of Students' Responses for Items 41 to 44 (n = 61)

	Figure Drawn	No Figure
Reasoning Attempted	21	10
No Reasoning Attempted	17	13

The ten students who appeared to attempt to give generalised responses, may have drawn figures elsewhere than on their test paper. This was so for S41 as shown in excerpts from the transcription of his interview for Item 41. When asked to work through the question, S41 replied *"Before I even look at the question, I'll draw the triangles."* Following the interview, S41, when asked whether he had drawn figures for the written paper, acknowledged that he had done so on spare paper. In the interviews, all six students drew figures before attempting to answer the questions.

Level 4 items were designed by Mayberry (1981, p.49) to show whether students were able to 1) supply reasons for steps in a proof, and 2) construct a proof. She translated this into questions requiring demonstration of ability to a) generalise in a proof, b) understand a given proof, c) distinguish between necessary and sufficient conditions, and d) construct a proof. The item estimates (thresholds) indicate that students found questions requiring the construction of a proof, especially when an explicit diagram is given as in Items 52 and 56, much less difficult than the other Level 4 questions.

Overall, analysis of the QUEST program statistics supports the original results detailed in Chapter 4, i.e., that most students were able to display competent van Hiele Level 2 reasoning, together with some awareness of Level 3 behavioural skills. Furthermore, the students have shown a greater familiarity with certain features, and types of questions.

Analysis of van Hiele Levels

An examination of the item estimate maps and figures raises a number of questions. These include:

1. Do the levels show a hierarchical order, and are they discrete?
2. Are the questions distributed equally across the levels?
3. Do the Mayberry questions measure the level for which each was designed?

Each of these questions is addressed individually below.

Do the levels show a hierarchical order, and are they discrete?

That the sets of questions designed to test for understanding of each of the van Hiele levels tend to be sequential is highlighted by the colour coding in the Item estimate maps (Figures 5-3 and 5-4). The increase in the mean difficulty threshold of the questions for each level supports other research evidence that the levels form a hierarchy. The overlap of the difficulty thresholds shown in the Item estimate maps (Figures 5-3 and 5-4) for the sets of questions designed by Mayberry to test each van Hiele level suggests that the levels are not discrete. This may be due, in part, to an incorrect allocation of questions to levels by Mayberry. Figures 5-5 and 5-6 suggest a discontinuity between Levels 2 and 3. However, this difference in difficulty thresholds is more likely an indication of the large proportion of students able to demonstrate Level 2 knowledge of properties, but unable to reason with the properties relationally (Level 3).

Overall, although the levels are shown to be hierarchical in nature, the results support the notion that, rather than being of a static nature, the levels are, in fact, of the more dynamic nature, and that the students, as suggested by Gutiérrez, Jaime and Fortuny (1991), might be growing in several levels at the one time.

Are the questions distributed equally across the levels?

The colour coding in the Item estimate maps (Figures 5-3 and 5-4) and the vertical lines indicating the distribution of questions across levels in the threshold mapping in Figures 5-5 and 5-6 both show that Mayberry has designed many more questions to test for Level 3 understanding than for the other levels. More than half of the seventy questions set for each paper are designed for testing Level 3. It may be that Mayberry, in consideration of the geometric background of her subjects, had the notion that most of them would have some degree of Level 3 understanding, and weighted her test design to accommodate this.

Do the Mayberry questions measure the level for which each was designed?

The hierarchical structure demonstrated in the Item estimate maps (Figures 5-3 and 5-4) and the clustering of questions in the item threshold Figures 5-5 and 5-6 suggest that, in general, the items are measuring the van Hiele levels for which they were designed. However, there are some anomalies which merit further investigation.

1. All except one (Item 44d) of the Level 3 circle questions have a negative threshold, similar to those of the Level 2 questions. This shows clearly on the item threshold mapping (Figure 5-5) as a cluster of questions noticeably below the thresholds of the rest of the Level 3 questions. This suggests that these questions may not be measuring Level 3 for which they were designed.

2. Both Item estimate maps (Figures 5-3 and 5-4) show a noticeable overlap between the difficulty thresholds for Level 3 and Level 4. In particular, the positions (and the thresholds) of Items 52 and 56 indicate that these items may not be measuring van Hiele Level 4 for which they were designed.

3. Several of the items designed to test Level 3 understanding are of the yes/no type (e.g., Item 31b) or require recall of factual information (e.g., Item 30a). A cluster of these questions is identified on the item threshold map (Figure 5-6), where their negative thresholds place them well below the thresholds for the rest of the Level 3 items. The low thresholds support the notion that these question types, in not requiring students to explain their responses, offer a weak student the opportunity of recalling information learnt without understanding, and a fifty per cent chance of selecting the correct option.

4. Item 11 (Paper II) shows on the Item estimates map (Figure 5-4) as having a greater difficulty than other Level 1 questions. Additionally, Item 11 falls outside the dotted lines which indicate the limits of fit statistic in the Infit mean square map (Figure 5-2), suggesting that this item could well be testing something different to that for which it was designed. An inspection of the question shows that it requires the students to identify from a group of four, the two isosceles triangles, one of which is equilateral. Mayberry's marking scheme requires that both triangles be identified.

In summary, analysis of the QUEST program statistics supports the notion that the van Hiele levels of understanding geometry are hierarchical, but suggests that the levels are of a dynamic nature rather than discrete. The maps clearly display the inequality in distribution of items across the levels. The QUEST analysis not only confirms the results detailed in Chapter 4, it also helps in the understanding of the anomalies appearing in the earlier results, indicating that not all the Mayberry questions may be testing the van Hiele level for which they were designed.

Conclusion

This chapter employed a (relatively) new technique developed by ACER to provide a detailed quantitative analysis of the data. It involved use of the Rasch partial credit modelling process provided by Masters (1982). This allowed on a single scale, a measure of item difficulty, and of an individual student's ability to achieve success.

The purpose of this work was to undertake a deeper analysis of the test items and the student responses than were available to Mayberry, and not previously undertaken in the area of geometry. Also, the analysis was designed to complement and extend the analysis of results in Chapter 4 by giving a sharper quantitative edge to what had been primarily a qualitative procedure. Of significance are three features. These are:

1. The quantitative analysis confirmed the trends in the results which emerged in the Chapter 4 analysis. These included confirmation that most students were able to demonstrate Levels 1 and 2 understanding, but not higher levels of reasoning, and that not all questions were testing the level for which Mayberry had designed them.

2. Patterns in the responses which had not been perceived, or were not particularly clear in the Chapter 4 analysis, emerged. These included the positioning of many of the Level 1 items outside acceptable limits in the infit mean square maps (Figures 5-1 and 5-2), indicating that higher performing students may have had difficulty in interpreting the thrust of these questions. A pattern made clearer in the QUEST analysis was that the more generalised the Level 3 questions became, the greater was the degree of difficulty experienced by the students.

3. The focus of the analysis in Chapter 4 on how students performed on a written test version of the Mayberry questions led to the identification of inconsistencies in the reasoning of some students. Many of the inconsistencies were associated with the nineteen response pattern errors. The QUEST analysis confirmed the occurrence of these inconsistencies.

The next chapter explores whether there is any pattern to the response errors, and, also, whether there are any unexpected patterns in the results. This leads to an evaluation of the validity of Mayberry's subjective success criteria for each concept and level.

Chapter 6

ANALYSIS OF RESULTS*

The results of student performances on the written version of the Mayberry test were recorded in Chapter 4, and the quantitative analysis of those results, in Chapter 5. Overall, the performance of the students was found to be similar to that of the Mayberry students. However, when collating the results in the study, inconsistencies in some of the students' reasoning emerged. While the Rasch analysis confirmed the inconsistencies, interviews did not appear to clarify them. A search for patterns which could suggest an explanation of the inconsistencies led to the identification of four main features which had the potential to result in an incorrect assessment of a student's level of understanding in geometry. It is this issue which is at the heart of the third research theme identified in Chapter 2, namely, do the Mayberry test items measure the van Hiele levels for which they were designed, and, are the Mayberry success criteria valid? This theme is addressed through the investigation into the four main features detailed below, i.e., the features which have the potential to lead to an incorrect assessment of a student's level of understanding in geometry.They are:

1. incorrect assignation of a level to certain items;
2. unequal treatment of concepts across levels;
3. uneven distribution of questions across levels; and
4. unbalanced distribution of question focus within levels.

A detailed examination of the Mayberry results showed that, although the sample was smaller, the patterns of inconsistencies could frequently be found also in those results. Additionally, many of the response pattern errors appeared to be associated with these features, indicating a probable linkage.

The purpose of this chapter is to focus on these four features identified above. To assist the investigation, these features were translated into four key research questions. These are

1. Do all items measure the level for which they were designed?
2. Is there equivalence in the treatment of the seven concepts?
3. Is there an even distribution of items across the four levels?
4. Is there a balanced focus on skills within each level?

---

* Aspects of this work have already been published, see Lawrie (1993), Lawrie and Pegg (1997).

The nineteen response pattern errors identified in Chapter 4 are discussed in association with the above research questions. Finally, the implications arising from the analysis are discussed, and an amended version of the Mayberry test is detailed.

Research Question 1:
Do all items measure the level for which they were designed?

In designing each of her items to measure a particular van Hiele level of understanding, Mayberry (1983, p.68) remarked that the choice of representative questions for each thought level was difficult. To minimise this difficulty, she had her original questions validated by a number of experts, among them Pierre van Hiele. When later writing about her study and its results, Mayberry (p.68) felt that the representativeness of her revised questions needed still further refinement. In this section, the representativeness of the questions for each thought level are investigated.

When assessing the responses in this study, the replies of the students to each question were compared to the responses to similar questions at the same level, both within that concept, and for the other concepts. Assuming that the students had given their best in answering the questions, this provided an indication of the representativeness of each question with regard to the van Hiele level for which it had been designed.

Most of the items appeared to measure the level which they were testing. This notion is supported by the consistency of the responses, especially within the first three levels, and also by the low occurrence (2%) of response pattern errors. It is also supported by the hierarchical structure demonstrated in the difficulty thresholds for the items and levels shown in the Tables of Item Estimates (Figures 5-3 and 5-4) which measure the degree of difficulty of each question on a logit scale.

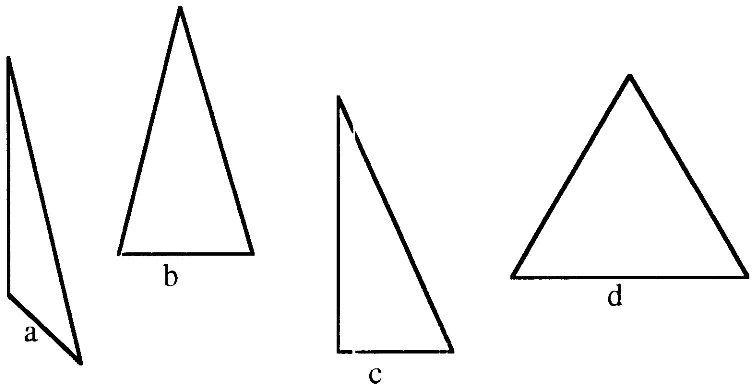
However, some of the items do not appear to be measuring the level for which they have been designed. The main problems appear to be associated with specific items from three levels, namely, Levels 1, 3 and 4. These are Item 11 testing for Level 1 recognition, parts of Items 25, 29, 30, 31, 34 and 35, Level 3 questions which require only simple unqualified responses, Items 38 and 39 which test for Level 3 understanding of parallel lines, and Items 52 and 56 testing for Level 4 understanding of formal proof.

Item 11 (Level 1)

Mayberry (1981, p.76) designed Item 11 to test whether students could discriminate isosceles triangles in a group of examples and non-examples. Identification of both triangles 'b' and 'd' is required for an acceptable response to this item. Level 3 knowledge of class inclusion is implicit to the identification of equilateral triangle 'd' as an isosceles triangle.

This indication of a problem with Item 11 is supported by the Rasch analysis. In Figure 5-6, the position of Item 11 suggests that it is much more difficult than other Level 1 items. The Item estimates map (Figure 5-4) confirms this, showing a difficulty threshold of -0.65, greater than for all other Level I questions. Additionally, Item 11 falls outside the dotted lines which indicate the limits of fit statistic in the Infit mean square map (Figure 5-2), suggesting that this item is testing something different to that for which it was designed.

Item 11



Which of these figures appear to be isosceles triangles?

Fifteen students identified the correct triangles 'b' and 'd', ten students identified only the acute-angled isosceles triangle 'b', and the other five students identified non-isosceles triangles. Of the fifteen students giving the correct response, a few were not able to give any properties associated with an isosceles triangle, indicating a limited knowledge of the concept. However, four of the fifteen demonstrated, later in their responses to the Level 3 questions, an awareness that an equilateral triangle is also isosceles, thus indicating the reasoning behind their identification of figure 'd' as an isosceles triangle.

Most of the ten students identifying triangle 'b' only, were able to reach Level 2 criterion, and gave no indication of knowledge of class inclusion for isosceles triangles. Although this suggests that the group of students saw isosceles and equilateral triangles as distinctly different, it did not necessarily follow. For example, two students identifying figure 'b' only, demonstrated understanding of class inclusion in their responses to Item 31. This is shown in S39's response:

Item 31

Triangle DEF has three congruent sides. Is it an isosceles triangle? Why or why not?

S39 *Yes, it is an equilateral* (triangle) *as it has 3 equal sides. However it also has the properties of an isosceles* (triangle).

The depth of understanding of class inclusion shown in this response raises the issue of why does a capable student have difficulty answering this Level 1 item. Has she interpreted it as a question requiring only a low-level response? This incorrect answering of Level 1 questions by capable students could well lead to response pattern errors. However, this is counteracted by Mayberry's setting the criteria for all Level 1 concepts at a low value of 50%.

Items 25, 29, 30, 31, 34, 35, (Level 3)

Several question parts which were designed by Mayberry to test for Level 3 understanding, namely, 25a, 29, 30(i), 31(ii)a, b, 34a, c, d, and 35(1), (2), (3), require only a simple yes/no response, or the recall of factual information. Understanding at Level 3, as defined in Chapter 1, and by Mayberry (1981, pp.4-5), includes perceiving the significance of the properties established at Level 2. This understanding of the significance of properties cannot be demonstrated in a simple response without explanation of reasoning.

That not all students giving a correct response to such question parts were working at Level 3 is shown in the Rasch analysis. The colour coding in Figures 5-3 and 5-4 identifies them as Level 3 questions (black) with difficulty thresholds which are similar to those of the Level 2 questions (blue) among which they are situated, towards the lower part of the charts. In Figure 5-5, they are shown as a cluster of questions indicated by a circle, and in Figure 5-6, they are shown as the cluster of questions identified by an ellipse. In all four figures (5-3 to 5-6), their difficulty thresholds are similar to those of the Level 2 questions, and well below the difficulty thresholds of the other Level 3 items. This indicates that correct responses to these question parts were given, not only by students able to display Level 3 reasoning, but also by students able to list only sets of properties.

Question parts requesting recall of properties without explanation of relationships, for example, Item 25a. Item 30 (i), and Item 35 (1), (2) and (3), allow for acceptable responses from students who do not understand necessarily the significance of the properties they have listed. A similar difficulty arises with requests for a definition. Dr. Cooney (Mayberry 1981, p.53) maintains that the stating of a definition (e.g., Item 29) may not "represent "real" progress in geometric thinking", i.e., a student has the opportunity of recalling information learnt without understanding. Again, questions which require a yes/no or true/false response (e.g., Item 31(ii) a and b, and Item 34 a, c and d) allow for students to give a correct response without demonstration of understanding of relationships and implications. Additionally, these questions offer a weak student a 50% chance of selecting the correct answer.

The question parts are associated with the concepts square, isosceles triangle, circle and parallel lines. The effect of items testing for the concept circle are discussed specifically in the next research question. For other concepts, for example, the isosceles triangle, although the design of the questions allows for some Level 3 credit that may not be justified, the weak questions are balanced by several questions requiring explanation of reasoning. This is associated, in each case with a discriminating criterion of 67%, and, hence, incorrect allocation of Level 3 has not occurred.

Items 38 and 39 (Level 3)

Mayberry included the condition that parallel lines are coplanar in her requirements for an acceptable response to Item 39 (Level 3), stating "Thirteen of the fifteen students who missed the question about parallel lines omitted 'in the same plane'" (1981, p.83). As recorded in Chapter 4, in the written paper no students included condition (c) in their responses.

Item 39

Circle the smallest combination of the following which guarantee that two lines are parallel. a) They are everywhere the same distance apart. b) They have no points in common. c) They are in the same plane. d) They never meet.

The wording of parts A, B and E in Item 38 implies the same expectation, awareness by students that parallel lines are coplanar. In the responses, no student showed awareness of skew lines.

Item 38

Are these lines or line segments parallel, a) always b) sometimes c) never? Give reasons for your answers. A Two lines which do not intersect. B Two lines which are perpendicular to the same line. C Two line segments in a square. D Two line segments in a triangle. E Two line segments which do not intersect.

Although Mayberry has not given any other indication of her expectations for this level and concept, the coplanar condition could well have been a requirement for all the Level 3 questions for parallel lines (Items 36 to 39).

In her behavioural definition for Level 3 (1981, p.48), Mayberry stated that students would be able to perceive relationships between the properties identified at Level 2. There is no indication in Mayberry's test that properties are to be considered in other than two dimensions. All other items in the test are situated in two-dimensional geometry. Hence, the expectation that students be aware of skew lines without some indication that the 3-D environment should be taken into account, is not reasonable.

This is supported by the responses to Item 39 of the six interviewed students, together with their reactions to prompting with the notion of skew lines. Initially, none of the six students included the coplanar condition. On being shown examples of skew lines within the room, three were not able to consider a third dimension, and two, although acknowledging that skew lines were not in the same plane, could not incorporate this in their responses. The remaining student (S14) was able to recognise the significance of the coplanar condition, reorganise her thinking, and, consequently, change her response. In her written paper, S14 displayed early Level 4 reasoning frequently in her responses, suggesting that Items 38 and 39 are testing Level 4 understanding.

In Chapter 2 levels were described as representing discrete stages of major knowledge reorganisation, and that the learning or knowledge able to be displayed by a student at a new level differs from the knowledge displayed by the student at the preceding level in that the earlier knowledge will have been reorganised The restructuring of knowledge of properties from a two-dimensional environment to three dimensions is a major reorganisation requiring different reasoning. Analysis of Items 38 and 39, together with the students' responses to the interview prompting on skew lines, indicates that the inclusion of the condition of parallel lines being coplanar is not true to Level 3 descriptions, and that Items 38 and 39 are testing the level above the one nominated, i.e. Level 4 instead of Level 3.

Items 56 and 52 (Level 4)

To gain consistency in the questions testing for each level, Mayberry (1981, pp.47-49) developed a description of the level in behavioural terms. For Level 4, this was:

Deduction is meaningful. Reasoning logically and reflecting on the significance of deduction and what it means to organize proofs logically are achieved. Constructing proofs, understanding the role of axioms and definitions and the meaning of necessary and sufficient conditions are achieved. A student on this level should

1) Supply reasons for steps in a proof;
2) Construct a proof.

Mayberry (1983, p.60) then designed three types of questions to evaluate the above two skills, namely:

1. Give reasons for steps in a proof.
2. Steps are given and questions asked about what has been proved.
3. Do a simple proof.

Successful completion of such questions should be an indication that a student can manipulate the intrinsic characteristics of relationships. The suitability of Items 56 and 52 to test for Level 4 skills is evaluated through a comparison with similar items in relation to the above behavioural terms. This comparison is supported by the responses given by the students in the written papers, and, where appropriate, by transcriptions from interviews. The occurrence in the results for the concept congruency at Level 4 of seven of the nineteen response pattern errors is also considered.

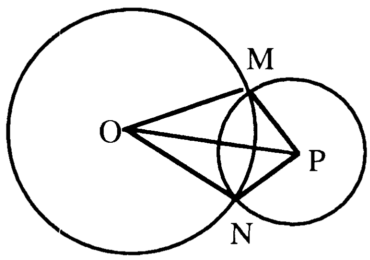
Item 56

The following discussion focuses initially on Item 56. However, to appreciate the full significance of the identified problem, a comparison between the students' performances on Items 56 and 55 is included.

Mayberry designed two tasks to test for Level 4 thinking, in the concept congruence, is measured by the two tasks, Items 56 and 55 provided below. Students are deemed to be reasoning at Level 4 if they meet the criterion of one correct solution. Both items require demonstration of the ability to construct a proof. Results support the notion that there is a marked difference in the degree of difficulty between these items.

This is also supported by the Rasch analysis of the degree of difficulty of the items as measured on a logit scale. The difficulty threshold for Item 56 was 1.11, similar to the degree of difficulty for many Level 3 items, while the threshold for Item 55 was greater at 2.39. This is shown in Figure 5-3.

Item 56

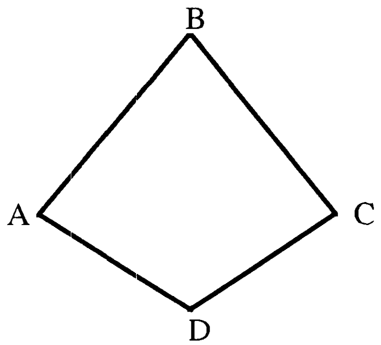


These circles with centres O and P intersect at M and N.

Prove: $\Delta OMP \equiv \Delta ONP$.

The above question includes an explicit diagram delineating clearly the two triangles OMP and ONP to be proven congruent. The solution requires the identification of the two pairs of equal radii (OM and ON, PM and PN) which, together with a common side (OP), result in three pairs of corresponding sides, and, hence, proof of congruency. In the written test (Paper I), nine of the thirty-one students supplied complete solutions for Item 56.

Item 55



In this figure AB and CB are the same length. AD and CD are the same length.

Will $\angle A$ and $\angle C$ be the same size? Why or why not?

There are several ways of solving Item 55. One method includes the use of congruency as a tool, and involves the spontaneous recognition of the need to construct a pair of triangles (by joining BD). The identification of the two pairs of equal sides from the data (AB and CB, AD and CD), together with the common side (BD) gives congruency (three pairs of corresponding sides). This leads to the concluding deductive statement that since corresponding angles in congruent triangles are equal, $\angle A$ equals $\angle C$. Four students gave correct solutions for Item 55 in the written test.

Item 55 was included in the interviews. The responses of the students interviewed supports Mayberry's inclusion of the item as a measurement of Level 4 ability. In the initial request for the students to solve the problem, all six students recorded a response which corresponded with the level of thinking which they had displayed in the written test. Two students again supplied a correct proof, while the remaining students either quoted the properties of a kite or gave a response which implied a visual comparison. Such a response was given by S08:

S08 *Yes, I'd say they'd be the same size because they both have the same lines going into them - they are the same lines. You could measure the sides - the measurements would be the same.*

The weaker four students were then prompted with "*If BD is joined, can you now give an answer?*" It was considered that this prompt should reduce the level of the original question. Three of the students could make no further contribution to the depth of their earlier attempt, demonstrating an inability to appreciate the implications of having a pair of triangles. However, the fourth student, S31 responded "*Congruent triangles. From the question, two sides are the same and BD is common to both.*" He thus demonstrated very clearly that the inclusion of significant prompts reduced the degree of difficulty for the question. During the interview, S31 consistently demonstrated confident Level 2 knowledge, and some Level 3 ability, even though he had not reached the Level 3 criterion for any concept in the written responses. His inability to recognise the need to construct triangles for the original question confirmed that he was unable to reason at Level 4. However, he demonstrated his partial grasp of Level 3 in his immediate 'solution' when prompted. This contrast between S31's inability to attempt Item 55 in its original form, and his solution once prompted with the construction of the triangles, illustrates one difference between questions measuring Levels 4 and 3, and supports the notion that Item 56 with its comprehensive diagram is not testing Level 4.

The inclusion of significant visual prompts, as in the explicit diagram in Item 56, allows only for simple deductions to be undertaken. The effect of such a significant prompt was demonstrated by the increased success rate for Item 55 following prompting during the interviews. By comparison, Item 55, in its original form for the written test, requires a decision on which deductive tool to use in the development of the proof of equality of the angles. This raises the question of whether Item 56 has sufficient complexity to measure Level 4 skills. That there is a very real difference between using an idea when it is apparent and recognising the need for the use of the idea is borne out by the results of the students. Of the nine students who solved Item 56 correctly, only four could solve Item 55 also. With a Mayberry requirement of one correct solution to demonstrate ability for this level for the concept congruency, nine students (29%) registered Level 4 success. This is more than twice the overall success rate (11%) for Level 4 across all concepts.

Seven of these nine students failed to reach Mayberry's criterion for Level 3 in the concept congruency, and, hence, recorded a response pattern error (1,1,0,1), indicating an inconsistency in performance between Levels 3 and 4. All the responses of the seven students were analysed in depth. Only one of the seven students showed consistent Level 4 reasoning across most concepts. The other six students did not show ability to use deductive skills for any other concept. This suggests that the incorrect assigning to Item 56 of ability to measure Level 4 deductive skills could be the reason for the occurrence of six of these seven response pattern errors.

That Item 56 is measuring Level 3 rather than Level 4 is supported by van Hiele's statement (1957, in Fuys *et al* 1984, p.239) that a student will have reached the [third] level of thinking if, for example, "on the strength of general congruence theorems, he is able to deduce the equality of angles or linear segments of specific figures."

Item 52

Item 52 was designed by Mayberry to measure Level 4 skills for two concepts, isosceles triangle and circle. This means that for this study the item was included in both Papers I and II, i.e., all sixty-one students attempted Item 52. Overall, fourteen students (23%) supplied correct proofs for the item, well above the average success rate for the fifteen Level 4 items of 9%. In the Mayberry results, nine of the nineteen students (47%) were successful with Item 52 (Mayberry's Level 4 overall success rate was 25%).

The Rasch analysis supports the notion above that, excluding Item 56, Item 52 is less difficult than other Level 4 items. The difficulty threshold of Item 52 in Paper I is given as 2.39, and in Paper II at 0.26 (Figures 5-3 and 5-4). The higher comparative difficulty for the question in Paper I could well result from its comparison with the difficulty of other questions designed to test for understanding of the circle. Another factor, discussed in the next section, investigates a problem associated with the Mayberry circle items.

Item 52

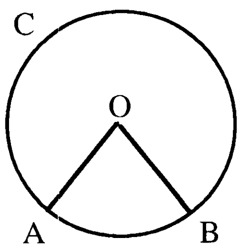


Figure C is a circle. O is the centre.

Prove that ΔAOB is isosceles.

Proof for the item requires two steps: the identification of equal radii AO, OB; and, a concluding statement that a pair of equal sides determines an isosceles triangle. Of the fourteen students assessed as completing this proof successfully, only two provided satisfactory solutions to other Level 4 questions for these concepts. Another two students demonstrated Level 4 skills for other concepts. The high degree of success with Item 52 compared to the general Level 4 success rate suggests that Item 52 may not have as great a degree of difficulty as other questions testing Level 4 skills.

A comparison between the success rates for the questions testing Level 4 skills for the two concepts, isosceles triangle and circle, supports the notion that these questions may have a lesser degree of difficulty than other items coded as Level 4. Three questions, Items 48, 50 and 52 test the concept isosceles triangle and two questions, Items 51 and 52, test the concept circle. As mentioned in Chapter 4, Item 48 requires the construction of a multi-step proof, whereas Item 50 gives the essence of a proof, for which, full understanding is necessary for a student to give a correct response. Three students answered Item 48 correctly, one gave a correct response to Item 50 and ten students were correct for Item 52. This range of results highlights the difference in the degree of difficulty between Item 52 and the other two questions. For the concept circle, Item 51 is again a question of the type which supplies the essence of the proof then asks "What have we proved?" One student gave a correct response for Item 51 and four for Item 52. While not showing as great a difference as between the results for the three isosceles triangle questions, the success rate still fits the pattern of results, supporting the notion that Item 52 is insufficiently abstract to measure Level 4 skills.

The weakness of Item 52 does not show in the response pattern errors. For the concept isosceles triangle, the criterion is two correct responses for the three questions. The greater degree of difficulty of Items 48 and 50 ensures that only students with Level 4 skills will achieve a successful score of more than one correct response. For the concept circle, the criterion for Level 4 requires a correct solution for just one of the two questions. This allows students who are able to give a correct response for only the less demanding question, Item 52, to register success for this concept and level. This would normally give rise to the potential for students who did not have sound Level 3 thinking for the concept circle to record response pattern errors. However, as mentioned above, there is another feature associated with the treatment of the circle concept which overrides the effect of the weakness of Item 52. This is discussed in the next research question.

As noted by Mayberry, the choice of representative questions for each level of thought is difficult. The consistency in the responses given by the students to questions designed to test a particular level supports the notion that most of the items measure the level for which they were designed. However, responses to Item 11, to some Level 3 question parts, and to Items 52 and 56 indicate that they are not necessarily measuring the levels for which they were designed. In particular, the responses to various Level 4 questions suggest that whereas items, such as 48, 50, 51 and 55, are measuring Level 4 skills, Item 52 and 56 do not appear to do so, and, hence, need further refinement. Van Hiele asserted that "A [fourth] level must be connected with the possibility of comparing, transposing, and operating with relations" (1986, p.44), and that a student "will reach the [fourth] level of thinking when he starts manipulating the intrinsic characteristics of relations" (1957, in Fuys *et al* 1984, p.240).

This raises the question of whether Mayberry's interpretation of her behavioural definition for Level 4 allows for the comparing, transposing and operating with relationships. She appears to have overlooked the fact that the construction of a simple proof is demonstration of a Level 3 skill, especially if it is empirically based, or if it does not require operating with relationships. All the items designed to test Level 4 fit Mayberry's behavioural terms above, yet it appears that weaknesses in some of the items, together with a success criterion that has been set too generously, cloud the meaning of the levels.

Research Question 2:
Is there equivalence in the treatment of the seven concepts?

Results for the circle suggested that the students had a much greater understanding of this concept than of any of the other six concepts, 68% of the students reaching criteria for Levels 3 or better, compared with the result across all concepts of 24%. An examination of the Mayberry students' results showed a similar trend, 68% demonstrating success with the higher levels (3 and 4) for the circle compared to a success rate of 49% for all concepts. Table 6-1 shows the comparison between the circle results and the results for all concepts for both sets of students.

Table 6-1
Comparison Between Results for the Circle and Results for All Concepts
in the Australian and Mayberry Studies

	Australian		Mayberry	
	Circle n=31	All Concepts n=244	Circle n=19	All Concepts n=133
van Hiele Levels 0, 1 2	10(32%)	185(76%)	6(32%)	68(51%)
van Hiele Levels 3 and 4	21(68%)	59(24%)	13(68%)	65(49%)

It has been noted already in Chapter 4 that in the interviews, the responses of the students confirmed the results of the written test. Hence, the levels of reasoning demonstrated in the interviews by the three students who attempted the Level 3 circle question, Item 35 in the written paper, were also confirmed. Additionally, the interview responses indicated that the students were more competent in answering the circle question than in responding to the Level 3 questions for the other concepts, five of the six students giving complete and correct answers to most of the question parts in the item.

This raises the issue: Why should the students record higher van Hiele levels of performance for the circle than for the other concepts tested? Is it that the circle is an easier concept to understand or, if this is not so, is the circle taught more effectively and with greater depth than the rest of plane geometry? If neither of these notions are correct, then the Mayberry test needs to be examined for possible reasons. Do the Mayberry circle questions test the level for which they have been designed, and/or are the success criteria realistic? If either of these factors is found to be in doubt, then is there an underlying reason for this?

As shown above, the students demonstrated in both the written test and the interviews, as well as in the Mayberry study, a much greater success rate with the Levels 3 and 4 circle questions than with similar questions for any of the other six concepts investigated. A close analysis of the students' responses, in particular the responses of the interviewed students, provided some insight into whether the students had a greater understanding of the concept circle, or whether their better than expected results occurred because the questions did not test the level for which they were designed. In addition to this, an examination of the settings of the Mayberry success criteria showed whether they were realistic.

When investigating whether the Mayberry items measure the level for which they were designed, and whether the critical values for registering success with levels are valid, for the concept circle, two levels are involved, namely, Levels 3 and 4. The levels need to be analysed separately. There are a few reasons for this. First, the different skills which need to be displayed for each level require questions of a different nature. Second, the inclusion of the circle question, Item 35 in the interviews, gives a deeper insight into the testing of Level 3.

The circle, Level 3

Mayberry used ten question parts distributed through Items 33, 34, 35 and 44 to test for Level 3 knowledge (criterion of 6 out of 10). Items 34 and 35 each contain four question parts, and are the main determinants for the attainment of the criterion of six out of ten to register success at Level 3. An analysis of these two items follows.

Item 34

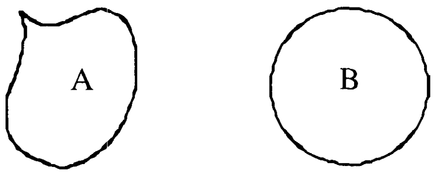


Figure A is a simple closed curve.
Figure B is a circle.

Is figure B a simple closed curve?
How are these figures alike? How are they different?

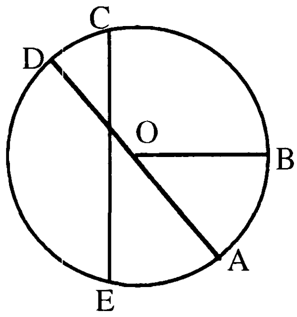
(T—F) All simple closed curves are circles.
(T—F) All circles are simple closed curves.

Item 34 is a repetitive question, asking in four different ways whether the student understands the concept that a circle is part of the class of figures which are simple closed curves. The placement of so much emphasis on a single notion has the potential to lead to an unbalanced assessment of Level 3 knowledge. This difficulty is compounded by the design of three of the question parts. The first part requires a yes/no answer, and the final two parts are true/false questions. Hence, each of the three parts offers the student a 50% chance of success, whether or not the question is understood. Twenty-two of the thirty-one students attempting this question scored for these three question parts. This degree of success is well above the overall success rate for Level 3 questions. The weakness in the design of this item is illustrated by examining the results of one of the interviewed students, S31.

His three correct answers to the above parts (a), (c) and (d) of Item 31 were his only complete and correct responses for the Level 3 questions for the concept circle. In addition, S31 failed to reach criteria for Level 3 for any concept. Responses of this student in the interview reinforced the assessment of his performance in the written paper in that he showed consistent knowledge of properties, and displayed occasionally some attempt at Level 3 reasoning. It can be argued that the weaknesses in the design of Item 34 contributed to the better-than expected success rate for Level 3 in the concept circle.

The selection of the students for interview from three performance strata meant that two of the six had displayed skills in their written responses which were consistently of Level 3 standard or better. The other four students generally responded with answers which indicated Level 1 or 2 reasoning, the two more able of the four students occasionally showing some beginning exploration of Level 3 understanding. However, in the interviews, five of the six students responded correctly to most of the parts of Item 35, indicating that each had the potential to reach Mayberry's Level 3 criterion of six. Correspondingly, this suggests that the students had an equal or greater understanding of the concept circle compared to the other concepts. An analysis of the six sets of interview responses follows. Where appropriate, the interview responses are compared to the written responses for Item 35.

Item 35



This figure is a circle with centre O. Would the following be: a) certain, b) possible, c) impossible. Give reasons for your answer.

1) OB=OA 2) OD=OA

3) 2OB=AD 4) AD=EC

As mentioned in Chapter 3, the six students were selected for interviews so that three had completed the written test for Paper I and three for Paper II. The responses of the students who had completed Paper 2, and, hence, had not previously seen the circle questions, are examined first. As expected, the most able student, S41 confirmed his overall depth of Level 3^+ reasoning in giving comprehensive responses to the four parts of Item 35. This is illustrated in his response to part 4.

S41 *Impossible. The reason is because AD is a diameter and a diameter is the largest line we can draw within a circle, so EC must be smaller.*

In contrast, a somewhat less comprehensive, yet correct response, was given by S33.

S33 *Definitely impossible ... this line here (AD) is going down through the centre, and this one here (EC), well for a start it doesn't go through the centre so it has to be less than the diameter, so it can't be equal.*

In her written paper, S33 consistently demonstrated confident knowledge of properties for all concepts but only partial possession of Level 3 skills. According to the Mayberry test, her four correct answers to Item 35 given in the interview suggest confident Level 3 skills for the concept circle. The third student, S59, who had not seen the circle questions previously, was the least able of the students interviewed and was the only student interviewed who did not give acceptable responses to Item 35. In the interview, she readily identified the circle and displayed knowledge of some of its properties. Her language was informal in keeping with her display of weak Level 2 skills. She referred to a radius as being *"half a circle."* When attempting to explain why two radii were equal in length to a diameter, she said *"You've got two of these* (radii), *they come in close halves. They're equal parts."* This degree of understanding of the properties of the circle matched the level of reasoning (Level 1 to 2) revealed in her responses for the four concepts tested in her written paper in which she displayed repeatedly an incomplete knowledge of the properties of the concepts. Her understanding of angle was particularly weak and led to her making, in some instances, confused statements about angles, and on other occasions, led to her ignoring angle properties completely.

The three students interviewed, who had completed Paper I, and hence, had previously attempted Item 35, were S14 (Level 3+), S31 (Level 2+) and S08 (Level 2). In the interview, S14 gave responses identical to her written answers for Item 35, confirming that she was confident in working with Level 3 geometric reasoning. In the written paper, S14 demonstrated her skill in relating properties of the circle with a comprehensive response to part (d) of Item 44 (testing Level 3):

Item 44

(d)	Will two circles each with a 10cm chord be congruent always, sometimes or never? Why?

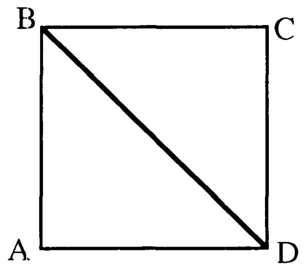
S14 *Sometimes, as different circles* (i.e., with different radii) *may have equal chords.*

Her lack of depth in understanding of Level 4 proofs led to her giving an incomplete response for Item 51. The much easier Item 52 gave S14 no difficulty.

In the written test, S31 achieved Level 2 criteria for all four concepts, including the circle. His interview responses for Item 35 are detailed in Chapter 4. Throughout the interview, S31 confirmed his knowledge of the properties of the concepts, frequently qualifying with statements, such as *"I'm just trying to put a picture in my head",* or allocating 45° values to angles. He was not able to explain clearly any of his statements about properties. This, together with his preference for working visually and/or quantitatively, confirmed his van Hiele level of geometric understanding (Level 2) as assessed in his written test.

The last student, S08 had completed a senior high school course (but no geometry) in the USA before coming to Australia for his tertiary studies. In the written test, S08 reached Level 2 criteria for the concepts square, right triangle and congruency, displaying awareness of most of the properties associated with these concepts. He made no attempt to relate the properties or to demonstrate knowledge of class inclusion. This possession of Level 2 skills, but not Level 3, is shown in his attempt with Item 23, an item testing Levels 2 and 3 for the square.

Item 23



ABCD is a square, BD is a diagonal.

Name an angle congruent to $\angle ABD$.

How do you know?

S08 *How do I know the angles are equal? The square — all the sides are equal and if you put a line in the middle of it, it will make both sides the same, so the angles should be the same.*

This response was given in both the written test and the interview. Neither prompting nor probing elicited any further explanation. This was the best answer he could give. However, for the concept circle, he answered not only all Level 1 and 2 questions correctly, he also reached Mayberry's criteria for Level 3 in scoring seven out of ten. His responses in the interview to the four parts of Item 35 were similar to those he gave in the written test. In answer to part 4 he replied:

S08 *Impossible, because AD goes right through the middle so it would be a lot further than CE because it (CE) goes to the side.*

S08's response to Item 35 shows many similarities to his response to Item 23. Both responses are a statement of known properties. Additionally, both answers imply a visual comparison. In Item 23, S08 found that the two triangles created by the diagonal appear to be "the same", i.e., equal, and in Item 35, S08 appears to have compared the length of the lines AD and CD visually, and found that AD has the greater length. In neither answer is there any attempt to give a logical explanation. This suggests that in the four parts of the circle question, Item 35 can be answered correctly by a student solely possessing Level 2 skills.

As stated above, Items 34 and 35 when considered together, supply eight of the ten question parts used by Mayberry to assess whether or not a student has Level 3 skills for the concept circle. With the criterion set at six correct answers, it can be argued that the weaknesses in the design of Item 34, together with the incorrect assigning of Level 3 to Item 35, give rise to the notion that students can be assessed incorrectly as reaching the Level 3 criterion for the concept circle. This notion is supported in that, of the twenty-one students registering success for Level 3 or better for the concept circle, only four were able to respond correctly to question 44 (d), a question of the type verified by van Hiele (Mayberry, 1981, p.57) as testing knowledge of Level 3 relations between properties.

The circle, Level 4

Five students (16%) reached criterion for Level 4 in the concept circle. Although this is not much greater than the average number of students to reach Level 4 overall (11%), analysis of the responses indicates that these results have the potential to lead to an incorrect prediction of the highest level of which some of the students are capable. Mayberry designed two questions, Items 52 and 51, to test whether students could demonstrate Level 4 skills, i.e., whether they could supply reasons for steps in a proof and whether they could construct a proof. The criterion was reached if a student gave a correct response to either one of the items. Following the detailing of the solutions for each item, the degree of difficulty for each item is discussed in relation to the success rate for each item. Student responses are used to illustrate the main points.

Item 52 requires that a student construct a proof. As shown in the first section of this chapter, there are only two steps in this proof. It is considered that the solution, the identification of equal radii, and the statement that a pair of equal sides determines an isosceles triangle, incorporates the recognition of the relationship between properties (Level 3). There is insufficient abstractness in the problem to allow a student to display ability to construct proof.

Item 52

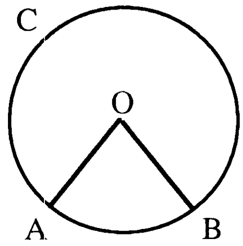


Figure C is a circle. O is the centre.

Prove that ΔAOB is isosceles.

In contrast to the solution for Item 52, a correct response to Item 51 requires a student to understand the essence of deduction so as to be able to follow through the given steps of the proof.

Item 51

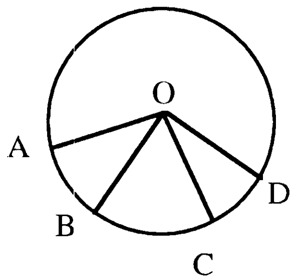


Figure O is a circle, O is the centre.
$\angle AOB = \angle COD$.
Since AO, BO, CO and DO are radii,
$\Delta AOB \equiv \Delta COD$, so AB=CD.

What have we proved?

Only one student was considered to have given an acceptable response for this item.

S19 *We have proved that the chords are equal due to having the same angle at the centre of the circle.*

While S19 did not complete her proof for Item 52, none of the four students who gave acceptable proofs for the item showed any indication that the essence of proof was understood in Item 51. One student (S25) failed to offer any answer, while S14 responded *"These Δs are all isosceles."* S16 and S29 both gave the answer *"that ΔAOB and ΔCOD are congruent."* Of these four students, only one showed evidence of mastery of deductive skills on other items.

There is a very real difference in the degree of difficulty between a question whose solution is the stating of a simple two-step proof, and a question for which giving a correct response requires the understanding of the essence of proof. Only one of the five students successfully attempting the Level 4 questions for the concept circle appeared to understand the intent of Item 51. This is supported by the measurements of the degree of difficulty for each question in the Rasch analysis. The difficulty threshold for Item 51 was 4.02 while the threshold for Item 52 was lower, at 2.39. It can therefore be seen that the prediction that students can demonstrate deductive skills for the concept circle on the basis of a correct solution for Item 52 alone may not be correct. Thus, it can be argued that a small number of questions testing for the understanding at a level can lead to an incorrect prediction of a student's performance for a level and concept. This weakness is magnified if, as in this instance, the criterion is set at a low value, facilitating the possibility of an incorrect assessment of a level.

The argument above has shown that the high percentage of students reaching Levels 3 and 4 in the concept circle is not an indication that they have a greater understanding of that concept. Rather, it is an incorrect assessment of their van Hiele level of geometric understanding. Analysis of the students' responses showed that the knowledge exhibited in their responses for the circle questions was consistent with the level of their responses for the other concepts, i.e., most of the students were giving van Hiele Level 2 answers. This supports the notion that the items designed to test knowledge of the circle were not all true to level descriptions. In particular, many of the questions designed to test for Level 3 understanding did not allow the students to display their knowledge of Level 3 skills, such as the relationships between properties. The analysis led to the identification of three particular design errors in the Mayberry circle questions All were associated with the testing of Levels 3 and 4 knowledge. These design errors are:

1. Repetitive testing of a notion can lead to an imbalance in the assessment of a level. This design fault is illustrated in Item 34.

2. The incorrect allocation of a level to a question can prevent students from being assessed for their true level of reasoning. Items 35 and 52 do not measure the levels for which they were designed.

3. Testing a level with a small number of questions can cause difficulty in assessing a student's level of understanding. This difficulty can be compounded by setting the success criterion too low. Two items, 51 and 52, test the circle for Level 4 knowledge, and one correct solution alone is required to reach criterion.

These design errors indicate that Mayberry may have had difficulty in finding enough questions to test adequately the understanding of the concept circle at the higher levels. This is highlighted particularly in her difficulty in designing questions true to Level 3. Analysis of responses and comparison of performance across all concepts suggest that of all the Mayberry items designed to test Level 3, Item 44 part (d) may be the only question part to do so. That this question does in fact test Level 3 skills is supported in that the students who answered this question correctly, also demonstrated Level 3 skills for the other concepts. Additionally, it has the support of van Hiele (Mayberry, 1981, p.57), who endorsed the inclusion of questions, such as Items 41 to 44, in the determination of Level 3 skills provided diagrams are not included.

In summary, when collating the results for the concept circle in assessing students' van Hiele levels, an inconsistency emerged. The students were coded at a higher level of understanding for the circle than for all other concepts tested. This could not be rationalised in terms of greater experience or familiarity with circles. Analysis of the students' responses together with the Mayberry questions led to the conclusion that the items designed for the circle were not true to level descriptions, i.e., many of the questions were testing the level below the one nominated. In particular, there appeared to be a paucity of Level 3 questions. The incorrect allocation of levels within this concept contrasts with Mayberry's treatment of the other six concepts in which the questions were more consistent with the level descriptions. It indicates that Mayberry may not have been able to find enough questions of an acceptable standard for the circle, and that she had difficulty in comparing the allocation of questions across concepts.

Research Question 3:
Is there an even distribution of items across the four levels?

If a test, designed to assess a student's van Hiele level of understanding geometry across several concepts, is to be valid, certain design factors need to be incorporated. Three of these factors are:

1. The questions need to test each level for each concept.
2. The number and type of questions allocated within a level, together with the success criterion should be consistent.
3. The number of questions and their success criteria across levels should be comparable.

These factors are addressed under the following two sub-headings; the number of questions and criterion for each concept and level, and the response pattern errors.

Number of questions and criterion for each concept and level

In designing her study, Mayberry used a matrix grid for developing tasks by level and by concept ensuring that one or more test items were designed for each cell in the matrix, and that the questions covering the same level would have parallel forms (Mayberry, 1981, p.51). However, the questions are not evenly distributed by level and by concept. Table 6-2 shows the number of test items (and scoring parts) designed by Mayberry for each cell.

Although the number of test items designed for each cell varies between one and seven, many items contain several scoring parts. This has given rise to a wide range in the number of scoring parts per cell. For example, a single question tests the concept similarity at Level 4 while thirteen scoring question parts are used to assess the same concept at Level 3. Overall, for the first four levels, Mayberry (1983, p.60) designed 124 scoring question parts across the seven concepts, 14 at Level 1, 25 at Level 2, 70 at Level 3, and 15 at Level 4. So as to achieve consistency in the testing within a level, Mayberry defined certain question types which were to be included in the test for each concept. "As far as possible, the same type of question was used across concepts" (pp.60-61).

Table 6-2

Matrix Grid Showing Distribution of Mayberry's Questions (and scoring parts) by Level and Concept

	Level 1	Level 2	Level 3	Level 4
Square	2(2)	2(3)	6(9)	2(2)
Right Triangle	2(2)	1(4)	4(7)	3(3)
Isosceles Triangle	2(2)	1(2)*	7(12)	3(3)
Circle	2(2)	2(6)	4(10)	2(2)
Parallel Lines	2(2)	1(2)*	4(9)	2(2)
Congruency	2(2)	1(4)	3(10)	2(2)
Similarity	2(2)	1(4)	4(13)	1(1)*

*criterion is 100%

The variance in the number of questions per cell parallels, in part, the number of question types defined by Mayberry. The question types, together with Mayberry's success criteria are summarised below. The values for the success criteria cited by Mayberry (p.60) are arithmetic mean values. For all except Level 1, Mayberry varied the proportion of correct responses required. For example, Mayberry (p.60) gives the Level 2 mean criterion as 80%, whereas, across the concepts, the individual criteria range from 67% to 100%. The range in the criteria set for the concepts for a level is given after the mean value.

Level 1 (Success criterion: 50% in each concept)
1. Name the figure.
2. Discriminate figures.

Level 2 (Success criterion: 80%, range 67%-100%)
1. Describe properties.

Level 3 (Success criterion: 65%, range 60%-71%)
1. Select properties from a list to guarantee a figure (square, isosceles triangle, etc.).
2. Recognise class inclusion.
3. Identify relations between properties.
4. Describe implications.

Level 4 (Success criterion: 60%, range 50%-100%)
1. Give reasons for steps in a proof.
2. State what has been proved.
3. Construct a simple proof.

Mayberry not only defined more question types for Level 3 than for other levels, she also designed more questions to assess the level, as shown in Table 6-2, resulting in a broader testing of Level 3 than of the other levels.

The general pattern of setting only one or two questions to assess Levels 1, 2 and 4 raises the issue of whether a small number of questions allows students to display adequately the knowledge they might possess for that level. It also raises the notion that the uneven distribution of questions across a concept has the potential to lead to an imbalance between levels within the concept. The setting of the success criterion with a small number of questions may also merit further consideration.

Response pattern errors

Nineteen response pattern errors occurred in the results tabled in Chapter 4. Their distribution, given in Table 6-3, shows that all errors occurred between Levels 2 and 3, and between Levels 3 and 4. No errors were registered between Levels 1 and 2, i.e., all students who achieved Mayberry's criteria for Level 2, had also been successful for Level 1 for that concept.

Table 6-3
Occurrence of Response Pattern Errors Between Levels

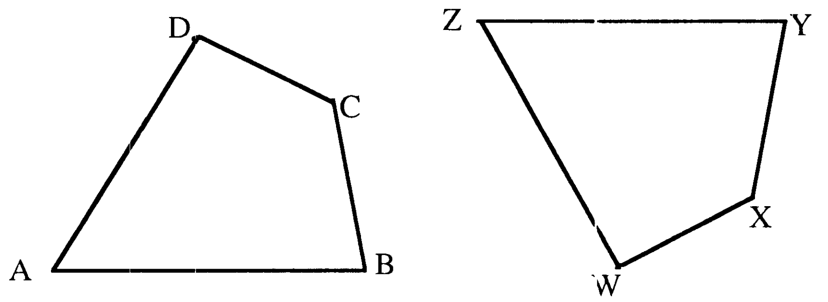
	Between Levels 1 and 2	Between Levels 2 and 3	Between Levels 3 and 4
Square	0	0	3
Right triangle	0	1	0
Isosceles triangle	0	1	0
Circle	0	3	0
Parallel lines	0	0	1
Similarity	0	1	1
Congruence	0	1	7

Many of these errors are associated with occasions when a concept and level is tested by two questions, with a criterion of 50%. However, although for every concept, Level 1 is tested in this way, no response pattern errors occur. This suggests that, while the use of two questions (success criterion: 50%) appears to be adequate in assessing recognition of a concept (Level 1), it does not allow for sufficient comparison across Levels 2, 3 and 4. There is insufficient information to make any further judgement on the testing of Level 1.

Response pattern errors between Levels 2 and 3

Seven students recorded response pattern errors, i.e., although they were unable to reach Level 2 criterion for a particular concept, they answered successfully the Level 3 questions. In each case, the students incorporated their knowledge of properties in giving correct Level 3 responses, knowledge which they failed to display in their responses to the Level 2 questions. For example, S06 missed the descriptive side and angle parts of Item 22 (Level 2), yet her later responses to Item 43 (Level 3) showed knowledge of both side and angle properties, as well as the relationships between the properties:

Item 22



These are congruent figures.

(a)	What is true about their sides?	S06	*they are unequal*
(c)	What is true about their angles?	S06	*they add to 360°*

Item 43

ΔABC is congruent to ΔDEF (in that order).

Are the following a) certain b) possible, or c) impossible?

Give reasons for your answers.

b)	$\angle A = \angle E$	S06	*(b,*	*yes if equilateral*
d)	AB = EF	S06	*(b,*	*only if equilateral*

Analysis of the answers of the students who recorded response pattern errors between Levels 2 and 3 suggests that the errors fit into a pattern. The combination of a small number of Level 2 questions, together with a high success criterion can lead to an incorrect assessment, especially if the student has misunderstood the context of the Level 2 question.

To gain a greater insight into the issue of the degree to which the context of the Level 2 questions had been misunderstood, the responses of all the students were re-examined. The analysis showed that there were fifty-one occasions when success was registered with Level 1 but not Level 2. Among these, on twenty-three occasions students showed in their Level 3 attempts that they possessed some knowledge of properties, i.e., out of 244 possible results (61 students across four concepts), there were indications that on twenty three occasions the context of the Level 2 questions may have been misunderstood. For example, the Level 3 questions, such as Item 28, asking for an indication of which properties of those listed guaranteed a figure, allowed students to register their knowledge of properties.

Item 28

Circle the smallest combination of the following which guarantees a figure to be an isosceles triangle? a It has two congruent angles. b It is a triangle. c It has two congruent sides. d An altitude bisects the opposite side. e The measure of the angles add up to 180°.

Several students who were unsuccessful with the Level 2 questions, marked parts (a) and (c) in their response to Item 28, suggesting knowledge of both angle and side properties. Some of these students failed to show knowledge of angle properties elsewhere, indicating that their response to this item may have been prompted by the wording of the question. However, other students displayed a sound knowledge of properties throughout their responses to Level 3 questions.

The potential to assess incorrectly a student's level of thinking because of a small number of questions may be compounded by the setting of a high criterion. Mayberry set a Level 2 criterion of 100% for two of the concepts, isosceles triangle and parallel lines. Twelve of the fourteen students who failed to reach criteria for one or both of the concepts later displayed some knowledge of the properties. Also, a criterion of 100% does not allow for an incomplete understanding of a level, or for the misinterpretation of the intent of a question. Although one of the aims of the preliminary study was to ensure clarity of the written questions, scoring of the main study indicated the need for a further modification in the wording of a few questions. An investigation into a contributing reason for the anomalies in the responses led to the identification of a lack of clarity in the phrasing in some of the questions for a small sample of students. This is illustrated in the comparison of the following pair of parallel items:

Item 17

Does a right triangle always have a longest side (largest angle)? If so, which one?

Item 18

What is true about the sides (angles) of an isosceles triangle?

The intention of Item 17 is clear, namely, is the hypotenuse the longest side (or is a 90° angle the largest angle) in a right triangle? However, the phrasing of Item 18 allows for nonscoring responses, such as that given by S57: *"It is obtuse"* and that of S35: *"They all add up to 180°."*

The setting of a success criterion of more than 50% for such questions also requires that students be able to display knowledge of both side and angle properties for a figure. This raises the question: How much knowledge of properties is required for a student to be deemed to have van Hiele Level 2 understanding? An analysis of the results for Mayberry's study showed that they followed a pattern similar to the Australian results above. Of the ten Mayberry students not successful with Item 18, six gave five or more correct responses (criterion 8 out of 12) to Level 3 items for the concept isosceles triangle, demonstrating that these students may also have misunderstood the intent of Item 18. Two of these six students reached the criterion for Level 3, and, as a result, registered response pattern errors.

Response pattern errors between Levels 3 and 4

Twelve response pattern errors were registered between Levels 3 and 4. Eleven of the errors occurred for concepts where only two items (with a criterion of 50%) had been set to test for Level 4 understanding. This suggests that being able to give successfully a single formal proof may not be sufficient demonstration of Level 4 reasoning. Additionally, it has already been shown that some of the questions designed by Mayberry to test Level 4 do not appear to be sufficiently abstract. The seven response pattern errors occurring between Levels 3 and 4 for congruency appear to result from the combination in Level 4 of a question of insufficient abstractness, the allocation of only two testing questions, and a success criterion of 50%. Three of the Mayberry response pattern errors appear to have resulted from the same Level 4 combination. However, there are other factors associated with the occurrence of the response pattern errors between Levels 3 and 4. These are discussed in Chapter 7.

Conclusion

The similarity of the results in the two studies reinforces the notion that a small number of questions may not be a sufficient measure of whether or not a student can perform at a given level, particularly when combined with a low or high success criterion. Although Mayberry used a matrix grid for developing tasks by level and by concept, thus ensuring that not only would one or more questions be developed for each cell in the matrix, and, also, that the questions covering the same level would have parallel forms, the test items were not evenly distributed throughout the cells. This has resulted in an imbalance between levels within a concept which has the potential to result in an incorrect assessment of a student's van Hiele level of performance.

Research Question 4:

Is there a balanced focus on skills within each level?

The examination of the van Hiele articles by Mayberry led to her identification of various behavioural characteristics associated with the levels. She then compiled the test questions, each of which was designed to focus on a characteristic. Where several characteristics have been identified for a particular level, as with Level 3, the distribution of the questions is not always balanced, i.e., an over-emphasis on a particular characteristic, e.g., class inclusion, sometimes occurs. Such an imbalance can result in a student who lacks exposure to that particular aspect, to be affected adversely by the structure of the test.

The behavioural characteristics which Mayberry identified for Level 3 were: giving definitions; recognising and naming relationships; and recognising class inclusions and implications (1981, p.48). The final questions testing these behavioural characteristics were of the following types.

1. Which of a given list of properties assumes that you have a square (circle, isosceles triangle, etc.)?
2. Class inclusions (Can a right triangle be isosceles? Why? Two similar triangles will be congruent always, sometimes, never? Why?)
3. Relations (Square A is always, sometimes, never similar (congruent) to square B? Why? Right triangle A with 10-cm hypotenuse will always, sometimes, never be similar (congruent) to right triangle B with a 10-cm hypotenuse? Why?)
4. Implications (In triangle AQB, angle Q is a right angle. What does that tell you about angles A and B? If Q is less than 90°, could the triangle be a right triangle? Why?)

(Mayberry 1983, p.60)

The distribution of the questions across the characteristics is not always balanced. Table 6-4 shows how the scoring parts are apportioned among the question types.

Table 6-4

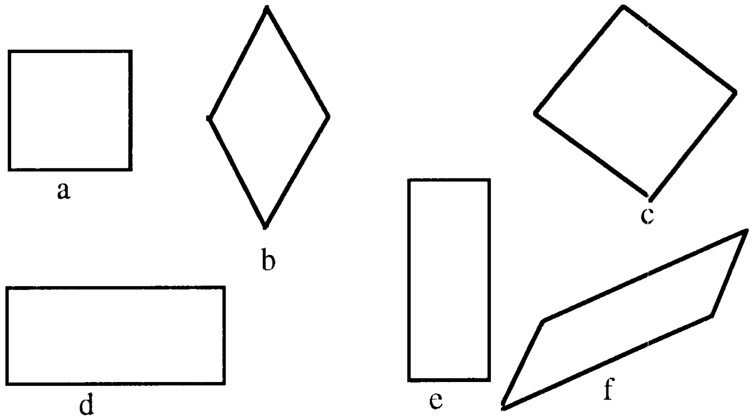
Distribution of Level 3 Question Types for Mayberry Items

	Criterion of Questions	Definition	Class Inclusion	Relations	Implications
Square	6 of 9	1	4	3	1
Right triangle	5 of 7	1	2	1	3
Isosceles triangle	8 of 12	2	5	0	5
Circle	6 of 10	1	4	5	0
Parallel lines	6 of 9	2	0	5	2
Congruency	6 of 10	0	1	4	5
Similarity	8 of 13	1	1	11	0

A comparison of the students' responses with their van Hiele assessments suggests that the imbalance of question types shown above, adversely affected the students' assessment in the concept square. The testing of this concept shows an over-emphasis on the characteristic, class inclusion. Significantly, students' results did not appear to be influenced by a similar ratio of class-inclusion questions to criterion for the concept isosceles triangle. Students' responses showed they were more aware that both equilateral triangles and right triangles can be isosceles than they were aware that a square is a rectangle. This suggests a few possibilities. These include the ideas: that class inclusion for triangles but not for quadrilaterals is included in the teaching/learning process; that, in the 'text-book' position, there is a strong similarity in the physical appearances of equilateral triangles and isosceles triangles; and, that the emphasis which is placed in primary school on the recognition of squares and rectangles as different figures is not always fully refuted effectively in the secondary school.

For the concept square, not only do three of the four class-inclusion question parts (Q9b and Q25b and c) test whether a student is aware that a square is a member of the family of rectangles, the implications question part (Q42d) is dependent also on knowledge of the same notion. These question parts are given below.

Item 9



(b) List all of these figures which are rectangles.

Item 25

(a) Name some ways in which squares and rectangles are alike?
(b) Are all squares also rectangles? Why?
(c) Are all rectangles also squares? Why?

Item 42

Will figures A and B be similar I-always II-sometimes or III-never?
Give reasons for your answers.

(d) A - a rectangle B - a square

Allowing that Item 25(a) is asking for lists of properties (hence is really a Level 2 question), there are still four questions for which a correct answer is dependent on the knowledge that a square is also a rectangle. A student cannot reach the critical score of six out of nine for the concept square without explicitly stating understanding of class inclusion.

In analysing the students' responses to the Level 3 questions for the square, no student demonstrated comprehensive mastery of Level 3 skills, only eight students reaching criteria for Levels 3 or 4. Forty-five students who failed to reach the critical score for Level 3 for the concept square, failed to demonstrate awareness of class inclusion. Of these, ten students displayed appreciable knowledge of the other Level 3 skills, scoring four or five out of the possible nine. This underpins the danger in placing such an emphasis on one skill, such as class inclusion, giving rise to the potential for an under-evaluation of a student's van Hiele level of reasoning. In addition, the very nature of class inclusion as a Level 3 concept is easily restricted with a simple question and response, such as is offered in this test. This is discussed later as one of the implications.

The responses of S58 who recorded a response pattern error between Levels 3 and 4 for the concept square, illustrate this feature, namely, the danger associated with over-emphasis of a particular characteristic. Although failing to identify all rectangles in Item 9(b) and denying that squares are also rectangles in questions 25 and 42(d), S58 indicated Level 3 thinking in several of her other responses:

Item 23(b) *ABCD is a square ∴ all sides are equal. The triangles are congruent and have sides and angles equal in relation to each other.*

Item 24 *A figure is guaranteed to be a square if (c) it has right angles and (e) the adjacent sides are equal in length.*

Item 42(a) *Two squares are always similar because they have the same angles.*

Additionally, S58 gave an almost complete proof for Item 46, deducing that a parallelogram with adjacent sides equal and one angle a right angle is a square. It is significant that although S58 denied on every occasion that a square was also a rectangle, she had no difficulty with the class relationship between a parallelogram and a square.

Altogether, three students, S29 and S41 as well as S58, recorded response pattern errors between Levels 3 and 4 for the concept square. While S58's error has been shown above, to result from lack of awareness of class inclusion, the response pattern errors recorded by S29 and S41 resulted from several of their Level 3 answers being incomplete, i.e., there was insufficient explanation. Both S29 and S41 demonstrated general Level 4 proficiency in reaching criteria for several concepts. The occurrence of these two response pattern errors suggests that a criterion of six correct responses for nine questions may be too high for a written paper, and again raises the issue of how much understanding of the behavioural aspects for a particular van Hiele level is necessary for a student to be said to be at that level. Analysis of the results of the Mayberry study suggest similar cases. Six of the nineteen students, although showing some degree of ability with Level 3 skills (at least three correct responses), have failed with all four of the questions on class inclusion, suggesting they may not have met the concept that a square is a rectangle. Three of the students showed strength in all other Level 3 skills, achieving perfect scores for the remainder of the questions.

These results show that several students from both studies demonstrated appreciable Level 3 skills for the concept square, although being unable to reach the set success criterion. This appears to result most frequently from an undue emphasis on the behavioural characteristic of class inclusion. There is no indication why Mayberry placed so much emphasis on this characteristic. It may well have been that the characteristic is easy to test. Class inclusion is only one of several Level 3 skills and may not be included in the teaching/learning process. Should a student not have been exposed to a particular aspect of a form of reasoning, the Mayberry scoring through an unbalanced focusing on an aspect could assess that student as not having mastery of a level, i.e., over-emphasis of one particular skill has the potential to lead to an incorrect assessment of a student's van Hiele level of reasoning.

Implications

In the preceding four sections, three features stood out, namely:

1. There are no indications that at the higher levels, the concept circle is better understood by students than other concepts, and that, at present there is a vagueness on how the circle fits into the van Hiele levels.

2. The lack of awareness of a behavioural characteristic is not an indication that a student does not possess the ability to reason at that particular van Hiele level.

3. The success criteria for each concept and level do not always appear to be set at a value which allows for the correct assessment of a student's ability to reason at that van Hiele level.

The implications of these three features are discussed below.

1. The circle

There is no evidence to support the notion that the better-than-expected results for the concept circle at Levels 3 and 4 are an indication that students find the concept easier to understand than other common 2-dimensional shapes. Rather, the results have been shown to be attributable to weaknesses in the design of several of the Level 3 and 4 questions. This indicates that Mayberry may have had difficulty in finding questions of an acceptable standard at those levels.

Further, an examination of the topic circle in the mathematics syllabuses for New South Wales that the majority of the students would have followed, suggests that the circle is not given the same emphasis as is given to the other geometric shapes. It is generally studied separately from other 2-D shapes, and at a later stage. For example, in the Space strand of the K-6 mathematics syllabus (NSW Department of Education 1989, pp. 69-94), when the development of the students' awareness of the properties of 2-D shapes is commenced (Unit 7), the emphasis is on the recognition of sides and angles, i.e., on the shapes which have straight sides. The circle is used as a comparison, but enquiry into the circle is not commenced until much later, in the last section, Unit 25, after angles, straight lines and parallelism have all been investigated. This positioning allows for the notion to arise that the circle is not a member of the set of 2-D shapes. Additionally, it suggests that, rather than being an easily understood topic, the circle requires a greater degree of maturity before the student is ready to investigate it. A similar sequence of study is seen in the topic tessellations. Again, the early emphasis is on shapes which have straight sides. Later, the concept of creating tessellations through the modification of the shape of a circle is introduced.

This trend in the positioning of the circle in the order of study of topics can be seen also in the secondary syllabuses. In the geometry section of the syllabus for Years 7 and 8 (Board of Secondary Education 1988, pp.7-43), the circle is considered equally with other 2-D shapes in the recognition and visualisation of shapes. However, the only further investigation into the circle, the identification and naming of the centre, circumference, arc, radius, diameter and sector, is introduced after a detailed investigation of many polygons. The remainder of the content of the syllabus includes classification of triangles, quadrilaterals and solids, an introduction into the concepts of congruency and similarity, and exercises involving deductive reasoning. None of the detailing of the teaching of these suggests the inclusion of the concept of the circle.

In Years 9 and 10 in NSW, provision has been made for the study of mathematics in three levels or courses, General (now Standard), Intermediate and Advanced. The General and Intermediate courses (Secondary Schools Board 1983, pp. 11 and 18) include the revision of parts of the circle, and the discovery of chord, angle and tangent properties of the circle. The Intermediate course includes also the application of properties to simple numerical deductive exercises for plane shapes, among which is the circle. These discovery and exploration segments would all be a development of van Hiele Level 2 and possibly, early Level 3 skills. Circle questions at the Intermediate level have not been included in the external examination for several years. This has resulted frequently in the omission of the teaching of circle properties. The Advanced course contains the formal proof of the geometry of the circle, i.e., is directed at van Hiele Level 4. This follows the Years 7 and 8 course's exploration of a circle, which is primarily directed at van Hiele Level 2. There does not appear to be any recommendation for further exploration of properties of a circle which could lead students into the investigation of relations and implications of the properties, i.e., directed at van Hiele Level 3. The Advanced Course section in Notes on the Syllabus (Secondary Schools Board 1982, p.29) expresses this:

> It is intended that informal practical work in geometry Years 7 and 8 would exclude investigation of properties of the circle. Circle properties should be met initially by students attempting their courses in the context of a set of proofs involving the use of deductive reasoning. All proofs should follow the standard setting out of diagram, data, aim, construction (if needed) and proof.

The investigation into the NSW school syllabuses showed there was less emphasis in NSW schools on the teaching of the circle than of other 2-D shapes. The circle is treated separately from, and explored to a lesser degree than other plane shapes, such as the square, the rhombus and the kite. In particular, there is no provision for the development of Level 3 skills. For example, the NSW Advanced Course (Years 9 and 10) recommends that circle properties initially be met in the context of a set of proofs (Level 4). Overall, this treatment of the circle is likely to result in students being less familiar with the concept.

2. Behavioural characteristics

Mayberry identified clearly the behavioural characteristics for each van Hiele level. However, in designing questions to test for these characteristics, she has not always given a comparable emphasis to the number of questions for each characteristic. There are occasions, especially in the testing for Level 3 understanding, when an unbalanced focus has been placed on some particular characteristic. If, in the teaching/learning process, a student has not met that characteristic, Mayberry's marking scheme can lead to the non-recognition of the student's Level 3 reasoning abilities. Mayberry's research identified that students do not necessarily attain the same level of thinking across all concepts (1981, p.74). An analysis of the results of both this research and of the Mayberry research suggests additionally, that within a concept and level, not all students demonstrate the same degree of understanding across all behavioural characteristics. A lack of recognition of one characteristic should not be taken as an indication of inability to reason at that level. Pegg (1992, p.24) in his investigation of recent research into properties of levels, discusses this issue with regard to the aspect of class inclusion at Level 3:

> It is not sufficient to say that a student is not at Level 3 if he/she does not believe a square is a rectangle. Class inclusion is not simply a part of a natural mathematical development. It is linked very closely to a teaching/learning process. It depends upon what has been established as properties. ...The main feature of Level 3 should not, in my view, be the acceptance of class inclusion but the willingness, ability and the perceived need to discuss the issue.

In not allocating a comparable emphasis across behavioural characteristics within a level and concept, Mayberry's marking scheme has the potential to fail to acknowledge a student's general ability to reason at a level. This perceived weakness gives rise to the question: Are there more reliable forms of assessing students' van Hiele levels of understanding? Such a question encourages the need to investigate alternative methods, such as that developed by Gutiérrez, Jaime and Fortuny (1991). This issue is taken up in the next chapter.

3. The success criteria

In her summary, Mayberry (1983, p.68) stated that her study was, in part, limited by the subjective determination of the success criterion for a given topic and level. Mayberry (1981, p.98) commented on the setting of the criteria:

> If the test were perfect, a student should answer all questions for a concept on a given level to be given credit for achieving that level. However, when the questions were examined, the investigator and the graduate assistant decided that credit for attaining levels should be given by meeting certain criteria for each concept and level.

The success criteria set by Mayberry for each level and concept range from 50% to 100%, most commonly lying between 60% and 80%. The results indicate that most of the criteria have been set at a realistic value. This corresponds with the use in science of 70% as a commonly accepted figure in criteria-referenced testing. This research has identified two features concerning the setting of the success criteria by Mayberry, which have the potential to lead to the incorrect assessment of a student's van Hiele level of understanding of geometry.

1. The setting of a very low or very high criterion may not lead to the correct assessment of a student's van Hiele level, particularly if the low criterion is associated with the asking of too few questions for the concept at that level.

2. There are indications that sometimes the success criterion for a written paper should be different from that for an interview situation.

The effect of success criteria set at extreme levels

The hierarchical nature of the van Hiele levels suggests that the higher a level is, the greater its abstractness is likely to be. This is supported by the level characteristic that features which are implicit at one level become explicit at the next. Mayberry's behavioural descriptions for Levels 1 to 4 show an increasing degree of abstractness. That the levels should not be treated equally is reflected in her acknowledgment that the choice of items for the lower levels is easier (Mayberry 1981, p.97). This is also shown in that there are indications that the choice of the numbers of questions and the success criterion necessary to satisfactorily test each level also differ.

A success criterion of 50% (one correct response for the two set questions) has been set for all of the concepts at Level 1, as well as for many of the concepts at Level 4. The results for Level 1, although not showing any incorrect assigning of cognitive levels to students, are not considered sufficient to justify a general conclusion. However, the analysis of the results in this chapter indicates that Mayberry found difficulty in choosing Level 4 questions which are true to the level description, and, in conjunction with the setting of the low criterion of 50% (with two items), has resulted in the occurrence of several response pattern errors. This indicates that the setting of one or two questions for Level 4 does not necessarily allow students to demonstrate adequately the knowledge they possess. Hence, this cannot be considered to give a reliable indication of Level 4 ability. In comparison, the setting of three questions, together with a criterion of at least two correct responses (67%) for the testing of each of the triangle concepts, appears to have resulted in a more reliable assessment of a student's ability to reason at Level 4.

A criterion of 100% has been set twice at Level 2 and once at Level 4. Mayberry has argued (1981, p.98) that if the questions designed to test for a level are perfect, then a student working at that level should score 100%. However, she also acknowledged that:

> There is always the possibility that different questions from the domain of possible questions or different wording of the actual questions would have elicited a different response pattern. (pp. 97-98)

The analysis of the various responses supports the notion that one student may interpret the wording of a question differently from another. This is illustrated in the testing at Level 2 of the isosceles triangle. The allocation of a criterion of 100% for the concept and level appears to have led, through students not perceiving the examiner's intention in the question parts, to the incorrect assigning of a level for some of the students. Hence, the setting of a very high or very low criterion may lead to the incorrect assessment of a student's van Hiele level of understanding of geometry.

Success criterion for a written paper versus an interview situation

Before the students in this study commenced the test, it was emphasised that it was important for them to show in their responses what they were thinking, and they were urged to give their best attempt. The interviews confirmed that, in general, the responses the students had given in the written test were a good expression of their thinking at that particular level.

However, there were occasions on which the answers were incomplete. For example, S29 and S41 both registered response pattern errors between Levels 3 and 4 through failure to give all details in their answers to Level 3 questions for the concept square. However, both showed competency with Level 4 questions for several concepts including the square.

An analysis of the responses of all the students showed that at some concepts and levels, several students showed appreciable knowledge of the skills being tested, yet did not record all they knew for all questions, and, hence, did not reach Mayberry's criterion. This was particularly evident in the responses to the Level 3 questions. This gives rise to the notion that, whereas in an interview situation one can probe for a complete response, in a written test one can only evaluate what the student is prepared to write. Students may not be fully aware that they have not recorded all their thoughts. This suggests that, for questions requiring lengthier explanations of reasoning as occur when testing at the higher levels, a lower criterion may be more appropriate for a written test than for an interview situation.

Overall, the comparison of the detailed analysis of the responses of students with their van Hiele level of reasoning as determined using Mayberry's method of assessment, indicates that there is need for:

1. further knowledge of the teaching/learning processes in the circle, particularly for van Hiele Level 3 skills;
2. further understanding of what is meant by a student's ability to discuss an issue at a level, and how to assess that ability;
3. further investigation into the balance between achievement and non-achievement for each level and concept under various conditions.

Mayberry (1981, p.100), in part, also acknowledged these needs for further knowledge, suggesting:

> A study could be done with fewer topics and more questions per level. This procedure would avoid the problem of assigning a student to a lower cognitive level than the student actually possesses because she just happened not to know answers to one or two questions worded in a particular way.

Summary and Amended Version of the Mayberry Test.

When collating results in the assessment of van Hiele levels in the study, inconsistencies in some of the student's reasoning emerged. Many of the response pattern errors seemed to parallel the inconsistencies, indicating a probable linkage. Interviews did not appear to clarify these inconsistencies. An analysis of the responses resulted in the identification of patterns of inconsistencies within and across the different concepts and levels. This led to the notion that certain aspects of the Mayberry items had the potential to lead to the incorrect assessment of a student's level of understanding in geometry. The aspects appeared also to be associated with two of the design problems which Mayberry (1983, p.68) acknowledged as needing further investigation, viz., the need to refine the choice of representative questions, and the need to evaluate the success criteria. In particular, four main features were found to account for major problems to the test validity. They were:

1. incorrect assignation of a level to certain items;
2. unequal treatment of concepts across levels;
3. uneven distribution of questions across levels;
4 unbalanced distribution of question focus within levels.

In this chapter these four features have been analysed in depth. Patterns in the students' responses have been detailed, links with the response pattern errors have been identified, similarities with the Mayberry results have been noted, and support from the interviews documented. Implications arising from the analysis have been discussed. This has resulted in the explication of five design problems. These are:

1. Certain Mayberry items do not appear to be measuring the level for which they were designed; either the items need further refinement, or new items need to be designed to replace them.

2. The allocation of levels for the circle questions contrasts with the treatment of the other concepts. The difficulty Mayberry experienced in designing suitable questions may be due to there being a lack of knowledge of how the circle concept fits into the van Hiele levels.

3. Incorrect assessment of van Hiele levels may result from the distribution of questions across levels. A small number of questions may not be a sufficient measure of whether a student is at a certain level. This can be compounded by the setting of an unrealistically high or low success criterion.
4. That a student has not been exposed to an idea can be compounded by an unbalanced distribution of question focus within a level, leading to an incorrect evaluation of the student's understanding of that topic. Assessment of a student should be whether he/she can reason at a level and discuss the issues arising therein.

5. Although the Mayberry items have translated into a satisfactory written test, the assessment of written responses needs to be considered further. Criteria which are satisfactory in an interview situation are not necessarily correct for a written paper.

Accommodation of design problems 3, and 4 as stated above, would require a major reconstruction of the Mayberry test and the writing of new items. Problem 5 needs further investigation. However, problems 1 and 2 can be accommodated through the following two sets of changes to the original test instrument. The application of these changes led to the development of an amended Mayberry test and marking scheme.

1. Items and question parts for which there has been shown to be an incorrect assignation of a level were re-classified. This involved the re-assigning of question part 25(a) and Item 35 (all four parts) from Levels 3 to 2, and of Items 52 and 56 from Level 4 to testing for Level 3 understanding. The assessment of the isosceles triangle at Level 1 was adjusted, allowing credit for the correct identification of one triangle in Item 11.

2. Repetitive and non-assessable items and question parts were eliminated. Ambiguous questions were restructured or eliminated. This resulted in the elimination of question parts 9(b) (ambiguous), 25(c), 31(b) and 34(d) (all repetitive). Items 1, 29 and 49(parts b and c) were eliminated since they were not assessable. Items 26 and 28 (ambiguous) and 58 were reworded.

The resultant amended Mayberry Test (see Appendix G) retains the essence of the original Mayberry test instrument, yet eliminates the difficulties caused by two of the design faults. Marking sheets were developed to match the amended test (see Appendix H). The students' original responses were then re-assessed allowing for the changes above. The full amended results are listed in Appendix J. A summary of the results is given below in Table 6-5.

The summary of the amended results shows that highest van Hiele level of understanding displayed by most of the students is still Level 2.

Table 6-5

Amended Results Summarised as the Highest Level Reached by Students for Each Concept

Concept (n)	No Level	Level 1	Level 2	Level 3	Level 4
Square (61)	0(0%)	2(3%)	51(84%)#	4(6.5%)#	4(6.5%)
Right Triangle (31)	1(3%)	6(20%)	17(54%)	6(20%)	1(3%)
Isosceles Triangle (30)*	2(7%)	8(27%)	12(40%)#	5(17%)#	3(10%)#
Circle (31)	0(0%)	5(16%)#	13(42%)#	12(39%)#	1(3%)#
Parallel Lines (30)	0(0%)	5(17%)	24(80%)	0(0%)	1(3%)
Congruency (31)*	0(0%)	11(36%)#	14(45%)#	2(7%)#	4(13%)#
Similarity (30)	0(0%)	13(43%)	12(40%)	3(10%)	2(7%)

Change in students' results

* Total of 101% due to effect of rounding

Although five design problems were identified, only two were able to be accommodated in the amended Mayberry test and re-assessment. This meant that seven of the original response pattern errors were eliminated. Table 6-5 shows the effect of the changes. However, twelve incorrect assessments were not able to be remedied since other problems could not be accommodated in the amendments.

Table 6-6 lists the occurrence of both the original and the remaining response pattern errors after the adjustment of the Mayberry items.

Table 6-6

Comparison of Occurrence of Response Pattern Errors Between Levels for Main Study results and Amended Set of Results.

	Between Levels 1 and 2		Between Levels 2 and 3		Between Levels 3 and 4	
	Main study	new set	Main study	new set	Main study	new set
Square	0	0	0	0	3	3
Right triangle	0	0	1	1	0	0
Isosceles triangle	0	0	1	1	0	0
Circle	0	0	**3**	**0**	0	0
Parallel lines	0	0	0	0	1	1
Congruence	0	0	**1**	**0**	**7**	**4**
Similarity	0	0	1	1	1	1

The remaining twelve response pattern errors were recorded by nine students. Three occurred between Levels 2 and 3 (two students), and nine occurred between Levels 3 and 4 (seven students). The errors occurring between Levels 2 and 3 were recorded by two capable students, both of whom appear to have misinterpreted the intent of the Level 2 items. Incomplete responses to Level 3 questions (particularly Items 41-44), together with an unbalanced question focus appear to have been the cause of the nine errors occurring between Levels 3 and 4.

In summary, addressing the first two of the five identified design weaknesses, approximately one third of the response pattern errors were accounted for. While there are indications of reasons for the other twelve errors, they were not able to be resolved even with an amended version of the Mayberry test. This issue is picked up specifically in the next chapter.

Conclusion

In designing her study, Mayberry first defined each van Hiele level operationally. From this description of the types of thought, she hypothesised that

> It should be possible to construct a series of tasks which students functioning on a given level could perform, and students functioning on a lower level could not perform (1981, p.8)

Thus the questions were designed to test a specified level. Students were assessed to be reasoning at a van Hiele level if they gave sufficient correct answers to reach the success criterion set for the level and topic. However, the analysis carried out in this chapter has shown that the assessment is not always correct. While some of the incorrect assessments have resulted from weaknesses in the design of items, others have resulted from characteristics inherent in the evaluation format. Many of the students' responses showed reasoning for a level other than that for which the item had been designed. However, with the Mayberry system of evaluation, this knowledge could not be used in the assessment of the students' levels. This raises the question of whether there is an alternative method of determining van Hiele levels of geometric reasoning; a method which allows for the evaluation of each response for the van Hiele levels reflected in it. Such a method would eliminate the difficulty associated with the matching of items to levels. It could also include the measurement of a student's degree of acquisition of each level, thus removing the need to set success criteria.

Overall the detailed analysis of the validity of the test items not only gives a clearer perspective about the Mayberry test and the results, but also provides further insight into the van Hiele Theory itself. In particular, it provides further empirical evidence about what it means to work at a particular level. It also indicates a need for the investigation of another form of interpreting results. Chapter 7 examines an alternative method presented by Gutiérrez, Jaime and Fortuny (1991), in which each response is evaluated for the van Hiele levels it reflects. The assessment of the responses using the Gutiérrez *et al* method is then compared to the results from the amended Mayberry evaluation, detailed in this chapter.