Chapter 7

AN ALTERATIVE ASSESSMENT: THE GUTIÉRREZ, JAIME AND FORTUNY TECHNIQUE*

In her research, Mayberry (1981) created a series of questions which could be used in the assessment of a student's van Hiele level of thinking. She also provided a method for analysing students' responses, and, hence, assessing their van Hiele level of reasoning. Mayberry's test items and method of evaluation are based on the key assumption that the van Hiele levels are discontinuous (p.22). This led Mayberry to set each test item to focus on a single van Hiele level, the response being assessed on whether it reflects that level of thinking. In the assessment, no credit is given, for example, for a response to a Level 3 question in which several properties of a geometric concept are listed (i.e., Level 2 features). In addition, no credit is given if some awareness of relationships between properties is displayed, but the answer is incorrect. Finally, there is no grading of the degree of difficulty of the questions within a level, nor of the depth of understanding of the level displayed in a response. For example, there is no distinction between a Level 4 question requiring a simple proof and one requiring a complex or generalised proof, or between a response to a Level 2 question which lists only one property and one which lists several properties For each group of questions for a given level and concept, Mayberry (p.60) set a success criterion (ranging from 50% to 100%) for that level and concept. A student is credited as thinking at a given van Hiele level if there are sufficient correct responses to reach Mayberry's success criterion.

An alternative paradigm for the evaluation of the acquisition of van Hiele levels by students has been presented by Gutiérrez, Jaime and Fortuny (1991). In contrast to Mayberry, they have based their research on the idea that the van Hiele levels are not discrete, rather they have a more dynamic nature, being continuous rather than static (Pegg 1992, p.25). Their (Gutiérrez, Jaime & Fortuny 1991, p.237) theory is based on observations that, when answering questions, most students show a dominant level of thinking, and frequently display some reflection typical of another level. It is this philosophical perspective which underpins this chapter.

In this chapter, the fourth research theme, namely, what implications can be drawn when an alternative assessment system, such as the method used by Gutiérrez, Jaime and Fortuny

* Aspects of this work have already been published, see Lawrie (1997), Lawrie (in press).

which does not have the discreteness of levels as a significant feature to influence the findings, is addressed. An analysis of the evaluation of the students' responses to the Mayberry items using the alternative method as propounded by Gutiérrez *et al* (1991) is presented. The assessments are then compared to those resulting from the amended Mayberry evaluation, given in Chapter 6, and the effect of the new evaluation on the remaining response pattern errors is discussed. Finally, unusual behaviour patterns which emerged during the final analysis are examined. Before beginning the analysis it is appropriate to review important aspects of the Gutiérrez team's method of evaluation.

Background

The method of evaluation results in a qualitative assessment of a student's degree of reasoning in each of the four levels. Gutiérrez *et al* stated that initially students are not aware of the need to think at a level (1991, pp.238-239). They have no acquisition of that level. As they become aware of the new level, an attempt to work at the level is made and a low degree of acquisition is acquired. Continual growth in awareness is shown in an increasing degree of thinking by the students at this level, through an intermediate degree of acquisition, a high degree, until they have complete acquisition of the thinking at that level.

Several steps are necessary in evaluating a student's van Hiele levels using the method of Gutiérrez *et al.* First, in considering a response, the highest level of reflection displayed in the response needs to be determined in order to give the student full credit for the understanding displayed. In making this decision, it can be necessary to consider the response in conjunction with the student's other answers. For example, a response which appears to express the necessary and sufficient conditions sought in Mayberry Items 24, 26, 28 and 39, can be an attempt at giving these conditions (Level 4), or it can be a statement of the few properties known by the student for that concept (Level 2). The level of reasoning used by a student in making such a response can only be decided by considering the student's other statements.

When assessing a student's level of understanding, it is not always necessary to consider other statements or responses. Many students, in their responses, show consistently a dominant level of thinking. With such students, their statements constantly re-confirm their most common level of reasoning. In comparison, students who are beginning to explore a higher level for some aspect of a concept may give a better than expected response if the question is focusing on that aspect or characteristic. Conversely, students who have been attempting to work at a new level of reasoning, sometimes, in a response, revert to a lower level which is more familiar to them.

Having decided on a level, the response is now assigned one of eight types of answer (see Appendix M). This categorisation depends on the degree of mathematical accuracy, and on how complete is the solution to the question. Gutiérrez *et al* (pp.239-240) stated:

> To determine which type an answer belongs to, it is necessary to consider it from the point of view of the van Hiele level it reflects, since an answer can be adequate according to the criteria of a given thinking level but not valid according to the criteria of a higher level.

These two steps result in an answer being assigned a vector (*l, t*), which shows the highest level (*l*) the answer reflects, together with the type (*t*) of answer according to its correctness and completeness. The responses are then quantified according to each vector.

An example of the process of allocating a level and vector, and the resultant values to a response, is demonstrated in the evaluation of S39's response to Item 45.

Item 45

ABCD is a four sided figure. Suppose we know that opposite sides are parallel. What are the fewest facts necessary to prove that ABCD is a square? *S39 All sides are equal, and have 90° angles.*

This response could be Level 2, i.e., an expression of the only known properties, or Level 3 in that the student knows a lot more properties, and she is attempting to minimise them. To determine the level of the response, S39's responses to other questions needed to be examined. Her responses to the three parts of Item 25 which tests for Level 3 knowledge are given.

Item 25

A.	Name some ways in which squares and rectangles are alike?
S39	*All angles equal 90°, it is a parallelogram, opposite sides parallel, opposite sides equal in length.*
B.	Are all squares also rectangles? Why?
S39	*Yes - because all properties of a rectangle fit that of a square, but rectangles are not squares.*
C.	Are all rectangles also squares? Why?
S39	*No because some properties of squares, i.e., adjacent sides are equal, are not true for the rectangle.*

In these responses to Item 25, S39 demonstrated not only that she knows several properties of squares, but also, she demonstrated understanding of class inclusion, a Level 3 skill. This

indicated that her dominant level of reasoning is probably Level 3. Hence, her response to Item 45 was assigned to Level 3 because S39 showed a lack of understanding of minimum conditions (Level 4), and to Type 6 because, according to the student's thinking level, the response was correct but incomplete. The question could have been answered at Levels 4, 3 or 2, and so attracts a value of zero for Level 4, 80 for Level 3, and 100 for Level 2 since a Level 3 response implies complete acquisition of Level 2 (Gutiérrez, Jaime & Fortuny 1991, p.246).

The student's degree of acquisition of each van Hiele level is determined by "calculating the arithmetic mean of the values of the student's answers to those items that could have been answered at that level" (Gutiérrez, Jaime, Shaughnessy & Burger 1991, p.109). Finally, the student is assigned a qualitative degree of acquisition when the arithmetic mean is converted to a subjective interpretation of No acquisition (0-15%), Low (15-40%), Intermediate (40-60%), High (60-85%), or Complete (85-100%) acquisition.

In formulating their alternative paradigm for the evaluation of the acquisition of van Hiele levels by students, Gutiérrez, Jaime and Fortuny (1991, p.239) started with some assumptions. These are:

- that it is more important to observe the students' type of reasoning than their ability to solve certain problems correctly in a set time,
- that a partially correct (or even a totally incorrect) answer may also afford information, and
- that an incorrect answer, when considered in conjunction with other answers, may give more than a negligible amount of information.

The Gutiérrez *et al* assessment method combines the several steps described above, with these three assumptions. Together they result in a qualitative assessment of a student's degree of mastery of understanding in geometry in each of the first four van Hiele levels. That all responses, whether complete or correct, attract a scoring vector, not only makes the evaluation more realistic, but also renders unnecessary Mayberry's assigning of levels to questions, and the setting of level criteria.

Results

To facilitate the Gutiérrez *et al* evaluation, marking formats were designed (see Appendix K). The degrees of acquisition of the van Hiele levels of reasoning attained by the students for each of the seven concepts are summarised in Table 7-1, and detailed results are given in Appendix L. The distribution of the degrees of acquisition shows a similar balance to that

found by Gutiérrez *et al* (p.247), and agrees with the hierarchical structure of the van Hiele levels. Within each concept, the higher the level, the lower the degree of acquisition. The results confirm the trend shown in the Mayberry results, namely, the majority of students have a high to complete degree of acquisition of Levels 1 and 2, and a low or no degree of acquisition of the higher levels.

Table 7-1

Number of Students Attaining Degrees of Acquisition of Each van Hiele Level

		Degree of Acquisition				
Concept	van Hiele Level	No acquisition	Low	Intermediate	High	Complete
Square n=61	1	0	0	0	0	61
	2	0	2	7	24	28
	3	37	9	13	2	0
	4	59	2	0	0	0
Right triangle n=31	1	0	[illegible]	0	1	29
	2	0	2	2	13	14
	3	14	10	2	4	1
	4	30	[illegible]	0	0	0
Isosceles triangle n=30	1	0	0	2	0	28
	2	0	2	1	8	19
	3	15	6	5	4	0
	4	29	1	0	0	0
Circle n=31	1	0	0	0	0	31
	2	0	4	0	11	16
	3	17	8	3	2	1
	4	29	2	0	0	0
Parallel lines n=30	1	0	0	1	1	28
	2	0	2	3	7	18
	3	23	6	1	0	0
	4	30	0	0	0	0
Congruency n=31	1	0	0	0	7	24
	2	2	2	5	10	12
	3	16	7	5	3	0
	4	27	4	0	0	0
Similarity n=30	1	0	0	1	0	29
	2	0	5	4	3	18
	3	18	7	5	0	0
	4*	26	3	0	0	0

*Similarity Level 4, n = 29

In general, it was found that an intermediate, or better, degree of acquisition in the Gutiérrez *et al* results corresponded with achievement of mastery of a level in the Mayberry evaluation, while an intermediate or less degree of acquisition corresponded with failure to reach Mayberry's criterion. The overlap in the qualitative assessment of an intermediate degree of acquisition results from two factors. First, the term 'intermediate' covers arithmetic means

from 40% to 60%. Second, Mayberry's criteria for each concept and level range from 50% to 100%.

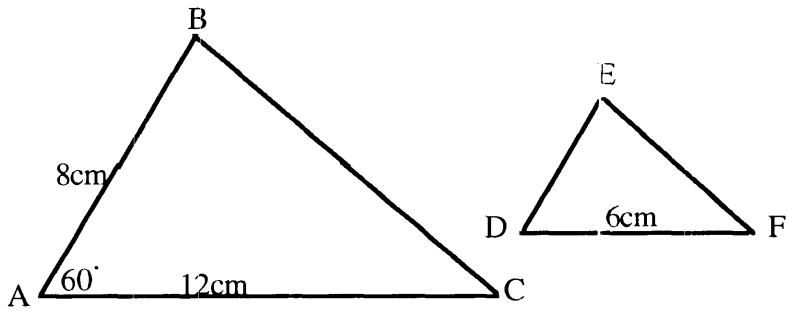
Although there was no time limit for the test, a few students limited the time they were prepared to devote to the test to 50 minutes. For these students, the time limitation did not appear to alter the level of reasoning they were using, i.e., each student tended to give responses at the same level or combination of levels for any one topic. However, on occasions it appeared to affect the completeness of answers. In particular, those students who limited themselves in time gave several incomplete answers for some of the more complex Level 3 questions (e.g., Mayberry Items 41-44). It could be that the incomplete responses were an indication of lack of mastery of the topic. Since all responses, whether wholly or partially correct, or incorrect, contribute to a Gutiérrez *et al* assessment, these students received credit for the knowledge displayed in the responses. This contrasts with Mayberry's method in which incomplete responses are assessed as incorrect. The incomplete responses to the Level 3 items were frequently the cause of response pattern errors in the Mayberry assessment. These response pattern errors occurring between Levels 3 and 4 are discussed in the next section.

A comparison between the Gutiérrez *et al* results for each student (see Appendix L) and the amended Mayberry results (see Appendix J) illustrates some of the differences between the two methods of evaluation. The alternative paradigm described by Gutiérrez *et al* (1991) gives a more flexible interpretation of the reasoning of students, compared to the method developed by Mayberry. This more complete view results from the evaluation of all responses, thus taking into account for every statement, a student's capacity to use each one of the van Hiele levels. The following advantages were identified:

1. a student can be shown to be developing in two consecutive levels of reasoning at the same time;
2. the incorrect assignation of a level to a question is of minimal significance;
3. the effect of unequal distribution of questions across levels is minimised;
4. the effect of unbalanced focus on a specific characteristic is minimised;
5. the inequalities associated with success criteria are eliminated;
6. incorrect assessment resulting from 'lucky' guesses such as in true/false questions, from weak and from misinterpreted questions is minimised; and
7. all responses, including incomplete and incorrect responses contribute to the overall evaluation.

However, there appears to be a disadvantage/weakness in the evaluation method. The automatic allocation of 100% acquisition for the level below that reflected in the response, was not always found to be justified. This is demonstrated in the responses of S32 to Item 21 (Level 2) and Item 41 (Level 3), for the concept similarity.

Item 21



Δ ABC is similar to Δ DEF

(a)	How long is ED?	*S32*	*4 cm*
(b)	How do you know?	*S32*	*DF/AC = DE/AB 6/12 = 1/2 = x/8* *8 = 2x x = 4cm*
(c)	What is the size of ∠EDF?	*S32*	*30°*
(d)	How do you know?	*S32*	*Ratio of 1:2*

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	S32	*possible - may have one side the same length.*
b)	AB > DE	S32	*possible - depends on which* (triangle)*is bigger.*
c)	∠A = ∠E	S32	*possible - depends on the size of the triangles. The corresponding angle are in the same ratio as the sides.*

In Item 21, although displaying understanding of the proportionality of sides of similar figures in her answer to relevant question parts, S32 gave incorrect responses to the questions about properties of angles. This confidence with the properties of sides, together with the lack of understanding about angle properties were again evident in her responses to Item 41. As in the above responses, there is no indication of understanding of angle equality in similar figures in any of S32's responses for similarity questions.

For the Mayberry scoring, S32 reached the criterion for similarity for Level 1 only. The application of the Gutiérrez *et al* evaluation means the allocation of a Level 2 score of 100 for every question attempted at Level 3. S32's responses to Items 41 and 58 reflected attempts at Level 3 reasoning, and, hence, were allocated 100% acquisition of Level 2. This resulted in a Gutiérrez *et al* assessment for S32 of complete acquisition of Level 1 (average = 100), complete acquisition of Level 2 (average = 93), and no acquisition of Level 3 (average = 9) and Level 4 (average = 0). Considering that S32's responses indicate only partial understanding of the properties of similarity, this is not an accurate representation of the student's incomplete understanding of Level 2, and raises the question of whether an automatic allocation of a score of 100 for Level n-1 is always valid when a student is attempting to work in a new van Hiele level.

Conclusion

The assessment of the students' responses using the method developed by Gutiérrez *et al* agreed with the hierarchical structure of the van Hiele levels, and, in general, supported the findings of the earlier determination of students' van Hiele levels of thinking. Additionally, it is of interest that while the results support the notion of hierarchy of the levels, they support also, as seen in S32's responses above, the issue, raised by Pegg and Davey (1989, p.25), of whether there is a hierarchy of properties, and whether "the absence of knowledge of some properties may not exclude growth to a higher thought level." The majority of students had a high to complete mastery of Levels 1 and 2, and a low or no mastery of Levels 3 and 4. However, because every response was assessed for the level displayed therein, the results provided a more realistic picture of each student's van Hiele level of understanding in geometry. While several advantages in the Gutiérrez *et al* method of assessment were identified, there were indications that the automatic assumption of complete mastery of level n-1 was not always justified.

Response Pattern Errors

Originally there were nineteen response pattern errors. Seven of these were eliminated in the amended Mayberry assessment, following the removal of repetitive and non-assessable question parts, and the resetting of incorrectly assigned van Hiele levels for some of the Mayberry questions. Thus, there still remained twelve response pattern errors that needed to be addressed. Three of these errors occurred between Levels 2 and 3, and nine between Levels 3 and 4.

Response Pattern Errors between Levels 2 and 3

The three response pattern errors occurring between Levels 2 and 3 all resulted from incorrect responses to Level 2 question(s), followed by sufficient acceptable Level 3 responses to

achieve Mayberry's criteria. They were registered by two students, S14 in the concept right triangle, and S57 in the concepts isosceles triangle and similarity. Each student's failure to achieve Mayberry's criterion resulted from an apparent misinterpretation of some of the Level 2 question parts. For example, the responses of S14 to Item 17 which investigates whether a student has knowledge of the side and angle properties of the right triangle, show incorrect answers to parts (c) and (d) concerning the angle. They are:

Item 17

a)	Does a right triangle always have a longest side?	S14	*Yes*
b)	If so, which one?	S14	*The hypotenuse.*
c)	Does a right triangle always have a largest angle?	S14	*No*
d)	If so, which one?	(This question was left blank)	

The Mayberry criterion of 3 out of 4 for this level and concept means that S14 was assessed as not possessing knowledge of the properties of a right triangle. However, her responses to later Level 3 questions, while showing an understanding of relations and implications, also demonstrate knowledge of angle, as well as side properties of right triangles. Her responses to the three parts of Item 27 illustrate this depth of understanding:

Item 27

QAB is a triangle.

a) Suppose angle Q is a right angle. Does that tell you anything about angles A and B? If so, what?

S14 *They are acute as they have to add to 180°.*

b) Suppose angle Q is less than 90°. Could the triangle be a right triangle? Why?

S14 *Yes as one of the other angles may be 90°.*

c) Suppose angle Q is more than 90°. Could the triangle be a right triangle? Why?

S14 *No as the other angles would be less than 90°.*

In answering the Mayberry Level 3 questions across all concepts, both S14 and S57 displayed knowledge of a broad range of properties (Level 2), as well as perceiving relationships between the properties (Level 3). Evaluation by the Gutiérrez *et al* method showed that S14's responses to the Level 3 questions for the concept right triangle displayed high to complete understanding of Level 3, while S57's Level 3 responses for the concepts isosceles triangle and similarity displayed a depth of understanding which varied from low to high. All the Level 3 questions could have been answered at Level 2, hence both students automatically

scored 100 for Level 2 for those questions answered at Level 3. The resultant arithmetic means for the right triangle indicated that S14 had a complete degree of acquisition of Levels 1, 2 and 3, and no acquisition of Level 4, while for both the concepts isosceles triangle and similarity, S57 had complete acquisition of Levels 1 and 2, intermediate acquisition of Level 3 and no acquisition of Level 4. The Gutiérrez *et al* evaluation not only removed the response pattern errors resulting from the misinterpretation of a question, but also gave a more complete and more accurate description of the ability of these students to use each of the van Hiele levels of reasoning.

As has been shown, misinterpretation of the context of Level 2 questions by students can lead to an under-evaluation of their understanding of properties, and, hence, to the recording of response pattern errors when assessment is made using the Mayberry method. In each of the three occasions described, re-evaluation using the Gutiérrez *et al* alternative paradigm gave credit for knowledge of properties displayed in responses to Level 3 questions, and, hence, eliminated these errors. This resulted in a more comprehensive assessment of the students' levels of geometric understanding.

Response Pattern Errors between Levels 3 and 4

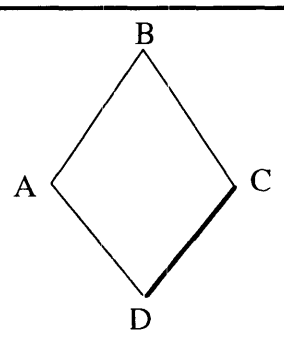
Insufficient correctly completed Level 3 responses, together with at least one acceptable Level 4 proof (criterion of 1) appeared to be the main cause of the nine error patterns occurring between Levels 3 and 4 . These were registered by seven students (two of whom each scored two errors), across four concepts. Four of the errors were for the concept congruency, three were for the concept square, and there was one each for parallel lines and similarity. Analysis of the responses of the students showed that, while there were other factors, e.g., S58 was not aware of class inclusion for the concept square, all had given incomplete responses to Level 3 questions, failing to detail sufficient reasoning to reach the Mayberry criteria. In addition, all seven students appeared to be exploring the early stages of formal proof. Each student had given at least one acceptable response for a Level 4 question. A Mayberry criterion for the relevant concepts of one correct response credited these students with mastery of Level 4 in the respective concepts. This, in association with failure to reach the Level 3 criterion, led in each case, to a response pattern error.

Level 4 responses

Examination of all the Level 4 responses revealed that none of the seven students showed more than an early awareness of Level 4 reasoning for any of the concepts. At each attempt, although the students had given an acceptable answer to a Level 4 problem, this often required only a simple proof. This is illustrated in the following responses of two of the students, S16 and S41.

S16's response to Item 55 was considered to be her only response at Level 4, all other attempts at Level 4 items being Level 3 responses.

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?

On the diagram, S16 marked equal lengths AB, CB and AD, CD. She also marked in the bisector, BD. Her written proof was:

S16 *Construct a bisector*
ΔABD and ΔCBD are congruent giving 3 equal sides BD, AB = CB, AD = CD
Therefore $\angle A = \angle C$ giving corresponding $\angle$'s.

This response, while considered acceptable in the Mayberry assessment, was assigned within the Gutiérrez *et al* coding, as a Level 4 Type 3, since it showed recognition of the need to make a construction together with indication of some ability to construct a proof. However, the proof was incomplete in that it contained insufficient explanations.

S41 attempted all Level 4 questions. His responses to Items, 45, 48, 50 and 52 were considered acceptable Level 4 responses in the Mayberry assessment. He registered a response pattern error between Levels 3 and 4 for the concept square. As S41 was one of the subjects interviewed, he had given both written and interview responses to Item 45. His two attempts are below:

Item 45

ABCD is a four sided figure. Suppose we know that opposite sides are parallel. What are the fewest facts necessary to prove that ABCD is a square?

S41 (written response)
Line AB = Line BC. 90° angles.

S41 (interview response)
All sides are equal in length. One angle equals 90°.

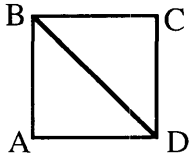
S41's first response provided too much information on angles, while the interview response mentioned a single angle but added more information than is needed about the sides. Prompting did not induce S41 to reduce both side and angle properties to a minimum, confirming his inability to control all of the elements for necessary and sufficient conditions. Using the Gutiérrez *et al* coding this response was also coded as Level 4 Type 3, since the attempt to minimise the conditions is incomplete. His response to the second Level 4 question for the square, Item 46, lacked rigor, and, hence, was judged to be at Level 3.

As illustrated in the above two responses, none of the seven students recording response pattern errors between Levels 3 and 4 were able to demonstrate comprehensive mastery of Level 4 skills, i.e., the students did not demonstrate ability to reflect on the significance of deduction, understanding of the meaning of necessary and sufficient conditions, and understanding of the role of axioms and definitions.

Level 3 responses

Analysis of the seven students' Level 3 responses showed that all, to varying degrees, had failed to justify statements. Ensuring that students explain all their knowledge is a more difficult prospect in a written test than in an interview situation, when it is possible to prompt for hidden knowledge. This is illustrated by the differences between the two sets of responses, written and interview, given by S41 to the Level 3 Item 23(b) for the concept square. In the written test S41 recorded five correct responses, failing to reach Mayberry's Level 3 success criterion for the square (6 out of 9), yet gave the acceptable Level 4 response above for Item 45. When prompted in his interview, his more complete Level 3 answer to Item 23(b) indicates ability to achieve the Level 3 success criterion.

Item 23



ABCD is a square, BD is a diagonal.

(a) Name an angle congruent to ABD. (Level 2)
(b) How do you know? (Level 3)

In both the written test and the interview S41 nominated $\angle$CBD as his response to part (a). In answer to part (b):

S41 (written test) *It is the angle that provides a sum of 90° when added to $\angle ABD$.*

S41 (interview) *$\angle DBC$ forms the other half of 90°.*

CL (Prompt) *Can you tell me more?*

S41 *A square has four 90° angles. The diagonals will split their angles in half so $\angle ABD = 45°$, so $\angle DBC$ must be 45° as well which makes the angle equal or congruent.*

The response in the written test was considered insufficient to merit Level 3 credit, whereas, the interview response displayed understanding of the implications of the symmetry of a square. This is reflected in the Gutiérrez *et al* evaluation which assesses S41 as having a high acquisition of Level 3, and no acquisition of Level 4.

Analysis of the responses to Level 3 items of the seven students showed that their failure to give complete answers occurred mostly with the four similarly-designed, multi-part questions, Items 41 to 44. Incomplete answers are considered unsuccessful in the Mayberry assessment, whereas, in the Gutiérrez *et al* method, all responses attract a score, accurate and more complete responses scoring more highly. Items 41 and 42 test for knowledge of similarity of figures, and Items 43 and 44 test for knowledge of congruency of various figures. Three of the questions occur in each written paper. Two of the questions (Items 41 and 43) tended to be better answered than the other two. This was particularly so for the seven students registering the nine Level 3/4 response pattern errors. The only difference between the two questions attracting the more complete responses, Items 41 and 43, is the

concept tested [indicated in square brackets]. The two items identified specific figures, and were phrased in concise sentences:

Item 41 [Item 43]

Triangle ABC is similar [*Item 43 replaces the word similar with the word congruent*] to triangle DEF (in that order). Are the following a) certain b) possible c) impossible? Give reasons for your answers.

In comparison, Items 42 and 44 were less direct and less specific. Again, the only difference between the items is the concept tested [indicated in square brackets]:

Item 44 [Item 42]

Will figures A and B be congruent [*Item 42 replaces the word congruent with the word similar*] I - always, II - sometimes, or III - never? Give reasons for your answers.

The responses to these questions by S16 (whose answer to Item 55 is given above) illustrate clearly the difference in the completeness of answers. S16 failed to provide any reasons for her attempts at Items 42 and 44, yet gave the following acceptable responses to Item 43:

Item 43

	Δ ABC is congruent to Δ DEF (in that order). Are the following a) certain b) possible, or c) impossible? Give reasons for your answers.	
(a)	AB = DE	*certain, corresponding sides*
(b)	∠A = ∠E	*possible, if both are isosceles Δs*
(c)	∠A < ∠D	*impossible, corresponding angles*
(d)	AB = EF	*possible, if both are equilateral Δs*

Evaluation of S16's responses for the congruency items using the Gutiérrez *et al* method changed the assessment from recording a response pattern error between Levels 3 and 4, to the Gutiérrez *et al* result of complete acquisition of Levels 1 and 2, high acquisition of Level 3 and low acquisition of Level 4.

The comparison of the above questions demonstrates that there are many factors which together contribute to the suitability of a question. That the students were more likely to give

incomplete responses, and, hence, score less with more indirect questions could be an indication of incomplete mastery of the level rather than an indication of lack of application.

Overview

The Gutiérrez *et al* evaluation of the responses of the seven students recording these nine response pattern errors supports the notion that students often have partial mastery of a level, and that, additionally, they may have commenced exploring the next higher van Hiele level before mastering the earlier level. In each of the nine occasions when a response pattern error occurred between Levels 3 and 4 in the Mayberry evaluation, the student was re-assessed (using Gutiérrez *et al*'s coding) as demonstrating an intermediate to high degree of acquisition of Level 3. Also, for each of the seven students, the results showed that Level 3 was their dominant level of reasoning. In assessing their Level 4 understanding as low or no degree of acquisition, the Gutiérrez *et al* method of evaluation provides a truer understanding of their degree of acquisition of Level 4. The alternative method of evaluation used by the Gutiérrez team eliminated all of the response pattern errors between Levels 3 and 4 which had resulted from Mayberry's method of evaluation.

Conclusion

The re-assessment of the students' responses using the alternative method of evaluation designed by Gutiérrez *et al* has led to a more comprehensive description of the students' degrees of acquisition of each van Hiele level. Excepting errors resulting from the automatic allocating of 100% knowledge of lower levels, these assessments appear to be more realistic. Not only have all the remaining response pattern errors that occurred when evaluating the responses using Mayberry's method, been eliminated, but also, the analysis has suggested reasons for the occurrence of the errors.

Behaviour Patterns

Although the Gutiérrez *et al* assessment eliminated the remaining response pattern errors which arose in the initial assessment of the students' responses, there were indications that there was still some under-evaluation of students' understanding for the lower van Hiele levels, i.e., Levels 1 and 2. This shows clearly when comparisons are made between students' performances for consecutive levels.

Unusual behaviour patterns between Levels 1 and 2

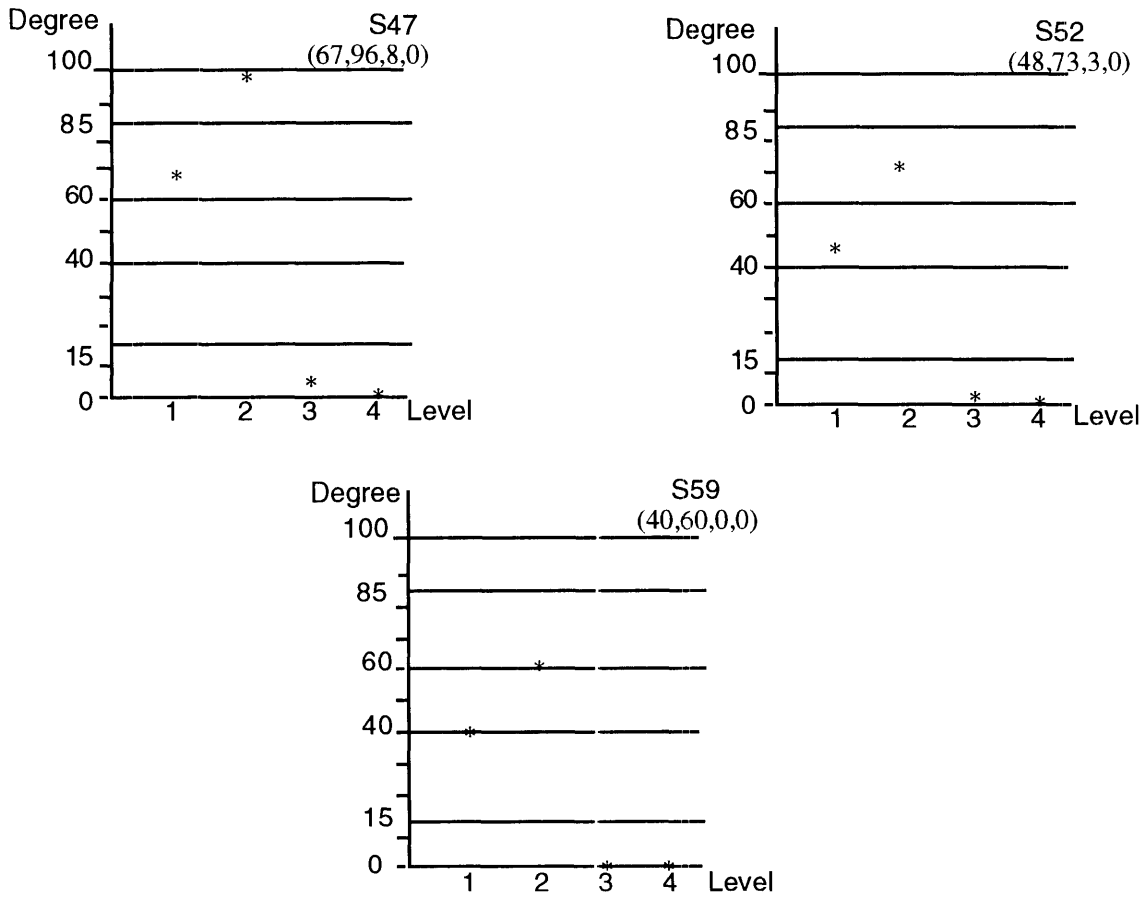
An examination of the quantitative results obtained in the re-evaluation of the students' responses using the Gutiérrez *et al* method revealed that not all results agreed with the hierarchical structure of the van Hiele levels. In twenty-six (2.7%) of the nine-hundred-and-

seventy-six assessments (sixty-one students for four levels across four concepts), the degree of acquisition of Level n was not less than the degree of acquisition of Level n-1. In every case, the non-hierarchical behaviour occurred between Levels 1 and 2, the degree of acquisition of Level 1 measuring below that of Level 2. However, many of the pattern errors in these quantitative results are considered to be trivial, the value of Level 1 being less than 10 points below the value of Level 2. For example, S05 scored averages of 75 points (high) for Level 1, 83 points (high) for Level 2, 29 points (low) for Level 3 and 0 points (no degree of acquisition) for Level 4.

Three results involving students S47, S52 and S59, are considered worthy of further examination, the degree of acquisition for Level 1 being at least twenty points below the degree of acquisition for Level 2.

Figure 7-1

Unusual Behaviour Patterns Between Levels 1 and 2 for S47, S52 and S59



In two cases (S47 and S52), neither the qualitative nor the quantitative assessment fit the hierarchical pattern. S47 was assessed as showing a high degree of acquisition of Level 1, complete acquisition of Level 2 and no acquisition of Levels 3 and 4, while S52 was assessed as an intermediate degree of acquisition of Level 1, a high acquisition of Level 2 and no acquisition of Levels 3 and 4. In the third case (S59), while the qualitative assessment showed an intermediate degree of acquisition for both Levels 1 and 2 and no acquisition of Levels 3 and 4, the quantitative measurement for Level 1 was an average of 40, and for Level 2, the average was 60. The three patterns of unusual behaviour between Levels 1 and 2 are graphed in Figure 7-1.

The occurrence of these patterns in which the degree of acquisition of Level 1 is lower than it should be is suggestive of two factors: (a) that the Mayberry questions designed to measure Level 1 are not always clear in their intention, and (b) that the criteria used to measure the acquisition of levels in the Gutiérrez *et al* alternative method may more suited to responses demonstrating reasoning at Levels 2, 3 and 4, rather than the visual identification of Level 1.

Profiles of students' reasoning

Gutiérrez *et al* (1991, pp.248-249) identified six profiles of students' reasoning that correspond to various stages in the process of a student's intellectual development. Their table (Table 5, p.249) showing the characteristics of the profiles, is reproduced below in Table 7-2. Table 7-3 shows the number of sets of results for this study which fit each profile.

Table 7-2

Characteristics of Observed Profiles of the Acquisition of van Hiele Levels

(Gutiérrez, Jaime & Fortuny, 1991)

	Degree of Acquisition			
Profile	Level 1	Level 2	Level 3	Level 4
1	Complete	Complete	Complete	≤Low
2	Complete	Complete	≤High	≤Low
3	Complete	High	≤Intermediate	≤Low
4	Complete	≤Intermediate	≤Low	No Acquisition
5	High/Intermediate	≤Low	No Acquisition	No Acquisition
6	Low	No Acquisition	No Acquisition	No Acquisition

Table 7-3

Distribution of Students' Degrees of Acquisition According to the Profiles

	Concept							
Profile	Square	Right Δ	Isosc. Δ	Circle	Parallel	Congru.	Similar.	Total
1	0	1	0	1	0	0	0	2
2	28	14	19	15	18[a]	13	18	125
3	24	12[a]	8[a]	11	7	9[c]	3	74
4	9	3	3[a]	4	5[a]	7[b]	8	39
5	0	1[a]	0	0	0	2	1	4
6	0	0	0	0	0	0	0	0

[a] One of the students did not fulfil the conditions of Level 1

[b] Two of the students did not fulfil the conditions of Level 1

[c] Three of the students did not fulfil the conditions of Level 1

Of the 244 sets of results for this study, 233 fit the profiles perfectly, and 11 fit some profile except for one level. Of the eleven results which did not fit the profiles perfectly, all indicate an under-evaluation of Level 1, supporting the notion above that Mayberry's test questions and the Gutiérrez *et al* evaluation method do not necessarily measure Level 1 adequately.

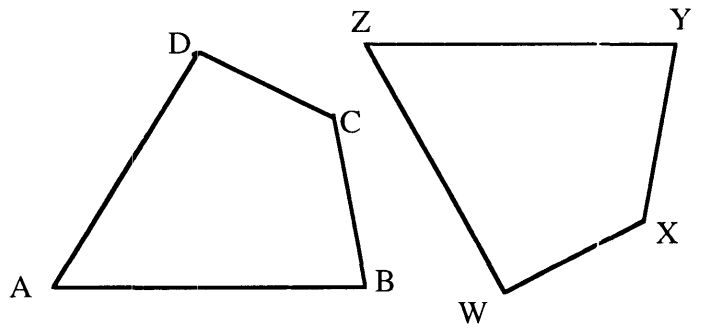
In summary, the results of this research indicate an under-evaluation for some of the students, of their understanding of Level 1. However, the non-hierarchical results between Levels 1 and 2 discussed above did not show in the results of Gutiérrez, Jaime and Fortuny (1991). There are two possible reasons for this. First, instead of a written paper with diagrams of shapes, the test is by interview, with the students allowed to manipulate models of the shapes. Second, Activity 1(a) is the only question testing solely for identification of a shape. In other questions, shape identification is implied through the display of knowledge of properties, and, hence, attracts a score of 100. This structure of the questions may obscure any problem associated with Level 1 assessment.

Unusual behaviour patterns between Levels 2 and 3

Although Gutiérrez *et al* (p.248) found four of their fifty students recording a Level 2 degree of acquisition lower than for Level 3, this did not occur in the Gutiérrez *et al* evaluation for this study. However, several students who showed comprehensive knowledge of properties in their responses to the Level 3 questions appeared not to understand the direction of the Level 2 questions, and, hence, did not register fully their Level 2 degree of reasoning. As shown earlier, this was demonstrated particularly by the students recording response pattern errors between Levels 2 and 3 in the Mayberry assessment. For example, S06 demonstrated

just sufficient knowledge of the relationships between congruent figures to reach the Mayberry Level 3 criterion in the initial assessment, yet failed to reach the Level 2 criterion. In the amended Mayberry assessment, S06 was credited with achieving Level 1 only for congruency, the changes to Mayberry's test resulting in the elimination of the response pattern error, but not acknowledging her partial knowledge of properties. The Gutiérrez *et al* evaluation credited her with a high degree of acquisition of Level 2 and an intermediate degree of acquisition of Level 3. Below are her responses to the congruency questions, Items 22 (Level 2) and 43 (Level 3):

Item 22



These are congruent figures.

What is true about their sides?
AD = _____

What is true about their angles?
∠B = _____

S06 named the corresponding side WZ and the corresponding angle ∠Y correctly. Her responses to the other question parts are not incorrect, but are not acceptable for the Mayberry assessment:

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

Item 43

Δ ABC is congruent to Δ DEF (in that order).
Are the following a) certain b) possible, or c) impossible?
Give reasons for your answers.

b)	∠A = ∠E	*S06*	*(b) yes if equilateral*
d)	AB = EF	*S06*	*(b) only if equilateral*

While S06's responses to Item 22 were incomplete, the responses to Item 43 show that she has some awareness of relations between the properties, as well as some knowledge of properties. Responses by S06 to other items showed low Level 3 ability. Hence, the Gutiérrez *et al* assessment for S06 of complete, high, intermediate and no degree of

acquisition of the four levels, reflects more accurately the student's understanding of congruency.

Although the Gutiérrez *et al* assessment removed all non-hierarchical patterns between Levels 2 and 3 in the results, it did not provide any clarification of the problem of strong students answering higher level items better than lower level ones. This problem of the failure of strong students to perceive the context of questions set to test for knowledge of a lower level has been recorded by other researchers, for example, Gutiérrez and Jaime (1987), Mason (1989), Mayberry (1981), and Usiskin (1982).

Summary

Analysis of the results indicates that students do not always display their full knowledge in responses to questions testing for understanding of van Hiele Levels 1 and 2. While there are no indications of why more capable students fail to understand the thrust of Level 2 questions, the under-evaluation of Level 1 appears to result from the nature of Mayberry's test questions and the Gutiérrez *et al* evaluation.

Conclusion

In this chapter, the fourth research question has been investigated. It concerned the implications which can be drawn when an alternative assessment system, such as the method used by Gutiérrez, Jaime and Fortuny, which does not have the discreteness of levels as a significant feature to influence the findings, has been applied, has been investigated. Analysis of the changes resulting from the alternative assessment of the students' levels of understanding geometry led to the identification of seven features.

1. The Gutiérrez *et al* evaluation minimises many of the problems occurring in the Mayberry test items and method of assessment.

2. By evaluating the degree to which understanding of each level is expressed in every response, the assessment method developed by Gutiérrez *et al* measures the full potential of a question to elicit a student's degree of reasoning. Correspondingly, the Gutiérrez *et al* method gives insight into the quality of a question. For a meaningful Gutiérrez *et al* evaluation, questions should have the potential to be answered at various van Hiele levels.

3. Gutiérrez *et al* found that some students show a better acquisition of Level 3 than of Level 2. This study has shown that several of the students demonstrating a high degree of acquisition of Level 3, do not demonstrate fully, their understanding of Level 2. This needs further investigation.

4. The Gutiérrez *et al* method does not appear to measure the visual level (Level 1) accurately. It is unclear if this is a limitation in the method of evaluation, or of the quality of the questions.

5. The automatic allocation of a value of 100 for Level n-1 for an attempt at Level n needs further investigation.

6. This research supports the explanation of Gutiérrez *et al* of why it seems possible for a student to develop in two consecutive levels at the same time (1991, p.250). Students can apparently work simultaneously in at least two levels as observed by Burger and Shaughnessy (1986), Fuys *et al* (1985), Pegg and Davey (1989), and by Usiskin (1982). The Gutiérrez *et al* evaluation provides an explanation of what they are really doing, i.e., while still working in Level n, their dominant level with an intermediate to high degree of acquisition, they are exploring Level n+1 at an early level, showing a low or even intermediate degree of acquisition.

7. That the Gutiérrez *et al* method gives a much more detailed and accurate picture of a student's ability to work in each van Hiele level, makes it an excellent research tool. However, in its current form it is very time-consuming to administer, and, hence, appears to be not suitable for regular classroom use.

Assessment using the method developed by Gutiérrez, Jaime and Fortuny has addressed many of the inconsistencies which emerged in the Mayberry analysis, and accounted for the remaining response pattern errors. While the alternative method provides a more complete and realistic evaluation of the student's van Hiele level of understanding geometry, there are indications of need for further investigation. Additionally, there are indications that a compromise between the Mayberry questions and a simplified version of the Gutiérrez method of assessment could be worth developing.

Chapter 8

DISCUSSION OF RESULTS AND CONCLUSIONS

This chapter considers the findings of the study in relation to the four research themes investigated in Chapters 4 to 7. Initially, limitations imposed by the design of the study are discussed. An overview of the results as applied to the four research themes is then considered. This is followed by the implications of the findings in relation to testing for levels, for teaching, and in relation to the van Hiele theory. The final section of the chapter considers the implications of the findings of the study for further research.

Limitations of the Study

The results described in the preceding four chapters must be viewed in the light of possible limitations imposed by features of the design. This section reviews the strengths and weaknesses in these features. Three limitations are seen to be imposed by features of the design, namely, those imposed by a one-hour test, the drawing of students from one program within a university, and the number of students interviewed.

To guarantee an adequately large number of students who would be committed to give their best, the decision was taken to include the test in the course structure in an initial pre-service degree program as part of the basic mathematics competency segment involving the understanding of geometry. This decision meant that only two hours of class time were available, one hour for students to do the paper and one hour for feedback and discussion. The disadvantage of this decision was that it meant that it was not possible to test all students in all concepts in the one-hour time interval available. To maximise the breadth of coverage, the seven concepts were divided between two papers, with the questions testing the concept square being included in both papers. This meant that all students attempted the questions for the square, while only half the students were exposed to each of the remaining six concepts. The students were allocated randomly to one of the papers, each student, therefore, being assessed for understanding on four of the seven concepts tested.

Clearly, it would have been advantageous, in terms of number of responses, to have all students tested on all concepts. It is likely that some different issues may have arisen in these circumstances. However, the experiences with the square, in which all students participated, do not support this view as no different general aspect was noted in one group as opposed to

the other. In addition, the disadvantage of a less than optimum number of students undertaking the same concepts was seen to be balanced by the test being seen as part of the students' unit of study. This had the desired effect of students taking the test seriously, and their giving their best. Such an occurrence may not have been the case if the test had have been optional.

The second limitation lies in the group of subjects being drawn from students enrolled in one program of study, and from one university. The students tested were solely those training to be primary teachers, and their responses may, in some way, be atypical of the normal population. However, the students had a range of mathematical backgrounds, i.e., some students had undertaken advanced Year 12 mathematics courses, while others had not taken mathematics past Year 10. Additionally, the students came predominantly from two Australian states, and there were a number of mature-age students in the program as well as recent school leavers. All these factors contributed to a large diversity of ability and experience within the group. Since the purpose of the study was to explore each of the first four van Hiele levels, this diversity in the sample would seem to meet the research agenda more than adequately. Also, the purpose of the research was not to obtain population standards. Nevertheless, the final word on this issue can only be addressed if this work is replicated in a number of different settings.

The final possible limitation is related to the number of students interviewed. The purpose of the interviews was to provide additional qualitative information about the levels displayed by students in the written test, i.e., to provide a richer context in which to interpret the results. The students were selected purposely so as to give a broad representation of student thinking, across all concepts, and including both genders. The levels of understanding displayed by students in the interviews supported the levels obtained in the written test. While it might have been valuable to interview more students, the focus of the thesis was on whether it was possible to code the responses of students to the questions posed within the van Hiele theory, and to see whether the questions posed addressed the levels Mayberry suggested. The number of students chosen was adequate for this task.

Summary of Results

This summary represents a bringing together of the findings of the study. Initially, the test papers were given to thirteen students, in a preliminary study, to ascertain whether the written version of the Mayberry test was appropriate. The written papers were designed to test for understanding of van Hiele Levels 1 to 4 across all seven Mayberry topics. Interviews carried out with eight of the students confirmed the levels assessed in the written test, and suggested

necessary changes in the wording to make the identified test questions clearer. To achieve the integration of the findings of the study, the results in Chapter 4 are combined with the findings in Chapters 5, 6 and 7. The discussion below considers the findings under each of the four research themes.

Research Theme 1

The first research theme was: How does a sample of Australian students training to be primary teachers perform on a written test version of the Mayberry questions?

In Chapter 4, preliminary results were identified. From the analysis, various patterns in the students' responses emerged. While some of the patterns had been identified by Mayberry (1981, p.89) in her study, many were new. Overall, the patterns indicated that:

1 Some students had difficulty in recognising a figure in a non-standard position.
2 Properties of sides of figures were more readily identified than properties of angles.
3 Difficulties arose for some students in establishing the context for questions designed to test Levels 1 and 2.
4 Many students did not have an understanding of class inclusion, for example, they did not recognise a square as a rectangle.
5 Almost no student in the sample had understanding of the significance of necessary and sufficient conditions.
6 Some students working at Level 2, i.e., property identification, resorted to measuring sides or allocating values to angles when presented with a Level 3 problem.
7 Students who were unable to work with a generalised figure, drew geometric figures when answering questions.
8 The highest level reached by more than half of the students who had completed a senior secondary geometry course was Level 2.

While analysis of the students responses led to the identification of the eight points detailed above, there were also inconsistencies in the students' responses which did not fit these patterns. Interviews did not clarify these inconsistencies. Four features were identified which appeared to be the cause of the problems. These are discussed in Research Theme 3.

Research Theme 2

The second research theme was: Can a quantitative analysis of the results, using a partial credit model, offer further insight into the nature of the Mayberry test?

The application of the QUEST program (which utilises the Rasch measurement theory) to the results for the Mayberry assessment confirmed the findings of Chapter 4. This included the major patterns identified in the investigation of Research Theme 1, and the inconsistencies that had emerged in the Mayberry evaluation and which were not clarified by interviews. Various items which did not appear to be measuring the specified thought level were identified, giving insight into one of the limitations identified by Mayberry (1981, p.97), namely, the difficulty of choice of representative questions for each level. The analysis indicated also that some of the better students have difficulty in interpreting the intention of the Level 1 items. Overall, while supporting the hierarchical nature of the van Hiele levels, the analysis showed an overlap in the difficulty thresholds between the levels, suggesting that the levels are not discrete.

Research Theme 3

The third research theme was: Do the Mayberry items measure the van Hiele levels for which they have been designed? Are the success criteria, established by Mayberry, valid?

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 clarified the inconsistencies which had emerged in the analyses in Chapters 4 and 5. This gave a clearer perspective about the Mayberry test and results, and provided further insight into the van Hiele Theory itself. The features 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.

Analysis of the inconsistencies and the identification of the four features led to the development of an amended version of the Mayberry test. An amended assessment of the students' responses reflected a more accurate evaluation of their understanding of geometry, and eliminated seven of the initial nineteen response pattern errors. The highest van Hiele level of thinking of the majority of students remained unaltered, at Level 2. However, there were still twelve unaccountable response pattern errors. These appeared to result from the Mayberry method of evaluation which assumes discreteness of the levels. This indicated a need to investigate an alternative method for determining van Hiele levels.

Research Theme 4

The fourth research theme was: What implications can be drawn when an alternative assessment system, such as the method used by Gutiérrez, Jaime and Fortuny, which does not have the discreteness of levels as a significant feature to influence the findings, is applied?

The inconsistencies (response pattern errors) remaining after the amended Mayberry assessment appeared to result from students either misinterpreting easy questions, or failing to give complete responses to the more difficult questions. This indicated a need for the investigation of another method of interpreting results, one in which a student's total geometric understanding is taken into consideration. Hence, the alternative method of assessment designed by Gutiérrez, Jaime and Fortuny (1991) was examined. As well as determining a student's dominant level of understanding, the alternative assessment indicates also any degree of understanding in other van Hiele levels. The analysis re-confirmed that the dominant level of understanding for most students was Level 2. In considering the total geometric understanding of students, assessment by the Gutiérrez *et al* method eliminated all the remaining inconsistencies occurring in the initial assessment of students' responses.

However, the Gutiérrez *et al* method of assessing students' van Hiele levels of understanding of geometry, while minimising the problems associated with the Mayberry method of assessment, and, hence, giving a more realistic evaluation of the students' ability, appears to need further refinement. First, the evaluation of visual awareness (Level 1) does not always appear to be reliable. Second, automatic assumption of complete mastery of Level *n-1* does not necessarily represent the true degree of understanding of a student. Finally, the method of assessment, while giving a more complete picture of a student's ability, is very time-consuming to administer, and, hence, is not suitable for application in all situations. For a meaningful Gutiérrez *et al* evaluation, questions should have the potential to be answered at more than one van Hiele level.

Conclusion

Overall, the analysis of the students' results have shown that there are weaknesses associated with some of the Mayberry items, and, in part, with the method of assessment which assumes discreteness of the levels. These weaknesses can be addressed by assessing each response for the level of understanding it reflects, rather than assessing whether a response reflects a specified level. Having an assessment scheme where the underlying philosophy is on the continuous nature of levels allows also for a more comprehensive evaluation of a student's understanding.

Implications of Testing for Levels

The investigation of the four features which have the potential to result in an incorrect assessment of a student's level of understanding in geometry led to the identification of four specific weaknesses in the design of the Mayberry test. These are summarised below:

Incorrect assignation of a level to certain items

Mayberry first developed a behavioural definition for each level, then designed the items to these definitions. Errors appear to have occurred for two reasons. First, not all questions are designed to include behavioural characteristics solely of that level, and, second, not all behavioural definitions appear to interpret correctly van Hiele's level descriptions. For example, not all Level 1 items are constructed to Mayberry's behavioural definition. In Item 11, Mayberry requires students to identify an equilateral triangle as an example of an isosceles triangle. To do this, students need be aware of class inclusion, a Level 3 skill, and be aware that the apparently simple question is asking for demonstration of that knowledge.

In the second case, Mayberry's interpretation of behavioural expectations for Level 4 does not appear to match fully van Hiele's description of the level. For example, some students whose highest level of reasoning was shown to be Level 3, answered correctly Mayberry's Level 4 items which required only a two-step proof. In addition, the nature of these items does not allow students to demonstrate Level 4 ability, i.e., ability to compare, transpose and operate with relationships. Mayberry appears to have overlooked van Hiele's statement, that the construction of a simple proof, such as the deduction of side and angle equality using congruency theorems, is a Level 3 skill. Several incorrect assessments resulted from this factor.

In summary, errors have been shown to exist in some of the test items designed by Mayberry to measure understanding at particular van Hiele levels. The implication of these errors is that they have the potential to result in the incorrect assessment of students' understanding.

Unequal treatment of concepts across levels

When collating the results, the students were coded at a higher level of understanding for the circle than for all other concepts tested. Analysis of the students' responses together with the Mayberry items led to the conclusion that many of 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. Additionally, there were a number of repetitive questions, and several yes/no types of questions. In particular, there appeared to be a paucity of Level 3 questions. This did not occur with the other concepts tested.

The re-coding of the circle items in the amended version of the Mayberry test resulted in lower evaluations of students' levels of understanding of the circle, and resulted in level assessments for this concept which were more in keeping with the results for all other concepts. The unequal treatment of the concept circle in comparison to the treatment of the other six concepts in the original Mayberry test, has the potential for students to be assessed at a higher level than their responses would indicate.

Uneven distribution of questions across levels

In designing her study, Mayberry ensured that there were one or more questions set to test for each level, across every concept. However, the number of items designed for each section varies. For example, while there is only one question set to test for Level 4 knowledge with similarity, four items containing thirteen scoring parts test for Level 3 understanding for the same concept. Overall, more than half of the fifty-seven items were designed to test for possession of Level 3 skills. This may be because Mayberry expected most of her subjects to display Level 3 reasoning. However, this approach does lead to an uneven testing of a concept across the four levels. Additionally, Mayberry's success criteria set for each level and concept vary from 50% to 100%. The combination for a level and concept, of a low number of test items together with a very low or very high criterion resulted in incorrect assessment of some students' levels of understanding, resulting, on occasions, in response pattern errors. Capable students who misinterpreted some of the easier questions (Levels 1 and 2) were particularly susceptible to this design weakness. The response pattern errors were eliminated when the students' responses were re-assessed by the Gutiérrez *et al* method.

Not having an even distribution of items across levels, has the potential to lead to the incorrect assessment of a student's understanding of geometry. The effect of the design weakness is particularly apparent when a low number of test items is combined with a very high or very low success criterion.

Unbalanced distribution of question focus within levels

Mayberry designed items to focus on specific behavioural characteristics. The distribution of the items across the characteristics is not always balanced. This is particularly apparent with Level 3 items. Such an imbalance can result in a student who lacks exposure to a particular aspect, e.g., class inclusion, to be affected adversely by the structure of the test. This design fault, although leading to some response pattern errors, resulted more commonly, in the under-assessment of students. Response pattern errors occurring as a result of this design weakness were eliminated in the Gutiérrez *et al* assessment, a method in which a student's dominant level of reasoning was more consistently recognised.

The implication of a lack of balance in the distribution of question focus is that this allows for the under-assessment of a student who, while generally possessing understanding of a particular level, may not have had exposure to some particular aspect of the level.

Conclusion

When collating the results for the Mayberry test, several inconsistencies emerged which were not clarified by interviews. Investigation of the inconsistencies led to the identification of four main weaknesses in Mayberry's design, namely, incorrect assignation of levels to items, unequal treatment of concepts, uneven distribution of levels, and a lack of balance in question focus. The implications of these weaknesses is that they have the potential to result in the incorrect assessment of a student's van Hiele level of understanding geometry, particularly when associated with the setting of a very low or very high success criterion. Many of the response pattern errors appeared to parallel the inconsistencies. The development of an amended version of Mayberry's test, and the subsequent re-assessment of the students' responses, led to the elimination of some, but not all of the response pattern errors. There still appeared to be problems associated with the Mayberry assessment method. A further re-assessment of the students' responses using the method of Gutiérrez, Jaime and Fortuny, in which each response is evaluated for the level demonstrated therein, resulted in the elimination of the remainder of the response pattern errors. This alternative method identifies the dominant level of reasoning displayed by a student, together with any indication by the student, of exploratory attempts at another level. However, the Gutiérrez *et al* method of assessment is not without difficulties, especially in terms of the time necessary for a person to understand and apply the categories. This suggests that a combination of the Mayberry questions and a simplified version of the Gutiérrez et *al* method of assessment would provide a valuable practical tool.

Implications for Teaching

In analysing the students' responses, some behaviour patterns emerged. While some students had difficulty in recognising a figure in a non-standard position, the most difficult problem was associated with angles, in particular, (a) the recognition of angle properties (Level 2), (b) the inability to use angular measurement (Level 2), and (c) the inability to reason with non-quantified angles (Level 3).

A lesser familiarity with the properties of angles of figures than with side properties was apparent with the weaker students. This supports the notion that the angle concept is particularly difficult in that it requires the integration of a number of different facets. These include the recognition that an angle consists of only two line segments, that they join at a

point which represents the vertex, and that the length of the line segments is not significant (Mitchelmore 1993, p.405).

Some students who recognised angles appeared unable to use the angular measurement, the degree. The students gave a linear measurement for angular magnitude. Whereas these students are able to recognise the angle as the junction of two lines, in describing the magnitude of the angle as the distance between the two line segments forming the angle, they are demonstrating an inability to abstract the notion of rotation of a linear segment from the concept of linear distance between the arms of an angle.

Other students who appeared to be in the early stages of exploring Level 3 notions, quantified angles as an aid to searching for an answer. In particular, the non-90° angles in right triangles were labelled as 45°, even when the triangles were obviously not isosceles. This tendency to label angles with a familiar symbol indicates an incomplete understanding of the angular concept.

Overall, the results showed students to have a wide range of understanding of geometry, from students unable to identify figures (No Level) to students achieving Mayberry's criteria for Level 4. All students were able to be assessed, with the interviews supporting the levels shown in the written responses. However, the difficulty some students had in recognising figures in non-standard positions indicates that teachers and textbook authors need to be careful in presenting figures in all orientations. Additionally, teachers need to be aware that the angle concept is particularly difficult, requiring the integration of a number of different facets. These include the discernment of key features, and the mastering of a unique measurement system (Davey & Pegg 1991).

Implications in Relation to the van Hiele Theory

The analysis of the results allows further insight into the van Hiele theory. It provides further empirical evidence about some of the properties and features associated with the theory. In particular, these include: discreteness, hierarchy, language and levels, differences in perception at different levels, different levels for different concepts, and crisis in thinking.

Discreteness

The analysis supports the notion that the levels are not discrete. They are of a more dynamic nature. Mayberry, in assuming that the levels are discrete, designed each question to measure a student's understanding of a single van Hiele level. Implicated in this is the notion that a perfect student who is at level *n* would answer correctly all questions at and below that level,

but none above that level. Mayberry recognised the fallibility of this in reality, and, hence, set a criterion for each level and concept, the achievement of which was deemed to indicate understanding of the level, for the concept.

However, the assessment method did not account for all situations. Many students offered some responses at the next level above the highest level at which they achieved Mayberry's criterion for the concept. Other students achieved the criterion for a level, having clearly shown a lack of awareness of some feature of a lower level. For example, a student can achieve Mayberry's Level 3 criterion for similarity without answering correctly any questions about angles. The Gutiérrez *et al* method of assessment, based on the assumption that the levels are continuous, shows that most students, while demonstrating a dominant level of reasoning in their responses, often display some attempt at reasoning at the next highest level. There appears to be some critical degree of mastery of one level, before this exploration of the next level commences. This critical degree of mastery varies for different students, a few students indicating preference for complete mastery of understanding of one level before they are ready to consider the next level.

Hierarchy

All three analyses of the results, the Mayberry assessment, the Rasch analysis, and the Gutiérrez *et al* assessment show the organisation of the levels to be successive, i.e., they support the notion of the levels having a hierarchical structure. Response pattern errors are records of non-hierarchical results. In the initial analysis, i.e., using Mayberry's approach, only nineteen (2%) response pattern errors occurred out of the possible 976 results, giving a Guttman co-efficient of reliability of 0.98. This value is greater than Torgenson's acceptable value of 0.90, and, hence, supports the notion of the hierarchical structure of the levels.

The re-assessment of the students' responses, first using the amended version of Mayberry's test, and second, evaluating by the method of Gutiérrez *et al*, eliminated all response pattern errors. This indicated that the nineteen response pattern errors resulted from weaknesses in the Mayberry test and assessment method, and, hence, were not indications contrary to the notion of hierarchical structure of the levels.

Language and levels

Some of the students' responses exhibited a strong linkage between certain words and word combinations, and the various levels at which they were working. For example, the following visual descriptions were given by Level 1 students in response to a Level 2 request for properties of figures: *"does not meet"*, *"not a full shape"*, *"looks like a flat tyre"*,

"crosses in the centre", "looks like one (circle)", and *"it is oval"*. These visual descriptions compare with responses to the same Level 2 question, given by students working at Level 3: *"it is not a closed curve", "not all points equidistant from a central one", "it is not curved, it has distinct corners", "it has an infinite number of points equidistant from a central point"* and *"the curve doesn't have a common radius."* Students working at Level 1, and, hence, using the former descriptors, if being addressed by a teacher using Level 3 language, would have difficulty in understanding the lesson.

Differences in perception at different levels

In their responses, students interpreted questions according to their level of understanding. For example, until students have met class inclusion, they are unaware that a square satisfies the definition of a rectangle. They perceive squares and rectangles as two different figures. This notion is supported in the responses to Item 25(b), "Are all squares also rectangles? Why?" A student who was assessed as being at Level 2, and who appeared not to have met the notion of class inclusion, responded with *"No, squares have all sides the same, while rectangles have two sides longer than the others."* This response compares with the response of a Level 3 student appreciating class inclusion, *"Yes. they both have two pairs of opposite sides equal."* While all these responses are correct for the level at which each students is working, the comparison of these responses illustrates clearly the feature, that students at different levels have different perceptions.

Different levels for different concepts

Mayberry's null hypothesis, that the students should be at the same level in all geometric topics to which they have been exposed, was rejected by her. Since this study tested students across four, rather than seven concepts, the results are not as representative as those for Mayberry's study when considering the possibility of students displaying different levels for different concepts. However, for all three assessments, the responses of the majority of students indicated they were not at a single level for each concept.

The results of Mayberry's amended test, in eliminating many of the weaknesses in the original test, should give the best indication of whether there was consensus of levels across concepts. Forty-six of the sixty-one students failed to show the same level of understanding across the four concepts for which they were tested. The complexity of the Gutiérrez *et al* results, in giving a degree of understanding for each concept, reduces the likelihood of students showing identical degrees of mastery across all concepts. Only one student did so. Thus, any indication of consensus of levels of understanding should be interpreted according to the complexity of the results.

Crisis in thinking

For students to move from one level to the next, they must face a crisis in thinking. If they avoid the crisis, they will not develop in thinking, remaining at the same level. Written responses indicate the level at which a student is reasoning at the time, and for that item. Hence, written responses are unlikely to show if a student has faced a crisis in thinking in responding to the item. However, during interviews, students can be prompted to see whether they can be induced to demonstrate a higher level of thinking.

In this study, prompting in interviews seldom resulted in a student giving a response at a higher level. Inability to accept a crisis in thinking is illustrated in the following interview responses to Item 45 which tests for understanding of necessary and sufficient conditions (Level 4) for a square. In the interview, weaker students were probed for awareness of class inclusion. S33's answers to the written test showed strong Level 2 understanding, but little or no Level 3 reasoning. The original question allowed the student to demonstrate the depth of her Level 2 knowledge in her listing of properties of diagonals, as well as the more familiar properties of sides and angles. In response to the prompt "Can ABCD be a rectangle?", the student replied *"We know the four sides are equal and parallel, and the angles are 90°. I know you want me to see something more, but I can't work out what it is. I am confused. No, it can't be, the four sides of a rectangle aren't equal."* This response shows clearly that the student recognised there was a crisis to face, but was not ready to accept the crisis, and, hence, the student remained at Level 2.

In conclusion, the above discussions of some of the properties and features associated with the van Hiele theory show that the analysis of the students' responses not only gives us a clearer perspective about the Mayberry test and the results it generated, it also allows further insight into the van Hiele theory, its properties and its features.

Implications for Further Research

The findings of the study suggest many further research questions. Several of these have already been identified. Six basic research issues stand out.

First, research is required into the appropriate testing of Level 1. The results indicate that the better students may have had difficulty in interpreting the thrust of the Level 1 questions. In the Mayberry assessment, although several of the better students failed to give correct responses to a Level 1 question, a criterion of one correct response out of two avoided the occurrence of response pattern errors. The infit mean square maps from the Rasch analysis indicated that a larger than expected number of good students gave incorrect responses to the

Level 1 items. The occurrence in the final assessment (using the coding system of Gutiérrez *et al)* of non-hierarchical results, all between Levels 1 and 2, supports the notion that better students have difficulty with Level 1 questions. This raises the question, what is the appropriate way to test for visual recognition (Level 1) in a written format?

Second, research is required into the difficulties experienced by students who consistently performed at higher levels in interpreting the context of questions designed to explore lower levels. This feature showed in the results of this study, causing the occurrence of response pattern errors. The Gutiérrez *et al* assessment, in giving a more realistic evaluation of the ability of students, eliminated the response pattern errors, but does not offer any explanation of the difficulties experienced by the better students. The feature has been remarked upon by several other researchers, (e.g., Gutiérrez & Jaime 1987; Mason 1989; Mayberry 1981; Usiskin 1982).

Third, research is required into the development of understanding of geometric concepts by mature-age students. This study identified many mature-age students at Level 2. Have these mature-age students stayed at the same level of understanding over an extended period of time, broadening their experience, or have they achieved a higher level of understanding sometime in the past, and then regressed to Level 2?

Fourth, given that we can find levels at which the students are actually working, what are the major impediments that seem to be preventing the students from reaching higher levels? This is particularly important for students, as in this sample, who are training to be teachers. A minimum requirement for them could be expected to be Level 3. Many of them have been taught at this level, but they cannot operate there.

Fifth, research is required into how one interprets partial knowledge of a van Hiele level. For example, in the Mayberry analysis, students were assessed as reasoning at a given van Hiele level if they achieved the criterion set by Mayberry. For the square, students were deemed to be at Level 3 if they answered correctly six of the nine questions. However, how does one interpret a score of five out of nine? Is it sufficient to say that student is not reasoning at Level 3? The Gutiérrez *et al* assessment allows for this in part, in that a student scoring 60% for a level is said to have an intermediate degree of mastery of the level. However, the question still remains, what is meant by mastery of 60% of a level?

Sixth, having validated the levels of the test, it would be valuable to replicate the study with

a) equivalent samples of students at other universities, and
b) other cohorts of students who need a strong geometric background, e.g., secondary mathematics pre-service teachers.

Conclusion

The study has examined the determination of students' van Hiele levels of understanding geometry through the testing of four research themes. In carrying this out, the study has investigated the difficulties associated with the designing of appropriate test items, and the advantages of different methods of assessing the responses of students to the test items.

Initially, Mayberry's test was replicated using a written format. Results of the Mayberry assessment were analysed through the application of the QUEST program which utilises the Rasch measurement theory. This analysis not only confirmed the results, it also confirmed the emergence of inconsistencies in some of the students' reasoning. Interviews did not clarify the inconsistencies. Analysis of these inconsistencies led to the identification of four features in the Mayberry design which had the potential to lead to an incorrect assessment of a student's level of understanding.

The detailed analysis of the validity of the test items gave a clearer perspective about the Mayberry test and results. However, it also indicated the need for the investigation of another form of interpreting results. The responses of the students to the Mayberry items were, therefore, re-assessed using the alternative method presented by Gutiérrez, Jaime and Fortuny. This led to a more realistic evaluation of students' van Hiele levels of understanding geometry, a deeper interpretation of the inconsistencies in the Mayberry design, and it also provided a comparison of two methods of assessment, indicating the advantages of each.

Mayberry acknowledged four specific design problems associated with her investigation, namely, the need for a larger sample size, the need to select a more diverse sample of students, the need for further refinement of the choice of representative questions, and the need to evaluate the subjectivity of the success criterion. This study has addressed these four issues. It has also provided further insight into the van Hiele theory. In particular, it provides further empirical evidence about what it means to work at a particular van Hiele level.