An Evaluation of Two Coding Systems in Determining van Hiele Levels

Christine Lawrie
University of New England

In the early 80s Mayberry (1981) developed a diagnostic instrument to be used in an interview situation to assess the van Hiele levels of pre-service primary teachers. At the University of New England, a detailed testing and interview program was undertaken, replicating the Mayberry study. The students' responses to the Mayberry items were assessed using two different methods, first by Mayberry's method, and second, using the method developed by Gutiérrez, Jaime and Fortuny. This paper presents an evaluatiuon of the two coding systems.

The ability to instruct students at their level of understanding is dependent, in part, on the teacher being able to assess the students' levels of understanding. In order to make this assessment, there needs to be available a reliable diagnostic instrument. In the early 80s Mayberry (1981) in her work with pre-service primary teachers, developed such a diagnostic instrument that could be used in an interview situation. Mayberry's test items and method of evaluation are based on the key assumption that the van Hiele levels are discontinuous (Mayberry 1981, p.22). This led Mayberry to design each item to test for understanding of a specific van Hiele level, the response being assessed on whether it reflects that level of thinking. There is no grading of the degree of difficulty of the Mayberry items within a level, nor of the depth of understanding of the level displayed in a response.

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 that they are of a more dynamic nature, that they are continuous rather than static (Pegg, 1992, p.25). Their theory (Gutiérrez, Jaime and Fortuny 1991, p.237) is based on observations that, when answering questions, although most students show a dominant level of thinking, a response frequently displays some reflection typical of another level. This paper presents an evaluation of the two coding systems in determining van Hiele levels displayed by the students in their responses to the Mayberry items.

Background

The van Hiele Theory

In the 1950s, Pierre van Hiele and Dina van Hiele-Geldof completed companion PhDs which had evolved from the difficulties they had experienced as teachers of Geometry in secondary schools. Whereas Dina van Hiele-Geldof explored the teaching phases necessary in order to assist students to move from one level of understanding to the next, Pierre van Hiele's work developed the theory involving five levels of insight. A brief description of the first four van Hiele levels, the ones commonly displayed by secondary students and most relevant to this study, is given:

Level 1 Perception is visual only. A figure is seen as a total entity and as a specific shape. Properties play no explicit part in the recognition of the shape.
Level 2 The figure is now identified by its geometric properties rather than by its overall shape. However, the properties are seen in isolation.
Level 3 The significance of the properties is seen. Properties are ordered logically and relationships between the properties are recognised.
Level 4 Logical reasoning is developed. Geometric proofs are constructed with meaning. Necessary and sufficient conditions are used with understanding.

The van Hieles saw their levels as forming a hierarchy of growth. A student can only achieve understanding at a level if he/she has mastered the previous level(s) They also saw (i) the levels as discontinuous, i.e., students do not move through the levels smoothly, (ii) the need for a student to reach a 'crisis of thinking' before proceeding to a

new level, and (iii) students at different levels speaking a 'different language' and having a different mental organisation.

Mayberry's Research

JoAnne Mayberry's study (1981) investigated, in part, whether the van Hiele level, at which a student is functioning in geometry, can be discerned. To carry out this investigation, Mayberry created a diagnostic instrument consisting of 62 items (many of them containing separate question parts) designed to the operational definition of each of the five levels. The items covered seven geometric concepts, namely, square, right triangle, isosceles triangle, circle, parallel line, congruency, and similarity. These concepts all occur in the elementary curriculum in the USA. A matrix/grid was used to develop questions by level and concept so that the questions would have parallel forms. One or more questions were developed for each cell in the grid. Experts in the fields of mathematics and mathematics education, among them Pierre van Hiele, were asked to validate the items by judging whether the items satisfied certain criteria (Mayberry, 1981, p. 52). The final form of the diagnostic instrument was then used in an interview situation to investigate the understandings of 19 pre-service elementary education students at Georgia College, Milledgeville, Georgia. In her method of assessment, a credit point is given "if the subject's responses to questions . . . indicated that the subject was thinking on the given level" (p.60). However, in her study, Mayberry assessed the response to every question part equally, whether the question required a simple yes/no answer, or whether it required a complex explanation. A criterion was set for each concept and level (ranging from 50% to 100%). If a student has given sufficient correct answers to reach Mayberry's criterion, the student is credited with having mastered that van Hiele level.

The research of Gutiérrez, Jaime and Fortuny

The method of evaluation developed by Gutiérrez, Jaime and Fortuny, based on the premise that the levels are continuous, results in a qualitative assessment of a student's degree of reasoning in each of the four levels. Gutiérrez *et al* (1991, pp.238-239) maintain that initially students are not aware of the need to think at a level. 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 a complete acquisition of the thinking at that level.

Several steps are necessary in evaluating a student's van Hiele level(s) 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 (p.239). For example, a response which appears to be close to the necessary and sufficient conditions sought in Mayberry Item 24 can be an attempt at expressing minimum conditions (Level 4), or it can be a statement of the few properties known by the student for that topic (Level 2). Consideration of the student's other responses is necessary to determine which is the correct level.

Item 24
Circle the smallest combination of the following which guarantees a figure to be a square.

a. It is a parallelogram.
b. It is a rectangle.
c. It has right angles.
d. Opposite sides are parallel.
e. Adjacent sides are equal in length.
f. Opposite sides are equal in length.

It is not always necessary to consider other statements or responses. Many students in their responses consistently show a dominant level of thinking. With such students, their statements constantly re-confirm their most common level of reasoning. However, the assessor needs to be aware of students who are beginning to explore a

higher level for some aspect of a concept and may give a better than expected response if the question is focussing 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. This categorisation depends on the degree of mathematical accuracy, and on how complete the solution to the question is. Gutiérrez *et al* (pp.239-240) explained "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 completeness and correctness. The responses are then quantified according to each vector, and the student's degree of acquisition of each van Hiele level 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 and Burger 1991, p.109). If a response has been given at Level *n* when the question could have been answered at Level *(n+1)*, a zero score is given for Level *(n+1)*. *However,* if a response has been given at Level *n* and the question could also have been answered at Level *(n-1),* a score of 100 is allocated to Level *(n-1)*, since a response at Level *n* implies complete acquistion of Level *(n-1)* (p.246). 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.

Design

In order to consider Mayberry's work in an Australian context, a detailed study of the geometric understanding of 61 first-year primary-teacher trainees was carried out at the University of New England. The study aimed, in part, to provide a written test based on the Mayberry interview schedule. Conversion of the Mayberry items to a written test involved some modification of the wording to ensure that the intention of each question was clear. A preliminary study validated the reliability of the written questions. Level 5 items were omitted, hence the written test assessed van Hiele Levels 1 to 4 (Mayberry items 1 to 57). Follow-up interviews were conducted with students to validate the levels of thinking as determined in the written test.

The responses were assessed initially using Mayberry's method. However, there were indications that not all questions measured the level for which they were designed (Lawrie 1993). This led to an amended version of the Mayberry test and marking scheme (Lawrie, in press). The students' responses were then re-assessed using the amended version of the marking scheme. Every endeavour was taken to replicate Mayberry's evaluation of responses. Her thesis was examined in depth to ascertain her expectations in the responses to the items. Following the amended Mayberry assessment, the students' responses were again re-assessed, this time using the method developed by Gutiérrez *et al.* For this assessment, all response attempts by students, whether wholly or partially correct or incorrect, contributed to the overall

picture of their depth of understanding of a topic. Also, to allow for the unequal expectations for some of the Mayberry questions, e.g. the difference between the expectations of a yes/no type of question compared to a question requiring a complex explanation, each item or group of question parts in the Mayberry test were assessed as one complete item. The two coding systems were than evaluated.

Results

Table 1 shows the number of students demonstrating reasoning for each van Hiele level and for each of the seven concepts for both methods of assessment.

Table 1
Number of Students Attaining Each van Hiele Level

Concept	van Hiele Level	Amended Mayberry Results	Gutiérrez et. al. Results				
			Degree of Acquisition				
			None	Low	Intermed	High	Complete
Square	0	0					
n = 61	1	2	0	0	0	0	61
	2	51	0	2	7	24	28
	3	4	37	9	13	2	0
	4	4	59	2	0	0	0
Right Δ	0	1					
n = 31	1	6	0	1	0	1	29
	2	17	0	2	2	13	14
	3	6	14	10	2	4	1
	4	1	30	1	0	0	0
Isos Δ	0	2					
n = 30	1	8	0	0	2	0	28
	2	12	0	2	1	8	19
	3	5	15	6	5	4	0
	4	3	29	1	0	0	0
Circle	0	0					
n = 31	1	5	0	0	0	0	31
	2	13	0	4	0	11	16
	3	12	17	8	3	2	1
	4	1	29	2	0	0	0
Parl lines	0	0					
n = 30	1	7	0	0	1	1	28
	2	22	0	2	3	7	18
	3	1	23	6	1	0	0
	4	0	30	0	0	0	0
Congruenc	0	0					
n = 31	1	11	0	0	0	7	24
	2	14	2	2	5	10	12
	3	2	16	7	5	3	0
	4	4	27	4	0	0	0
Similarity	0	0					
n = 30	1	13	0	0	1	0	29
	2	12	0	5	4	3	18
	3	3	18	7	5	0	0
	4*	2	26	3	0	0	0

*Similarity Level 4, n = 29

The amended Mayberry results give the highest level achieved by a student for each concept, while the Gutiérrez *et al* results show the degree of acquisition reached by the students for each level. Both sets of results agree with the hierarchical structure of the van Hiele levels. The results from the Gutiérrez *et al* method of assessment indicate that the higher the level, the lower the degree of acquisition, and, in general, confirm the findings of the Mayberry assessment, that the majority of students have mastery of Levels 1 and 2, but little or no understanding of the higher levels.

A closer inspection of the two sets of results shows that a high or complete degree of acquisition in the Gutiérrez *et al* results corresponds with mastery of a level in

the Mayberry evaluation, while a low or no acquisition corresponds with failure to reach Mayberry's criterion. Results showing an intermediate degree of acquisition (40 to 60) tend to correspond with scores close to the Mayberry criteria, some failing to reach the criterion, others achieving mastery of the level. For example, for the concept square, eight students (4@L3 and 4@L4) have been assessed as demonstrating mastery of Level 3 or better in the Mayberry results, while the Gutiérrez *et al* results for Level 3 show only two students having a high or better degree of acquisition. The balance of six students showing Mayberry mastery are among the thirteen students registering an intermediate degree of acquisition. Again, for the triangle results, the Mayberry scoring shows one student failing to reach the criterion in the identification of the right triangle while being assessed as having a low degree of acquisition of Level 1 in the Gutiérrez *et al* assessment. Similarly, the two students failing to reach the criterion in identifying the isosceles triangle, are both shown as having an intermediate degree of acquisition for Level 1 for the later assessment. These results illustrate one difference between the two methods of assessment. With the Gutiérrez *et al* method of assessment, the taking into account all responses, whether complete and/or correct, makes the evaluation more realistic and gives a more accurate measure of a student's degree of understanding of geometry.

However, a comparison of the results for the two methods does not always show correspondingly similar results. In the assessments of the responses to the circle questions, the amended Mayberry evaluation credited thirteen students (12@L3 and 1@L4) with mastery of Level 3 or better. In contrast, the Gutiérrez *et al* method found only six students with a comparable degree of mastery, one having a complete degree of acquisition, two having a high degree and three having an intermediate degree of acquisition of the level. The remaining twenty-five were not credited with having better than a low degree of acquisition of the concept circle for Level 3. Inspection of the Level 3 items for the circle showed there were several question parts requiring only a yes/no type of answer. Answers to such questions can earn a score in the Mayberry assessment method without requiring demonstration of Level 3 reasoning. This supports the notion that the results of coding using the method of Gutiérrez et. al. are the more realistic.

In conclusion, the alternative paradigm described by Gutiérrez *et al*, because it measures a student's capacity to use each one of the van Hiele levels in every statement made, results in a more flexible interpretation of the reasoning of the student. In particular:

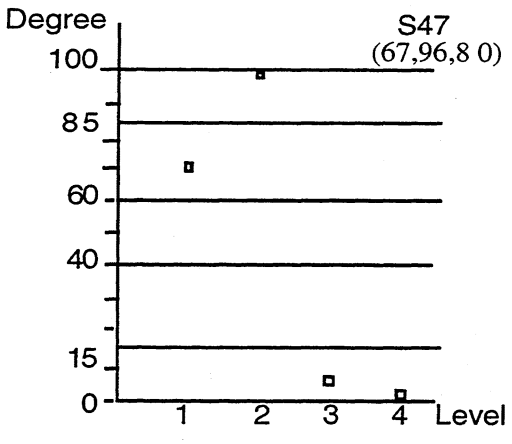
1. a student can be shown to be developing in two consecutive levels of reasoning at the same time;
2. the incorrect assignation of a level to a question is of minimal significance;
3. the effect of unequal distribution of questions across levels in minimised;
4. incorrect assessment resulting from 'lucky'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

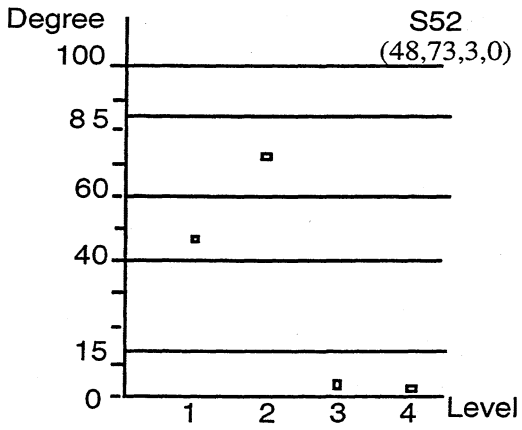
Unusual Behaviour Patterns

An inspection of the quantitative results obtained in the re-evaluation of the students' responses using the method of Gutiérrez *et al* reveals that not all results agreed with the hierarchical structure of the van Hiele levels. In twenty-six (11%) of the two hundred and forty-four assessments, the degree of acquisition of Level n is 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 indicated in responses at other levels. 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, one student (S05) scored 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. The results of two students (S47 and S52) illustrate cases in which 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. The two patterns of unusual behaviour are graphed below.

Graphs of Unusual Behaviour Patterns

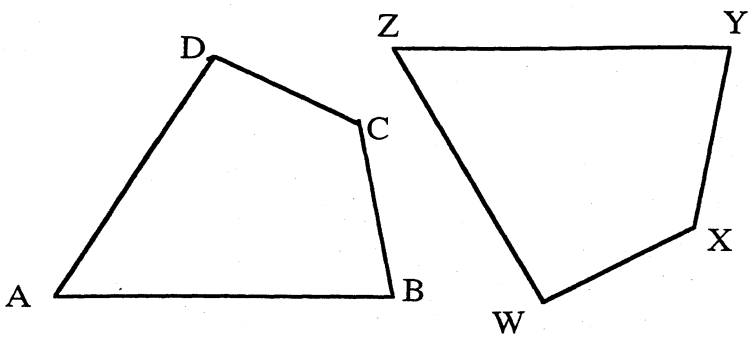




The occurrence of these patterns in which the degree of acquisition of Level 1 is lower than it should be is suggestive of two factors; (a) that the Mayberry questions designed to measure Level 1 are not always clear in their intention, and/or (b) that the criteria adopted to measure the acquisition of levels using the alternative paradigm of Gutiérrez *et al* may be more suited to responses demonstrating reasoning as with Levels 2, 3 and 4 rather than to the visual identification of Level 1. A closer inspection of the responses of the twenty-six students whose results showed the degree of acquisition of Level *n* being greater than the degree of acquisition of Level *(n-1)* confirmed the under-evaluation of the students' Level 1 understanding. This supports the notion above, that the Mayberry test questions and the Gutiérrez *et al* evaluation method do not necessarily measure Level 1 adequately.

Although Gutiérrez *et al* (1991, 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 evaluations 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 understanding. For example, one student (S06), although failing to reach Mayberry's Level 2 criterion, demonstrated sufficient knowledge of the relationships between congruent figures to reach the Level 3 criterion. 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. Her responses (below) to two of the congruence questions illustrate this feature.

Item 22 *(Level 2)*



These are congruent figures.

What is true about their sides?
AD = _____

What is true about their angles?
∠B = _____

S06 correctly named the corresponding side WZ and the corresponding angle, ∠Y. Her responses to the other question parts are correct but 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 *(Level 3)*

Δ ABC is similar 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*

As shown for Item 43, several of S06's responses to the Level 3 questions demonstrated an understanding of congruency relationships. Previous research has also found students who answer higher level items better than lower level ones (Mayberry 1981; Usiskin 1982).

Conclusion

Whereas the Mayberry coding system provides a more direct and quicker method of determing a student's van Hiele level of undertanding geometry, the method developed by Gutiérrez *et al*, in evaluating the degree to which understanding of each level is expressed in every response, measures more comprehensively, the van Hiele level of reasoning of a student. The latter provides a mechanism for measuring a student's degree of acquisition in two or more levels. It also provides insight into the quality of a question. It minimises many of the problems occurring in the Mayberry items and the associated method of assessment. However, the alternative paradigm, as developed by Gutiérrez *et al,* also needs further investigation and refinement.

1. Questions at van Hiele Level 1 are not always assessed accurately. Is there a limitation in the ability of the coding system of Gutiérrez *et al.* to evaluate visual recognition?
2. The automatic allocation of a credit of 100 for Level *(n-1)* for an attempt at Level *n* does not always seem justified and possibly presents a contradiction to other aspects of their coding system.
3. The coding system is very time-consuming. Whereas it gives a much more detailed and accurate picture of a student's ability to work in each van Hiele level, making it an excellent research tool, it is not suitable for use in the regular classroom in its present format.

Gutiérrez *et al* note that some students show a better acquisition of Level 3 than of Level 2. As with previous research (Mayberry, 1981; Usiskin, 1982), this study has found cases in which students demonstrating a high degree of acquisition of Level 3, do not show their full understanding of Level 2 in their responses to the lower level questions. This needs further investigation.

This research shows that there are many advantages in the coding system developed by Gutiérrez, Jaime and Fortuny, particularly when compared to Mayberry's coding system. However, it also demonstrates that further refinement is needed.

References

Lawrie, C. J. (1993). Some problems identified with Mayberry test items in assessing students' van Hiele levels. *Proceedings of the 16th Annual Conference of the Mathematics Education Group of Australasia,* Brisbane, 381-386.

Lawrie, C. J. (in press) An investigation into the assessment of a student's van Hiele level of understanding in geometry. University of New England, Doctoral thesis.

Gutiérrez, A., Jaime, A., & Fortuny, J. M. (1991) An alternative paradigm to evaluate the acquisition of the van Hiele levels. *Journal for Research in Mathematics Education,* 22, 237-251.

Gutiérrez. A., Jaime, A., Shaughnessy, J. M., &.Burger, W. F. (1991) A comparative analysis of two ways of assessing the van Hiele levels of thinking. *Trans. 15th Annual PME Conference.* Assissi (Italy). vol II, 109-116.

Mayberry, J. W. (1981). An Investigation of the van Hiele Levels of Geometric Thought in Undergraduate Preservice Teachers. Doctoral dissertation, University of Georgia. University microfilms no. 8123078.

Pegg, J. E. (1992). Students' Understanding of Geometry: Theoretical Perspectives. In B Southwell, B. Perry, K. Owens (Eds), *Space The First and Final Frontier* (pp.18-36). Sydney: Mathematics Education Research Group of Australasia.

Usiskin, Z (1982) *Van Hiele levels and achievement in secondary school geometry (Final report of the Cognitive Development and Achievement in Secondary School Geometry Project)* Chicago, IL. University of Chigago, Department of Education.

van Hiele, P. M. (1957) The problem of insight in connection with school children's insight into the subject-matter of geometry (Summary of doctoral dissertation, University of Utrecht) In D. Fuys, D. Geddes, and R. Tischler (Eds. and Trans.) *English translation of selected writings of Dina van Hiele-Geldof and Pierre van Hiele* (pp.237-241). New York: Brooklyn College, C.U.N.Y.

van Hiele, P. M. (1986) *Structure and Insight.* Florida: Academic Press, Inc.