THE DESIGN AND EVALUATION OF AN INSTRUMENT

FOR ASSESSING MASTERY VAN HIELE LEVELS

OF THINKING ABOUT QUADRILATERALS

by

Mary Lora Noffsinger Crowley

Dissertation submitted to the Faculty of the Graduate School of The University of Maryland in partial fulfillment of the requirements for the degree of Doctor of Philosophy
1989

Advisory Committee:

Associate Professor James Henkelman, Chairman/Advisor
Professor James Fey
Professor Martin Johnson
Professor Clayton Stunkard
Associate Professor Neil Davidson

ABSTRACT

Title of Dissertation: THE DESIGN AND EVALUATION OF AN INSTRUMENT FOR ASSESSING MASTERY VAN HIELE LEVELS OF THINKING ABOUT QUADRILATERALS

Mary Lora Noffsinger Crowley, Doctor of Philosophy, 1989

Dissertation directed by: James Henkelman, Associate Professor, Department of Curriculum and Instruction

The goal of this project was to create a 40 minute long multiple-choice instrument to assess an individual's dominant level of thinking, as described by the van Hiele model of the development of geometric thinking, on the topic of quadrilaterals. The study was composed of four stages: (a) item development, (b) pilot testing, (c) field testing and (d) final testing. Initially 53 items were developed and reviewed by a panel of experts. The revised items were then administered to 14 pilot study subjects, and, subsequently, to 113 field test subjects, both groups ranging in academic background from sixth grade to university. Item analysis comparing these subjects' choices of level specific responses and their dominant van Hiele level, as determined through the Burger and Shaughnessy interview, resulted in the identification of 19 items for the final instrument, the van Hiele Quadrilateral Test. For scoring purposes, the items on the test are considered as four subtests, with 4, 5, 6 and 4 items corresponding to Levels 1, 2, 3 and 4, respectively. The items

associated with Levels 2, 3 and 4 met all the item analysis criteria. Two interpretation schemes were identified.

The final instrument was administered to 50 subjects in ninth grade and 51 subjects in twelfth grade. Grade membership and performance on the Nova Scotia Achievement Mathematics Basic Concepts Test were compared to subtest performance and to the resulting mastery decisions. Chi squared statistics failed to support the independence of grade membership and van Hiele level. The correlation statistics, ϕ, indicated that there was a weak correlation between grade level and mastery of Levels 1 and 2, with stronger statistics associated with Levels 3 and 4. Little of the total variance in mastery designations ($\eta^2_{Y,X}$) was attributed to variance in grade level. Little to moderate variance ($\eta^2_{Y,X}$) in performance on the Achievement Tests was attributed to variance in the van Hiele level assignments. Two types of criterion-referenced reliability statistics, the agreement coefficient, p_o, and Cohen's Kappa, K, were also determined. These indices suggest that the subtests do not yield consistent results for these subjects. Until reliability can be established, the instrument is not appropriate for determing van Hiele mastery levels. The implications of these findings and suggestions for further research are considered.

Dedication

To my children, Amy and Adam Crowley, who have only known a "student" mother, and to my father, who has been there from the beginning.

Acknowledgements

There are many individuals who through their assistance and encouragement deserve to be acknowledged at the end of this research project. I hope that along the way I have let most of those individuals, particularly those closest to home, know how much their presence has meant to me.

At the University of Maryland, I wish to acknowledge the support and confidence shown in my long distance pursuit of a doctoral degree by the "mathematics education" group in the Department of Curriculum and Instruction. I am especially grateful to those Department members who served on my committee, Professors Fey, Davidson and Johnson. As well, I wish to acknowledge Professor Clayton Stunkard, from the Department of Measurement and Statistics, who was always available for statistical advice. Finally, I wish to thank my advisor, James Henkelman, who has consistently encouraged me to stretch my abilities to the limit.

I am also indebted to the scores of students, teachers, parents and administrators who allowed me to test and interview. Without their cooperation, this research could not have been accomplished.

TABLE OF CONTENTS

LIST OF TABLES

List of Figures

Chapter 1

INTRODUCTION TO THE STUDY

The role of geometry in the school curriculum is an on-going topic of debate amongst mathematics educators (Craine, 1985; Fey & Good, 1985; Gearhart, 1975; Hoffer, 1981; Lindquist & Shulte, 1987; Shaughnessy & Burger, 1985; Usiskin, 1987). At the heart of the controversy are the perceptions that the curriculum is inappropriate and that student performance is inadequate (Usiskin. 1987). Each of these views subdivides into further specific issues for consideration. When discussing curriculum, for example, questions arise over what content and emphasis are desirable: there are supporters for teaching Euclidean geometry from a "traditional" point of view (Gearhart, 1975); there are advocates for investigating other types of geometries and/or for teaching Euclidean concepts in non-traditional ways (Fey & Good, 1985; MacPherson, 1985). Debate has arisen over how formal the approach to geometry should be: some educators support a rigorous axiomatic treatment (Suydam, 1985), others favor an informal, intuitive approach (Shaughnessy & Burger, 1985). A few think that "formal" geometry should be abandoned altogether (Norris, 1981). From each of these perspectives, organizational questions arise: should geometry be a one-year course, taught as a half-year course for two consecutive years, or integrated into each year's curriculum (Cox, 1985; Craine, 1985; Gearhart, 1975; Shaughnessy & Burger, 1985)?

Within each of these contexts, the issue of audience also arises: do all students, or only some, need the content and logical reasoning skills potentially available from the study of geometry (Cox, 1985)?

Teachers, students and researchers report that students are having problems with the current curriculum (Gearhart, 1975; Usiskin, 1987). High school geometry teachers express dissatisfaction with the geometric abilities students demonstrate. They feel that students entering formal geometry courses do not have the necessary prerequisite background. They observe that students leaving the course have not grasped the nature of a deductive system nor have they seen the need for deductive reasoning (Williams, 1980). The teachers note that a majority of their students do not find geometry "exciting and enjoyable" (Gearhart, 1975, p. 489).

Teachers apppear to be correct in their estimates that students find geometry frustrating. Students report that geometry is difficult, irrelevant (Kerr, 1981) and uninteresting (Hoffer, 1981). Perhaps this impression explains in part why approximately one half of all North American high school students do not even begin a study of formal geometry (Kerr, 1981; Usiskin, 1987).

Nationwide American standardized test results corroborate what many teachers and students already know: students are not doing well in geometric situations which require higher order skills such as synthesis and analysis. For example, the geometry questions given in the Third National Assessment of Educational Progress (Carpenter, Lindquist, Matthews & Silver, 1983) showed that students did well on exercises where recognition, recall and manipulation were required. Some understanding of certain basic geometric concepts was also demonstrated. Little knowledge, however, of the properties associated with those concepts was evident. Little ability to apply the properties was demonstrated (Carpenter et al., 1983). In a mammoth undertaking by the Cognitive Development and Achievement in Secondary School Geometry (CDASSG) Project, "a rather low level of (student) achievement in writing proofs" (Senk, 1985, p. 448) was reported. Their data suggested that only about 1/3 of all students in a traditional one year geometry course reach a "75-percent mastery level in proof writing" (Senk, 1985, p. 453).

In the face of such frustration and difficulty, one might ask "Why teach geometry?" A casual review of the literature highlights the following reasons:

1. Geometry is practical. It can be used to describe the world around us. It can be used to solve real world problems.

2. Through the study of geometry, one can derive cultural and aesthetic pleasures. A knowledge of space, shape and form, for example, can help one in appreciating nature, art, and architecture.

3. Geometry can serve as an introduction to the deductive method. Logical reasoning and the ability to understand and formulate abstract arguments can be developed.

4. Geometry is a unifying theme in mathematics. For example, areas of rectangles can be used to demonstrate multiplication of

binomials or the derivative of a function can be seen as the slope of the tangent line to the graph of the function.

5. Geometry is a prerequisite for the study of other fields. Physics, crystalline structures, and mechanical drawing are examples.

6. The study of geometry provides opportunites to develop spatial perception and visual skills.

7. The study of geometry provides opportunities for problem solving.

8. Geometry is a traditional topic of study.

With a list such as the above, some educators think that there is no need to further justify geometry's place in the curriculum. Gustav Choquet typifies this when he says "'I shall not discuss here the need for teaching geometry; I shall simply consider the way in which it can be done'" (Willson, 1977, p. 13). Other educators, however, feel that the rationale and goals for teaching mathematics, including geometry, need to be re-examined periodically. Indeed, during the last ten years there have been three internationally prominent reviews of mathematics education: the Cockroft Report, England, 1982, An Agenda for Action by the National Council of Teachers of Mathematics, 1980, and the National Council of Supervisors of Mathematics position paper, 1978. Each reaffirmed geometry as an essential content area in the education of school children.

Upon examination, then, the picture which emerges about geometry is one where the importance of studying the subject is generally accepted, yet there is a problem with its teaching and learning. Geometry is, as Fey and Good (1985) declare, "a troubled strand" (p. 44). "Modifications of the course are needed... but there is no clear consensus on the form such modifications should take" (Gearhart, 1975, p.490). Given this situation, it seems strange that

> (c)ompared to the other main focus of mathematics, number, there has been little research in this area. . . .Whether this lack of attention reflects problems with geometry, with geometry education, or with research in geometry education is not clear at present, but the fact remains that mathematics educators do not have an extensive or comprehensive corpus of research from which they can draw ideas in tackling the issues surrounding the teaching of geometry. (Bishop, 1983, p.176)

One area in which educators are beginning to direct their inquiries, as they examine the learning and teaching of geometry, is that of learning theory. Over the last 10 years, the work of two Dutch educators, Pierre M. van Hiele and his wife, Dina van Hiele-Geldof, has gained the attention of researchers in North America. The couples' work describes the nature of insight in geometry, describes five sequential levels learners pass through as geometric thought matures and presents a guide to the development of lessons. The levels are labelled "visualization", "analysis", "abstraction", "deduction" and "rigor", from first to fifth, respectively (Burger & Shaughnessy, 1986). The instructional guide consists of five phases of learning which, according to the van

Hieles, when followed, result in movement through one level into the next. The components of the model are interrelated: the thought levels provide a means for both assessing student abilities and for helping students develop insight into geometry through instruction (van Hiele-Geldof, 1984/1957). Appendix A provides a detailed description of the levels of thinking and of the phases of learning.

During the 1980's, studies have been conducted with the intention of validating, developing and applying the theories. The hierarchical nature of the levels has been researched (Mayberry, 1981). Characteristics of learners at each level have been sought (Fuys, Geddes & Tischler, 1985, Shaughnessy & Burger, 1985). The levels have been used as a predictor of student performance (Usiskin, 1982). Educational materials based on the phases of learning have been created (Bobango, 1987, Fuys et al., 1985). Analyses of the van Hiele levels required of the reader of geometry textbooks have been conducted (Crowley, 1984; Fuys et al., 1985; Severin, 1987). In general, each of the studies supports the descriptive power of the model.

Assessment of an individual's van Hiele level has been an integral part of much of the van Hiele-based research. As a result, techniques for identifying at which van Hiele level an individual is functioning have been produced (Burger & Shaughnessy, 1986; Fuys, et al., 1985; Kay, 1986; Mayberry, 1981; Usiskin, 1982). The instruments developed by Burger and Shaughnessy, Mayberry, and Usiskin, because they are not linked to a particular instructional unit, have been used in a range of research situations. (Assaf, 1985; Bobango, 1987; Burger & Shaughnessy, 1986; Denis, 1987; Mayberry, 1981; Scally, 1987; Severin, 1987; Usiskin, 1982).

As part of a three year study into the van Hiele model, Burger and Shaughnessey (1986) developed an interview script with an accompanying analysis form and administered it to 45 students. A subset of these interviews was studied in detail. The researchers concluded:

> (1) that for the tasks that their study presented (polygonal only), the model is useful for describing students' thinking processes,
>
> (2) that it is possible to identify student behaviors typical of each van Hiele level and,
>
> (3) that interview procedures can be developed which reveal predominant levels of reasoning on specific geometry tasks. (Burger & Shaughnessy, 1986, p.47)

It is noteworthy that these researchers did not include in their set of tasks, activities corresponding with the highest van Hiele level. This level is acknowledged as undercharacterized (Fuys et al., 1985; Usiskin, 1982) and as beyond the level most individuals attain (Hoffer, personal communication, February 25, 1985). These circumstances, combined with the fact that the highest level of formal geometry instruction most people receive (high school geometry) requires, at most, thinking from the fourth level, are legitimate reasons for focussing initial research on the first four levels.

As part of the Cognitive Development and Achievement in Secondary School Geometry (CDASSG) project at the University of Chicago, Professor Zalman Usiskin and his team of researchers developed the VAN HIELE GEOMETRY TEST. They wanted a test which could be administered to a large number of students in order to "determine, if such a determination would be possible, the van Hiele level of the students" (Usiskin, 1982, p. 18). The result is a 25 item multiple-choice test which can be administered in one 35 minute sitting. There are six ways in which to interpret the raw scores. Two of the interpretation schemes result in level designations which range from Level 1 to Level 5. The other four interpretation schemes result in level designations corresponding with the first four levels only.

Usiskin indicates "that there has been a lot of interest in the van Hiele test we designed. It has been used around the world" (Usiskin, personal correspondence, September 4, 1987). Several important concerns arise, however, when interpreting the test results. One question at issue is which of the six schemes for interpreting the raw scores provides the most accurate assessment of van Hiele levels. A second concern is that reliability statistics associated with the Chicago project subjects' responses are low. A third concern is whether or not a test which predominately uses quadrilaterals and triangles in the items can claim to measure an individual's van Hiele level for "geometry". There is uncertainty as to whether or not an individual's van Hiele level is constant for all topics in geometry or whether it varies topic by topic (Burger & Shaughnessey, 1986; Denis, 1987; Mayberry, 1981).

A third instrument, one which assess only the first four van Hiele levels, was developed by Joanne Mayberry. This instrument combines both a multiple-choice approach and an interview technique. Intended to be administered one-on-one, the interviewer presents multiple-choice questions, then probes subjects about their reasons for each choice. The 62 item test contains level specific questions for seven geometric concepts: squares, right triangles, isosceles triangles, circles, parallel lines, similarity and congruence. She found that she could assign levels to her preservice elementary school teacher subjects. Those subjects, however, were not consistent across topics in the level of their responses (Mayberry, 1981).

Statement of Purpose

The van Hiele model of geometric thought development is currently receiving attention from researchers interested in investigating the learning and teaching of geometry. Essential to much of that research is the assessment of an individual's level of thinking about geometry. Presently, three instruments which purport to assess levels of geometric thought development are being used. Two of these instruments, those by Burger and Shaughnessy and by Mayberry, rely on interview techniques. This type of assessment is particularly effective when attempting to determine and clarify characteristics of thought, and when working with individuals. It is not, however, an efficient strategy when assessing large numbers of subjects. The one-on-one testing/observing format and the verbal probing required with interviewing make it difficult, if not impossible, to gather data in a traditional single testing session. A further drawback of the interview technique is that "scoring" requires the interpretation of observed actions and interview responses. These assessments are prone to subjectivity, varying from rater to rater, or, indeed, even intra-rater. In contrast, the third instrument, the CDASSG VAN HIELE GEOMETRY TEST, because of its timed multiple-choice format, can be easily administered to large groups of people at a single session. The responses are standardized and easily scored. With this particular instrument, however, there are some uncertainties about which interpretation scheme is the most useful and about what the test measures. Its empirical properties have not been clearly demonstrated.

Upon review, then, none of the major instruments designed to assess van Hiele levels meet the criteria of being easily administered to large groups, standardized, valid and reliable. Furthermore, other existing geometry instruments, those not specifically designed to measure the levels of thinking as described by the van Hieles', are not appropriate for assessing reasoning abilities. Almost without exception, they tend to measure achievement.

With these considerations in mind, this study will undertake to develop an instrument for assessing van Hiele levels of geometric thought, which is easily administered to large groups, reliable, valid, easily scored and easily interpreted. Specifically, the goal is to produce a multiple-choice test, covering the topic of quadrilaterals, which can be used to identify masters and nonmasters of each of the first four van Hiele levels. The test will be called the van Hiele Quadrilateral Test. A master of a level consistently demonstrates an understanding of the processes associated with that level, and applies those processs. A master of a level is ready for instruction at the next level. A nonmaster of a level does not demonstrate an understanding of, or utilize the processes associated with the level.

The research questions are:

(1) Can multiple-choice items, which discriminate between masters and nonmasters of a van Hiele level, on the topic of quadrilaterals, be developed?

(2) Assuming items can be identified and assembled into the van Hiele Quadrilateral Test, what is the reliability associated with the mastery decisions from the instrument?

(3) What validity is associated with the mastery decisions which result from the van Hiele Quadrilateral Test?

(4) Can the van Hiele Quadrilateral Test be easily administered?

(5) Can the van Hiele Quadrilateral Test be easily scored?

(6) Can the van Hiele Quadrilateral Test be easily interpreted?

The instrument parameters of question type, geometric topic, and van Hiele levels to be assessed were decided at the outset of the research. The fixed response mode, one where students choose responses from a provided list, was used because:

1. It is easy to administer.

2. Responses are standardized, thus facilitating interpretation of results, comparisons between individuals, and comparisons in test/retest situations.

3. Verbally unskilled subjects are not penalized for their lack of oral skills.

In particular, the multiple-choice format was chosen because it offered the opportunity to provide "correct" answer choices at several levels. The feasibility of questions where subjects could choose between level specific responses was of research interest.

Quadrilaterals were chosen as the content base for the instrument because:

1. Quadrilaterals are a core topic in the study of Euclidean geometry and as such are taught in most curricula, starting with elementary school and progressing through to high school. The fact that this concept, in some form, is taught at so many grade levels widens the instrument's applicability. It could be used with students from a wide age range, a wide instructional range, and a wide grade range.

2. Pierre van Hiele has stated (Mayberry, 1981) and research supports (Burger & Shaughnessy, 1986; Denis, 1987; Mayberry, 1981) that individuals may be at different levels of thinking for different content areas within geometry. Consequently several content areas should not be used to determine a "general" van Hiele level. Rather, each content area should be assessed individually.

3. In order to be a manageable length for in-class administration, the instrument should focus on a single content area.

As with several other instruments, the fifth van Hiele level was not assessed. The reasons for this decision were:

1. This is the least developed level in the theoretical framework. The descriptors for the level are not detailed, therefore it is difficult to design questions which evoke thought at this level.

2. The descriptors which do exist describe thinking at this level as the ability to view geometry in the abstract. It is, in a sense, independent of specific Euclidean concepts. Thus quadrilaterals are not an appropriate subject matter for consideration at this level of thinking.

3. The geometry taught in the secondary schools requires thinking associated with the first four levels, not higher. Thus, research at the elementary and secondary levels will focus on those levels. This instrument could serve those researchers.

Significance

Two elements which will contribute towards improved "van Hiele" based research, instruction, and learning are (a) accurate assessment tools and (b) a clear understanding of the van Hiele model. With these, for example, the methods of instruction, the content selection, the sequencing of materials and the other activities which occur in both the classroom and in interventionist research, could be matched to student capabilities. This research, therefore, has the potential to be significant in several ways, to those interested in the van Hiele model, particulary to those interested in determining van Hiele levels which correspond to an individual or to a group of individuals. The first, and most important, is that an empirically sound instrument, which is easily administered, easily scored and easily interpreted, would be available. Second, the design of the instrument -- its question and answer format, its scoring scheme, and its interpretation scheme -- may serve as a model for van Hiele based instruments covering other content areas. Third, the data collected will provide level specific information about students from each of the groups in the sample.

Summary

This chapter included a discussion of the importance of geometry in the school curriculum, outlined the van Hiele model of the development of geometric thinking, and introduced the assessment problem the research was designed to address. The next chapter provides a fuller discussion of the research into the model, with an emphasis on the research which has developed or used an assessment instrument for the purpose of determining an individual's van Hiele level. Subsequent chapters detail both the organization of, and the findings from, the four main production stages for the instrument: (a) writing the items, (b) piloting the items, (c) field testing the items and (d) the final test administration. In the final chapter, conclusions and recommendations based on the findings are offered.

CHAPTER 2

REVIEW OF THE LITERATURE

Throughout the 1960's and 1970's, the central focus for much of the research into children's understanding of spatial and geometric concepts was the work of Jean Piaget (Carpenter, 1980). By 1980, however, a new characterization of the development of geometric thought had come to the attention of North America educators. Thomas Carpenter, writing at that time in a book devoted to research in mathematics education, predicted that the work of the van Hiele's, "pick(s) up where Piaget leaves off....(and) provides a beginning framework for research in (geometry)" (1980, p. 174). He noted, however, that the model was untested in North America and suggested that research into the transportability of the model be conducted.

This chapter presents a summary of the van Hiele-based research reported in the literature. Studies into the validity of the model are presented first. This is followed by a discussion of the research which has applied the model, with a particular emphasis on the assessment instruments which have been developed and utilized.

Research on the van Hiele Model

Although first published, in Dutch, in the late 1950's, it was not until the mid-1970's, that the van Hiele model began to be mentioned in English language writings. The first such reference appeared in the book Mathematics as an Educational Task, published in 1973, by the van Hieles' mentor, the eminant Dutch mathematician and educator, Hans Freudenthal. He discussed the van Hieles' notion of learning as being structured by levels and he presented an application of the model in the form of a summary of the teaching experiment on which Dina van Hiele-Geldof based her doctoral work.

The first reference to the work by a North American came from Izzak Wirszup in 1976. Ironically, while describing the current state of mathematics education in the Soviet Union, Wirszup provided details about the Dutch theory. The Russians had first learned of the model through an 1959 article by Pierre van Hiele, written in French. Shortly after the publication of the article, the Soviets conducted validation studies, and, based on their confirmation of the theories, revised their national geometry curriculum.

Soon, other English language educators and mathematicians began to discuss the implications of the model. Coxford (1978), frustrated that Piagetian theories only described how students respond to certain geometric tasks, rather than the teaching and learning processes, suggested that the van Hiele model might be a more appropriate means to that end. In 1980, Carpenter, outlined the model and suggested, that if it was valid, it would have important implications for the instruction of geometry.

In the early 1980's, three large-scale and long-term American projects investigating model related issues were conducted. The range of topics collectively addressed by these studies--the validation of the model, applications of the model to instruction and instructional design, assessment of materials, and assessment of individuals--is representative of the van Hiele-based research in the 1980's. Seminal in their importance, an overview of the goals and methodology for each of the three projects is presented here. The results from these studies, and other van Hiele based research, will be integrated in the topical discussion of the research findings which follows.

The Large-Scale van Hiele-Based Projects, An Overview

The Cognitive Development and Achievement in Secondary School Geometry Project (CDASSG)

A research team at the University of Chicago, members of the Cognitive Development and Achievement in Secondary School Geometry project led by Zalman Usiskin, were the first of the large research projects to report findings (Usiskin, 1982). Funded by the National Institute of Education, the primary function of that project was to test "the ability of the van Hiele theory to describe and predict the performance of students in secondary school geometry" (Usiskin, 1982, p. 8). Using batteries of test, several developed by the researchers, students' van Hiele level and their understanding of geometric concepts were measured at the beginning and at the end of a traditional tenth grade geometry course. Over 2699 first year geometry students, from a range of socio-economic backgrounds and from across the United States, participated in the study.

The Brooklyn College Project

The three year research project conducted by the Brooklyn College researchers, David Fuys, Dorothy Geddes (principal investigator) and Rosamond Tischler, is the most comprehensive study about and with the van Hiele model to date. Conducted with National Science Foundation support, the researchers set out to identify behaviors specific to each van Hiele level; to develop, implement, and assess instructional modules, for Levels 1, 2 and 3, based on the tenets of the model; to investigate teachers' abilities to understand and utilize the model; to analyse, from a "van Hiele" perspective, the geometry strands for three American mathematics textbook series, kindergarten to eighth grade; and, to translate four of the van Hieles' works into English. Included in the transcriptions (Fuys et al., 1984) are Dina van Hiele-Geldof's dissertation, describing the teaching experiment she conducted with first year secondary school students, and Pierre van Hiele's illuminating 1959 article "A Child's Thought and Geometry". Subsequent research involving the model has been greatly facilitated by the availability of these primary sources.

The Oregon State University Project

Also funded by the National Science Foundation, Professors William F. Burger and J. Michael Shaughnessy, from Oregon State University, conducted a study to investigate three research questions:

> 1. Are the van Hiele levels useful in describing students' thinking processes on geometry tasks?
>
> 2. Can the levels be characterized operationally in terms of student behaviors?
>
> 3. Can an interview procedure be developed to reveal predominant levels of reasoning on specific tasks? (Burger & Shaughnessy, 1987, p. 32)

The responses of 45 students to project designed experimental tasks dealing with triangles and quadrilaterals were collected. Fourteen of those interviews, selected randomly but stratified by age groups to insure representativeness over the educational range from primary school to college mathematics majors, were analysed in detail.

The results from these three large-scale, federally funded projects are presented, topically, throughout the rest of this chapter.

Validation of the Model

Studies investigating the validity of the model have focused on the existence and description of the levels, and on the accuracy of the properties associated with the levels (Burger & Shaughnessy, 1986; Denis, 1987; Fuys et al., 1985; Mayberry, 1981; Usiskin, 1982, Wirszup, 1976). Research into these areas is often interrelated for, minimally, evidence supporting the level characteristics, by inference, also support the existence of the levels. Appendix A contains a detailed description of the levels of thinking and of the properties associated with the levels.

Existence and Descriptions of the Levels

As recounted by Wirzsup, the Russians first learned of the van Hiele model through Pierre van Hiele's article "A Child's Thought and Geometry". Once introduced to the model, the Russians "hastened to organize intensive research and experimentation on the levels of development outlined by van Hiele, and between 1960 and 1964 they verified the validity of his assertions and principles" (Wirszup, 1975, p. 77).

These Russian validity findings have two associated and important implications for the applicability of the model. The first is that, by using subjects from educational levels equivalent to North American grades 1 to 12, the Russian research extended the range of individuals to whom the levels of development might apply.

The van Hieles' didactical experiment and observations had focused only on secondary school students, aged 12 and up. The Russians found that the model was useful in describing the thinking of younger children, as well. The second contribution of the Soviet studies is that the context in which the model functions was expanded. Working in a cultural and educational setting, different from the Dutch environment, the Russians still found the levels accurate descriptors of development.

Similar validation results were found in the United States. Twenty years after the Russian research was initiated, Burger and Shaughnessy, while studying the responses to geometric tasks made by students ranging from kindergarten through college, observed that "behavior on these tasks was consistent with the van Hieles' original general description of the levels" (Burger & Shaughnessy, 1986, p.31). Again, the validity of the model was supported, for a wide range of individuals and in yet another cultural setting. They also compiled a list of specific behaviors characterizing individuals operating at the first four levels. This provided additional information about the levels, for the van Hieles made only occassional references to specific overt behaviors associated with each level.

Further support for the validity of the levels was provided by the findings from the Chicago group's research. In their final report they state that "in the form given by the van Hieles, Level 5 either does not exist or is not testable. All other levels are testable" (Usiskin, p. 79). The utility of the levels for describing geometric thought and development is not, however, compromised by this reservation about Level 5. The geometry taught in elementary and secondary school requires, at most, Level 4 thought (Hoffer, 1981).

Some uncertainties have, however, arisen around the processes associated with levels other than Level 5. Bobango recounts two such instances, relating to Level 3, which emerged from research conducted in South Africa. In a project designed "to determine if categories of geometric questions formed Guttman Scales and if they corresponded to the van Hiele levels" (1987, p. 47), it was suggested that one-step deductions "are possible at van Hiele levels lower than 3 or 4" (1987, p. 48). In a second South African study, after determining students' van Hiele levels through interviews, the researcher found that (a) students who had been identified as operating at levels lower than Level 3 demonstrated hierarchical skills, a process characterized by the van Hieles as Level 3, and (b) that students below Level 3 could reason deductively. It was hypothesized that "hierarchical class inclusion may develop independently from deductive thinking, and that one is not a prerequisite for the other" (Bobango, 1987, p. 49). The van Hieles' identified these two traits--accepting (and applying) class inclusion and simple deductive thinking--as characteristics of Level 3 thought (Van Hiele-Geldof, 1957/84) but did not offer any observations about their interrelationship.

In a study investigating how young children come to understand geometry, Cynthia Kay (1986) questions the accuracy of the first three levels. Working with 16 grade 1 students, she conducted a 10 day teaching experiment which was composed of ten 45-minute lessons. By introducing the figures from general-to-specific, rather than the more traditional order of specific-to-general, by focusing instruction on the characteristics and relationships for figures and classes of figures, and by labeling figures with hierarchical-based names, she observed that

> the van Hiele theory may not capture the full complexity of how young children come to understand geometric concepts. Specifically, the van Hiele theory may describe the development of concepts within a hierarchy when instruction proceeds from specific-to-general but not when instruction proceeds from general-to-specific. (p. ii)

In summary, the existence and description of the levels of the van Hiele model have been addressed directly by several studies. The findings from three of these, the Russian project, the Oregon project, and the Chicago project support the existence and accuracy of the first four levels. The fifth level remains problematic. Two South African studies, however, question the breadth of thinking combined in Level 3 (Bobango, 1987). Furthermore, a study conducted with very young children, suggests that the levels reflect the organization of the content, rather than parallel any "natural" development of the subject (Kay, 1986). It appears, then, that within the traditional North American pattern of geometry instruction, the accuracy of the van Hiele levels as descriptors of ways to think about geometry is generally supported by research.

Properties of the Levels

The properties associated with the van Hiele levels of thought have also been studied. Those properties are that (a) the levels are hierarchical, (b) movement through the levels is sequential, (c) movement from level to level is discontinuous, (d) advancement through the levels is promoted by instruction, (e) no learning occurs when there is a mismatch between learner and the teaching environment, (f) what is intrinsic at one level becomes extrinsic at the next and (g) each level has its own linguistic context. Much of the research into the levels has focused on these traits. Evidence supporting the validity of these properties provides further support for the existence of the level. The following section will discuss findings relating to each property.

Hierarchical Levels. Support for a hierarchical relationship amongst the levels has been found in studies conducted by Burger and Shaughnessy (1986), Denis (1987), Fuys et al. (1985) and Mayberry (1981). Mayberry assumed that if the levels described by the van Hieles' existed and were hierarchical, "it should be possible to construct a series of tasks which the students functioning on a given level could perform, and students functioning on a lower level could not perform" (1981, p. 8). To test this theory, she developed a 62 item evaluation instrument, in interview form, covering the geometric topics of squares, right triangles, isosceles triangle, circles, parallelism, congruence and similarity. For each content area, there were questions corresponding to each of the first four levels. She observed that since the fifth level is probably only reached in advanced mathematics courses, it was most unlikely that her subjects had been exposed to instruction at that level. Including that level on her instrument, she felt, might result in "artificially inflated statistics" (Mayberry, 1981, p. 64).

The responses to the items by the 19 preservice elementary school teachers in Mayberry's study were collected and analysed using the Guttman Scalogram Analysis technique. Mayberry found that the patterns of her subjects' responses, across the levels tested, formed a scale. From this she concluded that the first four levels of thinking form a hierarchy (Mayberry, 1981, p. 99).

The hierarchical nature of the levels has also been supported in other studies. Denis, investigating the relationship between Piagetian stages of cognitive development and the van Hieles' levels of thought, used Mayberry's interview questions to classify 156 students. She, too, found evidence to support the hierarchical nature of the model (Denis, 1987). Using their own materials, two other groups of researchers, Burger and Shaughnessy (1987) and the

Brooklyn College group, also reported similar findings (Fuys et al., 1985).

Sequential Movement between Levels. Investigation into the validity of the fixed sequence property was part of the large study conducted by the Brooklyn College group. A major aspect of their research involved developing three instructional units based upon the principles of the phases of learning. Focussing on Levels 1, 2 and 3, these modules covered (a) basic geometric concepts (parallelism, angles, congruence,...) and properties of quadrilaterals, (b) angle measurement and, (c) areas of triangles and quadrilaterals. The units were administered in clinical interviews, on a one-to-one basis, to 16 sixth graders and 16 ninth graders. Each student's performance on the modules was video taped and, subsequently, analysed for the student's level of thought, difficulties, language, learning style, etc.

The subjects in this study were observed over a period of time, while engaged in learning activities. This offered the possibility to study, directly, students moving through a level, as well as to study the hierarchical nature of the levels. Geddes and her colleagues found evidence supporting the fixed sequencing of the levels. Repeatedly, they found students who performed at Level "n" were also consistently successful performing at levels lower than "n". For a specific topic, students did not appear to "skip" a level as their thinking developed (Fuys et al., 1985).

In conjunction with their findings about sequencing, the Brooklyn group also concluded that "the highest level of thinking attained by a student on one concept was also attained by the student on other concepts" (Fuys et al., 1985). This stability appears to be in contradiction to other research findings. Burger and Shaughnessy (1986), Denis (1987) and Mayberry (1981) reported that they found students operating at different levels of understanding for different topics.

Fuys and his colleagues, addressing this apparent difference in findings, point out that they designated students' levels at two different stages of the research. An "entry" level was assigned before instruction began; a "potential" level was assigned after the instruction was completed. Their findings of level unanimity across topics are based on the second assignments, the "potential" levels. As the other researchers did not include an instructional component, their level designations can be considered as equivalent to "entry" level. From this perspective, the Brooklyn College results concur with the other findings. They found that it was often necessary for students to "fill in" lower levels "for topics which they had not yet studied" (Fuys et al., 1985, p. 233), but that with this, students then easily reached a consistent "top" level of performance across topics.

In 1981, Mayberry questioned van Hiele about the consistency of levels across concepts. He acknowleged that students might be functioning at different levels for different concepts. He cautioned, therefore, about aiming instruction in a "new" unit at the highest level of thinking a student has demonstrated. For each geometric concept, it is necessary to be guided through the levels, in sequence. Van Hiele suggests, however, that once a level is reached for one concept, it becomes easier, and requires less time, to reach that level when dealing with other concepts (Mayberry, 1981).

Discontinuity. The van Hieles' hypothesized that the levels are discrete, i.e., that learning is composed of plateaus traversed by jumps. The strategies of one level are utilized over a period of time, then a qualitative leap is made to the next level, where entirely new strategies replace the old ones. The results of research into this property are, however, "mixed on this point" (Fuys et al., 1985).

The Brooklyn researchers found that many students appeared to move between levels in "small steps" (Fuys et al., 1985, p. 234). These students often demonstrated strategies from two levels, reverting to the lower level when confronted with a new situation. The researchers conjectured that this apparent "continuity" between levels may have been a result, however, of the processes of instruction used in their modules. The constant interaction of the instructor with the student and the talk aloud strategies meant that students made "incremental progress in learning and using new concepts, and in processes such as testing if properties apply to unfamiliar shapes or summarizing a deductive arguement. But, at the same time, a gap still exists in their ability to spontaneously initiate those processes" (Fuys, 1985, p. 233). This description of the ability to self-initiate processes associated with a new level parallels the discontinuity of progress claimed by the van Hieles'.

Other researchers have noted that some students oscillate between levels when working on the same task, as well as when working in different content areas. The Burger group conjectured that "students may move back and forth between levels quite a few times while they are in transition from one level to the next" (Burger & Shaughnessy, 1986, p. 45). This observation led them to speculate that the levels are "dynamic rather than static and of a more continuous nature than their discrete descriptions would lead one to believe" (Burger & Shaughnessy, 1986, p. 45). Lowry also observed students who used strategies from several levels on a single task. Working with instructional units on area and perimeter, she noted that most of her 18 third and fourth grade subjects "appeared to always be in transition from one level to the next" (Lowry, 1987, p. 75). In the midst of consistently demonstrating thinking from one level, students would frequently "make an intuitive leap that would indicate movement to the next level. But upon probing, it was determined that the connection would be isolated; the child was not able to use the idea for further progress without instruction at that level" (Lowry, 1987, pp. 75 - 76).

Advancement. The paramount importance of instruction to advancement through the levels has also been supported by research. The Brooklyn College group, for example, strongly support the van Hieles' "contention that a student's level of thinking and progress through the levels are more dependent upon instructional experiences than on age or maturational factors and that instruction can foster (or impede) such progress" (Fuys, 1985, p. 238). Wirszup attributes similar results to the Russian validation studies. They found that "the development which leads to a higher geometric level proceeds basically under the influence of learning and therefore depends on the content and methods of instruction" (Wirszup, 1973, p. 79). Bobango observed that instruction based on the phases of learning had a "positive effect on raising students' van Hiele levels of thought" (1987, p. 168). Similarly, Lowry (1987) comments that her teaching protocols, based on the phases, promoted her subjects understanding and encouraged movement to the next level higher. Even the qualifications offered by Kay (1986), that the levels as described by the van Hieles may be dependent upon "specific-to-general" instruction, rather than arise from an inherent ordering of the content, support the importance of instruction to mastery of the thinking described by each level.

Mismatch. The van Hieles' claim that when instruction is offered at a level above that of the student, the student will not understand or master the content. While several projects have developed instructional units which take this property into consideration, the research findings on this property are, at best, indirect. For example, correlations in the Chicago study between the achievement results of the grade 10 geometry students and their level assignments, indicated that students are unlikely to succeed in a geometry course delivered at a level higher than the level on which the student is operating (Usiskin, 1983).

Mayberry approached the issue somewhat differently. Unlike Usiskin, she did not have achievement results from the geometry course to compare to her van Hiele level assignments. She noted, however, that 70% of the preservice elementary teachers in her project who had taken high school geometry, were classified as operating at a level below Level 4 (Mayberry, 1981). Assuming that the geometry courses taken by these students had required Level 4 thinking, Mayberry's observations support the mismatch property. Minimally, exposure to the course had not resulted in the acquisition, retention and demonstration of Level 4 thinking for those students.

Intrinsic/extrinsic. The van Hieles contend that the structures which underlie one level of thought become the objects of study at the next level. Only one study, that by the Brooklyn College researchers, has addressed this issue directly. They indicated that their findings supported this property, but caution that this might have occurred because the instructional modules were designed to incorporate this implicit-explict feature (Fuys et al., 1986). Nonetheless, it is noteworthy that the project was able to develop materials consistent with this characteristic.

Language. The van Hieles' proposed that each level has its own linguistic character. Subsequent research findings have supported the validity of this property and, "underscore the importance of language in doing geometry" (Fuys et al., 1985, p. 234). The Brooklyn group observed, however, that for many of the the Level 1 to 3 students participating in their research,

> the lack of familiarity with standard geometry language was striking, and this prevented many from progressing within a level or to a higher level. Many students had poor expressive language. Some were unable to communicate effectively about geometric aspects of shapes. For example, some needed to point to a shape when talking about a specific part or property. Others need considerable review of terms. (Fuys et al., 1985, pp. 234- 235)

They also found that students frequently had difficulty with the use of logical language such as "all", "some", "if-then", or "because".

The Brooklyn group noted that "for each level there might also be a language associated with the quality of thinking at that level" (Fuys et al., 1985, p. 235). Students, when working through the modules with an interviewer, used language which reflected the quality of thinking specific to their operational level, e.g. at

Level 2, "Oh, I see a pattern" or at Level 3, "I should prove this, right?"

Researchers also indicate that confusion often arises from the lack of precision in the use of language, particularly from the lack of consistency between colloquial language and mathematical language. Geddes and her co-researchers, for example, cite examples such as students using "straight line" to mean "parallel lines" and "space" to mean "area" (Fuys et al., 1985, p. 181). One of the South African studies reports of confusion arising from students interpreting the question "Is a square a rectangle" to mean "Are the two figures the same?" When, however, the question was reworded so that students were asked if a square is a special type of rectangle, this "helped students see that the question was asking about subsets and not equivalences" (Bobango, 1987, p. 49).

Summary of research into the properties. In general, research into the seven properties associated with the levels of thinking supports their validity. The prevalence of students who appear to use strategies from two adjacent levels, students sometimes labelled as "in transition" gives rise to some doubt, however, about the discontinuous nature of the movement between levels.

Application of the Model

Researchers have been interested in applying the van Hiele model to educational settings, as well as in conducting validation studies. For example, van Hiele based materials have been developed (Bobango, 1987; Fuys et al., 1985; Lowry, 1987), the utility of the levels as a predictor of student performance has been investigated (Usiskin, 1982) and assessment of materials, in terms of the van Hiele levels required by the user, have been conducted (Crowley, 1984; Fuys, et al., 1985; Lowry, 1987; Severin, 1986). Assessment of students' van Hiele levels has also been an integral element of much of this research.

Two styles of assessment for individuals have been used, interviews and written tests. Two studies, those by the Brooklyn researchers and by Kay, developed interview type assessment strategies particular to an instructional unit. Three other projects developed assessment instruments, independently from instructional units. These are the interview activites on quadrilateral and triangles designed by Burger and Shaughnessy, the multiple-choice test on geometry designed by the University of Chicago research team and the combination multiple-choice/ interview geometry instrument developed by Mayberry. Each of these assessment techniques is discussed in this section. The findings from studies which have used these assessment techniques is also presented.

Research where an Assessment Instrument was Produced

Brooklyn College Assessment Procedures

Assessment of students thinking about geometry was an integral aspect of the research conducted by the Brooklyn College team. Subjects worked with a trained interviewer on the three phase-based modules. Each unit contained assessment tasks keyed to specific level descriptors, ranging from Level 1 to Level 3. The students attended six to eight 45 minute sessions. Not all students completed all modules.

Each meeting was filmed on video tape. Using protocol forms developed by the project, these tapes were viewed by someone trained in the van Hiele model. Each student's level of thinking was determined and summmaries were written. Each analysis (and sometimes the video) was then further reviewed and validated by at least one other project member.

The nature of the teaching experiment allowed the researchers to identify a student's level of thinking at different times. Rather than think of these level assignments, however, in the traditional pre-intervention and post-intervention context, the researchers identified these levels as an "entry level" and a "potential level". The entry level was determined by student responses to questions at the beginning of each module. These questions allowed for responses at different levels. Little or no interviewer prompting occurred. The researchers felt, however, that such "static assessments" might not reflect a student's ability to think in geometry, particularly if the student had undergone little or no learning experiences with the topic involved. Consequently, responses were assessed as the student moved through the phase-based instruction, interacting with the interviewer, and a "potential level" determined (Fuys et al., 1985).

The students in the project were drawn from the sixth grade and the ninth grade. Of the 16 sixth grade students, at the end of the instruction, eight were designated as entering at Level 1, three made no progress, while five made progress into Level 2. The remaining eight entered at Level 2 and demonstrated "varying stages of transition" (Fuys et al., 1985, p. 112) towards Level 3. Of the 16 ninth grade students studied, two entered at Level 1 and remained there; seven entered at Level 1 and showed significant movement towards acquiring Level 2 thinking; the remaining seven entered at Level 2 and were demonstrating many of the Level 3 characteristics at the end of the instructional sequence.

The Brooklyn College project also provided training about the levels for teachers. After receiving this instruction, the teachers could identify, from observing the video taped sessions, students' van Hiele levels and could identify the van Hiele level required by the text materials (Fuys et al., 1985).

The assessment techniques used by the Brooklyn group, rich though the findings were, may not suit other research settings. For example, to use their materials, expertise in both interviewing and interpreting student activity, within the van Hiele framework, is required. Even with training, the Geddes team noted that their interviewers were often overdirective, were not responsive to student initiative and occasionally did not probe student responses carefully enough. Consequently, valuable interview information was not obtained (Fuys et al., 1985). As well, the responses used for assessment, linked as they are to the three instructional modules, take time to collect and time to evaluated.

Kay Interview

Kay (1986), working with 16 first grade students, developed a four part, structured interview which took into consideration the students' mathematical experience and the instruction they received from the researcher. The pre-instruction interview established whether or not the subjects where familiar with the number concepts of three and four. Kay felt that this was prerequisite knowledge for the understanding of the concepts of triangles and quadrilaterals. Next a student's abilities to name a given quadrilateral and to identify its characteristics were assessed, using manipulatives. Third, working with models of quadrilaterals, one at a time, the students' understanding of the characteristics of specific classes of quadrilaterals and the hierarchical relationship among classes of quadrilaterals was tested. Finally, working simultaneously with a group of seven shapes, some of which were not quadrilaterals, the students' understanding of the characteristics of specific classes of quadrilaterals and the hierarchical relationship among classes of quadrilaterals was again probed. For this initial interview, standard vocabulary was used by the researcher.

Over a 10 day period, Kay delivered an instructional unit on quadrilaterals, which she had developed, to the subjects. That instruction focussed on the use of questions, sequencing the presentation of the content from general-to-specific, using names for figures which reflected their hierarchical connections, the use of wire manipulative, repetition and review.

The instruction was immediately followed by a post-instruction administration of the interview. This time, however, part one was omitted, and for the remaining three parts, terminology developed during the 10 day instructional unit -- quadrilateral, rectangle-quadrilateral and square-rectangle-- was used by the interviewer. Based on these findings, Kay suggested that the van Hiele model is instruction driven, not a development which is inherent with the topic.

Burger and Shaughnessy Interviews

One of the goals of the Oregon project was that of developing interview procedures which would, for Levels 1 to 4, "reveal predominant levels of reasoning on specific geometry tasks" (Burger & Shaughnessy, 1986, p.32). The procedures consist of experimental activities, an interview script, and an analysis protocol for each of two content areas, triangles and quadrilaterals. There are three triangle activities, (a) drawing triangles, (b) identifying and defining triangles and (3) sorting cutouts of triangles. There are five quadrilateral activities (a) drawing quadrilaterals, (b) identifying and defining quadrilaterals (c) sorting cutouts of quadrilaterals, (4) what's my shape (using a set of verbal clues to identify a figure) and (5) working with equivalent definitions of "parallelogram". The drawing, identifying and sorting tasks were designed to elicit responses corresponding to thinking at Levels 1 to 3; the what's my shape activity and the equivalence activity were designed to gather information about thinking on Levels 3 and 4. The interview packages took over a year to develop, involving three pilot interviewing phases and three subsequent revisions.

Designed for easy administration by either teachers or researchers, the interviews can be used with subjects of all ages. Analysis of student responses is guided by the project developed analysis protocols. These culminate in a profile, in vector form, of the predominant level each student displayed on the eight

Burger and Shaughnessy Interviews

One of the goals of the Oregon project was that of developing interview procedures which would, for Levels 1 to 4, "reveal predominant levels of reasoning on specific geometry tasks" (Burger & Shaughnessy, 1986, p.32). The procedures consist of experimental activities, an interview script, and an analysis protocol for each of two content areas, triangles and quadrilaterals. There are three triangle activities, (a) drawing triangles, (b) identifying and defining triangles and (3) sorting cutouts of triangles. There are five quadrilateral activities (a) drawing quadrilaterals. (b) identifying and defining quadrilaterals (c) sorting cutouts of quadrilaterals, (4) what's my shape (using a set of verbal clues to identify a figure) and (5) working with equivalent definitions of "parallelogram". The drawing, identifying and sorting tasks were designed to elicit responses corresponding to thinking at Levels 1 to 3; the what's my shape activity and the equivalence activity were designed to gather information about thinking on Levels 3 and 4. The interview packages took over a year to develop, involving three pilot interviewing phases and three subsequent revisions.

Designed for easy administration by either teachers or researchers, the interviews can be used with subjects of all ages. Analysis of student responses is guided by the project developed analysis protocols. These culminate in a profile, in vector form, of the predominant level each student displayed on the eight activities. From this, judgement can be made on the predominant overall level of reasoning displayed by the student.

With no set time limit, the interviews tend to require between 40 to 90 minutes. Like the other interview instrument, the analysis of responses must be completed by someone familiar with the model.

The Mayberry Assessment

One instrument has been designed which combines interviewing with written responses. As part of her doctoral work with pre-service elementary school teachers, Mayberry (1981) designed a 62 item test containing level specific questions for seven geometric concepts: squares, right triangles, isosceles triangles, circles, parallel lines, similarity and congruence. The instrument is designed to be administered in a one-on-one situation, a van Hiele trained interviewer with a subject. The interviewee responds in writing to multiple-choice questions, then is probed by the researcher about the reasons for each choice.

As one of the early researchers into the van Hiele model, Mayberry found it necessary to commission translations of the van Hieles' works into English. Working from the descriptions of thinking contained in those original sources, she produced descriptions, in behavioral terms, corresponding to each level of thought. Questions were then written to correspond to the level specific behaviors. She comments that only a few questions were developed for the fifth level, as it is topic-free. Indeed, in the final instrument, she only tests for the first four levels. The level descriptors in behavioral terms and the questions were sent to 13 mathematicians and mathematics educators, including Pierre van Hiele, for review. She asked them to respond to the following requests: "1) Is this question suitable (yes, no), 2) Does this questions appear to test the given van Hiele level? 3) Does any aspect of the question seem to test a higher level? 4) What comments or suggestions can you give to help with evaluation, clarity, reformulation?" (1981, p.52). Based on their responses, Mayberry revised her item bank, then selected 62 items for the final interview. The distribution of the final questions by content area and level is given in Table 2.1. Some items cover more than one content area.

The criteria Mayberry used for "success" at a level ranged from answering 50% to 100% of the questions, depending on how many items there were per level. A subject's performance at each level was recorded in a 5 element matrix, one for each level, where a 1 indicated successfully meeting the criteria for that level and a 0 indicated lack of success. (Level 5, however, was not tested.) The operating level of the subject was then designated as the highest level for which the criteria were met <u>and</u> for which the criteria on every lower level had also been met.

Table 2.1

Number of Items, Mayberry Item Bank, by Content and Level

Content	Level 1	2	3	4	Total
Square	2	2	7	2	13
Right triangle	2	1	4	3	10
Isoscele triangle	2	1	7	3	13
Circle	2	2	4	1	9
Parallel lines	2	1	4	1	8
Similarity	2	1	4	1	8
Congruence	2	1	3	1	7
Total	14	9	33	12	68

Mayberry found that, using the results from her instrument, she could assign levels to her subjects, although the subjects' level designations were not always consistent across topics. She recommended that a similar study be undertaken, where fewer topics with more questions per level be tested. She notes, without saying why, that a multiple-choice test would be very difficult to develop and analyze. She goes on to say "the type of test which requires the student to give weights to each choice according to his confidence in that choice might bear investigation" (Mayberry, 1981, p. 101).

The University of Chicago Multiple-Choice Instrument

Professor Zalman Usiskin and his team of researchers at the Cognitive Development and Achievement in Secondary School Geometry (CDASSG) project at the University of Chicago developed, as part of their study about the relationship between the van Hiele theory and the performance of students in secondary schoool geometry, the VAN HIELE GEOMETRY TEST. It is a 25 item multiple-choice test designed to be administered in one 35 minute sitting. No expertise in the van Hiele model is required to administer the test; no expertise is required to score the test.

For purposes of interpretation, the test is considered as having 5 subsections, each containing five questions. Each set of five questions require a unique minimal van Hiele thought level in order to be answered correctly. The sets of questions are arranged in the same sequential order as are the levels in the model (e.g. the first set of questions, questions 1 - 5, were designed to elicit responses at Level 1; the second set of questions, questions 6 - 10, were chosen to elicit responses at Level 2, etc.) There are six different ways in which to interpret the raw scores. The differences hinge on whether or not the results from the highest (fifth) level are considered when assigning classifications, whether 60% or 80% mastery is required in order to demonstrate ability at a level, and whether or not to be designated as operating at Level "n", every previous level, e.g. 1,2...(n-1), must also be mastered. To be consistent with the model, the CDASSG group suggest that the last criteria should be required, i.e., mastery at <u>every</u> previous level must be demonstrated. Following that suggestion, the research report discusses results from four of the scoring schemes. These are "classical strong" (all 5 levels are considered, 80% mastery required), "classical weak (all 5 levels are considered, 60% mastery required), "modified strong" (4 levels are considered, 80% mastery required), "modified weak" (4 levels are considered, 60% mastery required).

The items developed for the instrument were based on quotes found in the van Hieles' writings (Usiskin, 1982). The items, covering a range of geometric concepts -- triangles, quadrilaterals, parallelism, circles -- were tested with students in an interview situation. Based on those student responses, a 25 item test was assembled. This test was then administered to entire classes to ensure that it could be completed in 35 minutes. The final test is "essentially the same as that piloted with the entire classes" (Usiskin, 1982, p.19).

Reliability statistics were calculated twice by the Usiskin group using the norm-referenced Kuder-Richardson Formula 20, once in the fall of the academic year and once in the spring of that same academic year. For each of the five subsections, the reliablity coefficients are low-- 0.31, 0.44, 0.49, 0.13, 0.10, respectively for Levels 1 to 5-- on their fall administration, with slightly higher figures for the spring testing (Usiskin, 1982). The research group indicates concern over these statistics and suggests these figures may stem from the small number of items in each subtest. (In an analysis of the reliablity of the VAN HIELE GEOMETRY TEST, Crowley (in press) observes that criterion-referenced reliability techniques are more appropriate to use with the instrument than are norm-referenced techniques.)

Usiskin's group administered the test, once in the fall and once in the spring, to over 2000 students enrolled in a one year geometry course. The research findings indicated that the levels assigned to the students, even though those levels often varied according to the scoring criteria used, were "a good descriptor of concurrent student performance in geometry and a reasonably good descriptor of later performance" (Usiskin, 1982, p. 89). In particular, students designated at the lower van Hiele levels did not do well when tested on geometry content or proof writing (Usiskin, 1982).

Research Incorporating Student Assessment into the Design

Each of the research projects cited above had the production of an assessment technique as a major goal. As these tools became available, other research projects involving the model began to use them. The Burger and Shaughnessy interviews were used by Bobango (1987) in a study using phase-based curriculum. Scalley (1987), in a project involving angles, designed interview tasks for that topic based on the Burger and Shaughnessy format and analysis techniques. Lowry made an "age-appropriate adaptation" (1987, p. 33) of the area and perimeter materials from the Brooklyn College project. Assaf (1985) and Bobango (1987) each used the CSASSG multiple choice instrument. The Mayberry interview was used by Denis (1987) in her investigation of van Hiele levels and Piagetian stages.

As part of an investigation into whether or not van Hiele phase-based instruction could provide "a geometric foundation for students before they were asked to construct proofs" (Bobango, 1987, p. 52), Janet Bobango designed a van Hiele based unit on quadrilaterals and triangles. A component of the instruction involved the students exploring figures using the quadrilateral and triangle software in the Geometric Supposer series. Bobango reported that the month long phase-based instruction had a positive effect on raising the tenth grade students' van Hiele levels but that the instruction, perhaps because of its short duration, "did not lead to significantly greater achievement in the standard content and in proof-writing success" (1987, p. 177).

Bobango based her observations about the students' van Hiele levels, on performances obtained from two assessment techniques. The first was a comparison of the pre-test and a post-test performances of the 40 subjects in her control group and the 32 subjects in her experimental group on the VAN HIELE GEOMETRY TEST. She also conducted interviews using the Burger and Shaughnessy interviews. Before instruction began, sixteen students, for whom van Hiele levels had been determined by the multiple-choice test, were administered the interviews on triangles. The researcher and two trained evaluators assessed van Hiele levels from these interviews. Although there were differences in opinions, the correlation values for the van Hiele levels as determined by the evaluators of the interviews and as determined by the Chicago test were 0.62. At the end of the instruction period, another sixteen students were administered the Burger interviews on quadrilaterals. Again, van Hiele levels were determined by the researcher and two other evaluators. When these level assignments were compared with the students' scores on the post administration of the Chicago test, the correlation coefficient was 0.84 (Bobango, 1987). As a result of her study, Bobango suggests that "a refinement of the measure for assessing student's van Hiele levels of geometric thought is needed" (1987, p. 182).

Two other van Hiele based research projects also involved the computer. In both of those, students were instructed in the use of the LOGO language's turtle graphics. As part of a project, whose overall purpose was to investigate the effects of a Logo environment on ninth grade subjects' understanding of geometric relationships, Susan Paalz Scally developed interview items on the topic of angles. These items were very closely modelled after the quadrilateral and triangle activities produced by Burger and Shaughnessy (Scally, 1987). The pre-instruction and post-instruction interview responses of 20 ninth grade subjects were analysed. The instructional unit was a 16 week course in Turtle geometry.

Scally identified two types of movement between the two interview situations, "gain" and "moderate gain". "Gain" was noted when a subject progressed from one level to the next level. or within levels when the student was able to provide additional information within several given tasks, demonstrate the use of new strategies or demonstrate a facility with level vocabulary from a new level. "Moderate gain" was noted primarily when a subject "engaged a task, perhaps with limited success, that s/he was unable to engage on the first interview, or when s/he employed a previously used strategy more successfully on the post-interview" (Scally, 1987, p. 2). Based on a qualitative analysis of the student's progress, using these two movement descriptors, she reports that a Logo learning environment "very well may" (Scally, 1987, p. 7) enhance students' understanding of geometric relations.

Working with nine third grade students and nine fourth grade students, all nine years of age, Joyce Lowry (1987) investigated whether the van Hiele model could be used (a) to assess a subject's concepts of area and perimeter and (b) to inform instruction which would promote the acquisition and application of those concepts. To achieve this, she ~~made~~ adapted materials from the Brooklyn College Project. Her unit consists of 8 activities. The first two activities assess the subject's initial operating level on area and perimeter. The next 5 activites present phase-based instruction on the area and perimeter of rectangles, right triangles, parallelograms, triangles and trapezoids. These activities combine instruction and assessment. The final activity, one on linear measure, was included to test if there was a relationship between a subject's understanding of linear measurement concepts and their van Hiele level for area and perimeter. With one exception, all activites were attempted by all students. The exception was the activity on the trapezoid. Only students who demonstrated an understanding of the area of parallelograms were given this unit. She used one-on-one clinical interviews, running approximately 40 minutes a session, over the course of several weeks. Each session was video-taped. Each session was reviewed by the researcher and two other individuals familiar with the model.

Lowry found that the "van Hiele model can indeed provide a useful structure in planning assessment activities for area and perimeter" (1987, p. 75) and that the teaching protocol she used "was successful in expanding these children's understanding and encouraged movement to higher levels of thought" (1987, p. 75). She also noted that there were differences between the initial levels of the third and fourth graders thinking on area and perimeter and conjectures that these differences were due to previous instruction. Support for this hypothesis is given by the fact that seven of the nine fourth graders tried to apply, from memory, a "rule" for area and perimeter. This is an example of van Hiele's reduction of level. No third grader appeared to have the rote formula tool. Once the subjects commenced instruction. however, little difference was observed in their final progress. Most of the subjects in each group "demonstrated readiness for instruction that would lead them to the next higher level of thought" (Lowry, 1987, p. 92).

In addition to the above findings, Lowry also observed that the classroom teachers of her subjects tended to present area and perimeter material only at Level 1, that the textbooks used in these classes were predominately at Level 1 and that all the children had a good working knowledge of linear measure, thus the correlation between this concept and any difference in the progress with area and perimeter concepts could not be determined.

Assaf designed and conducted a study which investigated "the effects of using Logo turtle graphics on the way students respond to questions at different van Hiele levels" (1985, p. 19). For one month, 22 students in an experimental group used researcher produced Logo activities, designed to introduce concepts from the eigth grade geometry curriculum. A control group of 26 subjects followed the normal curriculum. Using pre-test and post-test results, obtained from administering the University of Chicago's VAN HIELE GEOMETRY TEST to both groups, Assaf observed that the students who used Logo "were able to anwser questions at a relatively higher levels [sic] than those" (p. 159) who did not use Logo. To further explore the nature of the changes, he selected 9 items from that instrument and, using those, interviewed 16 subjects, asking them to think aloud as they answered each item. He found that students using Logo showed a tendency to respond at a relatively high van Hiele level, that they became less dependent on the irrelevant features of geometric shapes, that they were able to extract properties for geometric shapes and see relations between shapes more readily using Logo.

In a dissertation study conducted in 1986, Livia Denis investigated the relationships between the van Hiele levels of thought and the Piagetian stages of cognitive development. Puerto Rican adolescents, age 15 to 19, all of whom had completed a high school Euclidean geometry course, were administered two tests. The first, designed to assess an individual's Piagetian stage of operation, was the Test of Logical Thinking. Based on their performances on that test, two groups of students were identified, those designated as functioning at the concrete operational stage and those operating at the formal operations stage. Twenty students from each group were then administered the circle, congruence, right triangle and square questions from the Mayberry interview.

Denis states that her findings "clearly indicate that the Piagetian stages were found to be a possible predictor of the potentiality for geometric development of subjects in van Hiele terms" (1987, p. 91). In particular it was observed that there is a greater probability that students who are functioning at Piaget's formal operational stage, as opposed to those at the concrete-operational stage, will reach the higher van Hiele levels.

Van Hiele-Based Evaluation of Materials

Teaching materials have also been evaluated from a van Hiele model perspective. The Brooklyn College group, for example, examined three American textbook series from kindergarten to eight grade. They found a pattern where what little Level 3 thinking was required began in grade 8, where Level 2 thinking started to become necessary from grade 3 on but where in general, "average students do not need to think above level (1) for almost all of their geometry experience through grade 8" (1985, p. 221) in order to complete the geometry based exercises and test questions. In examining the books for didactic consistency with the level, they found many questions which required only memory (reduction of level); emphasis on application of formulas, not understanding; little emphasis on interrelations between concepts; and a lack of emphasis on underlying structures. In summary, the level required for successful performance was low; reduction of level was common, and the phases of learning were not reflected.

Lowry (1987), while examining only two texts, one for third grade and one for fourth grade, and from different publishers, found similar results. The predominant level required to deal with the material was Level 1. When "Level 2 thinking could be encouraged, the correct answer could be obtained with Level 1 thinking" (Lowry, 1987, pp. 71-72). As well, reduction of level in the form of encouraging formula memorization was in evidence.

In an analysis of the exposition and exercises in the geometry strand of two Canadian textbook series over two grade levels, 9 and 10, Crowley (1984) found that when van Hiele levels of thinking are required to understand text and/or answer questions, the modal frequencies followed the sequencing of the levels of thinking. The majority of geometry work at the grade 9 level required Level 2 thinking while the Grade 10 work required primarily Level 3 work, with a minimum of Level 4 work. While there was no accompanying information on the operating level of the students in the courses using these books, the emphasis on Level 3 work at tenth grade may reflect a shift in emphasis from the American paradigm of a Level 4 geometry course in Grade 10. Crowley observed, however, that there was no evidence that the text materials were used or that they promoted the acquisition of thought. Furthermore, many exercises required no level of geometric thought to correctly answer or, similar to the findings of the Geddes group, accepted as correct. answers which could merely be memorized.

In another Canadian study (Severin, 1987), one which analysed the Grade 9 geometry curriculum in Ontario, four textbooks, provincal and school board curriculum guides and 320 students were assessed for operational van Hiele levels. The students were tested using the items associated with the first four levels from the CDASSG VAN HIELE GEOMETRY TEST. Three academic strands were considered: Basic, General and Advanced, where Basic is the least demanding academically, where General is the norm, and where Advanced is for accelerated students. The study found that the textbooks required higher thinking skills than the intended curriculum in two of the three cases, the Basic and the Advanced, while matching in the General case. The modal van Hiele level of thinking of students, however, in each setting was lower than the texts in each case. According to the theory, mismatches such as these will cause learning difficulties.

Although not subjected to the rigorous testing of the research projects mentioned above, a van Hiele based high school geometry text has also been published: Geometry, A Model of the Universe by

Alan Hoffer. It corresponds in spirit and format with the model. The organization the three of the four major sections of the text parallels the sequencing of the levels. Hoffer starts with an emphasis on visual characteristics, then begins to emphasize analysis and ordering. Each section also includes laboratory activities for the student. It is not until a point approximately half-way through the book that the concept of a deductive system is introduced. (The last section, provides alternative ways to view geometric concepts: vectors, transformation and coordinate geometry and is highly numerical in its approach.)

Summary

Over the last decade, English speaking educators have begun to explore the potential of the van Hiele model of the development of geometric thinking for providing assistance in the development of educational activites, and, concommitantly, for assessing student potential and progress. Studies into the levels of thinking and their properties, in general, support the model's validity. Research into the relationship of the levels to student success in geometry suggest that there is a correspondence. Van Hiele based assessments of students have been an important part of much of that research, and as such, a range of assessment instruments have emerged.

CHAPTER 3

METHODOLOGY

The goal of this study was to create a 40 minute multiple-choice instrument which will assess an individual's dominant level of thinking, as described by the van Hiele model of the development of geometric thinking, on the topic of quadrilaterals. The individual is said to be a "master" of the dominant level, and a "nonmaster" of the higher levels. The sequential nature of the levels implies that masters of a given level have also, in the past, been masters of each of the lower van Hiele levels. The instrument is called the van Hiele Quadrilateral Test.

A discussion of the methodology associated with the development of the van Hiele Quadrilateral Test is presented in this chapter. The first section focuses on the procedures used to develop the instrument: writing the items, validating the items, constructing the test, administering the test, and assessing the reliability and validity of the test results. The discussion is organized around the four research phases: developing the items, the pilot study, the field testing and the final testing. The second section discusses the selection processes and the subjects selected for the project. The chapter concludes with a description of the measures, other than the van Hiele Quadrilateral Test, which were used as part of the research.

Procedures

The development of the van Hiele Quadrilateral Test proceeded through four sequential stages: developing the items, a pilot study, field testing and a final testing. In this section, each of those phases is discussed.

Developing the Items

As an instrument designed to describe an "examinee's behavior repertoire, rather than an examineee's ability relative to other examinees" (Nitko, 1984, p. 9), the van Hiele Quadrilateral Test is said to be a criterion-referenced instrument. It was necessary, therefore to identify in detail the criteria against which each subject's performance was to be measured. For this instrument, those criteria are the level specific behaviors associated with each van Hiele level.

An inventory of the level behaviors was compiled from the van Hiele based literature (Burger & Shaughnessy, 1986; Fuys et al., 1985; Hoffer, 1981; Usiskin, 1982; van Hiele-Geldof, 1984). These behaviors are called the Level Indicators and are listed in Appendix B. Question and answer combinations were then written to correspond with the indicators. As well, to assure that the content area for which the van Hiele levels were being identified was well represented, a list of quadrilaterals, their properties, and the traditional quadrilateral theorems encountered in the study of Euclidean geometry was also assembled (see Appendix C). These mathematical concepts were the basis of the geometry content contained in the items.

Both types of guidelines were used to ensure that the set of items constructed for the initial item pool was representative across levels, within levels and across geometric topic. At this stage in the instrument development, the goal was to have at least one item for each indicator and a relatively equal balance amongst the shapes referred to in the items.

Initial item pool. The initial item pool consisted of 53 multiple-choice questions, each with five answer choices (see Appendix B). For review purposes, the answers to each question were keyed to indicate which level descriptor their choice might reflect. In order to gain maximum information from each item, some questions were constructed so that more than one answer choice corresponded to a specified, distinct level. For example, in the original item 14, presented below, both options C and D were intended as "correct" answers, each corresponding to different levels of thinking.

14. Which combination of statements is the shortest list needed to guarantee that a four sided closed figure is a rectangle.

Statement 1: two long sides, two short sides.
Statement 2: opposite sides the same length.
Statement 3: opposite sides parallel.
Statement 4: one angle is a right angle.
Statement 5: all 4 angles are right angles.

(A) 1
(B) 2, 3
(C) 3, 4 (2.14)
(D) 1, 2, 3, 5 (1.11)
(E) None of these combinations describe a rectangle.

By level, 10 items in the pool corresponded to the first level, 16 items corresponded to the second level, 20 items corresponded to the third level, and 12 items corresponded to the fourth level. Of the 53 items, 5 questions had answer choices corresponding to more than one level.

Panel of experts. To assess the validity of the items, the questions, with their answers keyed to specific level indicators, and the level indicators were sent to five experts on the van Hiele model. (Although all had agreed to review the materials, one, in fact, did not respond.) The respondents were Dr. Janet Bobango (University of Cincinnati), Dr. Michael Shaughnessy (University of Oregon), Dr. Rosalind Tischler (Brooklyn College) and Dr. Pierre M. van Hiele (Voorburg, The Netherlands). The panel was asked to review the level indicators for their breadth and accuracy, and to comment on the appropriateness of the questions and answers for eliciting the indicated level-specific responses. Appendix B contains a complete copy of the information mailed to these experts.

The panel's comments on the level indicators and on the potential of the question and answer combinations to reflect level specific thinking were evaluated. The list of indicators was revised (see Appendix D). The items were revised, where possible. In general, if more than one reviewer felt that a question/answer combination was unacceptable, that item was dropped.

Pilot Study

In order to test the feasibility of the project, in particular the likelihood of identifying items which corresponded with the van Hiele levels, and of identifying subjects who operate at these levels, a pilot study was conducted. This phase focussed on assessing the performance of a group of individuals, for each of whom a van Hiele mastery level was known, on the revised item pool items.

Administration of the revised items. The revised items were administered to the 14 subjects participating in the pilot study, at one common sitting. Students were supplied with scrap paper, pencils, rulers, protractors and a copy of the items. (See Appendix E for the items.) They were instructed to indicate their answer choices directly on the test copy as a separate answer sheet was not provided. There was no time-limit for completing the items, since it was not important to know how much work could be accomplished in a fixed time period. Rather, the objective was to ascertain the congruence between a student's response to an item and that student's van Hiele mastery level. At the completion of the test, each student was also asked to comment on several structural facets of the test, such as the reading level, the content, the diagrams, any items which seemed unclear, inappropriate vocabulary, and so on. Their suggestions were incorporated into the next version of the instrument.

Establishing a van Hiele mastery level for each subject: Burger and Shaughnessy interviews. In order to assess each participant's dominant van Hiele level, independently from the responses to the written items, the interview procedures developed by William F. Burger and J. Michael Shaughnessy on quadrilaterals was administered by the researcher. The interviews were conducted in private, in a one-on-one environment, and with no time limit. Each interview was audio-taped.

The interview tapes were listened to twice by the researcher, once on the day of the interview and again at least a week later. Using the coding system developed by Burger and Shaughnessey, and with the level indicators as a guide, the interviewee's preferred level of reasoning on each task was identified. From those, an overall van Hiele level was assigned. The subject was then classified as an interview master of that level.

Two administrative questions arose from the decision to interview: (a) should the interviews be conducted before or after the students responded to the written items and (b) how much time should elapse between administering the two instruments? While the decision as to which procedure (written test or interview) should be administered first might appear to be arbitrary, the concern was that the interviews, because of their verbal and concrete nature, might act as an instructional influence to a greater extent than the paper-and-pencil test. In addition, as the final van Hiele Quadrilateral Test probably would not be administered after an instructional event similar to the interviews, it was decided that the written instrument should be administered first. Thus, each interview was conducted after each individual had written the multiple-choice instrument. The responses to the multiple-choice tests, however, were not scored until after the students were interviewed. This sequence was intended to ensure that an individual's performance on the test in no way influenced the level assigned to a student as a result of the interviewing.

It was also important to set boundaries on the time which elapsed between each evaluation situation. Testing twice on the same material, even with the interviews placed second, might result in a higher rating the second time. To lessen the possible impact of this "testing effect", at least 10 days elapsed between administering the written test and administering the interview. The interviews, however, were completed within 15 days of the written test. This was done in an effort to try to minimize the likelihood that students would acquire (or lose) geometric skills and knowledge between the two testing events. None of the students in the pilot study were receiving any mathematical instruction concurrent with the testing/interview period. This removed the possibility that they would receive further formal instruction in the area of quadrilaterals, although incidental learning could occur.

Identifying items for the draft instrument. In order to investigate whether or not the responses to each item tended to differentiate between those who were masters and those who were nonmasters of a level, an analysis of each item, relative to the interview mastery status of the subjects was performed. This involved an evaluation of the examinees' answer selections from the fixed choice responses, as well as an assessment of the written responses which were requested in some instances. Advice on the mechanical effectiveness of the items--wording, diagrams, etc.--was also solicited from the subjects. Using the results from these analyses, a draft instrument composed of the items which appeared to discriminate between masters and nonmasters was assembled. Directions for the examinee and an answer sheet were also developed to accompany the draft instrument. (See Appendix F for all of the draft instrument documents.)

Field Testing

The activities of the field test phase of the research were of two types. The first related to the identification of items from the draft instrument which appeared to discriminate between masters and nonmasters of the van Hiele levels. Once those items were identified, the reliability of the level assignments associated with the response patterns to that collection of items was explored. The goal was to have, at the end of this phase, an instrument and an interpretation scheme which could associate with a subject's responses on the test, the highest van Hiele level that individual had mastered.

Administering the draft instrument. The draft instrument was administered to 113 students from five mathematics classes in grades 6, 10, 11, 12 and university. The date for each administration was established in consultation with each classroom teacher. Approximately one week before the test was to be given, a permission slip was distributed to each student. This requested parental permission, where appropriate, for the student's participation in both the writing of the test and the interview. Copies of the permission form and the accompanying letter to the parents are contained in Appendix G.

The test was administered by the researcher to each class during their regular mathematics period. In order to meet the time allocations provided by the schedules of the schools from which the students were selected, a time limit of 60 minutes was imposed. The students were supplied with scrap paper, pencils, rulers, protractors, a test booklet and a separate answer sheet on which to mark their responses (see Appendix F).

Determining van Hiele levels. In order to investigate the response patterns of the field test participants in relation to their van Hiele mastery levels, the Burger and Shaughnessy quadrilateral interview protocols were administered by the researcher. The procedure described for the pilot group was followed, with the written test being administered before the interviews and scored after the interviews. The interviews began at least a week after the written test was completed and were completed within three weeks of an individual's writing the draft instrument. The interviews were administered on a one-to-one basis, away from the classroom in a quiet setting. No time limit was imposed on the interview. All interviews were audio-taped. In every instance, no instruction in geometry occurred in the regular classes between the time the written test was given and the last interview occurred.

One hundred interviews were completed. Although an attempt was made to interview all 113 students who wrote the draft instrument, this was not possible. The major reasons students did not participate in the interviews were:

(1) the inability to find a mutually agreeable "free" time to conduct the interview. (For all but the sixth grade students, interviews were conducted outside of class time. Some students had no free periods and/or worked before or after school.)

(2) students failing to show up for interviews due to sickness, forgetfulness, or whatever.

To determine the mastery assignments, each interview was listened to twice by the researcher, once on the day of the interview and again at least a week later. When both assessments agreed, the subject was assigned that mastery level. If, after listening twice, there was a difference in the mastery level assigned to an individual, the interview was listened to a third time, and a final decision made. In three cases, the researcher was unable to assign a mastery level with confidence. Those subjects' results were discarded. Confidence in the assignment of levels might have been further enhanced if the interviews had also been assessed by someone other than the researcher. Given, however, that no trained observer was available, that no likely candidate for such training was available and that considerable time would be required to train such an individual, once identified, an independent evaluation was not feasible.

Item analysis. In order to judge whether or not each item differentiated between masters and nonmasters and to identify structural flaws, an item analysis was performed. The students' collective performances on each item were analysed relative to their interview mastery levels. Items which appeared to discriminate between levels were identified. As well, a choice analysis was conducted to determine whether or not the distractors were functioning.

Selecting items for the final instrument. Up to this point, the majority of the research had focused on identifying items which appeared to correspond to particular van Hiele levels of thinking. Once such items were identified, the final instrument was assembled. As only 15 items emerged as corresponding to the levels, 5 items at Level 2, 6 items at Level 3 and 4 items at Level 4, all these items were selected for the instrument. In addition, the 4 Level 1 items which corresponded with the "strongest" discrimination statistics from the item analysis were also retained. Thus, 19 items where chosen for the van Hiele Quadrilateral Test. These items can be grouped and considered as four subtests, one corresponding to each van Hiele level. The items in the subtests corresponding to Levels 2, 3 and 4 have meet all the item selection criteria.

One of the criteria for the final instrument was that it be administerable within a 40 minute period. As the 60 minutes allotted for the 37 item draft instrument was sufficient for the field testing, it was felt that the 19 item final instrument could be completed in 40 minutes.

Selecting an interpretation scheme to convert raw scores into mastery decisions. The raw scores recorded on this instrument can be reported in two ways. The first is the overall number of correct answers. The second is the number of correct reponses, by subtest. The latter approach results in four scores being reported, a score for the Level 1 subtest, a score for the Level 2 subtest, etc. By using the interview mastery assignments for the field test subjects, and by considering their performance on the 19 items selected for the final instrument, scoring schemes based on each type of raw score were investigated. This exploration also addressed the issue of the reliability of the mastery decisions.

A limitation of calculating the reliability statistics with the field test subjects, however, is that their responses were also used to determine which items would be selected for the final instrument. Calculating test score reliability statistics from these responses may, therefore, appear to be a guarantee of obtaining a high reliability index. It is possible, however, that a collection of items which _individually_ discriminate between masters and non-masters, might not, when interpreted _collectively_ differentiate between masters and nonmasters. Minimally, then, calculating reliability statistics for this group could provide information which would, if the statistics were low, indicate the case described above, i.e., that there is some question about the interpretation of the items when viewed collectively. If, however, the reliability statistics are high, this would be additional support, though not conclusive, that the items, when viewed collectively, are functioning as intended.

Final Testing

This component of the research focused on the reliability of the mastery decisons obtained with the final instrument and on the validation of those mastery decisions. To study these issues, the instrument was administered to two criterion groups, subjects from the ninth grade and the twelfth grade. Students from these academic levels were chosen for two reasons. One was the differences in the geometry schooling each group had experienced. The twelfth grade students had completed their secondary school geometry education. The ninth grade students were only half-way through, and, as such, had not begun their study of deductive reasoning. Consequently, it was informally hypothesized that the performance of the two groups on the instrument would differ. The other reason that ninth and twelfth graders were chosen was that test scores from an external measure, the Basics Concepts section of the 1988-89 Nova Scotia Achievement Test were available for each group. As there was a strong geometry component on each test, the relationship between students' performances on this test and the van Hiele Quadrilateral test could be studied.

The van Hiele Quadrilateral Test was administered to 101 students, 51 students in the twelfth grade and 50 students in the ninth grade. The dates for the administration of the van Hiele Quadrilateral Test were decided in consultation with the cooperating teachers and the school board. Permission slips were sent home, approximately a week in advance of the testing date, requesting parental approval for subjects to participate in the testing. The permission form and the accompanying letter were similar to that of the field test subjects. (See Appendix H for copies of these documents).

The van Hiele test was written during the students' regular mathematics period, with a 40 minute time limit. The examinees were provided with an answer sheet, a test booklet, and a pencil. (The field test subjects indicated that they had not needed a straight-edge or a protractor.) The two twelfth grade classes wrote the test on the same day. The test was administered to the first class by the researcher. The classroom teacher, having observed the researcher administer the test to the first class, administered the test to the second class. Involving the teacher was necessitated by the fact that the researcher was administering the test to one of the junior high school classes at the same time that the second twelfth grade class was scheduled to write the test. The fourth class was administered the test, by the researcher, three days later. Both administrators followed the instructions which accompanied the instrument. (See Appendix I for copies of the instrument and the instructions.)

The Nova Scotia Achievement Tests had been written four months prior to the administration of the van Hiele Quadrilateral Test. During the interim period, however, neither the twelfth grade students nor the ninth grade students had studied geometry. The contents of the standardized test are specific to the curriculum for each grade level. The scores, therefore, from the Basic Concepts Test were used to make comparisons of the students' performance, <u>within</u> a grade, on the van Hiele Quadrilateral Test.

Comparisons between the performance of the members of the two grades were also conducted. These included the calculation of Chi squared statistics and of correlation indices. The first provided a measure of the independence between mastery assignments and grade level. The second provided information about the relationship between grade level and mastery decisions and about the relationship between grade level and performance on each subtest.

Subjects

The van Hiele Quadrilateral Test is designed to identify the van Hiele mastery level of students at the secondary school level, the seventh to the twelfth grade. This group was chosen because (a) the majority of the school-based geometry instruction occurs during this period and (b) students across this range of schooling have had varying exposure to and success with the topic of quadrilaterals. The effect of the latter is that students, often within the same class, display a range of geometric knowledge and a range of geometric skills. Information about how individuals and groups of students perceive geometric concepts can, therefore, assist with the development and delivery of appropriate instruction. Thus, the participation of subjects who represented the range of academic training provided in the secondary curriculum and the range of thinking skills reflected in the first four van Hiele levels was required for this research.

The subjects participating in the three phases of the test development which involved students -- the pilot study, the field testing and the final testing -- are discussed in the following sections.

Pilot Study Subjects

One of the purposes of the pilot study was to provided, for subsequent phases of the research, insight into which grades masters of each of the four van Hiele levels could be located. As the model indicates that instruction, not maturation, is the key element in attaining van Hiele levels, the pilot study subjects were chosen to represent a wide range of mathematical schooling. It was anticipated that individuals from each van Hiele level would be included in a group determined by this academic breadth.

Fourteen volunteers participated in the pilot study. Because this component of the research occurred in July, the students had just completed the school year. When school resumed in the fall, 2 subjects would be entering seventh grade, 2 subjects would be entering ninth grade, 3 subjects would be entering tenth grade, 3 subjects would be entering twelfth grade, 1 subject would be entering the first year of university with no declared major, 1 subject would be entering the third year of university as a biology major and 2 subjects would be entering their fourth year of university as mathematics education majors. There were 4 males and 10 females. The subjects ranged from 11 to 31 years of age.

The pilot subjects were enrolled in public schools located in a medium-sized coastal Canadian city. The pre-university students attended schools within the same affluent urban school district. Of the 2 students entering seventh grade, one, although a native

English speaker, attended a french immersion school. (This was the only subject in the research educated in a language other than English.) The 2 students entering ninth grade and the 3 students entering tenth grade attended the same junior high school during the academic year which had just concluded. Both groups had been taught that year by the same teacher. The 3 students in the twelfth grade attended the same high school, although they were each taught mathematics by a different teacher. The university students, with the exception of the first year student, attended the same local institution but came from different high schools in the metropolitan region. The first year student, having just completed ninth grade mainly through home schooling, was entering a different local university as a special student.

All the subjects had studied mathematics during each year of their schooling. The students entering seventh grade had received instruction in elementary school on (a) identifying, by name, geometric shapes--including triangles, quadrilaterals, other polygons and circles, (b) identifying components of figures, (c) using geometric instruments such as the compass and protractor, and (d) measurement (length, area, volume). As well, the elementary school curriculum included an introduction to the concepts of congruence, similarity, lines of symmetry and simple isometries. The students entering ninth grade had studied (a) types of polygons and their properties, including classification of shapes, (b) had engaged in exploratory work to learn about the isometry transformations and about dilatations, and (c) had used geometric instruments for constructions. The students entering tenth grade had also studied (a) properties of isometries, (b) algebraic descriptions of isometries, (c) properties of figures of plane geometry, explored through constructions and transformations, and (d)congruence through empirical approaches. The students entering twelfth grade had studied (a) deductive reasoning, (b) the traditional theorems of plane geometry (quadrilaterals, triangles, circles, parallelism, congruence, similarity) and (c) coordinate geometry, including proof using coordinates. The geometry in the high school setting was integrated into the mathematics course over two years, rather than presented as a one year course.

Field Testing Subjects

A central component of the field testing phase was the identification of masters and nonmasters for each of the four van Hiele levels. Using the known mastery groups, the ability of an item to elicit an appropriate response from each criteria group could be analysed.

Setting. Based on the results from the pilot, five educational settings were identified as likely sites from which to draw the subjects for the field testing. These were the sixth, ninth, eleventh and twelfth grades, and university mathematics courses for mathematics majors. It was anticipated that students from these settings would display the range of van Hiele levels required for this study.

Once the educational levels from which to draw the subjects had been identified, the selection of the participating classes was based on four factors: (a) identifying mathematics teachers and school administrators who were willing to let their students participate in the research, (b) identifying settings where students would have sufficient time to complete the draft instrument, (c) identifying school schedules which would allow students to have free time during the regular school day to participate in the interview used to identify van Hiele levels and (d) identifying groups large enough to provide the number of masters and nonmasters required for the research.

There was a minimum of difficulty in meeting the four requirements. Six schools, each of which the researcher had previous worked with in a professional capacity, were approached. All the principals and teachers expressed an interest in allowing their students to participate in the project. Finding a ninth grade setting, however, where students were in class longer than 40 minutes and where students had free time during the day for interviews was not possible. For this reason, no junior high school class was used at this stage. Instead, a tenth grade transition mathematics class was selected for the project. These students had not completed the junior high school mathematics curriculum, yet were in a high school setting. While they were older than ninth grade students and while they had been studying mathematics for 10, not 9, years, it was felt that they could still be included in the study. The perceptions these student had about geometry were more likely to be parallel to traditional ninth grade students, than to their tenth grade peers. The essential factor at this time was to identify "masters" and "nonmasters" of van Hiele levels, regardless of how much schooling those individuals had experienced.

The schools used for the field testing were located in the same school system as the schools from which the pilot students came. The sixth grade students attended an urban K-6 school in an affluent university neighborhood. For the most part, they had been taught by the same teachers each of the previous six years. The senior high school students all attended the same suburban three year (10th - 12th grade) high school. One class from each grade level participated, the tenth grade transition class described previously, a university oriented eleventh grade mathematics class, and an accelerated twelfth grade class. The university students were members of a seminar for honors mathematics majors. No grade was assigned for this class. It served an organizational function, providing a scheduled meeting each week for announcements, guest lecturers, field trips, etc., rather than an instructional function. Minimally, however, all the students participating in this class had completed a full year's study of calculus and either completed, or were taking, a course in matrix algebra. The elementary and secondary school students had studied the same topics, in the same sequence, as those described for the pilot subjects. (The accelerated twelfth grade class had studied the same topics as their non-accelerated peers, but in more detail).

With the exception of the sixth grade class, the participating classes were identified through the recommendation of the department head within the respective schools. This decision was made following discussions with the researcher about the goals of the field testing. As there was only one sixth grade class in the school selected for the research, once the teacher's approval was obtained, no further selection procedures were required.

Sample size. The minimum sample size sought for each van Hiele level at this stage was 21 masters and 21 nonmasters. With this sample size, a minimum of 105 subjects were required for the field testing, 21 each for the nonmasters of Level 1, the masters of Level 1, the masters of Level 2, the masters of Level 3, and the masters of Level 4. (For this selection, masters of Level n were not also considered as masters of Level n-1).

The decision about the size of the sample was based on five assumptions:

1. The binomial distribution was used to represent the theoretical distribution of scores for the masters and nonmasters of a given level.

2. The difference between the masters and nonmasters success rates was estimated to be, minimally, 25%. Based on the distinct nature of the levels, it might be reasonable to predict that masters would have a consistently high success rate on items based on that level and that the nonmasters would have a consistently low success rate on those items, say, for example 90% and 20%, respectively. This would result in a large difference between success rates, 70% in this case. In practice, however, these extreme rates may not correspond to master and nonmaster performance. As the rates demonstrated by the two distinct groups may be less divergent, the more conservative 25% figure was selected. Accordingly, master and nonmaster "success" rates were calculated, respectfully, at 66% and 41%, at 70% and 45%, at 75% and 50%, at 80% and 55%, at 85% and 60%, and at 90% and 65%, for each sample size tested. This provided a broad range of rates for evaluation.

3. The power of the test, denoted $1-\beta$ was set at the nominal level of $1-\beta \geq 0.80$. This statistic is the probability of making a correct rejection of the null hypothesis, that is, the power to detect the alternative hypothesis. In this instance, the null and alternative hypotheses would be:

$$H_o : \mu_m - \mu_{nm} \leq 0$$

$$H_A : \mu_m - \mu_{nm} > 0$$

where μ_m is the mean of the masters scores and μ_{nm} is the mean of the nonmasters. Setting the power statistic, in turn, establishes beta, $\beta \leq 0.20$. Beta is interpreted as the probability of failing to reject the null hypothesis when it is false. This failure is called a Type II error.

4. Rejecting the null hypothesis when it is true is called a Type I error. In this instance, the maximum probability of making a Type I error, α, was kept close to the nominal value of 0.05. By convention, this is the largest risk an experimenter is willing to take of rejecting a true null hypothesis. When slightly higher values of α were considered, the justification lay with the fact that "...it might be desirable to set the value of α at .10 or perhaps .20...in preliminary stages of test construction, when it is more important to discover items of possible value than to be certain of eliminating 'duds'" (Minium, 1978, p. 271).

5. As a directional prediction was being made, a one-tailed test was considered.

Table 3.1 presents, for sample sizes of 20, 21, 22 and 23, over a range of success rates for masters and nonmasters, each differing by 25%, values of α, β, and $1-\beta$ which correspond to the research criteria. Twenty-one was the smallest sample size where the criteria for α and β were simultaneous met for the range of success rate tested. Using the broader range for α suggested by Minium as acceptable in the developmental stages,

Table 3.1

Selected Critical Values for Sample Sizes 20, 21 22,and 23 when Success Rates Differ by 25%

n	p(m)	p(nm)	c	α	β	$1-\beta$
20	.66	.41	10	.1032	.1480	.8520
	.70	.45	11	.1133	.1308	.8692
	.75	.50	12	.1018	.1316	.8684
	.85	.60	14	.0673	.1256	.8744
	.90	.65	15	.0432	.1182	.8818
21	.66	.41	10	.0637	.2000	.8000
	.70	.45	11	.0676	.1841	.8159
	.75	.50	12	.0561	.1917	.8083
	.80	.55	13	.0431	.1971	.8029
	.85	.60	14	.0287	.2002	.7998
	.90	.65	16	.0522	.0924	.9076
22	.66	.41	11	.0893	.1415	.8585
	.70	.45	12	.0916	.1328	.8672
	.75	.50	13	.0746	.1431	.8569
	.80	.55	14	.0561	.1518	.8482
	.85	.60	15	.0368	.1584	.8416
	.90	.65	16	.0182	.1629	.8371
23	.66	.41	11	.0555	.1895	.8105
	.70	.45	12	.0546	.1836	.8164
	.75	.50	13	.0408	.2024	.7976
	.80	.55	15	.0715	.1152	.8848
	.85	.60	16	.0463	.1240	.8760
	.90	.65	17	.0226	.1309	.8691

Note. n = size of samples
p(m) = success rate for masters
p(nm) = success rate for nonmasters
α = probability of making a Type I error
c = value at which α occurs (critical value)
β = probability of making a Type II error at critical value "c"
$1-\beta$ = probability of making a correct rejection of the null hypothesis at critical value "c"

sample sizes of 20 would also have been sufficient. (See Appendix J for the binomial expansions using sample sizes 20 to 23, with a range of success rates.) When the spread between the masters' and nonmasters' rates of successfully answering is more than the 25% assumed above as the minimum, a critical value can be found where, simultaneously, the probability of making a Type I error and a Type II error is reduced. This is demonstrated in Table 3.2 using several values for n = 21.

Sample subjects. Of the 113 students who wrote the draft test in the field testing phase, 24 were in the sixth grade, 25 were in the tenth grade, 28 were in the eleventh grade, 20 were in the twelfth grade, and 16 were in the university honors mathematics seminar. The examinees ranged from a minimum of age 10 to a maximum of age 30. Distribution by gender was approximately equal within grade levels and across the sample (Table 3.3).

Final Testing Subjects

The final set of items, assembled into the van Hiele Quadrilateral Test, was administered to 50 students in the ninth grade and 51 students in the twelfth grade. These students were enrolled in schools in the same province as the subjects involved in the earlier stages of the research, but the schools were located in a different city. This meant that the school curriculum, year by year, was the same as described previously, but there were local variations within the sequencing of topics.

Table 3.2

Critical values for selected success rates which differ by 25%, 30%, 35% and 40% when sample size is 21

p(m)	p(nm)	p(n) - p(nm)	c	α	β	$1-\beta$
.70	.45	.25	11	.0676	.1841	.8159
.70	.40	.30	10	.0264	.1744	.8256
			11	.0676	.0849	.9151
.70	.35	.35	9	.0087	.1723	.8377
			10	.0264	.0772	.9228
			11	.0676	.0314	.9686
.70	.30	.40	8	.0024	.1477	.8523
			9	.0087	.0676	.9324
			10	.0264	.0264	.9736
			11	.0676	.0087	.9913

Note. p(m) = success rate for masters

p(nm) = success rate for nonmasters

α = probability of making a Type I error

c = value at which α occurs (critical value)

β = probability of making a Type II error at critical value "c"

$1-\beta$ = probability of making a correct rejection of the null hypothesis at critical value "c"

Table 3.3

Distribution of Field Testing Subjects by Gender and Grade

	Grade				
Gender	6	10	11	12	University
Male	11	11	13	10	12
Female	13	14	15	10	4

The two academic levels, ninth grade and twelfth grade, were chosen on the basis of the diversity of the geometry instruction which the students had received. Because of the variation, it was anticipated that the performance of these two groups on the geometry test would be different. The ninth grade students, given their academic background, would be unlikely to have mastered the concepts associated with Level 3 and even more unlikely to have encountered, much less mastered, the concepts associated with Level 4. The grade 12 students, having completed the study of formal geometry, might be expected to have mastered Level 3 thinking, and in many cases, to have mastered Level 4 thought.

The participating classes were assigned to the researcher by the school system, in response to the request to work with a minimum of 50 students at each level. The twelfth grade subjects were members of two mathematics classes, taught by the same mathematics teacher. The three year high school (10th - 12th grades) they attended was located in a lower middle class urban neighborhood. As one answer sheet was spoiled, only 50 responses were considered. Of these, there were 22 males, 28 females. With the exception of one student who was 20, these subjects were 17 or 18 years of age.

The Grade 9 subjects were members of mathematics classes in two different schools. One school was a feeder school for the high school used in this stage. The other school, located in a modest middle class urban neighborhood, was a feeder school for a different high school in the same city. Twenty-five members of each class were present on the day the test was administered, for a total of 50 subjects from the ninth grade. Of those, 21 were male and 29 were female. All but two of these students were either 14 or 15 years of age, the age expected for this grade level. The exceptions were older, with one 16 years old and the other 17 years old.

The Measures

Two measures, other than the van Hiele Quadrilateral Test, were used in the research, the Burger and Shaughnessy Interview on quadrilaterals and the Basic Concepts Test from the Nova Scotia Achievement Test. Each of those is described in this section.

Burger and Shaughnessy Interview On Quadrilaterals

The interview procedures developed by William F. Burger and J. Michael Shaughnessy (1986) for quadrilaterals were used to assess the dominant van Hiele level of the participants in the pilot phase and in the field test phase of the research. The developers' goal was to design an interview script and analysis protocols which could easily be administered by teachers and researchers. Their interview addressed two content areas, quadrilaterals and triangles, in separate collections of activities. Only the quadrilateral activities were used in this research. These activities were designed to be used in a one-on-one situation with no time limit. They can be used to reveal predominant van Hiele levels of reasoning, over Levels 1 to 4.

The interview material consists of three parts: (1) the interview activities, (2) the interview script, (3) the analysis coding packet. The quadrilateral activities involve five sequential tasks, (a) drawing, (b) identifying and defining, (c) sorting, (d) inference, and (e) axioms, theorems and proofs. Supplied with pencils, straight edge, paper, and compasses, students manipulate, draw, sort, and respond to the interviewer's scripted questions in these five areas. As an example, students are presented with a set of 9 cutout quadrilaterals of various shapes. The subject is asked to put some shapes together that are alike in some way. The researcher then probes the basis on which the student identified the figures as being alike. The responses to each question (captured on audio-tape), the student's drawings and the interviewer's notes are analysed according to the response categories in the analysis protocols. The predominant level of thinking displayed by the subject on each task is determined. From these, an overall van Hiele level of reasoning is assigned.

The interview materials are the result of three cycles of piloting and revisions, each conducted by the researchers, in a project investigating the van Hiele levels. Once developed, the interviews were used by their developers with 45 students from kindergarten age through university. When the five quadrilateral tasks and the three triangular tasks were administered, they found that the time required for completion ranged from 40 minutes to 90 minutes. Three researchers analysed the responses of 14 subjects, for each of the 8 interview activities, then assigned an overall level of thought for each individual. Interrater consensus studies were conducted on these results.

The Nova Scotia Achievement Tests

The Nova Scotia Achievement Tests are a series of tests measuring knowledge and the ability to use knowledge in each of seven subjects areas: social studies, science, mathematics computation, mathematics basic concepts, reading, mechanics of writing and english expression. The tests are designed to "help determine the extent to which provincial, district, school and individual classroom objectives are being met" (Nova Scotia Department of Education, 1989, p. 11).

The tests were developed cooperatively between Applied Measurement Services, Mount Holly, New Jersey and the Nova Scotia Curriculum and Research Sections. The twelfth grade tests were first administered in 1972; the ninth grade tests began in 1976. Each year approximately 25% of the questions are revised. The items have been constructed to parallel the curriculum, texts and teaching guides used in the provincial courses. The items have also been reviewed by a panel consisting of the relevant provincial curriculum supervisor and teachers from a range of grade levels.

This study used the results from the Level 9 and Level 12 Mathematics Basic Concepts Test. The objectives of these two tests are to measure application, comprehension, evaluation and inference skills. The content areas covered are (1) geometry, measurement and logic, (2) number facts and operations, (3) ratio, proportion, probability and statistics, and (4) relationships and sets. The geometry section is 40% of the ninth grade test and 38% of the twelfth grade test.

Each test is administered by the school. There is a 60 minute time limit on each of the 50 four-choice item tests. While no example of a test item was made available to the researcher, the literature published by the province cites as an example of a "comprehension of concepts question", a question which tests comprehension of the concept of reflection (Nova Scotia Department of Education, 1989).

Students receive both a standard score and a percentile rank for each test. For the 1988-89 school year, the statistics about the Basic Concepts Test presented in Table 3.4 were reported by the Nova Scotia Department of Education, Research Section (personal communication, May 17, 1989).

Table 3.4

Nova Scotia Achievement Basic Concepts Test Statistics

	Grade	
Statistic	Ninth	Twelfth
Mean Raw Score	21.90	23.54
Standard Deviation	8.04	7.91
Alpha Reliability	0.85	0.84
Standard Error of Measure	3.12	3.13

The performance of each subject in the research on the geometry questions only was not made available for the research. Thus, a limitation of using the results on this instrument, to make comparisons with van Hiele mastery level assignments, is that this standardized instrument tests topics other than geometry.

Summary

The procedures followed for the development of the van Hiele Quadrilateral Test were presented in this chapter. Included were a description of the stages of development, of the subjects and of the instruments used for collecting data. In chapters 4, 5, 6, and 7, the findings from each of the production stages--developing the items, the pilot study, the field testing and the final testing--respectively, are presented. Chapter 8 draws conclusions from those findings and makes suggestions for further research in the area of assessment.

Chapter 4

DEVELOPING THE ITEMS

The development of the van Hiele Quadrilateral Test involved four interrelated stages. In the initial phase, the goals of the assessment were identified and items with the potential to correspond with these goals were assembled. Next, the items were administered to students. First a small group of subjects, the pilot study subjects, responded to the items. After revisions, the items were assembled into a draft instrument and administered, as part of the field test study, to a larger group. Finally, based on the responses of the field test subjects, the final instrument was assembled and tested with another group of subjects. This chapter presents the findings from the first phase in the development of the van Hiele Quadrilateral Test.

Writing the Initial Items

To assist in the development of the multiple-choice questions and answers for the van Hiele Quadrdilateral Test, an inventory of characteristics displayed by individuals operating at each van Hiele level was assembled. These behaviors, called the "level indicators", are contained in Appendix B. Using these as a guide, items were written to correspond with each level. The 53 items in the inital item pool corresponded with 53 (72%) of the 74 original indicators. Table 4.1 presents the distribution of the level indicators across the item pool, with the level indicators in numerical order. Table 4.2 presents the same information, but with the items listed in numerical order.

In general, the descriptors for which multiple-choice items were not written were (a) those calling for observations of the students interacting with concrete objects, (b) those calling for verbal descriptions, and (c) those involving the monitoring of multi-stepped strategies. The multiple-choice format, in combination with the requirement that the instrument be easily administered to a large number of examinees in a single session, would not allow examinees to interact with materials in a context which an evaluator can observe. Instead, the examinees are required to react, selecting an acceptable answer from predetermined written choices. They are not able to generate their own responses, written or verbal. They are not able to demonstrate the interim strategies they have used to arrive at solutions.

The items in the initial item pool were also categorized by geometric concepts, particularly quadrilaterals. The distribution of the items by shape is presented in Table 4.3. Items in the "general" category mainly require a knowledge of components of figures, rather than of specific shapes, or emphasize the nature of deductive principles, independently of the geometric figures. (Item 37 introduces a "new" shape and expects the subjects to make some simple deductions. This type of problem represents an attempt

Table 4.1

Correspondence between Initial Item Pool Items and Original Level Indicators (with indicators in numerical order)

Level Indicator	Item
0.01	1, 3, 6, 7
0.02	2
0.03	2,3,4
0.04	4
0.05	-
0.06	-
0.07	5
0.08	8
0.09	1, 3, 6, 7
0.10	6
0.11	15
0.12	6
0.13	7, 11
0.14	8
1.01	9, 18
1.02	11, 20
1.03	9, 12, 18, 21
1.04	9, 13, 17, 21
1.05	-
1.06	-
1.07	-
1.08	10, 12
1.09	9, 10, 13, 18
1.10	9
1.11	14, 24
1.12	10, 20, 22
1.13	15, 18
1.14	17
1.15	8, 16
1.16	-
1.17	-
1.18	11, 19, 20
2.01	25, 36
2.02	-
2.03	-
2.04	27, 32, 35
2.05	-
2.06	-
2.07	32, 39, 40
2.08	23, 26, 31, 36
2.09	-
2.10	29, 30, 52
2.11	41
2.12	28
2.13	28, 32
2.14	14, 24
2.15	27, 31, 33, 35
2.16	37
2.17	34
2.18	38
2.19	-
2.20	29
2.21	30
2.22	38
2.23	23, 25, 28
3.01	53
3.02	53
3.03	-
3.04	46, 47, 51
3.05	-
3.06	-
3.07	42, 43, 44
3.08	48
3.09	45, 50, 52
3.10	51
3.11	44
3.12	49, 51
3.13	45
3.14	-
3.15	-
3.16	-
3.17	49
3.18	-
3.19	-

Table 4.2

Correspondence between Initial Item Pool Items and Original Level Indicators (with items in numerical order)

Item	Level Indicator	Item	Level Indicator
1	0.01, 0.09	27	2.04, 2.15
2	0.02, 0.03	28	2.12, 2.13, 2.23
3	0.01, 0.03, 0.09	29	2.10, 2.20
4	0.03, 0.04	30	2.10, 2.21
5	0.07	31	2.08, 2.15
6	0.01, 0.09, 0.10, 0.12	32	2.04, 2.07, 2.13
7	0.01, 0.09, 0.13	33	2.15
8	0.08, 0.14, 1.15	34	2.17
9	1.01, 1.03, 1.04, 1.09, 1.10	35	2.04. 2.15
10	1.08, 1.09, 1.12	36	2.01, 2.08
11	0.13, 1.02, 1.18	37	2.16
12	1.03	38	2.18. 2.22
13	1.04, 1.09	39	2.07
14	1.11, 2.14	40	2.07
15	0.11, 1.13	41	2.11
16	1.15	42	3.07
17	1.04, 1.14	43	3.07
18	1.01, 1.03, 1.09, 1.13	44	3.07, 3.11
19	1.18	45	3.09. 3.13
20	1.02, 1.12, 1.18	46	3.04
21	1.03, 1.04	47	3.04
22	1.08, 1.12	48	3.08
23	2.08, 2.23	49	3.12, 3.17
24	1.11, 2.14	50	3.09,
25	2.01, 2.23	51	3.04, 3.10, 3.12
26	2.08	52	2.10, 3.09
		53	3.01, 3.02

Table 4.3

Distribution of Original Item Pool Items Across Geometric Concepts

Concept	Item (by Number)	Totals
Kite	30	1
Parallelogram	6,7,17,19,21,24,27,29,33,36,40,41,46,47,51	15
Quarilateral	3,15,22,34,35,45,46,48,49,53	10
Rectangle	4,5,8,10,13,14,23,25,26,27,31,36,37,42,45,	15
Rhombus	9,16,20,26,29,49	6
Square	1,11,17,22,23,26,27,31	8
Trapezoid	28	1
General	2,4,12,18,32,37,38,39,41,43,44,50,52	13
	Total	69[a]

[a] Some items are listed more than once

to avoid problems which could be solved by memory, rather than through understanding.)

Validating the Level Indicators and the Items

To assess item validity, both in terms of the level reflected in the question/answer choices and the geometry content, the item pool and the level indicators were sent to five experts on the van Hiele model. Each person was asked to review the level indicators for their breadth and accuracy, and to comment on the appropriateness of the questions and answers for eliciting the indicated level specific responses. Four of the five individuals who initially agreed to review the materials responded.

Level indicators

In general, the experts agreed with the level indicators. Four strategic comments, however, were made:

(1) One expert felt that the indicators at the first level were too sophisticated.

(2) Another individual, in response to a request issued to all the experts, replied that the decision to identify "the ability to accept equivalent definitions" at the third level, was appropriate. (This was the only direct reference to the request.)

(3) One expert inquired about the numbering system for the levels, wondering which choice -- identifying levels as 0, 1, 2 and 3, or as 1, 2, 3 and 4 -- would be the more appropriate.

(4) The non-returning expert, in a telephone conversation, suggested checking the indicators against Pierre van Hiele's 1986 book, Structure and Insight: A theory of mathematical education.

In response to the experts' replies, several revisions were made to the indicators. The first eliminated redundant descriptors, particularly those in the visual and logical categories. As those phenomena can be observed only through action, they could be subsumed into other categories. For example, the original indicator 1.01, "notices properties of a figure", is inferred when an individual writes about, speaks about or otherwise indicates a property. In this instance, the presence of the actions of indicators 1.03, 1.04, 1.05, 1.13, or 1.14, could be interpreted as evidence of indicator 1.01. The consolidation of the descriptors also addressed, in part, the issue of the sophisticated nature of the first level descriptors which one expert had raised.

Another revision was the renumbering of the levels. The designations of the levels used by P. M. van Hiele in 1959 were:

Level 0: Base level

Level 1: Aspect of geometry

Level 2: Essence of geometry or aspect of mathematics

Level 3: Discernment of geometry or essence of mathematics

Level 4: Discernment in mathematics

In a recent discussion of the level designations, Professor van Hiele indicated that originally the model did not concern itself with what occurred before the "aspect of geometry" (van Hiele, 1986). Thus, "aspect of geometry" was treated as the first level. Subsequent work with the model, however, emphasized the importance of understanding and clarifying the stage preceding the "aspect of geometry". This has resulted in an elaboration of the behaviors associated with the initial level, and a subsequent renumbering of the levels.

To be consistent, then, with the most recent thinking by van Hiele, the level designation used in this research were renumbered. The renumbered levels and current lable designations used henceforth are:

Level 1: Visualization

Level 2: Analysis

Level 3: Abstraction

Level 4: Deduction

Level 5: Rigor

The level indicators, renumbered and revised, are presented in Appendix D.

Items

The experts' responses to the questions and answers aggregated into two categories: (a) comments particular to the goal of eliciting van Hiele based responses and (b) comments about the structure of the question and answer combinations. The first group of concerns, those which were model based, have significance for the validity of the items.

The nature of the experts' concerns, arising from their familiarity with the model, were (a) whether predetermined answer choices were representative of student thinking, (b) whether the reason an answer was selected was consistent with the proposed type and level of thinking, (c) what prerequisite vocabulary and concepts students would bring to the testing situation, (d) the emphasis given in the items to familiarity with vocabulary, (e) the inclusion of extraneous concepts, particularly those of a numeric or algebraic nature, and (f) the use of diagrams. Representative examples of the panel's comments in these six areas are presented below. The circumstance prompting the comment is indicated in parenthesis.

(a) Whether predetermined answer choices are representative of student thinking:

> Clever pupils have their own solutions and therefore they will not come up to the standards. (A general comment)
>
> There are several of your questions for which I do not feel you can decide the level of reasoning solely from the choice...I see no harm in including space on the exam for some items to ask students why they picked the answer they did. For me, the students' reasoning often can only be made explict if they are asked to talk about it in some way. (A general comment)
>
> What if a student comes up with the answers 2,4 or 2,5 and so answers E? Such a student could be Level 3. (Item 14)

(b) Whether the reason an answer was selected was consistent with the proposed type and level of thinking:

It is a question of remembrance, not of insight. (Item 21)

Could be a reduction of level. (Response "e", item 8)

Not Level 2 is tested but Level 1. Deduction is not necessary. (Item 38)

I think this is more than Level 2. It depends on knowledge of, or ability to explore, varied definitions. (Item 29)

It seems to me the question evaluates whether they can recognize a definition and theorem, but not the need for definition & theorems. (Item 42)

(c) What prerequisite concepts and vocabulary students would bring to the testing situation:

I'm not sure about this! Doesn't correctness of answer, at Level 3, depend on how one sets things up? (Item 42)

Is there a way to use the same level indicator with a more common term?--I've had many students who just didn't know the meaning of adjacent. (Item 12)

Should you also tell students what diagonals are? (Item 9)

Could be a lower level if the concept has been learned correctly... (Item 36)

(d) the emphasis given in the items to familiarity with vocabulary:

The asking of names is unfit to decide about levels. (Item 2)

Too much attention to standard vocabulary.... (Item 4)

(e) the inclusion of extraneous concepts, particularly those numeric and algebraic in nature:

The question is mixed up with algebra. (Item 19)

The inclusion of length here changes the objective? Might students have a solid concept of rectangle, but count intersection points to get length? (Item 5)

(f) the use of diagrams:

The drawings reduce the level. (Item 18)

At this high level figures are not allowed. (Item 53)

I realize that you are trying not to 'give away' answers to several questions by supplying figures. On the other hand, it seems like there are a number of items for which a figure would enhance the clarity of the question, and make it easier to understand. I believe you should supply more figures for them. This is geometry, not reading. (General comment)

The panel's comments on the potential of the items to reflect level specific thinking were coded into three categories. One category corresponed with agreement; the expert felt the item matched the proposed level. One category corresponded with rejection; the expert felt the item did not match the proposed level. One category corresponded with uncertainty; the expert was unsure about whether or not the item matched the objective, or the expert suggested that revisions be made in order for the item to correspond to the proposed level. Table 4.4 contains an item by item profile of those responses.

Based on the experts' suggestions, items were reviewed, revised, retained, or rejected. With one exception, which is discussed below, an item was rejected if more than one reviewer felt that it did not correspond to the van Hiele model. The items which were not rejected were reviewed. The suggestions from the experts on how to revise items were considered and adopted, where possible. For example, the apparently contradictory advice regarding diagrams given by experts which was cited earlier was addressed by not including diagrams at the highest level and by also re-evaluating each item in terms of its clarity of meaning.

Table 4.4

Distribution of the Experts' Responses to the Items in the Original Item Pool

Item Number	Expert's Responses 1[a]	0[b]	-1[c]
1	1	2	1
2	1	1	2
3	0	2	2
4	2	0	2
5	1	2	1
6	1	2	1
7	1	3	0
8	2	1	1
9	3	1	0
10	3	1	0
11	1	1	2
12	1	0	3
13	2	1	1
14	2	1	1
15	3	1	0
16	1	2	1
17	2	2	0
18	2	1	1
19	2	0	2
20	3	1	0
21	2	1	1
22	2	2	0
23	3	1	0
24	3	1	0
25	2	0	2
26	3	1	0
27	2	2	0
28	2	2	0
29	0	1	3
30	1	2	1
31	3	1	0
32	2	2	0
33	2	1	0
34	1	1	2
35	3	1	0
36	2	0	2
37	3	1	0
38	2	1	2
39	1	3	0
40	2	2	0
41	2	1	1
42	0	2	2
43	3	1	0
44	4	0	0
45	2	2	0
46	3	1	0
47	2	1	1
48	2	1	1
49	2	1	1
50	4	0	0
51	1	1	2
52	2	1	1
53	2	1	1

[a] 1 = Item matches the designated van Hiele level.
[b] 0 = Item needs revision or uncertainty exists about whether item matches the designated van Hiele level.
[c] -1 = Item does not meet the designated van Hiele level.

Another suggestion from a reviewing expert was that, in some instances, the format of the items be expanded. He felt that, in the development stage, it would be of value to have examinees explain, in writing, why they selected an answer. This might reveal (a) whether or not a student's reasoning corresponded with the developer's level-designations of the answer choices, (b) if not, why not and (c) structural flaws in the items (misleading diagrams, etc). Items where at least one reviewing expert indicated interest in knowing more about how the "correct" answer was determined, were revised to elicit this type of response. As well, one item (#25), which more than one reviewer had rejected, was given this format and included with the revised items. For this item, confirmation of the experts' rationale for rejection was being sought.

The Revised Item Pool

The revised item pool consisted of 45 items. There were 9 questions with answers corresponding to Level 1, 15 questions with answers corresponding to Level 2, 17 questions with answers corresponding to Level 3, 8 questions with answers corresponding to Level 4. Of the 45 items, 4 had answer choices corresponding to two levels and 17 requested a written response, in addition to the multiple-choice response, explaining the reasoning used when an answer choice was selected. The relationship of the revised item pool items to the original item pool items is shown in Table 4.5.

Table 4.5

Items Retained From the Original Item Pool for the Pilot Study

Original Number	Pilot Number	Original Number	Pilot Number
1	1, 2[a]	31	28
2	-	32	30
3	-	33	31
4	-	34	-
5	-	35	32
6	3[a], 5[a], 6[a], 18	36	33[a]
7	7	37	34[a]
8	-	38	21
9	17	39	-
10	10	40	35
11	-	41	36
12	-	42	-
13	19, 20	43	38
14	11[a]	44	49
15	8[a], 12	45	40
16	-	46	41
17	13, 14	47	42
18	-	48	43
19	-	49	37[a]
20	15[a], 22[a]	50	44
21	-	51	-
22	16[a]	52	45
23	27[a]	53	-
24	23[a]		
25	26[a]		
26	24[a]		
27	25[a]		
28	-		
29	-		
30	29		

[a] These items requested a written explanation.

Summary

The initial step in the development of the items for the van Hiele Quadrilateral Test was that of identifying the guidelines for writing the items: the level indicators and the quadrilateral facts. Once the items were written, the level indicators and the items were sent to a panel of experts for review. Based on the responses from the panel, revisions were made in both the indicators and the items. The revised item pool was used in the next phase of the test development, the pilot study. The results of the pilot study are presented in the next chapter.

Chapter 5

PILOT STUDY

The pilot study was conducted to (a) provide insight into the correspondence between an individual's answer selections on the revised item pool items and the individual's van Hiele level, (b) to suggest future research groups and (c) to uncover structural flaws in the items. Fourteen students, chosen from a range of mathematical schooling, were administered the items in the revised item pool. As well, each subject's van Hiele mastery level was determined using the Burger and Shaughnessy quadrilateral interview. Comparisons were then made between subjects' performances on the items and their interview performance. The mechanics of the items were also investigated through a choice analysis and from students' comments. A discussion of the findings from these studies is presented in this chapter.

Item Analysis

Using the mastery levels assigned to each subject through the interviews--two subjects were masters of Level 1, three were masters of Level 2, five were masters of Level 3, and four were masters of Level 4--difficulty indices for masters and nonmasters were calculated for each item. From those, a discrimination index for each item was also obtained.

Item Difficulty Index

An item difficulty index indicates the proportion of individuals in a designated category who correctly answer the item under consideration. As such, the index, usually presented in decimal form, ranges in value from zero to one. An index value close to one indicates that a majority of the individuals in the category successfully answered the item. The item was an "easy item" for that group. An index close to zero indicates that very few individuals in the category successfully answered the item. The item was a "hard item" for that group. A difficulty index of 0.50 indicates that half of the individuals in the group answered correctly, while half of them did not.

For each item used in the pilot testing, two types of difficulty indices were calculated. One considered the responses of the individuals who had mastered the level associated with the item. The other considered the responses of the individuals who had not mastered the level associated with the item.

For this analysis, mastery and nonmastery were defined on the basis of the interview mastery designations. To calculate an index for an item corresponding to Level "n", the master's category was composed of all individuals who were designated by the interview as masters of Level n or of any level higher than Level n. This grouping is referred to as "all masters". For example, "all masters" of Level 3 are those individuals who, through the interview procedure, were designated masters of either Level 3 or Level 4. Similarly, for this calculation, the individuals who had not yet mastered Level n, or a higher level, were considered as nonmasters of Level n. This group was referred to as "all nonmasters". The "all nonmasters" of Level 3, for example, were those individuals who were interview masters of Level 2, interview masters of Level 1, or those individuals who had not mastered Level 1. Table 5.1 shows the level by level correspondence between interview mastery designations and the "all masters and nonmasters" grouping. The rationale supporting the combination of the interview masters into these two larger categories comes from the sequential property of the van Hiele model: to have mastered Level n + 1, one has also to have mastered Level n.

For the pilot study, when an item had two or more answers which corresponded to different levels, a difficulty index was generated for the response at each level. For the response corresponding with the highest level, masters and nonmasters were determined in the same way as for the other items. One answer was considered as correct; all other answers were considered as incorrect. When the next lower level response was being considered, however, selecting either the response for that lower level, or the higher level response, was considered as "correct". (This meant that when the "lower" level was being considered, the item had a higher probability of randomly being answered correctly

Table 5.1

All Mastery Assignments, By Level

	Interview Mastery Level				
	Pre-1	1	2	3	4
Level 1					
All masters		X	X	X	X
All Nonmasters	X				
Level 2					
All masters			X	X	X
All nonmasters	X	X			
Level 3					
All Masters				X	X
All Nonmasters	X	X	X		
Level 4					
All masters					X
All nonmasters	X	X	X	X	

that did the single answer items. For example, with two correct responses available, the probability of randomly guessing a correct response would be 0.40, as opposed to 0.20 for an item with a single correct answer.)

Item Discrimination Index

An item discrimination index measures the difference between the performance of two groups. For the pilot study, this statistic was calculated by subtracting the difficulty index of the "all nonmaster" group from the difficulty index of the "all master" group.

$$\text{Discrimination Index} = \text{Difficulty Index "all masters"} - \text{Difficulty Index "all nonmasters"}$$

The maximum value for the item discrimination index is 1.00. This occurs when all of the masters answer the item correctly and none of the nonmasters answer the item correctly. An item with an index of one would be considered as discriminating well between masters and nonmasters. A discrimination index of 0.00 occurs when an equal percentage of both groups answered the item correctly. No discrimination between masters and nonmasters appears to result from an item with this index. A negative index occurs when the nonmasters answer the item in a greater proportion than the masters, usually an undesirable result.

Decision Criteria

When an item from the pilot testing registered a positive discrimination index, it was identified as a potential item for inclusion at the next research stage. As well, where available, the students' written responses explaining why they selected their answer choice was considered. The van Hiele level corresponding to these explanations had to be consistent with the intended level for the item in order for the item to proceed to the next stage.

When an item registered a discrimination index of 0.00 or lower, it was reviewed. The re-assessment included consideration of the written responses from the students, when available, and an assessment of the difficulty indices for the masters and nonmasters. When this analyses indicated that "non-level" reasoning was consistently leading to correct answers, or that nonmasters of a level were consistently selecting answer choices associated with that level, the comments from the panel of experts was again consulted. If all three factors indicated there was weak support for an item, it was eliminated. If, however, at least two of the analyses techniques supported the item's potential to identify masters or nonmasters, the item was retained for further analysis at the next stage.

Statistical Findings

The difficulty indices and the discrimination index for each item in the pilot study are presented in Table 5.2. As there were no nonmasters of Level 1 amongst the pilot subjects (i.e., everyone was a master of some level), no difficulty index for nonmasters could be calculated for items associated with Level 1. Consequently, no discrimination index could be found. For the remaining levels, however, indices are available for each item.

On the basis of the discrimination indices, eight items (8, 10, 11, 13, 14, 16, 21 and 22) which did not appear to discriminate between masters and nonmasters of Levels 2, 3 and 4 were identified. Items 8, 10 and 22 were not retained. Items 11, 13, 14, 16 and 22 were retained. For three of these, 13, 14, and 16 (each a Level 2 item), the master's difficulty index was at least 0.50. While the nonmaster's difficulty indices were also high, those figures had been calculated on the responses of only 2 subjects. With such a small sample, the resulting nonmasters difficulty index might not be representative of the response patterns for nonmasters of this level. Therefore, even though the nonmasters indices were high, <u>and</u> because the master's difficulty indices were strong, it was decided to test these items with a larger group. (Items 10 and 22 also demonstrate this index pattern. The level descriptors associated with those items, however, were being tested by other items, and with more apparent success. Those two items were not, therefore, retained.) Item 11

Table 5.2

Analysis of Items from the Pilot Testing

Item # and Answer	Objective Measured	"All" Difficulty Index		Discrimination Index
		Masters	Nonmasters	
1c	1.06a, 1.07a	0.93	---	---
2b	1.06a, 1.07a	0.93	---	---
3e	1.07, 1.08	0.93	---	---
4	1.04	0.93	---	---
5a	1.06b, 1.07a	0.93	---	---
6c	1.06b, 1.07a	1.00	---	---
7d	1.06c	0.57	---	---
8a	1.08	0.93	---	---
8b	2.09	0.57	1.00	- 0.43
8d	3.05,3.17	0.71	0.14	+ 0.57
9c	2.10	0.66	0.00	+ 0.66
10d	2.10	0.83	1.00	- 0.17
11c	3.05	0.33	0.40	- 0.07
11d	2.14	0.83	0.00	+ 0.83
12a	1.07	0.50	0.00	+ 0.50
12e	2.11	0.83	0.00	+ 0.83
13e	2.11	0.84	1.00	- 0.16
14e	2.11	0.50	0.50	- 0.00

(table continues)

Item # and Answer	Objective Measured	"All" Difficulty Index		Discrimination Index
		Masters	Nonmasters	
15b	2.10, 2.15	0.50	0.00	+ 0.50
16a	2.15	0.91	1.00	- 0.09
17d	2.08	0.42	0.00	+ 0.42
18d	2.10, 2.15	0.67	0.50	+ 0.17
19	2.08	0.58	0.50	+ 0.08
20	2.08	0.75	0.50	+ 0.25
21d	2.15	0.16	0.50	- 0.34
22c	2.15	0.83	1.00	- 0.17
23c	2.14	0.66	0.50	+ 0.16
23d	3.05	0.55	0.00	+ 0.55
24a	3.07	0.67	0.20	+ 0.46
25a	3.06	0.88	0.00	+ 0.88
26b	3.17	0.88	0.44	+ 0.44
27b	3.07, 3.17	0.33	0.00	+ 0.33
28c	3.06	0.66	0.40	+ 0.26
29b	3.12	0.33	0.00	+ 0.33
30b	3.09d	0.66	0.40	+ 0.24
31c	3.06	0.33	0.00	+ 0.33
32b	3.06	0.88	0.40	+ 0.48
33a	3.07	0.88	0.60	+ 0.28

(table continues)

Item # and Answer	Objective Measured	"All" Difficulty Index		Discrimination Index
		Masters	Nonmasters	
34d	3.07	0.77	0.00	+ 0.77
35a	3.09e	0.44	0.40	+ 0.40
36d	3.15	0.55	0.20	+ 0.35
37c	3.05	0.44	0.00	+ 0.44
38e	4.07	1.00	0.00	+ 1.00
39d	4.07	0.25	0.00	+ 0.25
40d	4.08	0.75	0.10	+ 0.65
41c	4.05	0.50	0.20	+ 0.30
42d	4.05	1.00	0.30	+ 0.70
43d	4.08	0.50	0.20	+ 0.30
44c	4.08	0.75	0.20	+ 0.55
45b	4.08	0.75	0.70	+ 0.05
45e	3.12	1.00	0.40	+ 0.60

was retained because it was an item with correct responses from more than one level, and one response had a strong discrimation index associated with it. Item 21, was retained, because of the researcher's interest in seeing how a larger group of students might respond to it.

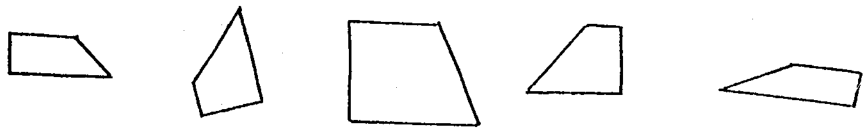
Written Responses

The examinees' written responses, describing "why" they chose their answers, were also studied. In most cases, when a "correct" answer was selected, the written response indicated reasoning at the van Hiele level associated with the response. Similarly, when an "incorrect" answer was selected, the written response indicated reasoning that was not compatible with the van Hiele level associated with the item. Mismatches did occur, however. Examples of these, as well as the research response to them, follows.

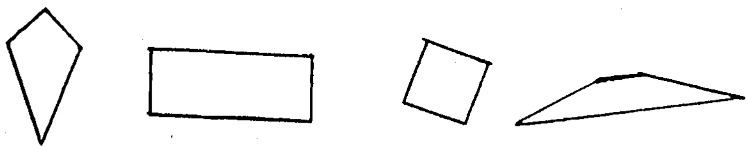
1. Correct reasoning leading to an answer choice designated as "incorrect":

In response to question 5 on the pilot, a Level 3 student selected "E" for her answer, rather than the answer choice "A" designated as the correct answer (Level 1) for this question.

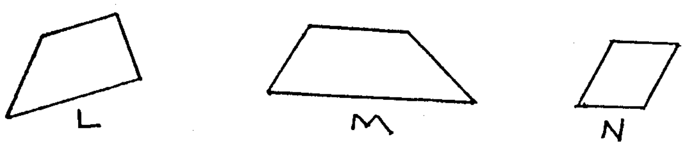
5. These are examples of a figure called a tetragon.



NONE of these figures is a tetragon.



Which of these appear to be a tetragon?



(A) L (1.06b, 1.07a)
(B) M
(C) N
(D) M and N
(E) L, M and N

In explaining why this choice was made, the student wrote:

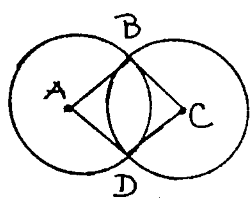
> A tetragon appears to be a figure that has four sides and is unsymmetrical. L, M & N could not be folded in half to fit perfectly.

The student identified properties of a tetragon and applied those properties correctly. Given the examples, her "definition" is correct. In order to avoid this unanticipated explanation, an example of a "non-tetragon" without line symmetry was included. (See Appendix F, draft instrument, item 6.)

2. Selecting a correct response using "inappropriate" reasoning:

For question 15, a Level 2 student selected the Level 2 answer, "B".

15. Two circles intersect in such a way that the figure ABCD is formed when the centers of the circles and the points of intersection are connected. AB=BC=CD=DA.



Which of the following could be used to show that BD is perpendicular to AC?

(A) Properties of a square
(B) Properties of a rhombus (2.10, 2.15)
(C) Properties of a rectangles
(D) Properties of a parallelogram
(E) None of these

Explain why you chose your answer:

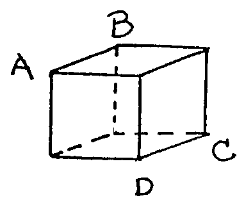
The student provided the following rationale.

> The diagonals of a rhombus connect opposite vertices of angles that are congruent. (I guessed)

In the interview which followed the testing, this student, and several others, indicated a lack of familiarity with several figures, including the rhombus, the kite and the trapezoid. Students often said they had heard of these figures but could not remember much about them, although the kite's picturesque name prompted students to be able to draw one. As a large number of students were to be tested at the next stage, thus (perhaps) increasing the likelihood that students were familiar with the topic, this item remained in the item pool for consideration. It did not, however, discriminate well with the field test examinees and it was not included on the final instrument.

Another example of "inappropriate" reasoning leading to a correct answer occured with item 26.

26. A cube is a 3-dimensional figure with 6 sides (faces), each of which is a square. The faces are perpendicular to each other. What would be the shape of the plane figure ABCD which results from cutting the cube through vertices A, B, C and D?



(A) Square
(B) Rectangle (3.17)
(C) Trapezoid
(D) Either A or B
(E) Not enough information

Explain why you chose your answer.

Intended to elicit Level 3 responses (informal deduction based on properties of a figure), this question consistently was answered by "appearance", a Level 1 response. For example, one student who selected answer "B", stated:

> I drew in the diagram, the figure ABCD and it appears to be a rectangle.

This type of response supported the concerns previously expressed by the experts. This item was not retained.

On the basis of the written responses, item 33 was not retained. The correct response for item 33, which was intended to correspond with Level 3 thinking, was chosen by every Level 2 student. (It is interesting to note that, correspondingly, the difficulty index for the "all nonmasters" on this item was very high, 0.60.) Regardless of interview mastery level, the "successful" students on this item all claimed to use the properties of a rectangle to make their decision. For example, one Level 2 student said "I chose (a) because they all have the properties of a rectangle." This student seemed to have no difficulty in "allowing" a square to also be a rectangle. Explanations for the uniformly high success rate might include the fact that the item was coupled with the wrong level, or that students had encountered the problem before and had memorized the answer. In any event, the item did not appear to be discriminating between masters of Level 3 and other masters, thus it was not retained.

Further Eliminations

Three additional items were dropped at this point, items 19, 29 and 42. Item 19 appeared equally attractive to masters and nonmasters. As item 20 was associated with the same level descriptor, and was apparently discriminating more effectively.

item 19 was not retained. Item 29 was dropped because, in reviewing all of the items for wording and clarity of meaning, the researcher felt the question was confusing to students. This occurred, in large part, because of the amount of reading which was required. Item 42 was dropped because it appeared to involved "word play", more than geometric thought.

Draft Instrument Items

Thirty-seven items were retained from the pilot test and assembled into a draft instrument for use in the subsequent field testing. (Table 5.3 indicates which items were retained.) There were 8 items corresponding to Level 1, 12 items corresponding to Level 2, 13 items corresponding to Level 3, and 7 items corresponding to level 4. Of the 37 items, 3 items had answers corresponding to two levels. All items were in the multiple-choice only format.

Future Research Settings

The academic range of the students used for the pilot study was also informative for the next stages of the study. The pilot phase demonstrated that students as young as the sixth grade could handle the multiple-choice format, read the instructions, follow directions, etc. As well, the spread of van Hiele levels demonstrated by the pilot group indicated that it would be important, during the next phase of the development of the

Table 5.3

Items Retained from the Pilot Study for the Draft Instrument

Number on the Pilot Instrument	Number on the Draft Instrument
1	1
2	2, 3
3	5
4	-
5	6
6	4
7	7
8	-
9	10
10	-
11	12
12	14
13	13
14	15
15	11
16	16
17	9
18	8
19	-
20	17
21	18
22	-
23	19
24	20
25	21
26	-
27	22
28	23
29	-
30	24
31	25
32	26
33	-
34	27
35	28
36	29

(table continues)

Number on the Pilot Instrument	Number on the Draft Instrument
37	30
38	32
39	31
40	34
41	35
42	-
43	36
44	37
45	33

instrument, to include students from the university and upper elementary school, in order to identify masters of the extreme levels.

Summary

The pilot study provided information about the discriminatory power of the 45 items in the revised item pool, about the structure of items, and about the range of academic settings from which to draw students in subsequent stages of the research. Using mastery level designations obtained from administering the Burger and Shaughnessy quadrilateral interview, an item analysis was conducted. As well, subjects were asked, for selected items, to describe the reasoning they used in answering an item. The 37 items which emerged from this stage were assembled into a draft instrument which was administered in the field test stage of the research. The next chapter, Chapter 6, describes the findings from the field testing.

Chapter 6

FIELD TEST STUDY

The goal of the field test phase was to have at its completion an instrument which could be used to assign an individual a van Hiele mastery level. To achieve this goal, 113 field test subjects, from sixth, tenth, eleventh and twelfth grade, as well as university, were administered the draft instrument. While each subject was also scheduled to participate in the quadrilateral interview, only 100 were able to attend. The results from the draft instrument and the interview were used to determine questions which would be used on the final instrument, to explore scoring schemes and to investigate the reliability of the decisions made by applying the scoring schemes to the final items. The findings associated with those decisions are discussed in this chapter.

Mastery Assignments

Interview Masters and Nonmasters

The Burger and Shaughnessy interview activities and analysis protocols for quadrilaterals were administered to 100 of the students who had participated in the testing using the draft instrument. A dominant van Hiele level was determined for 88 of them. As in the pilot study, this level was called the subject's "interview mastery level". Of the 12 remaining subjects, 9 had not yet mastered level one thinking, and 3 gave a range of responses from which no predominant level could be identified. These individuals are classified as "pre-Level 1" and "undecided", respectively. The distribution of the mastery assignments for the 100 subjects interviewed was:

Pre-Level 1	Level 1	Level 2	Level 3	Level 4	Undecided
9	24	22	21	21	3

The two groupings of subjects used for analysis at this stage were based on the interview mastery designations. These groups, the "all masters and nonmasters" and the "exact masters and nonmasters" are described in the following sections.

All Masters and Nonmasters

The "all masters and nonmasters" grouping scheme used the responses from all of the field test subjects. To be designated a master of Level n with this organization of the subjects, an individual had to be an interview master of Level n or any level higher. To be designated a nonmaster of Level n, an individual had to be an interview master of some level lower than Level n. This grouping is identical to the classification used with the item analysis which was conducted using the pilot subjects' responses (see Table 5.1).

<u>Exact Masters and Nonmasters</u>

With the "all" mastery grouping, a range of <u>interview</u> mastery (and nonmastery) levels is associated with each van Hiele level. This range could result in misleading or inflated results. For example, it might be possible for an item to have "all" indices associated with it which meet some minimum criteria. At the same time, however, the response patterns for the item, when just the interview masters at the level and those at the level immediately below are considered, might not reflect similar index strength. Specifically, an item at Level n might not discriminate between masters of Level n and masters of Level n-1, even though, when the "all masters and nonmasters" are considered, the item appears to do so. To counteract the distortion which might arise from using the blended "all" mastery group, a second criteria grouping, the "exact masters and nonmasters", was identified.

The "exact masters and nonmasters" grouping involved a subset of the interviewed subjects. Here, when Level n questions were investigated, the responses of the interview masters of that level, only, were considered as "masters' responses". Similarly, only the responses of the interview masters of Level n-1, the level immediately below Level n, were considered as "nonmasters responses" for Level n. As an example, when analysing Level 3 questions with this organization, the responses of the Level 3 interview masters, only, would be considered as the "masters' responses". The "nonmasters' responses" for Level 3, with this grouping, would be those of the interview masters for Level 2, only. This grouping of the subjects by adjacent interview mastery levels is referred to as the "exact masters and nonmasters". The relationship of "exact" masters and nonmasters to the interview masters is displayed in Table 6.1.

The distributions of the subjects, by mastery and nonmastery designations, for the "all" grouping and the "exact" grouping are presented in Table 6.2. For each level, the number of subjects in each grouping and the percentage of the group which that number represents are given. Only 9 pre-Level 1 subjects were identified. Thus the nonmasters of Level 1 group does not meet the minimum sample size required to control Type I and Type II errors at a level of α = .05 and β =.20, respectively. This is a limitation of the study.

Item Analysis

An item analysis was conducted in order to judge whether or not each item tended to differentiate between masters and nonmasters. Difficulty indices and discrimination indices, the same as those used in the pilot study, were calculated for each item. As well, an additional discrimination index, ϕ , the Pearson product-moment correlation for dichotomous data was calculated. The findings using each measure are discussed in the following sections.

Table 6.1

Exact Masters And Nonmasters Designation, By Level

	Interview Mastery Designations				
Exact Mastery Designations	Pre-1	1	2	3	4
Level 1					
Exact masters		X			
Exact nonmasters	X				
Level 2					
Exact masters			X		
Exact nonmasters		X			
Level 3					
Exact masters				X	
Exact nonmasters			X		
Level 4					
Exact masters					X
Exact nonmasters				X	

Table 6.2

Number (%) of Subjects Classified at Each van Hiele Level, for each Mastery Grouping

Level	Masters	Nonmasters
	All Grouping	
1	88 (91%)	9 (9%)
2	64 (66%)	33 (34%)
3	42 (43%)	55 (57%)
4	21 (22%)	76 (78%)
	Exact Grouping	
1	24 (73%)	9 (27%)
2	22 (48%)	24 (52%)
3	21 (49%)	22 (51%)
4	21 (50%)	21 (50%)

Item Difficulty Indices

Using the two groupings of subjects, "all masters and nonmasters" and "exact masters and nonmasters", difficulty indices were calculated. Each item had 4 difficulty indices associated with it: all masters, all nonmasters, exact masters and exact nonmasters. These indices are presented, by level, in Tables 6.3, 6.4, 6.5, and 6.6.

The difficulty indices were used to identify questions where masters tended to select correct answers and, simultaneously, nonmasters tended to select incorrect answers. The criteria used to identify these items were a difficulty index for both types of masters, all and exact, which was greater than 0.60 and a difficulty index for both types of nonmasters, all and exact, which was less than .50. For a given item, this corresponded to masters selecting a correct answer more than 60% of the time. Correspondingly, the cutoff for nonmasters indicated that they selected the correct answer less than 50% of the time.

Discrimination Indices

Two types of discrimination indices were calculated. The first is defined as the difference between the difficulty indices for masters and nonmasters. This is the same discrimination index

Table 6.3

Item Analysis Results, Level 1

	Difficulty Indices		Discrimination Indices	
Item	Masters	Nonmasters	Difficulty Index Difference	ϕ
		Exact Masters and Nonmasters (n = 33)		
1	.88	.66	.22	.24
2	.95	.77	.18	.06
3	.81	.77	.04	-.07
4	.92	.66	.26	.31
5	.83	.66	.17	.18
6	.96	.88	.08	.28
7	.29	.33	-.04	-.04
		All masters and nonmasters (n = 97)		
1	.94	.66	.28	.29
2	.94	.77	.17	.19
3	.86	.77	.09	-.07
4	.95	.66	.29	.32
5	.91	.66	.25	.22
6	.96	.88	.08	.24
7	.59	.33	.26	.17

Table 6.4

Item Analysis Results, Level 2

Item	Difficulty Indices: Masters	Difficulty Indices: Nonmasters	Discrimination Indices: Difficulty Index Difference	Discrimination Indices: ϕ
		Exact Masters and Nonmasters (n = 46)		
8	.86	.58	.28	-.04
9	.27	.16	.11	.13
10	.77	.29	.48	.48
11	.32	.29	.04	.03
12(3)	.29	.23	.06	.06
12(2)	.68	.33	.35	.35
13	.73	.25	.48	.48
14(2)	.77	.37	.40	.40
14(1)	.46	.44	.02	.02
15	.18	.21	-.03	.03
16	.91	.29	.62	.62
17	.91	.50	.41	.44
18	.09	.08	.01	.01
19(3)	.48	.05	.43	.49
19(2)	.40	.42	-.02	-.01

(table continues)

	Difficulty Indices		Discrimination Indices	
Item	Masters	Nonmasters	Difficulty Index Difference	ϕ
		All masters and nonmasters ($\underline{n}$ = 97)		
8	.89	.55	.34	.39
9	.55	.24	.31	.29
10	.75	.33	.42	.40
11	.56	.27	.29	.25
12(3)	.48	.20	.28	.29
12(2)	.78	.39	.39	.38
13	.77	.24	.53	.50
14(2)	.77	.24	.53	.45
14(1)	.73	.44	.39	.18
15	.45	.24	.19	.21
16	.92	.33	.59	.62
17	.92	.50	.41	.57
18	.39	.16	.23	.28
19(3)	.55	.13	.42	.47
19(2)	.64	.29	.35	.23

Note. Item numbers followed by a parenthesis had responses which were appropriate for two different levels. The levels are indicated by the numeral in the bracket.

Table 6.5

Item Analysis Results, Level 3

	Difficulty Indices		Discrimination Indices	
Item	Masters	Nonmasters	Difficulty Index Difference	ϕ
		Exact Masters and Nonmasters (n = 43)		
20	.76	.41	.35	.36
21	.71	.05	.66	.66
22	.33	.09	.24	.29
23	.43	.22	.21	.21
24	.71	.36	.35	.35
25	.19	.00	.19	.32
26	1.00	.68	.32	.43
27	.81	.14	.67	.67
28	.62	.23	.39	.40
29	.76	.36	.40	.40
30	.19	.18	.01	.01

(table continues)

Item	Difficulty Indices		Discrimination Indices	
	Masters	Nonmasters	Difficulty Index Difference	ϕ
		All masters and nonmasters ($\underline{n}$ = 97)		
20	.79	.20	.59	.58
21	.76	.09	.67	.67
22	.52	.16	.36	.38
23	.62	.27	.35	.35
24	.83	.23	.60	.59
25	.33	.14	.19	.22
26	.98	.41	.57	.58
27	.79	.16	.63	.62
28	.62	.14	.48	.49
29	.81	.25	.56	.55
30	.36	.18	.18	.20

Table 6.6

Item Analysis Results, Level 4

	Difficulty Indices		Discrimination Indices	
Item	Masters	Nonmasters	Difficulty Index Difference	ϕ
		Exact Masters and Nonmasters (n = 42)		
31	.38	.05	.33	.40
32	.90	.38	.52	.55
33(4)	.90	.52	.38	.42
33(3)	.76	.54	.22	.22
34	.67	.38	.29	.29
35	.38	.19	.19	.21
36	.76	.43	.33	.34
37	.81	.38	.43	.44
		All masters and nonmasters (n = 97)		
31	.38	.03	.35	.48
32	.90	.21	.69	.60
33(4)	.90	.35	.55	.45
33(3)	.88	.37	.51	.50
34	.67	.19	.48	.42
35	.38	.14	.24	.24
36	.76	.21	.55	.49
37	.81	.25	.56	.48

Note. Item numbers followed by a parenthesis had responses which were appropriate for two different levels. The levels are indicated by the numeral in the bracket.

as used with the pilot subjects' responses. The second discrimination index was the Pearson Product-moment correlation coefficient for dichotomous data, phi (ϕ).

The Pearson product-moment correlation for dichotomous data, ϕ_J is a measure of the association between two variables, each of which can be designated in a "yes" or "no" fashion. At this stage in the research, phi was used to explore the relationship between the mastery assignments (masters/nonmasters) and answer selection for each item (correct/incorrect). A contingency table, such as the one below, was used to organize the information.

	Number of Nonmasters at Level n (0)	Number of Masters at Level n (1)	Totals
Number of subjects who Selected the Level n response (1)	a	b	a + b
Number of Subjects who did not select the Level n response (0)	c	d	c + d
Totals	a + c	b + d	a + b + c + d

Using the notation from the contingency table,

$$\phi = \frac{bc - ad}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

Phi ranges in value from +1.00 to -1.00. The value of 1 can only be obtained when a = d = 0. With the variables defined as they are in the contingency table, a value of 1 would mean that all the masters of Level n selected the Level n response, and only the Level n masters selected that response. The value of -1 can be obtained only if the distribution of the two assignments is reversed, with all the nonmasters of Level n and only the nonmasters of Level n selecting the Level n response, (i.e., b = c = 0). No relationship between mastery level and answer selection is reflected in a value of $\phi = 0$.

Each item had four discrimination indices associated with it: a difference discrimination index for the "all" grouping, a difference discrimination index for the "exact" grouping, ϕ for the "all" group, and ϕ for the "exact" group. These indices are presented by level, in Tables 6.3, 6.4, 6.5, and 6.6. A minimum index value of 0.25, for all the discrimination indices, was used to identify items with the potential for discriminating. This reflected a positive relationship, with masters answering the item proportionally more successfully than nonmasters.

Interpreting the Difficulty and Discrimination Indices Collectively

In order for an item to be considered for the final test, it had to simultaneously meet, for both grouping of subjects, "all masters and nonmasters" and "exact masters and nonmasters", the minimum criteria for all the indices:

(a) a difficulty indices for masters which was greater than 0.60,

(b) a difficulty indices for nonmasters which was less than 0.50, and

(c) discriminatory indices which were greater than or equal to 0.25.

Discrimination Findings

Evidence of discrimination, as defined by the decision criteria, was observed with draft instrument items associated with Levels 2, 3 and 4. Specifically, analysis of the responses indicated that (a) questions 10, 12, 13, 14, and 16 appeared to discriminate between masters and nonmasters of Level 2, (b) questions 20, 21, 24, 27, 28, and 29 appeared to discriminate between masters and nonmasters of Level 3 and (c) questions 32, 34, 36 and 37 appeared to discriminate between masters and nonmasters of Level 4. None of the items intended for Level 1 met the decision criteria. For each Level 1 item, both masters and nonmasters of the level were consistently successful.

This section discusses the decisions concerning the identification of discriminating items. First, a general discussion is presented. Then, the decisions relating to the items which had responses associated with two levels is discussed.

<u>Item discrimination</u>

A review of the items identified as meeting the minimum difficulty and discrimination indices criteria reveals that the statistics associated with the items are stronger than the minimal criteria. This section discusses the decisions made on the basis of those index values.

(1) All of the questions, except #28, had a difficulty index for masters (of both types) greater than or equal to 0.67. This value for the index can be interpreted to mean that two-thirds or more of the masters at the level corresponding to the question selected the correct answer. Question 28, on the other hand, had a master's difficulty index of 0.62 with both mastery classifications. This can be interpreted to mean that slightly fewer than two-thirds of the masters chose the correct answer to this question. On the basis of this statistic, question 28 may appear to be a less desirable question than the others. Taking into consideration, however, the low value of its nonmasters

difficulty indices, (0.23 and 0.14 for the "exact" and the "all" groups, respectively), and the associated strong discriminatory indices (0.39 and 0.40 for the "exact" group and 0.49 and 0.49 for the "all" group), this question was nevertheless retained.

(2) All the questions had a difficulty index for nonmasters equal to or less than 0.43. When only the results from the "all" subjects are considered, lower values, and thus, more desirable values, of the nonmasters difficulty indices occur. Every one of the 15 questions has an "all" nonmasters' difficulty index below 0.40. Indeed, 12 of the 15 questions, including all of the Level 3 and Level 4 questions, have nonmaster difficulty indices which are less than 0.30. If students were randomly selecting answers, the expected difficulty index for each question would be 0.20. Thus, the questions are demonstrating nonmaster difficulty indices in a desirable range.

(3) The discrimination index was generally larger when calculated from the results of the entire group, than when calculated from the subgroups. As confidence in statistics usually increases the larger the sample size, this trend appears to be desirable. This must, of course, be considered in light of the composition of the groups, and the discrepencies, especially inflations, which might result. Indeed, this is just why the "exact masters and nonmasters" group was identified.

Items with level responses from different levels

For items with responses at two levels, separate indices were calculated for both responses. The same groupings of subjects. "all" and "exact", were used. Masters and nonmasters were designated relative to the intended level of each response. For the higher level response, this meant that the calculations were completed as if there were no other "correct" responses. i.e. similar to the other questions on the instrument. For the calculations involving the lower level response, however, a subject who selected either level response was considered to have answered "correctly". This meant that for the lower level, 2 of the 5 responses were correct. Each two-level question, then, had 16 indices, eight for each level.

In order for these items to be considered as discriminating for both levels, the minimum criteria for the indices had to be met for each response. No item met the criteria for both levels.

On the basis of the item analysis indices, however, items 12 and 14, were retained and associated with a single level. For item 14, only the higher level response, the Level 2 response, met the discrimination criteria. On the final instrument, therefore, it was treated as if it had only one correct answer, the Level 2 response. Question 12 was somewhat more problematic. It appeared to discriminate for Level 2 when the Level 3 choice was also considered correct. Counting either of two answer choices as corrrect increases the likelihood of guessing the correct answer from 0.20 to 0.40. Nonetheless, the item was retained, and both answer choices were scored as correct. Subsequently, the response patterns of the subjects on the final test were used to further analyse the appropriateness of this item.

Choice Response Analysis

A choice response analysis was conducted to evaluate the response patterns of the distractors. For each item, the response patterns were organized by interview mastery levels, as shown below.

Item	Interview Mastery Level	Response Choices A	B	C	D	E	Omits	Total
#	4	*	*	*	*	*	*	***
	3	*	*	*	*	*	*	***
	2	*	*	*	*	*	*	***
	1	*	*	*	*	*	*	***
	none	*	*	*	*	*	*	***
	TOTAL	***	***	***	***	***	***	***

The number of individuals at each mastery level choosing each response was tabulated. From those, the total number of responses per answer choice was then obtained. For the Level 1 items, some distractors were not chosen by any of the subjects. This corresponds to the high success rate demonstrated by the subjects on these items...most examinees selected the correct response. For the higher level items, however, every distractor was chosen. The range of responses appeared to be attractive to the examinees. No distractors were changed.

The Final Instrument Items

The items selected for the final test from the draft instrument are grouped, by level, in Table 6.7. The 15 questions identified as meeting all of the item analysis criteria are included. Of these, 5 questions are associated with Level 2, 6 questions are associated with level 3, and 4 questions are associated with Level 4. In addition, the four Level 1 questions associated with the strongest discrimination indices from the field testing were also included on the final instrument. Although those Level 1 items did not met the research criteria, they were included for two reasons. The first was that as all subjects should do well on these items, encountering them at the beginning of the test might help students gain confidence in the testing environment. Secondly, poor performance on these items might serve as an indicator to a researcher that something went awry...students misunderstood directions, students were unfamiliar with the topic, items had been miskeyed, etc.

The level descriptors associated with each item were also reviewed to determine how representative the items at each level were. The range of level descriptors to which the final items

Table 6.7

Final Items: Level Descriptors and Draft Item Number

Descriptor	Item Number: Draft	Item Number: Final
1.06a, 1.07a	1	1
1.06a, 1.07a	2	2
1.06b, 1.07a	4	3
1.07, 1.08	5	4
2.10	10	5
2.14	12	6
2.11	13	7
2.11	14	8
2.15	16	9
3.07	20	10
3.06	21	11
3.09d	24	12
3.07	27	13
3.09e	28	14
3.15	29	15
4.07	32	16
4.08	34	17
4.08	36	18
4.08	37	19

corresponded was quite narrow (see Table 6.7). In all cases, the items were only associated with the "applied" desriptors for a level. Within that subcategory of descriptors, at any level, only a few descriptors were associated with final test items. This lack of representativeness may stem from the restrictions associated with a multiple-choice format, for it does not allow students to initiate activity. Or, it may indicate a weakness in the item development stage, i.e. items which correspond to the broad range of descriptors can be written but were not generated in this instance.

Once the items for the final instrument were identified, the field test subjects' responses on those items were used to establish a scoring scheme and to investigate the reliability associated with the responses to the instrument. The decisions corresponding to the selection of a scoring scheme are presented in the next section. That is followed by a discussion on reliability.

Interpretation Scheme

As the items on the van Hiele Quadrilateral Test can be grouped into four level specific subtests, raw scores for each subtest can be reported, as can a total raw score. In considering each organization as a possible base for making mastery decisions, questions such as the following were addressed: what mastery levels might be associated with the total scores?, what meaning might be associated with the raw scores from each subtest?, and how might each subtest score contribute to a final mastery decision? These issues were explored by comparing the field test subjects' performances on the 19 questions selected for the van Hiele Quadrilateral Test to their interview mastery designations.

Using the Total Raw Score

In order to develop a scoring scheme which would associate a mastery level assignment with a total score, the relationship between the field subjects' total scores on the van Hiele Quadrilateral Test and their known interview mastery levels was investigated. The distribution of the subjects by level and raw score performance is presented in Table 6.8. To measure the linear association between the two variables, Pearson's product moment correlation coefficient, r_{xy}, was calculated. For the field test data, $r_{xy} = 0.86$.

The strength of r_{xy} suggested that a linear line of best fit might be considered for predicting mastery levels from total raw scores. The linear regression line corresponding to the field test data was

$$\hat{Y} = 0.2532X - 0.5376,$$

where X is a raw score and Y is a van Hiele mastery level. Using this equation, the following associations between raw scores and mastery levels were obtained:

Table 6.8

Field Test Subjects' Interview Mastery Level and Raw Score Performance on the Nineteen Final Test Items

	Interview Level					
Raw Score	Pre-1	1	2	3	4	Total
3	1					1
4	1	2				3
5	2	6				8
6	1	7	2			10
7		3	2			5
8	3	4	3			10
9		1	2	1		4
10	1	1	2	1		5
11			5	2	1	8
12			4			4
13			2	5	1	8
14				3	3	6
15				1	5	6
16				5	2	7
17				3	3	6
18					3	3
19					3	3
Total	9	24	22	21	21	97

Note. Mean Raw Score = 10.88

raw scores (X)	0 - 4	5 - 8	9 - 11	12 - 15	16 - 19
predicted level (Y)	pre-1	1	2	3	4

Using this scale, the predicted mastery level for each field test subject was compared with their interview mastery designation. The distribution of those assignments is presented in Table 6.9. The percentages of the interview masters who were classified the same with the two techniques were:

pre-Level 1	Level 1	Level 2	Level 3	Level 4
22%	83%	41%	43%	52%

Overall, 52% of the subjects were classified the same using both techniques.

Using the Subtest Scores

Scoring subtests

The other scoring option considered was that of assessing each subtest separately, then combining those results. This required identifying a cutoff score for each subtest, where a cutoff score is the minimum number of items in a subtest which an examinee must answer correctly to be classified as "successful" on the subtest. For each subtest, a range of cutoff scores was investigated. The lowest cutoff score examined was 2. The highest cutoff score considered was the total number of questions in the subtest, that is, a perfect subtest score.

Table 6.9

Field Test Subjects' Mastery Level Designations by Interview and by Raw Score Prediction

	Predicted Mastery Level					
Interview Mastery Levels	Pre-1	1	2	3	4	Totals
Pre-1	2	6	1	-	-	9
1	2	20	2	-	-	24
2	-	7	9	6	-	22
3	-	-	4	9	8	21
4	-	-	1	9	11	21
Totals	4	33	17	24	19	97

Two statistics were used to explore the cutoff scores, the correlation coefficient, ϕ , and the correlation ratio, $\eta^2_{Y,X}$. Each of these statistics measured an aspect of the relationship between the "success" assignments resulting from the application of the cutoff scores and the known interview mastery assignments. The correlation coefficient, ϕ , measured the correlation between the mastery grouping and the success status on the subtest. The values of ϕ range from 1 to -1. A positive value of ϕ indicated that masters of the level were succeeding on the subtest, and that nonmasters were not. A negative value of ϕ indicated that masters of the level were not succeeding and that nonmasters of the level were succeeding.

The correlation ratio, $\eta^2_{Y,X}$, measured the proportion of the total variation in the mastery designations (Y) attributed to the variance in the "success" and "nonsuccess" of the subjects (X) on a level subtest. It is "a measure of the extent to which Y is predictable from X by a 'best-fitting' line that may be either straight or curved" (Glass and Stanley, 1970, p. 151). That line passes through the mean of the Y values for each value of X. In this case, with just two X values, the line is straight.

The correlation ratio has the following definitional form:

$$\eta^2_{Y,X} = 1 - \frac{SS_{within}}{SS_{total}}$$

where SS_{total} is the sum of squared deviations of each Y score from the mean of all Y scores and SS_{within} is the "sum of the squares within" for a one-factor analysis of variance with unequal n's.

To explore the effects of the range of cutoff scores on the performances of groups of different sizes and compositions, the two coefficients were calculated for both item analysis mastery groupings, "all masters and nonmasters" and the smaller subset of "exact masters and nonmasters". The resulting statistics for the "all" grouping are presented in Table 6.10. The resulting statistics for the "exact" grouping are presented in Table 6.11. Those subtest statistics are discussed, level by level, below.

Level 1. Cutoff scores of 2, 3 and 4 were applied to the 4 items on this subtest. Each cutoff corresponded with weak correlation coefficients and correlation ratios. This trend is consistent with the nondiscriminating nature of these Level 1 subtest items, as indicated in the item analysis. Most examinees were successful with these questions, regardless of their van Hiele mastery level. As no cutoff score performed strongly, the cutoff of 3 was selected for use with this subtest. Using this, rather than a cutoff of 4, allowed for some measurement error.

Level 2. Three was chosen as the cutoff score for this five item subtest. For both mastery groupings, the strongest values of the coefficients occured with 3 as the cutoff point. The

Table 6.10

Cutoff Score Statistics, All Masters and Nonmasters

	Subtest			
Index	Level 1	Level 2	Level 3	Level 4
		Cutoff of 2		
ϕ	0.3191	0.5275	0.6317	0.5806
$\eta^2_{y,x}$	0.1019	0.2783	0.3991	0.3371
		Cutoff of 3		
ϕ	0.3228	0.7906	0.7883	0.6486
$\eta^2_{y,x}$	0.1041	0.6251	0.6213	0.4206
		Cutoff of 4		
ϕ	0.2089	0.6384	0.7780	0.6409
$\eta^2_{y,x}$	0.0434	0.4075	0.6049	0.4107
		Cutoff of 5		
ϕ	------	0.6384	0.6561	-----
$\eta^2_{y,x}$	-----	0.4075	0.4305	-----
		Cutoff of 6		
ϕ	-----	-----	0.4502	-----
$\eta^2_{y,x}$	-----	-----	0.2026	-----

Note. *n* = 97

Table 6.11

Cutoff Score Statistics, Exact Masters and Nonmasters

Index	Subtest Level 1[a]	Level 2[b]	Level 3[c]	Level 4[d]
		Cutoff of 2		
ϕ	0.2887	0.5043	0.5728	0.4216
$\eta^2_{y,x}$	0.0833	0.2543	0.3281	0.1778
		Cutoff of 3		
ϕ	0.2406	0.7424	0.6949	0.4767
$\eta^2_{y,x}$	0.0579	0.5511	0.4828	0.2273
		Cutoff of 4		
ϕ	0.3218	0.5963	0.7879	0.4877
$\eta^2_{y,x}$	0.1036	0.3555	0.6208	0.2375
		Cutoff of 5		
ϕ	-----	0.5151	0.6001	-----
$\eta^2_{y,x}$	-----	0.2654	0.3601	-----
		Cutoff of 6		
ϕ	-----	-----	0.3278	-----
$\eta^2_{y,x}$	-----	-----	0.1074	-----

[a] $n = 33$
[b] $n = 46$
[c] $n = 43$
[d] $n = 42$

correlation coefficients were strongly positive at 0.79 (all) and 0.74 (exact). The correlation ratios, 0.63 (all) and 0.55 (exact), indicated that the source of most of the variation in the mastery assignments was attributable to the success assignments.

<u>Level 3</u>. Two potential cutoff scores emerged for the six item Level 3 subtest, a cutoff of 3 and a cutoff of 4. For the all grouping, the largest values of the coefficients occurred when 3 was the cutoff score. For the exact grouping, the largest values of the coefficients occurred when the cutoff score was 4. For both groupings, however, the values associated with the statistics which resulted from using either cutoff point were strong. On the basis of this comparability, both cutoffs scores were selected for use with this subtest.

<u>Level 4</u>. For this four item subtest, for each grouping, the statistics associated with the cutoff scores of 3 and 4 were quite similar. (They were also stronger than those for the cutoff of 2.) In the "all" group, the correlation coefficient, ϕ , for cutoffs of both 3 an 4, was moderately positive. The correlation ratios for the same cutoffs, however, are a change from the previous subtests. Here, the proportion of the variance in the mastery assignments associated with the success assignments, with both cutoffs, is slightly below 0.50. For the "exact′ group, statistics similar to the "all" group, but weaker, were obtained. In choosing between the two stronger cutoffs, 3, rather than 4, was selected for this subtest. This allowed for some measurement error.

Based on the level by level cutoff performances, two sets of cutoff scores, differing only at Level 3, emerged for the subtests. These were 3, 3, 3, 3 and 3, 3, 4, 3 for the Level 1, 2, 3 and 4 subtests, respectively. For the Level 2 and 3 subtests, the cutoffs scores were associated with strong measures of relationship between performance (success/nonsuccess) on the subtest and mastery/nonmastery of the level with which it was associated. Weaker associations existed with the Level 4 cutoff. As no cutoff emerged as strong for Level 1, the highest cutoff, without requiring a perfect performance was selected.

<u>Assigning mastery levels from subtests</u>

Once the subtest success criteria were determined, a means of converting a subject's subtest performances into a mastery level designation was sought. Two approaches were considered. The first designated the level of the "highest" subtest an examinee successfully completed as that subject's van Hiele mastery level. This designation was made regardless of how the subject performed on any lower levels. With the second technique, the level assignment was based on a pattern of sequential successes for the subtests. Using this sequential approach, the mastery level was the level of the "highest" subtest for which the success criteria had been met <u>and</u> for which all the lower level subtests had also been answered successfully. A subject, for example, who successfully answered subtest 1, 2 and 4, would be designated as a master of Level 4 by the first (highest) technique but only as a master of Level 2 by the second (sequential) technique.

The distribution (in percent) of the subjects whose subtest classifications were identical to their interview mastery classifications is given in Table 6.12. Overall, using the highest subtest and the 3, 3, 3, 3 and 3, 3, 4, 3 cutoffs, 62% and 66% of the subjects were classified the same as their interview designations. Using the highest sequential subtest with both the 3, 3, 3, 3 and the 3, 3, 4, 3 cutoffs, 67% of the subjects were classified in each case the same as their interview mastery designations. The complete set of distributions by interview designation and both of the subtest scoring schemes is presented in Tables 6.13 and 6.14.

Level by level, the percentages of assignments which resulted in mastery designations identical to the interview designations, are similar for the two subtest techniques. This would happen if the highest subtest each subject successfully answered was consistently the highest subtest in a sequence of successfully answered subtest, i.e., the subtests on which a subject is successful form a sequence.

To test whether or not the successful response patterns on the subtests form a sequence, the Guttman Scalogram Analysis (Guttman, 1944) was used. The response pattern of each subject, by subtest, was described using a 1 x 4 vector, where the first

Table 6.12

Distribution (%) of Subjects With Identical Mastery Assignments from the Interview and from a Subtest Scoring Scheme

Level	Subtest Scoring Scheme			
	Highest Subtest		Highest Sequential Subtest	
	3,3,3,3 cutoffs	3,3,4,3 cutoffs	3,3,3,3 cutoffs	3,3,4,3 cutoffs
Pre-1	0	0	33	33
Level 1	67	75	75	75
Level 2	64	86	64	86
Level 3	67	52	66	52
Level 4	76	76	76	67

Table 6.13

Distribution of Mastery Level Designations, Interview and "Highest" Subtest Interpretation Scheme

Mastery Levels from Interviews	Mastery Level Designations from Subtest: Pre-1	1	2	3	4	Totals
	3, 3, 3, 3 cutoff criteria					
Pre - 1	0	6	2	1	0	9
1	2	16	4	2	0	24
2	0	2	14	5	1	22
3	0	0	1	14	6	21
4	0	0	0	5	16	21
Total	2	24	21	27	23	97
	3, 3, 4, 3 cutoff criteria					
Pre - 1	0	6	2	1	0	9
1	2	18	4	2	0	24
2	0	2	19	0	1	22
3	0	0	4	11	6	21
4	0	0	3	2	16	21
Total	2	26	32	14	23	97

Table 6.14

Distribution of Mastery Level Designations, Interview and "Highest Sequential" Subtest Interpretation Scheme

Mastery Levels from Interviews	Mastery Level Designations from Subtest					
	Pre-1	1	2	3	4	Total
	3, 3, 3, 3 cutoff criteria					
Pre - 1	3	5	1	0	0	9
1	3	18	3	0	0	24
2	1	2	14	5	5	22
3	0	1	1	14	5	21
4	0	0	0	5	16	21
Totals	7	26	19	24	21	97
	3, 3, 4, 3 cutoff criteria					
Pre - 1	3	5	1	0	0	9
1	3	18	3	0	0	24
2	1	2	19	0	0	22
3	0	1	5	11	4	21
4	0	0	4	3	14	21
Total	6	26	32	14	18	97

position represented Level 1, the second position represented Level 2, etc. Meeting the success criteria for a subtest was indicated by placing a "1" in the subtest position; not meeting the success criteria for a subtest was indicated by entering a "0" in the subtest position. For example, the vector (1, 1, 0, 1) represents an examinee who met the success criteria for Levels 1, 2 and 4, but did not meet the criteria for Level 3. This subject is said to have one error, because one success (at Level 3) is required to form an unbroken sequence. A subject with the response pattern, (0, 0, 1, 0), has 2 errors because a one in the first position and a one in the second position are required to form an unbroken sequence.

Using the performances of all the subjects in the field testing, represented in vector form, the coefficient of reproducibility (Rep) was calculated.

$$\text{Reproducibility Coefficient} = 1 - \frac{\text{Total number of errors}}{\text{number of subjects x vector magnitude}}$$

For this data, the index reflected the likelihood with which a subjects' success pattern on the subtests could be reproduced from knowing only the highest subtest on which the subject was successful. A value of 1.00 indicates all subjects performed in perfect sequences. It has been suggested that the minimum reproducibility coefficient associated with sequential response patterns is 0.90 (Mayberry, 1981, p. 13).

The values of the reproducibility coefficient, calculated from the field test subject's performances, as determined by the two cutoff schemes, 3, 3, 3, 3 and 3, 3, 4, 3, were both .98. This implies that for both sets of success criteria, the majority of the subjects' responses formed an unbroken sequence. Therefore, the results from assigning levels from the two subtest techniques, highest subtest and highest sequential subtest, would be expected to be quite similar.

Interpretation Scheme for the Final Instrument

Based on a comparison of the percentage of subjects who were classified identically by the interview and by one of the scoring schemes, the subtest schemes performed with more accuracy than did the total raw score scheme. Of the two subtest interpretation schemes, the highest sequential subtest scoring scheme was chosen for the final instrument, despite the similarity in its performance to the highest subtest scheme. As the van Hiele model claims that an individual operating on Level n has mastered all the levels below that one (hierarchical and fixed sequence property), an underlying assumption of the evaluation process is that masters at Level n, while perhaps not preferring them, when confronted in a fixed response format where the only "correct" choice is from a "lower" level, will choose that response. Consequently, a master of Level n should demonstrate mastery of each lower level on the instrument.

Reliability

Unlike norm-referenced tests, where items are selected to produce a maximum of variation amongst examinees, criterion-referenced tests often result in little variation in scores. This is because criterion-referenced tests frequently contain questions, any one of which, the majority of examinees can answer. Therefore, reliability indices which are predicated on variability (those traditionally used with norm-referenced instruments) are not necessarily appropriate for criterion-referenced instruments (Popham and Husek, 1969).

For criterion-referenced tests in which mastery/nonmastery status is determined by a cutoff score, two types of reliability measures can be considered. The first type, threshold loss function, focus on the consistency of <u>the mastery decisions</u> across repeated forms or parallel forms of a test. The second type, squared-error loss function, focus on the consistency of <u>the test scores</u> across repeated forms or parallel forms of a test (Berk, 1984). With the latter, misclassification of students whose scores are far above or below the cutoff point are viewed as more serious than misclassifications from scores close to the cutoff (Berk, 1980). As this study is concerned with identifying mastery status, rather than degrees of mastery, the reliabily measures used belong to the threshold loss function family.

Two indices, the agreement coefficient and Cohen's Kappa coefficient, are used to discuss different aspects of threshold loss reliability (Berk, 1984). The first index focusses on the consistency of the classifications, regardless of the source of this consistency. The second index provides information about the degree of consistency gained by using the measurement procedure (Nitko, 1983). Both rely on two administrations of the instrument. Each, however, can be approximated from a single administration.

The Agreement Coefficient

The agreement coefficient gives the proportion of the examinees consistently classified as masters and nonmasters on two test administrations. The distribution of those mastery assignments can be represented in a contingency table such as the one shown here:

		Test One		
		Masters	Nonmasters	Totals
	Masters	a	b	a + b
Test Two	Nonmasters	c	d	c + d
	Totals	a + c	b + d	N

where
a = the number of examinees classified as a master on both administrations of the test,
b = the number of examinees classified as a nonmaster on the first test and a master on the second test,
c = the number of examiness classifed as a master on the first test and a nonmaster on the second test,
d = the number of examinees classified as a nonmaster

on both administrations of the test,
N = the total number of examinees in the group,
$a + b + c + d$

Using the designation from the contingency table, the agreement coefficient, p_o , is given by:

$$p_o = (a + d) / N$$

The upper bound of this coefficient is 1.00. This occurs when there is complete agreement between the assignment of masters and nonmasters, on both tests, for ALL examinees in the group. The lower bound of the coefficient is given by

$$p_{chance} = \frac{(a + b)(a + c) + (c + d)(b + d)}{N^2}$$

The lower bound "represents the proportion of consistent classifications expected by chance if mastery-nonmastery outcomes on the second administration were completely independent of outcomes on the first administration....p_{chance} will be greater than or equal to .50" (Subkoviak, 1988, p. 48).

The agreement index is affected by the cut-off score, the number of items on the test, and the mastery composition of the examined group. For a unimodal score distribution, the closer the cut-off is to the mean, the lower is p_o and vice-versa. (This tendency is not necessarily demonstrated with bimodal score distribution.) Increases in the value of p_o are associated with increases in the test length and with increases in score variability. Of these, the cut-off score has the most influence on p_o.

Cohen's Kappa Coefficient

Cohen's kappa coefficient "measures the test's contribution to the overall proportion of consistent classifications, that is, test consistency" (Berk, 1984, p. 241). Designated as κ, it is given by:

$$\kappa = (p_o - p_{chance}) / (1 - p_{chance})$$

where p_o and p_{chance} are defined as in the section above.

Kappa displays the following properties:

1. Kappa varies from 0 to 1, inclusively, with 1 indicating that outcomes from the two administrations of the test are identical and 0 indicating that the outcomes from the testing are completely independent of each other (Subkoviak, 1988).

2. Negative values of Kappa should be interpreted as 0 (Huynh, 1976).

3. Kappa increases as a function of test length.

4. Kappa is particulary responsive to test score variability (Huynh, 1976), and thus to the homogeneity of the

tested group. As variability increases, kappa increases and vice versa.

5. Kappa varies with the cutoff score, taking smaller values when the cutoff score is close to the extremes of the scoring range (Huynh, 1976).

Interpretation of the reliability indices

Little discussion occurs in the literature about which values of the agreement coefficient and Cohen's Kappa are appropriate for which functions (Subkoviak, 1988). Berk suggests, however, that p_o be used "where an absolute cut-off score is chosen and for other tests that may contain short subtests and/or yield low score variance" (1984, p. 243). He also indicates that the use of p_{chance} when calculating κ "make this index problematic" (Berk, 1984, p. 241) and urges caution in its use and interpretation. Subkoviak proposes that the indices be considered in context: how serious is the decision being made (for example, determining high school graduation or determining mastery of a unit of instruction) and what can "realistically be expected of a test" (Subkoviak, 1988, p. 51) given conditions such as time and test length. For teacher-made tests, used for relatively routine decisions and one period in length, Subkoviak (1988) suggests as minimal values, $p_o = .75$ and $\kappa = .35$. These are the decision criteria used for this research.

Calculations from a single administration of an instrument

Several methods for approximating the agreement coefficient and Cohen's Kappa from a single administration of a test have been proposed (Huynh, 1976; Subkoviak, 1976; Peng & Subkoviak, 1980.) In general, these techniques employ either complex statistical concepts or sophisticated computer software. Subkoviak (1988), however, has produced tables, based on the procedure developed by Peng and Subkoviak (1980), from which approximations of the agreement coefficient and the kappa coefficient can be read directly (see Appendix K).

To use the agreement coefficient or kappa coefficient tables, two instrument-based statistics are required: (1) a traditional reliability score such as Cronbach's alpha or the Kuder-Richardson Formulas 20 or 21 and (2) the raw cutoff score of the test, expressed as a standard score (z). For this research, Kuder-Richardson's Formula 20 (KR-20) was used because, consistent with the intent of the van Hiele Quadrilateral instrument, it treats answers as either right or wrong and it makes no assumptions about the relative difficulty of each item, within a level. The standard score, z, was calculated using the formula

$$z = (c - 0.5 - M)/\ S$$

where
c = raw cutoff score,
M = the mean of the scores,
S = the standard deviation of the scores.

The 0.5 value is "a correction for continuity" (Subvokiak, 1988, p. 49).

The responses of the subjects in the field test to the 19 questions selected for the final instrument were used to calculate the reliability coefficients with the Subkoviak approximation technique. Reliability statistics were calculated for each subtest using the success criteria determined previously, a cutoff score of 3 for each level, with Level 3 statistics also calculated for a cutoff score of 4. The reliability indices, calculated for each of the mastery grouping, "exact" and "all", are presented in Table 6.15.

The values of the statistics associated with the Level 2, 3 and 4 subtests, with one exception, meet the minimum criteria. For Level 2, and the "all grouping", the agreement coefficient, at 0.73, is slightly below the minimum research criteria of 0.75. The strength and consistency of the statistics associated with the Level 2, 3 and 4 subtests suggest that if these subtests were re-administered to this group of subjects, one could expect "success" patterns on each subtest to be similar to those already observed. They also suggest that the test is contributing to the consistency of the classifications. These are both desirable findings.

Table 6.15

Reliability Statistics from the Field Test

Index	Subtest Level (cutoff/total # of items)				
	1 (3/4)	2 (3/5)	3 (3/6)	3 (4/6)	4 (3/4)
			Exact Grouping		
p_o	0.73	0.83	0.76	0.76	0.86
κ	0.10	0.35	0.49	0.49	0.71
			All Grouping		
p_o	0.94	0.73	0.80	0.81	0.82
κ	0.08	0.40	0.59	0.58	0.46

Of the four statistics associated with the Level 1 subtest, only one meets the minimum criteria, the agreement coefficent for the "all" group. The same coefficient for the "exact" group, however, is just below the minimum criteria. This suggests that if this group of subjects rewrote the test, the distribution of masters and nonmasters at this level would be about the same. The very low values of Cohen's Kappa, however, suggest that little gain in consistency is realized by using the test, much beyond what would be expected by chance with a group of this composition. This might be explained by the fact that the ("known") interview mastery composition of the group indicates that 91% are masters of Level 1 or a higher level.

A limitation of calculating the reliability statistics associated with the instrument from the field testing subjects'responses is that their responses were also used to determine which questions would be selected for the instrument. This may appear to be a guarantee of obtaining a high reliability index. It is possible, however, that a collection of questions which <u>individually</u> discriminate between masters and non-masters, might not, when interpreted <u>collectively</u> differentiate between masters and nonmasters. Minimally, then, calculating reliability statistics for this group could provide information which would, if the statistics were low, indicate the case described above, i.e., that there is some question about the interpretation of the items when viewed collectively. If, however, as was the case with Levels

2, 3 and 4, the reliability statistics meet the minimum criteria, this would be additional support, though not conclusive, that the itmes, when viewed collectively, are functioning as intended.

Summary

This chapter has included a discussion of the findings associated with the administration of the draft instrument to 113 subjects. For 97 of those individuals, van Hiele mastery levels were determined, using the Burger and Shaughnessy interview on quadrilaterals. Comparisons of the examinees' performances on the draft instrument items and on the interviews resulted in the identification of 19 items for the van Hiele Quadrilateral Test. Grouped by level, 4 items corresponded with Level 1, 5 items corresponded with Level 2, 6 items corresponded with Level 3 and 4 items corresponded with Level 4. (The Level 1 items did not meet the minimum discrimination criteria; all other items did meet the minimum criteria.) An interpretation scheme for converting subtest performance into a mastery designation was selected. Reliability statistics were calculated for each subtest.

The final product of this stage was the van Hiele Quadrilateral Test. The next chapter presents a discussion of the findings associated with the administration of that instrument.

Chapter 7

FINAL TESTING

The 19 item van Hiele Quadrilateral Test, developed in the earlier stages of this research, was administered to 101 subjects, 50 students in the ninth grade and 51 students in the twelfth grade. Based on their performances, subjects were assigned a van Hiele mastery level, reliability statistics were calculated, the sequential nature of the subtest successes was explored and the success rates associated with each item were investigated. The relationship between the subjects' performances on the instrument and their grade membership was analysed. The relationship between the subjects' performances on the instrument and their performances on an external measure, the Basic Concepts Test of the Nova Scotia Achievement Test, was analysed. The findings from these investigations are discussed in this chapter.

Reliability

Two types of reliability indices, specific to criterion-referenced tests, were applied to the results obtained on the final instrument. These indices are the agreement coefficient (p_o) and Cohen's Kappa coefficient (κ). The first coefficient represents the proportion of examinees consistently classified on two administrations of a mastery test. The second coefficient quantifies the degree of consistency in assigning mastery and nonmastery status contributed by the measurement procedure, beyond the chance effects associated with the group's mastery composition (Nitko, 1983). As only one administration of the test was conducted with the final group of subjects, Subkoviak's (1988) approximation technique for p_o and κ, based on a single administration, was used. The nature of the interpretation scheme, i.e., considering the four subtests separately in order to determine a final level designation, meant that reliability statistics were calculated for each subtest, not for the test as a whole. It is therefore the consistency of the "success" decisions which is investigated, where success on a subtest was determined by answering correctly at least the number of items associated with the cutoff score. "Nonsuccess" meant the subject did not meet the cutoff score.

The performance of each subject, by subtest, is presented in Appendix L. Based on those scores, values of p_o and κ were calculated for each subtest. The consistency of success on each subtest was investigated for the combined group of ninth grade subjects and twelfth grade subjects and, separately, for each grade level (see Table 7.1). For the combined group, the minimum acceptable value for p_o of 0.75 was met for 3 subtests: Level 1, Level 3, when the cutoff of 4 is used, and Level 4. None of the corresponding values of κ, however, reached the minimum of 0.35 which Subkoviak proposed as acceptable. These statistics suggest that, while for some of the subtests the proportion of subjects who

Table 7.1

Reliability Indices, Agreement Coefficient (p_o) and Cohen's Kappa (κ), All Masters and Nonmasters, by Subtest

Reliability Index	Subtest Level (cutoff / # items per subtest)				
	1(3/4)	2(3/5)	3 (3/6)	3(4/6)	4(3/4)
	Combined Grade 9 and Grade 12				
p_o	0.95	0.63	0.67	0.77	0.81
κ	0.08	0.19	0.33	0.30	0.28
	Grade 9				
p_o	0.86	0.57	0.75	0.96	>0.96
κ	0.03	0.13	0.05	0.02	0.11
	Grade 12				
p_o	>0.96	0.72	0.68	0.68	0.63
κ	0.02	0.24	0.33	0.33	0.19

would be consistently successful is acceptable, the subtests are not contributing to the consistency of the success decisions, for these subjects, much beyond chance.

The pattern displayed by the combined group of subjects is repeated when the test results are analysed for each grade separately. The agreement coefficient for some of the subtests is greater than or equal to the minimum criteria: for the ninth grade subjects, at Level 1, Level 3 (both cutoffs) and Level 4, and for the twelfth grade subjects, at Level 1. Again, however, the values of κ for every subtest are less than the minimum acceptable for this research. The subtests do not appear to be contributing sufficiently to the overall consistency of the success classifications.

Sequential Nature of the Subtest Responses

To investigate whether or not the subjects' success patterns on the four subtests formed a sequence, the Guttman Scalogram Analysis technique (Guttman, 1944) was applied to the subtest performances. The resulting values for the coefficients of reproducibility are presented in Table 7.2. They are given for each interpretation scheme, for each grade level and for all the subjects.

Table 7.2

Coefficient of Reproducibility by Grade and by Interpretation Scheme

	Subtest scoring criteria	
Group	3, 3, 3, 3	3, 3, 4, 3
Grade 9	0.97	0.99
Grade 12	0.97	0.97
All Subjects	0.97	0.98

In each case, the reproducibility coefficient is greater than 0.90, the minimum value associated with a sequential pattern of responses. This implies that the pattern of successes on the subtests can be considered to form a sequence. Furthermore, these statistics indicate that there would be little difference between basing the mastery designation on the highest level subset a subject successfully answered or basing it on the highest level subtest, in a sequence, successfully answered.

Comparisons Between Grades

Subtest Findings

The correspondence between grade level and success on each subtest was also investigated. The distribution of subjects' successes for each subtest, by grade level, is shown is Table 7.3.

Table 7.3

Number of Subjects, by Grade, Successful on Each Subtest

	Subtest Level(cutoff/items per subtest)				
Grade	1(3/4)	2(3/5)	3(3/6)	3 (4/6)	4(3/4)
9	46	30	7	2	1
12	50	39	28	19	17

Note. *n* = 50 for each grade

A correlation coefficient (ϕ) and the correlation ratio ($\eta^2_{y,x}$) were calculated for each subtest using the subtest success status (success, nonsuccess) and grade (twelfth or ninth) information. The possible values for ϕ range from -1 to +1 with 1 indicating that all twelfth graders were successful and only twelfth graders were successful, -1 indicating that all ninth graders were successful and only ninth graders were successful and 0 indicating that there was no correlation between grade level and success status. The correlation ratio, $\eta^2_{y,x}$, ranges in value from 0 to 1. It was used to measure the proportion of the variation in the subtest success assignments (Y) which is attributed to the variance between the grade levels (X).

The values of ϕ and $\eta^2_{y,x}$ calculated using the responses from the subjects in the final testing phase of the research are presented in Table 7.4. The positive nature of ϕ indicates that the twelfth grade subjects met the success criteria proportionally more frequently than did the ninth grade subjects. The values obtained for the correlation coefficients for the Level 1 and Level 2 subtests, however, indicate that there is little correlation between grade membership and performance on these first two subtests. In fact, both groups were quite successful on these subtests, as Table 7.3 indicates. The values of $\eta^2_{y,x}$ for these same two levels, at close to 0.00, indicate that the proportion of the total variation in the performances on the subtests which is attributable to the variance in the grade levels is very small.

With the upper two levels, there is a stronger correlation, ϕ, between performance on the subtests and grade level. The corresponding values of $\eta^2_{y,x}$ indicate, however, that the proportion

Table 7.4
<u>Correlation Coefficients (ϕ) and ($\eta^2_{y,x}$) for Grade and Subtest Success</u>

Correlation Statistic	Subtest Level(cutoff/items per subtest)				
	1(3/4)	2(3/5)	3(3/6)	3 (4/6)	4(3/4)
ϕ	0.2041	0.1946	0.4403	0.4166	0.4166
$\eta^2_{y,x}$	0.0417	0.0379	0.1938	0.1736	0.1736

of the total variance in the performances on the subtests which is attributable to the variance between the grade levels, while greater than for Levels 1 and 2, is still not large.

Mastery Assignment Findings

Based on their subtest performances, each subject was assigned a van Hiele mastery level (see Appendix L). The distributions of the mastery levels, by grade, for each interpretation scheme are presented in Table 7.5. (Using the two sets of cutoff scores resulted in differences in the assignments of masters at Levels 2, 3 and 4. This is because the mastery assignments are based on a sequential pattern of successes at each level.) Using that data, the relationship between membership in a grade and mastery level was investigated.

As a measure of independence between the two variables, grade membership and mastery level, Chi squared (χ^2) was calculated. The null hypothesis was that membership in a grade and van Hiele level classifications were statistically independent. The resulting values for the the two interpretation schemes were $\chi^2 = 25.587$ for 3, 3, 3, 3 and $\chi^2 = 19.99$ for 3, 3, 4, 3, each with 4 degress of freedom. If the null hypothesis were true, the probability of χ^2 attaining either of these values is less than 0.01. ($\chi^2_{.01,4} = 13.277$) Thus, these values of χ^2 support the rejection of the null hypothesis.

Table 7.5

Assignments to Mastery Level by Grade and Interpretation Scheme

Group	Mastery Level Designation					
	Pre-1	1	2	3	4	Total
			3, 3, 3, 3 interpretation scheme			
Grade 9	4	18	24	4	0	50
Grade 12	0	11	14	11	14	50
Total	4	29	38	15	14	100
			3, 3, 4, 3 interpretation scheme			
Grade 9	4	18	26	2	0	50
Grade 12	0	11	21	7	11	50
Total	4	29	47	9	11	100

Some caution should be used in the interpretation of these statistics, however, as the expected frequencies for some of the cells is less than five. (The expected frequency for a cell may be determined by dividing the product of the cell's marginal totals by the total number of subjects.) Many statistical experts say that χ^2 should not be applied with cells smaller than five. Edwards, however, suggests that for a contingency table with more than 1 degree of freedom, if no more than 20% "of the expected numbers are less than 5, then a minimum expected number of 1 is allowable in using the χ^2 test of significance" (1973, p. 140).

The mastery designations for each grade were further analysed, level by level, using the correlation coefficient, ϕ. The variables were grade membership (twelfth or not twelfth, i.e, ninth) and mastery status for the level (master, nonmaster), when the "all masters and nonmasters" grouping was used. With this arrangement, values of ϕ close to one indicate a strong correlation between twelfth grade membership and mastery of a level. Values close to negative one indicate a strong correlation but with the distribution reversed, i.e., with ninth graders as masters. Values close to 0 indicate that no correlation exists between grade levels and mastery status for a level.

The values of ϕ calculated for the mastery designations at each level and for each set of interpretation schemes are shown in Table 7.6. The relatively small positive values of ϕ associated with Levels 1 and 2 indicate that, for these subjects, there is a

Table 7.6

Correlation Coefficient (ϕ) for All Masters and Nonmasters Grouping and Grade Membership

Interpretation Scheme	van Hiele Level 1	2	3	4
3, 3, 3, 3	0.2041	0.2339	0.4627	0.4035
3, 3, 4, 3	0.2041	0.2339	0.4000	0.3516

weak relationship between the mastery assignments and grade level, with twelfth grade subjects designated masters, proportionally, more frequently than ninth graders. At Levels 3 and 4, a stronger relationship exists. The larger values of ϕ indicate that there is a moderate correlation between the mastery designations and the grade level. Again, the twelfth grade subjects were designated masters, proportionally, more often than the ninth graders.

For each van Hiele level, the proportion of the total variance in the mastery designations attributed to the variance in the grade levels, $\eta^2_{y,x}$, was also calculated. The statistic ranges in value from 0 to 1. For dichotomous data, $\eta^2_{y,x} = \phi^2$. The values for $\eta^2_{y,x}$ by level, for each interperetation scheme, were small (see Table 7.7). This indicates that a small proportion of the total variance

Table 7.7

Correlation Ratio ($\eta^2_{y,x}$) for All Masters and Nonmasters Grouping and Grade

Interpretation Scheme	van Hiele Level 1	2	3	4
3, 3, 3, 3	0.0416	0.0547	0.2142	0.1628
3, 3, 4, 3	0.0416	0.0547	0.1600	0.1236

in the mastery designations, at each level, can be attributed to the variance between the grade levels.

Implications from the Findings Involving Grade Levels

Three types of statistics, chi squared , χ^2, the correlation coefficient, ϕ , and the correlation ratio, $\eta^2_{y,x}$, were used to explore the relationship between grade level membership (twelfth or ninth grade) and performance on the van Hiele Quadrilateral Test. Overall, the findings suggest that there is some association between the two variables:

(1) Chi Squared, χ^2. The chi squared statistics failed to support the independence of the grade level and mastery designation.

(2) Correlation Coefficient, ϕ. Level by level, for both performance on the subtests and mastery designations, as defined by the all grouping, correlation with grade level membership was low for the Level 1 and 2 subtests and higher for the Level 3 and 4 subtests. This is what one might expect, given the nature of the instruction each group has received. The ninth grade subjects, having not yet studied deduction, would not be expected to be successful on the subtests which correspond to abstraction and deduction, Levels 3 and 4. The twelfth grade subjects, on the other hand, having completed their study of Euclidean geometry, would be expected, as a group, to perform more strongly than the ninth graders on the upper two levels. Both groups, however, would be expected to do well on the subtests corresponding to the lower levels.

(3) Correlation ratio, $\eta^2_{y,x}$. The patterns of the correlation ratio suggest that grade level is not a particularly strong factor from which to predict either performance on the individual subtests or "all" mastery designations. If the instrument results correspond with an accurate description of the mastery distributions, this variability information might be seen to support the need for a van Hiele assessment technique. If grade level and van Hiele level were synonymous, there would be no need for such an assessment. Furthermore, these statistics might be interpreted to indicate that there is variability in the level assignments <u>within</u> each class, i.e. that there is a range of van

Hiele levels within each class. This would make it all the more important for the instructor to understand the range of levels, and to adjust curriculum and instruction accordingly.

Comparisions with the Nova Scotia Achievement Test

For the subjects in each grade, comparisons were made between the subjects' performances on the van Hiele Quadrilateral Test and their performances on the Nova Scotia Achievement Basic Concepts Test. Final mastery designations and subtest performances on the van Hiele Test were both used to investigate the source of the variation in the performances on the standardized test. It was informally hypothesized that students' van Hiele levels would correspond positively to performance on the Basic Concepts Test; the subjects with the higher van Hiele mastery levels would also have the higher test score.

Mastery Decisions

For each grade level, the proportion of the variance in the standard scores on the Basic Concept Test (Y) which was attributed to the van Hiele level mastery designation (X) was determined. The resulting values of $\eta^2_{Y/X}$ using the two interpretation schemes are presented in Table 7.8. A moderate amount of the variation in the twelfth grade subjects performances on the Nova Scotia Test is associated with the variance in their mastery levels. For the ninth graders, little of the variation in the test scores is

Table 7.8

Proportion of Variance ($\eta^2_{y,x}$) in the Nova Scotia Achievement Basic Concepts Test Scores (Y) Attributed to Variance in the Overall Mastery Designations (X)

	Interpretation Scheme	
Grade	3, 3, 3, 3	3, 3, 4, 3
9	0.1989	0.1440
12	0.4109	0.4419

associated with their van Hiele mastery levels. These patterns in the source of variance were repeated when the all mastery assignments, one level at a time, were compared to the Basic Concepts Test scores (see Table 7.9).

These statistics suggest that for both the overall mastery assignments and the level by level "all" mastery designations, knowledge of the ninth graders van Hiele mastery level does not in itself, appear to be highly predictive of the student's performance on the Basic Concepts Test. For the twelfth grade subjects, knowledge of the mastery designations at Level 3 or at Level 4 is a moderate predictor of performance on the Basic Concepts Test.

Table 7.9

Proportion of Variance (η^2_{yx}) in the Nova Scotia Achievement Basic Concepts Test Scores (Y) Attributed to Variance in the Mastery Assignments. (X)

Interpretation Scheme	van Hiele Mastery Level 1	2	3	4
		Ninth Grade		
3, 3, 3, 3	0.1233	0.0534	0.0965	------
3, 3, 4, 3	0.1233	0.0534	0.0232	------
		Twelfth Grade		
3, 3, 3, 3	------	0.1160	0.3299	0.3274
3, 3, 4, 3	------	0.1160	0.3244	0.3902

Subtests

Like the mastery decision findings, the proportion of variance in the Basic Concepts Test scores which was associated with either success or nonsuccess on a subtest, was higher for the twelfth grade subjects than for the ninth grade subject (see Table 7.10). For both groups, however, little of the variance in the Basic Concepts Test scores is attributable to the difference in performance on the subtest.

Table 7.10

Proportion of Variance ($\eta^2_{y,x}$) in the Nova Scotia Achievement Basic Concepts Test Scores (Y) Attributed to Variance in Subtest Success Status (X)

	Subtest Level (cutoff / # of items)				
Grade	1(3/4)	2(3/5)	3(3/6)	3(4/6)	4(3/4)
9	0.1832	0.0539	0.0154	0.0231	-----
12	-----	0.1160	0.3017	0.2665	0.2160

Implications of the Findings Associated with the Nova Scotia

Achievement Tests

For students in both grades, neither mastery level nor subtest performance were strong predictors of performance on their respective Basic Concepts Test. (For the twelfth grade subjects, however, mastery designations were a moderate predictor of performance on the Nova Scotia test.) These weak results might be attributable, however, to the composition of the Basic Concepts Tests. While it is known that the geometry content on each grade level test is approximately 40%, the proportion of that which deals with quadrilaterals, for which the van Hiele levels were being determined, was unavailable.

Item Analysis

An analysis of the response rates of the subjects in each grade to the 19 items on the test was conducted. The percentage of the responses to each item which were correct is presented in Table 7.11. A discussion of those response patterns, by level, follows.

Level 1 Subtest

The item analysis conducted in the previous stage, the field testing, indicated that the items included in this subtest did not discriminate between the field test masters of Level 1 and the field test nonmasters of Level 1. At that time, all subjects did

Table 7.11

Item Response Rate, Percent Correct, by Grade Level

Level	Item	Grade 9	Grade 12
1	1	88	88
	2	96	98
	3	80	96
	4	82	86
2	5	46	42
	6	64	72
	7	60	70
	8	50	72
	9	56	76
3	10	12	38
	11	12	42
	12	28	58
	13	18	54
	14	44	54
	15	32	58
4	16	10	44
	17	16	60
	18	14	56
	19	22	46

well on these items. This pattern was repeated in the final administration of the test. The items in this subset were consistently answered correctly.

Level 2 Subtest

The overall success rates on this subtest were 60% for the ninth grade subjects and 78% for the twelfth grade subjects. Item by item, the success rates for the ninth grade subjects were moderately consistent, ranging from 46% to 64% correct. The twelfth grade subjects, however, while correctly answering items 6 through 10 consistently in the 70% range, only demonstrated a 42% correct response rate for item 5. Of the 39 twelfth grade subjects who were successful on this subtest, only 20 of them selected the correct response for item 5. This item should be reviewed and analysed in terms of its usefulness for discriminating at this level. Considerations could include issues such as: Are students familiar with the shapes, properties and components described? Is the vocabulary appropriate? Is the way the item is presented confusing?

Item 6 is the only item on the final test with two "correct" answers. It was designed with a response corresponding to Level 2, and a response corresponding to Level 3. The item analysis conducted during the previous field testing stage, however, indicated that the two answer choices were not discriminating between the two levels. Nonetheless, because the item, when both answer choices were accepted, appeared to discriminate between masters of Level 2 and nonmasters of Level 2, it was retained for the final test, with both answer choices deemed acceptable. The distribution of the performances of the subjects from the final testing on the item 6 answer choices is presented in Table 7.12.

Table 7.12

Distribution of Answers Selected for Item 6

	Level of answer choice		
Grade	3	2	none
9	1	20	9
12	16	17	6

The differences in performance by grade level on item 6 suggest that this item should be reviewed. One consideration would be to change the answer corresponding to Level 3. It does not appear to be attractive to the ninth grade (the generally lower van Hiele level) subjects. If it was altered to become a choice which was not associated with any level, this would simplify the marking of the item. The corresponding effect of such a change on the twelfth grade (and generally higher van Hiele level) subjects' responses would also have to be pursued. An alternate consideration, however, would be to further investigate the effectiveness of the question to elicit responses at two levels. Could the wording of the question be altered in some way? Would a different combination of "statements" corresponding to Level 3 thinking be more attractive?, etc.

Level 3 Subtest

Two cutoff scores were considered for the Level 3 subtest, 3 out of 6 items and 4 out of 6 items. The success rates, by subtest, for the ninth grade subjects, using each cutoff score, were 14% and 4%, respectively. The corresponding rates for the twelfth graders were 56% and 38%.

The item by item performance of the ninth grade students corresponded with their overall "nonlevel" performance. On item 14, however, 22 (44%) of the ninth grade students, selected the correct response. This item appeared easier for this group than the other Level 3 items. When responses were further analyzed in terms of students who were not successful on the subtest, this "easiness" was corroborated. Of the 43 students who answered 2 or fewer of the items at this level correctly, 18 (42%) answered this item correctly. The ability of this item to discriminate between masters and nonmasters of the level should be further investigated. Administering it to larger numbers of subjects, for whom van Hiele levels were known, would assist in this. Also, as the item deals with "definitions", it might prove valuable to discuss with students at different levels their understanding of that concept. Perhaps, in fact, this item was not a Level 3 item? Perhaps the language in the item was inappropriate?

The twelfth grade students answered all the Level 3 items, except for items 10 and 11, correctly more than 50% of the time. For item 10, 19 (38%) of the twelfth grade subjects answered the item correctly. Of those individuals, no individual who answered exactly 3 items in this subtest correctly answered this item, while 17 of the 19 subjects answered 4 or more items in the subtest correctly. The lack of success on this item associated with those who scored 3 correct on the subtest, particularly since 4 was a cutoff score, suggests that the item should be reviewed and further information about the validity of the item collected.

Item 11 was answered correctly by 21 (42%) of the twelfth grade subjects. The response patterns for this item were not as distinctive as those cited for item 10. The distribution of answer choices by the 9 individuals who correctly answered 3 items on the subtest was equally distributed between correct and incorrect, with 4 choosing correctly and 5 choosing incorrectly. Of the 19 individuals who correctly answered 4 or more, 15 of them selected the correct response. Nonetheless, further information about the validity of the item should be gathered.

Level 4 Subtest

The performance of the ninth grade subjects on the Level 4 items was consistent with their "nonlevel" performance--only one subject answered more than 2 items correctly, no subjects were classified as masters of Level 4. There was no Level 4 item which appeared "easy".

Of the twelfth graders, only 34% of the subjects were able to answer at least 3 of the 4 items on this subtest, and thus be designated as successful on the Level 4 subtest. In light of that rate, the percentage of correctly answered items at this level, which range from 44% to 60%, might seem high. Further analysis indicated that 70% of the twelfth grade students correctly answered 2 of the 4 items. The high percentage of students who answered two of the items correctly suggests that, the Level 4 items should be tested further to see if they do require Level 4 thought. The criteria for success at Level 4 might also be re-evaluated.

Summary

An analysis of the performance of the 101 subjects in the ninth and twelfth grades who participated in the final administration of the van Hiele Quadrilateral Test was presented in this chapter. There is evidence from this analysis supporting some association between grade membership and van Hiele mastery level. The chi squared statistics failed to support the hypothesis of independence of those two variables. The correlation indices, as informally hypothesized, suggest there is a moderate relationship between membership in twelfth grade and being designated a master of the upper two levels. Mastery levels were a very moderate predictor of performance on the Nova Scotia Achievement Basic Concepts Test for the twelfth graders, and were a poor predictor of performance for members of the ninth grade on their equivalent standardized test. The importance of the results just described, however, is overshadowed by the reliability findings. The reliability indices suggest that the subtests do not yield consistent results for these subjects. Until reliability can be established, the instrument is not appropriate for determining van Hiele mastery levels. Conclusions based on these findings, as well as from the other development stages, are presented in Chapter 8.

THE DESIGN AND EVALUATION OF AN INSTRUMENT FOR ASSESSING MASTERY VAN HIELE LEVELS OF THINKING ABOUT QUADRILATERALS

by

Mary Lora Noffsinger Crowley

Dissertation submitted to the Faculty of the Graduate School of The
University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1989

Advisory Committee:

Associate Professor James Henkelman, Chairman/Advisor
Professor James Fey
Professor Martin Johnson
Professor Clayton Stunkard
Associate Professor Neil Davidson

Chapter 8

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Over the last decade, many of those interested in evaluating and improving geometry instruction have looked to the work of Pierre M. van Hiele and Dina van Hiele-Geldof for direction. The van Hieles' proposed a three part, interrelated model of geometric cognition. In it, they described the nature of insight, they outlined five sequential levels of geometric maturity, and they provided a description of the way an individual moves (learns) from one level to the next. Instruction, they say, not maturation, is the major factor in this progression.

If learning, as outlined by the van Hieles, is to occur, it is imperative that instruction be matched with the audience. Prior knowledge of a particular group of students' van Hiele levels, for example, could influence the content and methodology of an instructional activity or of a series of such activities. Assessment following such instruction could be used to chart students' progress through the van Hiele levels, and be used as a starting point for further instruction. Means for assessing the van Hiele level on which students are operating are therefore required.

To date, there are only a few instruments which can be used to assess an individual's van Hiele level. All but one of those involve one-on-one interviews, and as such, they are very time consumming. The other instrument, the 35 minute multiple-choice VAN HIELE GEOMETRY TEST, is designed to assess large numbers of students, at one time. The claim that the results from the test identify a van Hiele level for "geometry" is, however, somewhat problematic. Research into the model suggests that individuals may be at different levels for different content areas. The developers of the instrument also report that the test has poor norm - referenced reliablity statistics associated with it. Despite these drawbacks, as the only such instrument available which is easily administered to a large group of subjects and easily scored, the VAN HIELE GEOMETRY TEST is used.

This research undertook to develop an instrument for assessing dominant van Hiele levels of geometric reasoning, which is easily administered to large groups, easily scored, easily interpreted, and for which the test results are reliable and valid. To be consistent with the research findings indicating that the van Hiele levels of thought are not necessarily consistent across content areas, the instrument was limited to one topic, quadrilaterals. This topic was selected as it is a core topic in the study of Euclidean geometry. Items were written, reviewed by a panel of experts and, in revised form, piloted with 14 subjects. After further revisions, the items were field tested with 97 subjects, for whom van Hiele levels had been independently established. A final instrument, the van Hiele Quadrilateral Test, was assembled, based on the item analysis conducted during the field study. Using grade level membership, performance on the Nova Scotia Achievement Test and performance on the van Hiele Quadrilateral Test as variables, reliability and validity studies were conducted. This chapter will summarize the research findings, discuss the implications of the research results and suggest areas for further research.

Conclusions and Implications

In Chapter 1, five research questions were identified. The following summary of and conclusions from the research findings correspond to those questions.

(1) Can multiple-choice items, which discriminate between masters and nonmasters of a van Hiele level, on the topic of quadrilaterals, be developed?

Item validity was gauged using the responses from a panel of four van Hiele model experts, and using the responses to the items from students for whom van Hiele mastery levels were know. From an initial item bank of 53 multiple-choice items, 15 items eventually emerged as discriminating between masters and nonmasters of van Hiele levels. Five were associated with Level 2, six were associated with Level 3 and four were associated with Level 4. All of these items were selected for the final instrument. Further analysis of the 15 items, based on the responses of the subjects who wrote the final instrument, suggested that several items be reviewed. In general, however, it appears that items can be written to correspond with Levels 2 to 4.

No items were produced which appeared to discriminate between nonmasters of Level 1 and masters of Level 1 when the item discrimination indices were considered. Those item statistics indicated that both groups were correctly identifying the responses associated with Level 1 mastery. At least three possible explanations for this performance arise. One is that this result is a function of the small number of Level 1 nonmasters in the study. An item analysis conducted using the responses from a larger group of nonmasters of Level 1 on the Level 1 items from the draft instrument should be conducted.

Another possible explanation for the nonmasters success is that they might be in transition towards mastering Level 1. Other research has noted that some students appear to fluctuate in their use of strategies from adjacent levels. Possibly the items on the instrument associated with Level 1 thinking tested the characteristics which these "soon to be masters" of Level 1 had acquired. This raises the possibility that, if movement is not discrete between levels, some level characteristics are acquired before others. Further research might investigate whether or not the levels are nondiscrete, and if so, whether the characteristics associated with a level are acquired sequentially.

The item format may also be a contributing factor in the difficulty which existed in identifying items assoicated with Level 1. The reactive nature of the items may not permit identification of Level 1 thinking. Perhaps distinguishing Level 1 thought from other thinking requires student initiated activities or student corroboration. For example, activities, such as sorting, can be governed by Level 1, Level 2 or Level 3 thinking. To distinguish which, the subjects must signal, in some way, the reasoning behind the action. This is difficult to do with multiple-choice items. If, however, this format is to be maintained, perhaps the item "stem" and responses should be different from the rather traditional format used in this instrument. For example, a problem and a response to that problem might be detailed in the item stem. After reading those, the subjects might be asked to select on what basis the solution was determined. Possible answer choices might include typical responses from each level, such as "It looks like..." (Level 1), "The properties are..." (Level 2), "If...then..." (Level 3). Even with this type of approach, however, the subject is reacting, not generating responses.

The research also attempted to develop items which provided answer choices associated with several levels. (This research effort was encouraged by members of the panel of experts, on the basis of the assessment potential). No item, however, met the discrimination criteria for more than one level.

Conclusion: The research was able to develop multiple-choice items which appeared to discriminate between masters and nonmasters of Levels 2, 3 and 4. No items which discriminate between masters and nonmasters of Level 1 were produced.

Implications: If the instrument is to distinguish between masters and nonmasters of Level 1, items which discriminate "at" Level 1 must be identified. Minimally, the Level 1 items on this instrument, should be administered to a larger group of nonmasters than used with the field test, and an item analysis on their responses should be conducted. Additional Level 1 items might written and tested at the same time.

(2) What is the reliability associated with the mastery decisions from the instrument?

Two criterion-referenced reliability coefficients, the agreement coefficient and Cohen's Kappa, were calculated in both the field study and the final testing phase of the research. For the field study, the statistics were calculated using the subjects responses to the 19 items contained on the final instrument, only. In each setting, the statistics were calculated for each of the four subtest, rather than for the test as a whole. The mathematical requirements of the statistical techniques necessitated this level by level approach.

The reliability statistics from the two research settings do not support the consistency of the mastery decisions over repeated testing. The values obtained for the agreement coefficients, in both research stages, suggest that the overall mastery decisions which resulted from administering the van Hiele Quadrilateral Test would be, at best, consistent across several administrations of the instrument for the Level 1, Level 3 and Level 4 subtests, only. (Even for these levels, the reliability figures for the twelfth grade students when they are considered on their own are slightly below the minimum research criteria.)

For the two settings, the values obtained for Cohen's Kappa, describing the test consistency, are contradictory. In the field test they indicate that the subtests contribute to the mastery decisions, for those subjects, beyond chance. For the final administration of the instrument, however, the reliability coefficients suggest that the subtests contribute very little to the consistency of the decisions.

Conclusion: The reliability studies from the field testing and from the final testing are conflicting.

Implications: With inconclusive reliability statistics, the instrument cannot be used with confidence to determine van Hiele mastery levels. Additional reliability studies could be conducted with the items on this instrument. Any study of that nature should include subjects from a broad academic range. Upper elementary school children could provide non-Level 1 subjects. University students could provide Level 4 subjects.

The reliability of the instrument could also be enhanced by including additional items. The two reliability coefficients used in the research are both sensitive to the number of items on a test, and to the location of the cutoff score relative to that number. Stronger reliability statistics might be obtained if each subtest was lengthened. (Particular attention should be paid to obtaining valid items associated with Level 1.) Changing a subtest's length would, of course, also require a review of the cutoff score used to determine success on the subtest.

(3) <u>What validity is associated with the mastery decisions which result from the van Hiele Quadrilateral Test?</u>

Evidence corresponding to three types of test score validity -- content validity, criterion-related validity, and construct validity -- was collected in this study. Content validity, in this instance, is interpreted to mean the representativeness of the test items. The final items associated with the level subtests do not represent the range of level descriptors, even when just the descriptors in the "applied" category from which they are drawn is considered. An argument might be made that, as the model states that movement from one level to the next occurs in "leaps", evidence from an individual of any type of thinking asssociated with a level is therefore sufficient to say that individual has mastered the level. Further investigation, however, into the "absolute nature" of the acquisition of the thinking processes asssociated with a given level of thought should be conducted. This is, of course, related to the issue of individuals in transition which was identified in the discussion of the Level 1 items as an area for further study.

Conclusion (Content Validity): The items on the subtests in the van Hiele Quadrilateral Test do not correspond with a cross-section of the level descriptors.

Implications (Content Validity): The representativeness of the items should be tested further. Research of this nature might be associated with an investigation into the discreteness of each level.

The criterion-related validity studies investigated (a) the relationship between performance on the van Hiele Quadrilateral Test and membership in a grade, and (b) the relationship between performance on the van Hiele Quadrilateral Test and performance on the Nova Scotia Achievement Basic Concepts Test. The Chi squared statistics suggested that there was an association between grade membership and mastery designations. The correlation indices, however, suggested that the association was, at best, moderate for mastery/nonmaster decisions, and then only at Levels 3 and 4. For the lower two levels, the correlation indices could be interpreted to say that there was little association between mastery decisions and grade level. (This latter result corresponds with the fact that the majority of the subjects in each class met the success criteria for the subtests at the lower two levels). Furthermore, the correlation ratio statistics suggest that membership in Grade 9 or Grade 12 is not a strong predictor of a van Hiele mastery level.

The statistics obtained in the criterion-related validity studies might be seen to provide support for the notion that the van Hiele levels do not strictly correspond with grade levels. (If they did, there would be no need for an assessment instrument.) Furthermore, if the diversity of van Hiele levels identified by the instrument is, in fact, present within each grade level, the importance of both knowing that this range is present and knowing what the van Hiele profile of the class is reinforced. Students do not understand instruction requiring thinking from a higher level.

For the standardized Basic Concepts Test, the twelfth grade performances appeared to correspond moderately with performance on the van Hiele test. For the grade 9 subjects, performance on the van Hiele test was a poor predictor of performance on the standardized instrument. These weak associations might stem, however, from the nature of the content in the standardized instrument. Only 40% of the items dealt with geometry.

<u>Conclusion (Criterion-related validity)</u>: When comparing membership in Grade 9 or Grade 12 to the mastery decisions from the van Hiele Quadrilateral Test, there was an indication of some relationship

between grade and mastery level. Overall, however, grade level was not a good predictor of a subjects' mastery level. When comparing the mastery decision to performance on the Level 12 Nova Scotia Achievement Mathematics Basic Concepts Test, the mastery decisions for the twelfth grade subjects were, at best, moderate predictors of performance. The mastery decisions for the ninth grade subjects were poor predictors of performance on the Level 9 Nova Scotia Achievement Mathematics Basic Concepts Test.

Implications (Criterion-related validity): Further validity studies should be conducted. In particular, additional studies comparing performance on the van Hiele Quadrilateral Test and an independent measure of the van Hiele levels, for example the Burger and Shaughnessy interview, should be conducted. If additional studies are conducted where membership in a grade is considered as a variable, upper elementary school children, say, in fifth or sixth grade should be included. These subjects would be younger than those used in the last stage of this research. As such, they might provide a setting where information on nonmasters of Levels 1 and 2 could be collected. This would strengthen the validation studies.

Finally, the results from the Guttman scalogram analysis indicate that the subjects demonstrated a sequential pattern of success on the subtests. If success on each subtest is, in fact, associated with the mastery of the level with which it is associated, the results of the Guttman scalogram analysis support the hierarchical property of the model.

Conclusion (Construct validity): The subject's performances on the subtests, level by level, appear to support the construct that the levels are hierarchical in nature.

Implication: Further supporting evidence demonstrating that the success of a subtest does correspond with level mastery would increase confidence in these findings.

(4) Can the test be easily administered?

The van Hiele Quadrilateral Test can be administered within one 40 minute class period. The testing requires the students be issued copies of the test, a one page answer sheet, and a pencil. Instructions are provided for the subjects and require approximatley five minutes for the administrator to review with the subjects. Instructions for the administrator regarding equipment, timing, etc. are also provided.

Conclusion: The test can be easily administered.

(5) Can the test be easily interpreted?

The interpretion scheme which converts the raw subtest scores into mastery decisions is a three stage process. First the raw score on each subtest is obtained. Then an individual's subtest success record is determined by comparing the raw score to the cutoff score for each subtest. Finally, the mastery decision is made based on the sequence of subtest successes. Using this procedure to determine mastery levels is more cumbersome than translating an overall raw score into a mastery designation would be. A further complication of the research was that two different cutoff scores were applied to Level 3. This necessitated the compilation of two mastery designations, sometimes different, for each subject. It was a goal of the final stage of the research to identify which of the two cutoff schemes was associated with valid mastery decision. No such decision, however, was reached.

The results from the Guttman scalogram analysis also have implications for the interpretation scheme. The consistency of the subjects' successes on the subtests to form a sequence suggest that there would be little difference between assigning mastery to correspond with the highest subtest and assigning mastery from the highest subtest successfully answered in a sequence. Using the highest subtest regardless of sequencing, would simplify the interpretation procedures.

<u>Conclusions</u>: The interpretation scheme, while involving several stages, is not difficult to implement. Two interpretation schemes, however, were used and no decision was made regarding which scheme should be used to assign mastery decisions.

Implications: A single set of cutoff scores should be decided upon. As well, the viability of designating the highest subtest successfully answered as the mastery level should be explored. Both of these investigations, should be coupled with further studies into the reliability of the mastery decisions.

In summary, the criteria used to assess the product of this research, the van Hiele Quadrilateral Test, indicate that further developmental work needs to be completed before the test can be used to determine mastery levels.

Recommendations for Further Research

The research suggestions emerging from this study focus on two areas, investigations relating to the van Hiele model and investigations specific to the assessment issue. While recognizing that the research suggested in the first category would influence the second category, the two areas are discussed separately.

Research Relating to the Model

Two areas for further research relating to the tenets of the model were identified in this study. They are (a) the nature of level acquisition, discrete or continuous, and (b) the relationship between the objects of consideration at a level and the acquisition of the level.

One of the suggestions for why it was difficult to identify items which discriminate between Level 1 masters and nonmasters, centered on the issue of the manifestation of the acquisition of a level. It was suggested, as has some of the other research into the validity of the van Hiele model, that movement from level to level may not, as the van Hieles proposed, be discrete. Evidence about this point would influence the design of an assessment instrument. If progress is made by "leaps", then perhaps only a few items related to a level are sufficient for making mastery decisions. A subject either has all the skills associated with a level or none. If, on the other hand, movement from level to level is continuous in its nature, minimally, this would say that a much larger proportion of the activity associated with a level must be demonstrated before an individual is designated a master. Indeed. the amount of that "larger proportion" -- 100%, 90% etc, would also be a topic of investigation.

The second implication relating to the model emerging from this research pertains to whether or not an individual operates on the same van Hiele level for all geometric concepts or whether individuals might operate on different van Hiele levels for different topics. Other researchers have found evidence to suggesting the latter. In this research, which attempted to focus on one topic, it was observed that the objects of consideration at Level 4, and to some extent Level 3, are not confined to a single geometric shape or notion. For example, information about parallel lines or rotations is required for either an informal or a formal proof of some of the angle properties associated with quadrilaterals. Thus, functioning at Levels 3 and 4 would seem to require an equivalent level of thought on a range of interrelated topics. Further exploration into whether or not an individual has a "unique" van Hiele level for different geometric topics should be investigated. In particular, is mastery classification at Level 1 and Level 2 topic specific? Does being identified as a master of Level 3 or Level 4 for a certain topic, also indicate (require) a minimum mastery level for other related topics?

Research Relating to Assessment Issues

The second area identified for further research deals with the assessement of an individual's van Hiele mastery level. If the van Hiele Quadrilateral Test is to be refined, further evidence on the reliability of the instrument must be obtained. As well, (1) the existing Level 2 to Level 4 items should be reviewed for further evidence relating to their validity, (2) items which discriminate between masters of Level 1 and nonmasters of Level 1 must be identified, (3) more items for each level could be developed (this could increase the reliability and content validity associated with the instrument) and (4) the interpretation scheme would need to be further assessed. Any of these revisions should be accompanied by extensive field testing, preferably using subjects whose van Hiele levels have been determined by an external measure.

In a more general context, techniques for assessing van Hiele levels from written instruments might be further explored. If assessment is to be conducted with fixed choice responses, such as the multiple-choice questions used in this instrument, additional effort could be spent in trying to identify items which have responses associated with several levels. This might require rethinking what the "stem" of the item contains. Perhaps, as suggested earlier, a problem and a solution could be described, then students could indicate from a set of fixed choices the response which "best" explains why or how the solution was determined.

Another assessment approach, still using a written test, which might be considered is the use of items which are open-ended. One question type which might be appropriate is the format suggested above, where the stem describes a problem and a solution. The student could then describe, in his own, words why or how the solution was obtained. Or, a problem might be described and the student might be asked to describe how he would approach solving it. Evaluation with open-ended items, however, where the answers are not predetermined, requires a subjective judgement as to the van Hiele level with which the response is associated. Explicit guidelines for making such determinations would have to be provided, and even those would not be able to anticipate every "correct" response.

The efficacy of using written assessments to identify an individual's van Hiele mastery level might also be investigated. Is the time required to write and validate items worth the effort? Does this format lend itself to identifying individuals operating on some levels, better than other levels? Is it possible with a multiple-choice test for an individual to demonstrate insight?

Limitations

The limitations of this research include:

1. The choice of subject matter. Quadrilaterals, while an important content area in geometry, are a restricted field of study.

2. The nature of the multiple-choice test. This form limits the types of activities which can be used, and thus the level descriptors which can be assessed. As well, this type of question does not provide the examinee the opportunity to generate responses. Instead, answering requires recognition and reaction.

3. Using students who have been schooled using only one curriculum, the Nova Scotia mathematics curriculum. The generalizability of the findings to other jurisdictions should be established.

4. The limited educational range of the subjects. Only a few of the subjects were in grades lower than ninth, and none were below sixth grade.

5. The validity of the items in the Level 1 subtest has not been established. Too few nonmasters of Level 1 were identified in the field testing stage.

6. The researcher was the only judge for the mastery decisions which resulted from the interviews conducted at the pilot and field testing stages.

7. Only 40% of the items on the Nova Scotia Achievement Basic Concepts Tests were related to geometry.

Appendix A

The van Hiele Model of the Development of Geometric Thought

The van Hiele model of the development of geometric thought emerged in the late 1950's from the work of two Dutch school teachers, Dina van Hiele-Geldof and Pierre M. van Hiele. Concerned about their secondary school students' performances in geometry, and interested "in improving teaching outcomes" (van Hiele, 1986, p. vii), the van Hieles' doctoral dissertations studied complementary aspects of developing insight in geometry. Pierre van Hiele "formulated the scheme and psychological principles: D. van Hiele-Geldof focused on the didactics experiment to raise students' thought levels" (Hoffer, 1983, p. 207). The model consists of three major components: (1) the nature of insight, (2) the levels of thought, and (3) the phases of learning.

The Nature of Insight

In his doctoral disseration, Pierre van Hiele examined "the meaning and functions of (geometrical) insight during a process of learning" (van Hiele, 1957/1984a, p. 237). For him, insight is demonstrated when a person is able to perform <u>adequately and with intention in a new situation</u> (van Hiele, 1986). "(H)e acts according to the structure he perceives, corresponding to his mental structure, the structure of his expectations" (van Hiele, 1986, p. 24). Students with insight "understand what they are doing, why they are doing it, and when to do it. They can apply their knowledge in order to solve problems" (Hoffer, 1983, p. 205).

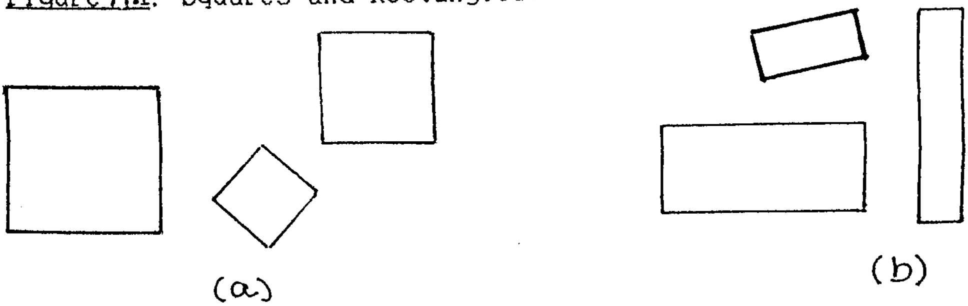
Levels of Thinking

The five levels of thinking developed in the model are descriptions of characteristics of the thinking process, i.e. of the mental structures which govern learning and insight. The theory asserts that the learner starts at the first level and, assisted by appropriate instructional experiences, moves sequentially along the levels. Elegant in their simplicity, a general description of the levels is provided below.

Level 1: Visualization

At this initial stage, students are aware of space only as something that exists around them. Geometric concepts are viewed as total entities rather than as having components or attributes. Geometric figures, for example, are recognized by their shape as a whole, that is by their physical appearance, not by their parts or properties. A person functioning at this level can learn geometric vocabulary, can identify specified shapes, and given a figure, can reproduce it. For example, given the diagrams in Figure A.1, a student at this level would be able to recognize that there are squares in (a) and rectangles in (b) because these are similar in shape to previously encountered squares and rectangles. Furthermore, given a geoboard or paper, the student could copy the shapes. A person at this stage, however, would not recognize that the figures have right angles or that opposite sides are parallel.

Figure A.1. Squares and Rectangles.



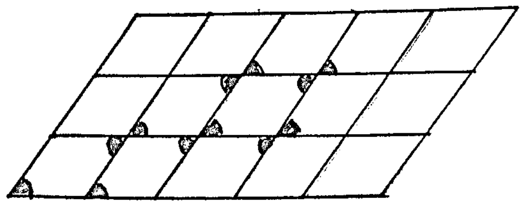
In the van Hieles' early writings, this level was refered to as the Base level, or Level 0, rather than as the first level. The levels following this one were the original first level, second level, etc. Van Hiele explains the initial designations as arising from "not having seen the importance of the visual level" (van Hiele, 1986, p. 41). As he now acknowledges, however, this initial level is integral to the model. This shift in emphasis has led to a confusion of numbering systems in the literature. Some systems start with Level 0 and end with Level 4, paralleling the original van Hiele designations; others run from Level 1 to Level 5. As the most recent work of P. M. van Hiele refers to the initial level as the first level, the former first level as the second level, etc., this research refers to the levels as Level 1 to Level 5.

There is also a lack of consensus amongst those writing about the levels concerning the verbal labelling of the levels. The van Hieles' original terminology for the five levels, i.e., base, aspect of geometry, essence of geometry, discernment of geometry and discernment in mathematics, respectively, have not been popular with English language writers. Hoffer (1983), for example, used "recognition", "analysis", "ordering", "deduction", and "rigor" to label the five levels. Burger and Shaughnessy (1986) described the levels as "visualization", "analysis", "abstraction", "deduction", and "rigor". Pierre van Hiele in his latest book Structure and Insight (1986) calls the levels "visual", "descriptive", "theoretical", "formal logic" and "the nature of logical laws". The titles suggested by Burger and Shaughnessy are used for the current research work. They most consistently describe the salient characteristic of the mental structures functional at each related level.

Level 2: Analysis

At Level 2, an analysis of geometric concepts begins. For example, through observation and experimentation students begin to discern the characteristics of figures. These emerging properties are then used to conceptualize classes of shapes. As a consequence, figures are recognized as having parts and are recognized by their parts. Given a grid of parallelograms such as those in Figure A.2, students could, by "coloring" the equal angles, "establish" that the opposite angles of parallelograms are equal. After using several such examples, students could make generalizations for the class of parallelograms. Relationships between properties, however, cannot yet be explained by students at this level, interrelationships between figures are still not seen, and definitions are not yet understood.

Figure A.2 Parallelogram Grid



Level 3: Abstraction

At this level, students can establish the interrelationships of properties both within figures (e.g., in a quadrilateral, opposite sides being parallel necessitates opposite angles being equal) and among figures (a square is a rectangle because it has all the properties of a rectangle). Consequently, they can deduce properties of a figure and recognize classes of figures. Class inclusion is understood. Definitions are meaningful. Informal arguments can be followed and given. The student at this level, however, does not comprehend the significance of deduction as a whole or the role of axioms. Empirically obtained results are often used in conjunction with deduction techniques. Formal proofs can be followed, but students do not see how the logical order could be altered nor do they see how to construct a proof starting from different or unfamiliar premises.

Level 4: Deduction

At this level, the significance of deduction as a way of establishing geometric theory within an axiomatic system is understood. The interrelationship and role of undefined terms. axioms, postulates, definitions, theorems and proof is seen. A person at this level can construct, not just memorize, proofs. The possibility of developing a proof in more than one way is seen. The legitimacy and impact of "arbitrarily" choosing certain criteria as the set of assumptions on which to build deductions is understood , i.e students "...understand that it depends from the starting point if a statement is a definition or a theorem" (van Hiele, personal communication, 22 March 1988). Concepts which emerge at this level include "the link between a theorem and its converse, why axioms and definitions are indispensable, when a condition is necessary and when sufficient" (van Hiele, 1958/1984b, p. 250).

Level 5: Rigor

This level is concerned with formal abstract aspects of deduction. At this stage the learner can work in a variety of axiomatic systems, that is, non-Euclidean geometries can be studied and different systems can be compared. Geometry is seen in the abstract. Few students are exposed to, much less reach, this level. "One cannot attain this ....level until one is sufficiently familiar with the procedures of mathematicians that one can do them automatically" (van Hiele, 1958/1984b, p.250).

This last level is the least developed in the original (and subsequent) works and has received little attention from researchers. Van Hiele points out that "in school we have to deal with Levels 2, 3, 4" (1986, p. 47). Indeed, the majority of high school geometry courses are taught at Level 4. Thus it is not surprising that most research has concentrated on the lower levels.

Properties of the Levels of Thinking

The van Hieles' also identified characteristics which link and illuminate the levels of thinking.

Hierarchical. The levels are arranged in a fixed order. Van Hiele (1986) presents an interesting discussion on whether or not. Implicit in this ordering, there is also the notion of the higher the level, the more valued the performance. He cited a Dutch colleague, Kees van Baalen, as having cautioned

> the theory makes use of an unstated assumption, namely that, whereas natural numbers are ethically indifferent, still in giving the names first level, second level, and so on, there is really an estimation of value. That means that the second level is valued higher than the first level. (van Baalen, 1980/1981, p. 429, cited in van Hiele, 1986, p. 41)

Indeed, van Hiele confesses to initially believing in the increasing "value" of the levels. Now, however, he claims to believe, as Kees van Baalen went on to suggest

> the order of succession of values has to be reversed. In this sense the first level is the highest and the other levels are subordinate to it.
>
> The first level is the level at which people (including pupils) think in their daily life, with which they have their experiences, and with which they make their decisions. The other levels (in my eyes lower levels) are those in which, from a limited perspective, parts of the matter used at the first level are chosen to make models as an aid for thinking and deciding at the first level. (van Baalen, 1980/1981, p. 429, cited in van Hiele, 1986, p. 42)

Sequential. Geometric thinking developes through the levels in order. To function successfully at a particular level, a learner must have acquired the strategies of all of the preceding levels and these levels are attained sequentially. Thus, not only are the levels hierarchical, e.g., they have a fixed order, but as well, progress through them occurs only by beginning at Level 1 and moving through each level in order. There is no "skipping" of levels.

Discontinuity. Movement between levels is a discontinuous process. As evidence of this, the van Hieles cite instances when the student seems to have stopped learning, only to later resume learning using the strategies of a new level. According to the van Hieles', these jumps in learning imply (1) the presence of levels and (2) that students operate on only one level at any one time. Indeed, when a level is attained, the strategies of the former level are superceded by the strategies of the new level.

Advancement. Progress (or lack of it) from level to level depends more on the content and methods of instruction received than on age or biological development. No method of instruction allows a student to skip a level. Some methods enhance progress: other methods delay or even prevent movement between levels. van Hiele points out that it is possible to teach "a skillful pupil abilities above his actual level, like one can train young children in the arithmetic of fractions without telling them what fractions mean, or older children in differentiating and integrating though they do not know what differential quotients and integrals are" (Freudenthal, 1973, p. 25). Geometric examples include the memorization of an area formula or relationships like "a square is a rectangle". In situations such as these, what has actually happened is that the subject matter has been reduced to a lower level and understanding has not occurred.

Mismatch. If the student is at one level and instruction is at a different level, the desired learning and progress may not occur. In particular, if the teacher, instructional materials, content, vocabulary and so on, are at a higher level than the learner, the student will not be able to follow the thought processes being used.

Intrinsic and extrinsic. The inherent objects at one level become the objects of study at the next level. For example, for an individual operating at Level 1, only the form of a figure is perceived. The figure is, of course, determined by its properties, but it is not until the individual moves to Level 2 that the figure is analyzed and it components and properties are discovered. At Level 3, the properties "recede" as the object of study and the focus shifts to the interrelationships between those properties.

Linguisitics "Each level has its own linguistic symbols and its own systems of relations connecting these symbols" (van Hiele, 1959/1984b, p. 246). Thus a relation that is "correct" at one level may be modified at another level. For example, a figure may have more than one name -- a square is also a parallelogram. A student at Level 2 does not conceptualize that this kind of nesting can occur. This type of notion and its accompanying language, however, are fundamental at Level 3. At each level the knowledge obtained during the previous level is reinterpreted and reconstructed. To accomplish this transition, new geometric and logical terms and symbols are required.

Phases of Learning

The van Hieles observed that the most significant factor influencing progress through the levels is instruction, not age or maturation. For them, the method and organization of instruction, as well as the content and materials used, was an important area of pedagogical concern. To address these issues, they proposed five sequential phases of learning: inquiry, bounded orientation, explicitation, free orientation and integration. They asserted that instruction developed according to this sequence would promote the acquisition of a level.

Phase 1: Information

At this initial stage, the teacher and students engage in conversation and activity about the objects of study for this level. Observations are made, questions are raised and level-specific vocabulary is introduced by the teacher (Hoffer, 1983). The purpose of these activities is two fold: (1) the teacher learns what prior knowledge the students have about the topic, and (2) the students learn what direction further study will take. The context of the study becomes clear.

Phase 2: Bounded Orientation

The students explore the topic of study through materials that the teacher has carefully sequenced. These activities should gradually reveal to the students the structures characteristic of this level. Thus, much of the material will be short tasks designed to elicit specific responses. These activities, when properly chosen, "form the proper basis of thinking on the higher level" (van Hiele, 1986, p. 97).

Phase 3: Explicitation

Building on their previous experiences, students express and exchange (make explicit) their emerging views about the structures that have been observed. Other than to assist students in using accurate and appropriate language, the teacher's role is minimal. It is during this phase that the level's system of relations begins to become apparent.

Phase 4: Free Orientation

The student knows "what their subject is about, they have read relations from concrete situations, they now know the relevant language symbols. The domain of their study is distinctly marked out" (van Hiele, 1956, p. 97). The student encounters more complex tasks -- tasks with manys steps, tasks that can be completed in several ways, and open-ended tasks. "They gain experience in finding their own way or resolving the tasks. By orienting themselves in the field of investigation, many relations between the objects of study become explicit to the students" (Hoffer, 1983, p. 208).

Phase 5: Integration

The students review and summarize what they have learned with the goal of forming an overview of the new network of objects and relations. The teacher can assist in this synthesis "by furnishing global surveys" (van Hiele, 1959/ 1984b, p. 247) of what the students have learned. It is important, however, that these summaries not present anything new.

At the end of the fifth phase, students have attained a new level of thinking. The new structure replaces the old, and student are ready to repeat the phases of learning at the next level.

Summary

The van Hiele model of thinking in geometry identifies three interrelated aspects of geometric activity: insight, levels of thinking, and phases of learning. Insight exists when a person performs competently, deliberately and consciously in a new situation. The nature of these actions is governed by the level of thinking an individual has attained. To acquire the "next" level of thought, instruction should be sequenced according to the phases of learning. Instruction, rather than biological maturation, is highlighted as the most significant factor contributing to the acquisition of a level of thought and of the "insights" which accompany that level.

Appendix B

Materials Sent to Panel of Experts

March 15, 1988

Professor J. Michael Shaughnessy
Department of Mathematics
Oregon State University
Corvallis, OR 97331
U.S.A.

Dear Professor Shaughnessy,

Thank you for agreeing to review the pool of van Hiele based questions which I have written. As I indicated to you on the phone, I am completing a Ph.D in mathematics education at the University of Maryland. My doctoral dissertation advisor is Professor James Henkelman. The other mathematics educators on the committee are Professors James Fey, Neil Davidson and Martin Johnson. I am grateful that you can take the time to react to these questions. Developing this multiple choice instrument is the major component of my dissertation.

As my most recent graduate and professional work has involved the van Hiele model of the development of geometric thought, I have had occasion to examine and use several of the instruments currently available for assessing an individual's level of geometric thinking. Of these, the multiple choice instrument developed by the Cognitive Development and Achievement in Secondary School Geometry (CDASSG) project at the University of Chicago appears to be the instrument of choice when trying to identify quickly a van Hiele profile for large groups. Several important concerns arise, however, when interpreting the results from this test. One issue centers around which of the five proposed scoring schemes provides the most accurate assessment of van Hiele levels. A second concern is that the reliability figures provided by the test designers are quite low. A third concern is that the test claims to assess a general level of geometric thinking, yet there is evidence from the research that an individual's van Hiele levels may vary across content areas.

Guided by these considerations, I am attempting to develop a new instrument for assessing the first four van Hiele levels of geometric reasoning. Specifically, I wish to develop a fixed choice response format test covering the topic of quadrilaterals.

One of the first steps in this process is developing a pool of questions. I realize that it is the individual who "has" a level not the material. I have therefore tried to create questions and answers which will elicit level specific thinking. To do this, I have compiled from the literature a list of "indicators" for each level. Within levels, I have subdivided the indicators by the type of geometric skills each indicator represents. Enclosed you will find a copy of these indicators (see blue sheets). There is also a set of questions and answers cross-referenced to the level indicators.

I very much appreciate it that you have indicated that you will read over these questions/answers and comment on their appropriateness. Enclosed you will find a form for responding to each question (see pink sheets). If this is not convenient, please adopt any format which suits you. I would also like your views on the level indicators (see comments attached to level indicator sections). Based on the responses I receive from you and several other experts, I will revise the questions appropriately, then design a prototype instrument for field testing. To assess construct validity and concurrent validity, I will also be administering interview protocols which you and William Burger developed.

After our phone conversation, I realize that you have only a very limited amount of time to spend at this task. If it is convenient, could you return the questions with your comments to me in the enclosed self-addressed envelop around April 30, 1988. Please take a little extra time if need be.

Thank you again for helping me with this research. I hope that this instrument will complement the work you have done, providing a general profile of groups where your interviews provide information about individuals.

Sincerely,

Mary L. Crowley

LEVEL INDICATORS

Attached you will find a list of level indicators. These reflect how an individual at each designated level reasons about geometric topics. This list has been compiled from the following sources:

Burger, W. F. and Shaughnessy, J. M. (1986). Characterizing the van Hiele levels of development in geometry. Journal for Research in Mathematics Education, (17), 31-48.

Geddes, D., Fuys, D & Tischler, R. (1985). An investigation of the van Hiele model of thinking in geometry among adolescents (Grant no. SED 7920640). Washington, D.C.: National Science Foundation.

Hoffer, A. (1981). Geometry is more than proof. Mathematics Teacher, (74), 11-18.

Usiskin, Z. (1982). Van Hiele levels and achievement in secondary school geometry. Chicago: University of Chicago, Cognitive Development and Achievement in Secondary School Geometry Project.

Van Hiele-Geldof, D. (1984). Dissertion of Dina van Hiele-Geldof entitled: Didactics of geometry in the lowest class of secondary school. In D.Geddes, D. Fuys & R. Tischler, An investigation of the van Hiele model of thinking in geometry among adolescents (Grant no. SED 7920640). Washington, D.C.: National Science Foundation.

The source of each indicator is designated by information in the parenthesis at the end of each statement. Within the parenthesis is the first letter of the last name of the source researcher. For example, as the first indicator for the Basic level is followed by an "H", it is cited by Hoffer. The abbreviation "B&S" indicates the Burger and Shaughnessy article: the abbreviation "G" indicates the Geddes et. al. research as the source. "U" and "vH-G" indicate Usiskin and D. van Hiele-Geldof, respectively.

In general there is very little conflict amongst sources. There is, however, one area of ambiguity about which I would like you to comment. This is the "equivalence of definitions". Geddes et al. (p. 76), on the strength of Dina van Hiele-Geldof's work, say that understanding equivalence of definitions is a level III characteristic. Pierre van Hiele is cited by Usiskin (p. 11) as stating that equivalence in a logical sense is level II. ('The understanding of implication, equivalence, negation of an implication belongs to the second thought level.') Burger and

Shaughnessy (p. 44) identify "the ability to accept equivalent forms of definitions" as a level II characteristic. I, too, have placed this in level II. What would you suggest?

With one variation, I have also adopted Hoffer's cross categorization of geometric skills for each level. He identifies five areas of basic geometric skills: visual, verbal, drawing, logical and applied. I changed "drawing" to "representational". I envision this latter skill as including drawing, working with models, measuring, etc.--all concrete activity. I feel that the sub-categorizing will be especially helpful when selecting representative questions for the instrument.

Would you look over these descriptors? Please feel free to comment on their wording, on their accuracy, and on any other aspect which in your opinion might help me.

Basic Level (Level 0): Visualization

The student reasons about basic geometric concepts, such as simple shapes, primarily by means of visual considerations of the concept as a whole without explicit regard to properties of its components. (B&S). The student does NOT think of properties as characterizing a concept. (G)

The student:

Visual

0.01. identifies instances of a figure by its appearance as a whole:in a simple drawing, diagram, or set of cutouts (e.g. squares, right angles). (H)

0.02. recognizes information labeled on a figure. (H)

Verbal

0.03. names or labels shapes and other geometric figures appropriately using standard and/or nonstandard names and labels. (H, G)

0.04. interprets sentences which describe figures. (H)

0.05. verbally describes shapes by their appearance as a whole (e.g. a rectangle "looks like a window", a parallelogram "looks like a slanty rectangle", an angle "looks like hands on a clock"). (G)

0.06. sometimes includes irrelevant attributes when identifying and describing shapes, such a orientation of the figure on the page. (B&S)

Representational

0.07. constructs, draws, or copies a shape (on a geoboard, on dot/graph/grid/plain paper).(G)

0.08. operates on shapes by folding, measuring, coloring, constructing, manipulating (e.g. making patterns with pattern blocks or by coloring a triangular gird; solving a geometric puzzle).(G)

Logical

0.09. realizes there are differences and similarites among figures. (H)

0.10. understands the conservation of the shape of figures in various positions. (H, G)

Applied

0.11. compares and sorts shapes on the basis of their appearance as a whole (e.g. on an "it looks like basis) (G, H): may be inconsistent, e.g. sorting by properties not shared by sorted type. (B&S)

0.12. recognizes shapes and other geometric figures in different positions/orientations.(H)

0.13. recognizes shapes and other geometric figures: (G,H)
  a. in a photograph or physical object;
  b. in a shape (e.g. angles in a quadrilateral or in two intersecting lines; shapes in a pattern of a triangular grid; edges, faces, vertices of a cube).

0.14. solves routine problems by operating on shape--using observation,measuring, counting, overlays, etc.,-- rather than by using properties which apply in general. (e.g. finds area of a shape by covering it with tiles or counting squares on a grid overlay; trial and error).(G)

Level I: (Analysis)

The student reasons about geometric concepts by means of an informal (empirical) analysis of component parts and attributes. Necessary properties of the concept are established. (B&S) Properties are used to solve problems. (The student does not see how properties are interrelated; does not formulate and use formal definitions; does not explain subclass relationships; does not see need for logical explanations of generalizations discovered empirically) (G)

The student:

Visual

1.01. notices properties of a figure. (H)

1.02. based on properties, identifies a figure as part of a larger, complex figure. (H)

Verbal

1.03. recalls and uses appropriate vocabulary for components and relationships (e.g. opposite sides, corresponding angles are congruent, diagonals bisect each other). (G)

1.04. describes a class of figures (e.g.parallelograms) in terms of its properties. (G)

1.05. may describe types of shapes by explicit use of their properties, rather than by type names, even if known. (B & S)

Representational

1.06. finds and tests relationships among components of a figure (e.g. congruence of opposite sides of a parallelogram; congruence of angles in a tiling pattern) by measuring, drawing, coloring (G); treats geometry as physics. (B&S)

1.07. interprets and uses a verbal description of a figure in terms of its properties and uses this description to draw/construct the figure. (H, G)

Logical

1.08. understands that figures can be classified into different types. (H)

1.09. realizes that figures have properties and that they can be used to distinguish figures. (H)

1.10. generalizes properties for a class of figures based on empirical discoveries (e.g. angle sum of a triangle is 180 by observing several examples). (G)

1.11. applies a list of necessary properties instead of determining sufficient properties when identifying shapes, explaining identifications, and deciding on unidentified shapes. (G)

<u>Applied</u>

1.12. identifies a shape given certain properties. (G)

1.13. sorts shapes (in different ways) according to certain properties; when sorting, usually uses a single attribute e.g. properties of sides while neglecting angles, symmetry, etc. (B&S)

1 14. identifies which properties used to characterize one class of figures also apply to another class of figures: compares classes of figures according to their properties (e.g. notes how a square and rectangle are alike and different in terms of sides and angles) (G,H) but prohibits class inclusion. (B&S)

1.15. interprets verbal or symbolic (e.g. a=bh) statements of rules and applies them. (G)

1.16. rejects textbook definitions of shapes in favor of personal characteristics. (B&S)

1.17. discovers properties of an unfamiliar class of figures. (G)

1.18. solves geometric problems by using known properties of figures or by insightful approaches. (G)

<u>Level II: (Abstraction)</u>

The student logically orders the properties of concepts, forms abstract definitions, can distinguish between the necessity and sufficiency of a set of properties in determining a concept. (B&S). The student does not grasp the meaning of proof in an axiomatic sense and cannot yet establish interrelationships between networks of theorems. (G)

The student:

<u>Visual</u>

2.01. recognizes interrelationships between different types of figures. (H)

<u>Verbal</u>

2.02. makes explicit references to definitions. (B&S)

2.03. formulates sentences showing interrelationships between figures. (H)

2.04. uses language of comparison, quantification and implication: "all", "some", "every", "none" "at least" (G) "if...then", "provided that", "since", "because", "so" (B&S, G)

<u>Representational</u>

2.05. given certain figures, is able to construct other figures related to the given ones. (H)

<u>Logical</u>

2.06. formulates complete definitions. (G, H)

2.07. recognizes equivalence of definitions. (B&S)

2.08. accepts logical partial ordering among types of shapes, including class inclusion. (B&S)

2.09. forms correct informal deductive arguments, generally supported with evidence obtained empirically (G); implicitly uses logical forms such as chain rule and modus ponens. (B&S)

2.10. follows simple deductive argument (G)

2.11. informally recognizes differences between a statement and its converse as opposites (G)

<u>Applied</u>

2.12. applies definitions (G); modifies definitions. (B&S)

2.13. immediately accepts and uses definitions of new concepts. (B&S)

2.14. identifies or gives minimum sets of properties which can characterize a concept. (G)

2.15. orders and interrelates properties (G); can deduce one property from another. (U)

2.16. uses properties to determine if one class of figures is contained in another class. (H)

2.17. sorts shapes according to a variety of mathematically precise attributes. (B&S)

2.18. gives informal arguments (using diagrams, cutouts shapes, other materials) (G); discovers new properties by simple deduction (usually based, at least partially, on empirical evidence). (G)

2.19. sometimes gives more than one correct explanation, argument. (G)

2.20. follows a simple deductive argument, perhaps supplying parts of the argument. (G)

2.21. summarizes or give a variation of a simple deductive arguement. (G)

2.22. on the strength of general theorems, can deduce facts. (DvH-G)

2.23. identifies and uses strategies of insightful reasoning to solve problems. (G)

Level III (Deduction)

The student reasons formally within the context of a mathematical system, complete with undefined terms, axioms, an underlying logical system, definitions and theorems. (B&S)

The student:

Visual

3.01. uses information about a figure to deduce more information. (H)

3.02. recognizes when and how to use auxiliary elements in a figure. (H)

Verbal

3.03. gives examples of undefined terms, definitions, postulates, and theorems; can explain interrelationships. (G)

3.04. recognizes what is given in a problem and what is required to find or do (H); clarifies ambiguous questions and rephrases problem tasks into precise language. (B&S)

3.05. conjectures frequently and attempts to verify conjectures deductively. (B&S)

Representational

3.06. deduces from given information how to draw or construct a specific figure. (H)

Logical

3.07. recognizes need for and structure of undefined terms, definitions, postulates, theorems (G); implicitly accepts postulates of Euclidean geometry. (B&S)

3.08. recognizes characteristics of a formal definition (e.g. necesssary and sufficient conditions)

3.09. uses rules of logic to develop proof. (H)

3.10. deduces consequences from given information. (H)

3.11. relies on proof as the final authority in deciding the truth of a mathematical proposition. (B&S)

Applied

3.12. deduces properties of objects from given or obtained information (H); (includes proving relationships which were explained informally on level II). (G)

3.13. proves relationships between a theorem and related statements (e.g. converse, inverse, contrapositive). (G)

3.14. establishes interrelationships among networks of theorems. (G)

3.15. establishes a general principle that unifies several different theorems. (G)

3.16. solves problems that relate objects. (H)

3.17. investigates the effects of changing an initial postulate in a logical sequence. (G)

3.18. creates proofs from simple sets of axioms frequently using a model to support arguments (G)

3.19. generates, compares and contrasts different proofs of theorems (G)

QUESTION POOL

Attached is a set of multiple choice questions. In most cases, I have written these question myself. The major exceptions are that I have included a few questions from the test developed by the CDASSG project at The University of Chicago. I am trying to identify with more specificity than that project provided which objectives these questions meet.

For all questions, following each "correct" answer are references to the level specific indicators I believe that response reflects. (A list of all level indicators should be enclosed and on pink paper.) The level of each answer is indicated by the digit in the units position; the indicator within that level is indicated by the digits following the decimal. Thus, for question #1, answer C reflects two indicators. These are both at level 0, the Basic level. The answer corresponds to the Basic level indicator 0.01 (identifies instances of a figure ...) and indicator 0.09 (realizes there are differences ...).

Basic Level

(0. 01, 0. 09)

indicator 1 indicator 9

In question #12, answer C corresponds to level I, indicator 3 (recalls and uses appropropriate ...)

Level I

(1. 03)

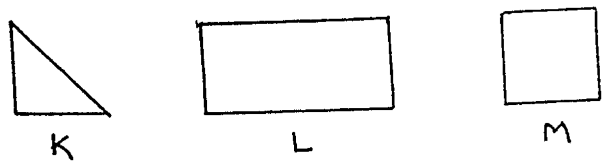
indicator 3

Some questions will have several answers which correspond to level indicators from different levels, e.g. question #8.

Enclosed you should find a set of pink papers. If you find it convenient, use these sheets to record you reaction to each question. I would like your opinion on whether or not these questions and answers require the thinking skills which I have designated. If you think that I have mislabeled the answer, please indicate what in your opinion is the correct corresponding indicator. If I have completedly misjudged a question/answer please indicate how. This will help me in making revisions.

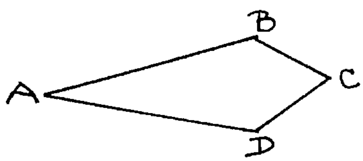
Thank you

1. Which of these are squares?



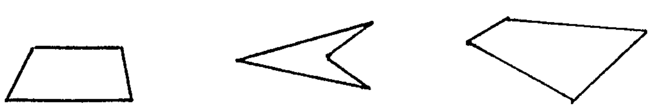
(A) K only
(B) L only
(C) M only (0.01, 0.09)
(D) L and M only
(E) All are squares.

2. In the figure ABCD, the part called $\overline{AB}$ is a



(A) Side (0.02, 0.03)
(B) Slant
(C) Corner
(D) Vertex
(E) Diagonal

3. Which term names all three shapes:

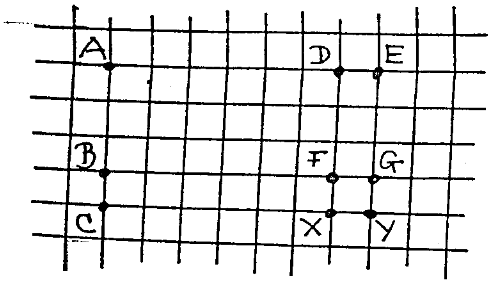


(A) Quadrilateral (0.01, 0.03, 0.09)
(B) Quadrangle
(C) Quadrant
(D) Quadruple
(E) None of (A) - (D) is correct.

4. In rectangle ABCD , where the vertices are labeled in clockwise order, what are the line segments AC and BD called?

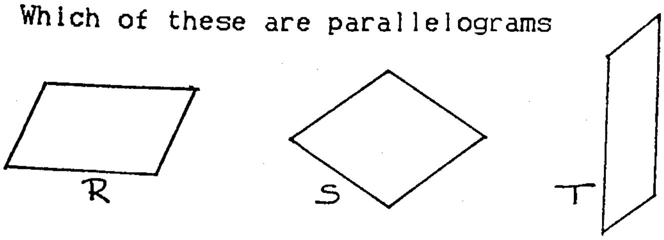
(A) Edges
(B) Slants
(C) Diagonals (0.03, 0.04)
(D) Intersectors
(E) Perpendiculars

5. When connected, which set of points result in a rectangle with side lengths of 4 and 7 units?



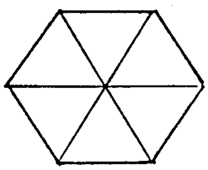
(A) ADFB
(B) ADXC
(C) AEYC (0.07)
(D) AEGB
(E) No set of points form the rectangle.

6. Which of these are parallelograms



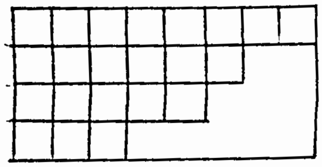
(A) R only
(B) R and S only
(C) R and T only
(D) All of these are parallelograms (0.01, 0.09, 0.10, 0.12)
(E) None of these are parallelograms

7. What 4 sided shape do you see in this figure?



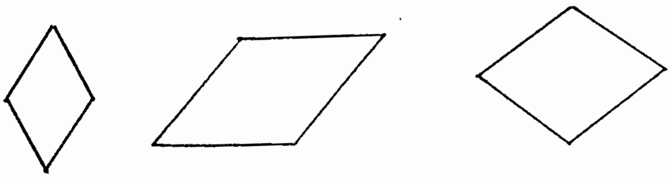
(A) Square
(B) Triangle
(C) Rectangle
(D) Parallellogram (0.01, 0.09, 0.13)
(E) None of the above.

8. To determine the area of the rectangle, someone has started to cover it with square tiles. How would you complete the task?



(A) Ask what area means.
(B) Apply the Laws of Pythagoras
(C) Cover the entire figure with tiles, then count them. (0.08, 0.14, 0.15)
(D) Add up the number of tiles it takes to go around the edges of the figure.
(E) Stop covering with tiles because there is enough information available to use the formula "Length x Width". (1.15)

9. A rhombus is a four sided figure with all sides the same length. Here are three examples.



Which of the statements (A) to (D) about the diagonals of any rhombus is false?

(A) The diagonals bisect each other.
(B) The diagonals are lines of symmetry.
(C) The two diagonals are perpendicular.
(D) The two diagonals have the same length. (1.01, 1.03, 1.04, 1.09, 1.10)
(E) Each diagonal bisects two angles of the rhombus.

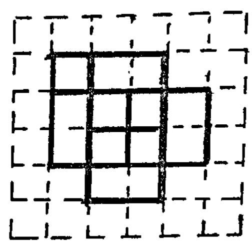
10. Consider the following properties of a four sided figure:

1. Opposite sides are equal.
2. Diagonals are equal.
3. Opposite angles are equal.

These properties are always true for which type of figure?

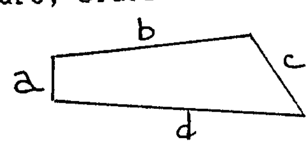
(A) Quadrilateral
(B) Parallelogram
(C) Rectangle (1.08, 1.09, 1.12)
(D) Kites
(E) Tetrahedron

11. How many squares are in this picture



(A) 5 (0.13)
(B) 9
(C) 10
(D) 11 (1.02, 1.18)
(E) 13

12. In the figure, sides a and b are



(A) images
(B) parallel
(C) adjacent (1.03)
(D) perpendicular
(E) corresponding

13. Which of (A) to (D) is false in some rectangles?

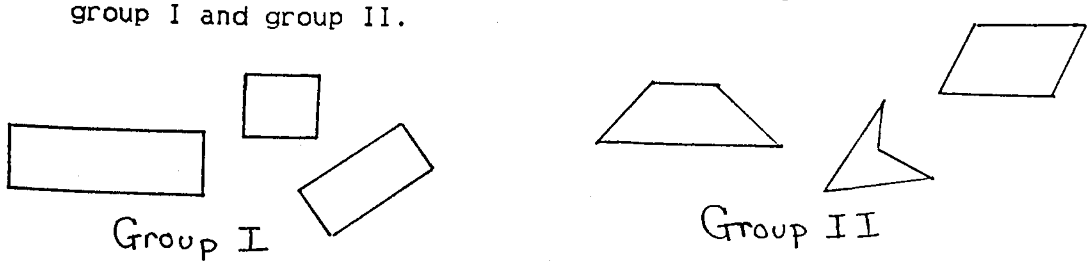
(A) There are four sides.
(B) There are four right angles.
(C) The diagonals have the same length.
(D) The opposite sides have the same length.
(E) All of the above are true in every rectangle.(1.04, 1.09)

14. Which combination of statements is the shortest list needed to guarantee that a four sided closed figure is a rectangle.

Statement 1: two long sides, two short sides
Statement 2: opposite sides the same length
Statement 3: opposite sides parallel
Statement 4: one angle is a right angle
Statement 5: all 4 angles are right angles.

(A) 1
(B) 2, 3
(C) 3, 4 (2.14)
(D) 1, 2, 3, 5 (1.11)
(E) None of these combinations describe a rectangle

15. A set of six shapes was sorted into the two groups shown here, group I and group II.



What characteristic can be used to describe why figures were put into group I.

(A) All the corners are even (0.11)
(B) Adjacent sides are equal
(C) The opposite sides are parallel
(D) All the figures are quadrilaterals
(E) No angle is greater than 90 degrees (1.13)

16. The area of a rhombus is calculated by

$$\text{Area} = 1/2\ (d_1 \times d_2)$$

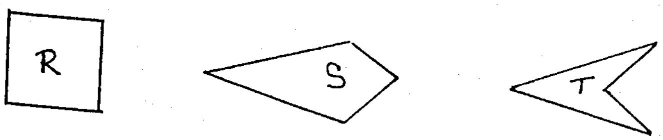
where d_1 and d_2 are the lengths of the diagonals. What is the area of a rhombus ABCD when AB = x, BC = x, AC = y and BD = z

(A) $1/2x^2$
(B) $1/2yz$ (1.15)
(C) $1/2xy$
(D) $1/2xz$
(E) There is not enough information

17. What do all squares have that some parallelograms do not have?

(A) Opposite sides equal
(B) Opposite angles equal
(C) Opposite sides parallel
(D) Diagonals bisect each other
(E) Both have all of the above (1.04., 1.14)

18. Which of the following figures have at least one set of adjacent sides congruent?

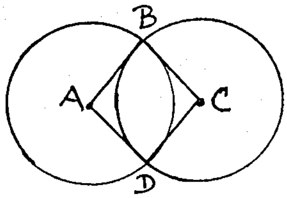


(A) R only
(B) S only
(C) T only
(D) R and S
(E) R, S and T (1.01, 1.03, 1.09, 1.13)

19. What is the measure of an angle in a parallelogram if it is 30 degrees less than twice its opposite angle.

(A) 15
(B) 30 (1.18)
(C) 60
(D) 90
(E) 150

20. Two circles intersect in such a way that the figure ABCD is formed when the centers of the circles and the points of intersection are connected. AB=BC=CD=DA.



Which of the following could be used to show that BD is perpendicular to AC?

(A) Properties of a square
(B) Properties of a rhombus (1.02, 1.12, 1.18)
(C) Properties of a rectangles
(D) Properties of a parallelogram
(E) None of these

21. Which of the following statements about parallelograms is always true:

(A) The diagonals are congruent.
(B) The diagonals are perpendicular.
(C) The adjacent sides are congruent.
(D) The opposite angles are congruent.(1.03, 1.04)
(E) The opposite angles are supplementary.

22. Which quadrilateral always has 3 sides equal?

(A) A kite
(B) A square (1.08, 1.12)
(C) A rectangle
(D) An equilateral triangle
(E) None of the above.

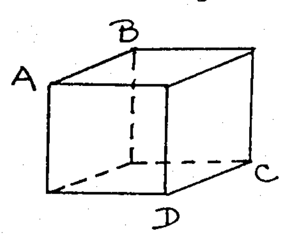
23. In rectangle PQRS, diagonal PR bisects angle SPQ. If PQ = 10, how long is PS?

(A) 5
(B) 10 (2.08, 2.23)
(C) 20
(D) $10\sqrt{2}$
(E) There is not enough information to determine this.

24. Which of the following is or are sufficient (enough) information to determine that a four sided figure is a parallelogram?

(A) Opposite sides are equal
(B) Opposite sides are parallel
(C) Both (A) and (B) are needed (1.11)
(D) Either (A) or (B) is sufficient (2.14)
(E) None of the above.

25. A cube is a 3-dimensional figure with 6 sides (faces), each of which is a square. The faces are perpendicular to each other. What would be the shape of the plane figure ABCD which results from cutting the cube through vertices A, B, C and D?



(A) Square
(B) Rectangle (2.01, 2.23)
(C) Trapezoid
(D) Either A or B
(E) Not enough information

26. What type of a figure can be called both a rhombus and a rectangle?

(A) Square (2.08)
(B) Rhombus
(C) Rectangle
(D) Parallelogram
(E) No figure

27. Which is true?

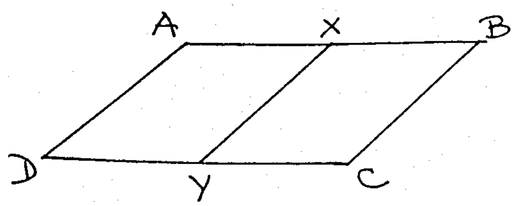
(A) All properties of rectangles are properties of all squares (2.04, 2.15)
(B) All properties of squares are properties of all rectangles
(C) All properties of rectangles are properties of all parallelograms
(D) All properties of squares are properties of all parallelograms
(E) None of (A) to (D) is true

28. An isosceles trapezoid is a quadrilateral in which exactly two sides are parallel and the other 2 sides are equal. The parallel sides are called the bases. Base angles of an isosceles trapezoid are the angles which share the same base as an arm (or side). The angles in each pair of base angles are congruent.

Question: If M is an angle in an isosceles trapezoid, what can be said about the measure (size) of an adjacent angle.

(A) It is supplementary to angle M
(B) It has the same measure as angle M.
(C) Not enough information to determine
(D) Either A or B (2.12 or 2.13, 2.23)
(E) Either B or C

29. On the basis of what is presented, choose which reason, (A) to (E) could most appropriately be used to justify step 6 in the following proof.



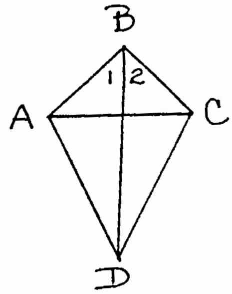
ABCD is a rhombus, X, Y are midpoints of AB and CD respectively
Show that AXYD is a parallelogram

1. ABCD is a rhombus	1. Given
2. AB=DC, AB //CD	2. Definition of a rhombus
3. X,Y midpoints of AB and CD	3. Given
4. AX = 1/2 AB, DY = 1/2 YB	4. Definition of midpoint
5. AX = DY	5. Halves of equals are equal
6. AXYD is a parallelogram	6.

(A) Given
(B) Both sets of opposite sides are parallel
(C) One set of sides is equal and parallel (2.10. 2.20)
(D) Both sets of opposite sides are equal
(E) None of the above

30. What property or properties of kites is established by this proof?

Given: ABCD is a kite



1. ABCD is a kite
2. AB = BC and AD = CD
3. BD = BD
4. $\triangle ABD \cong \triangle BCD$
5. $\angle 1 = \angle 2$
6. In $\triangle ABC$, $BD \perp AC$

1. Given
2. Definition of kite
3. Reflexive
4. SSS
5. CPCTE
6. Bisectors of vertex $\angle$ of isosceles $\triangle$'s

(A) A kite is a figure with two sets of adjacent sides congruent
(B) If a quadrilateral is a kite, the diagonals are perpendicular (2.10, 2.21)
(C) If the diagonals of a quadrilateral are perpendicular, the figure is a kite.
(D) If a figure contains two congruent triangles, the perpendiculars bisect.
(E) All of the above

31. Here are three properties of a figure

property D: It has diagonals of equal length
property S: It is a square
property R: It is a rectangle

Which is true:
(A) D implies S which implies R
(B) D implies R which implies S
(C) S implies R which implies D (2.08, 2.15)
(D) R implies D which implies S
(E) R implies S which implies D

32. Figure A is defined by definition A. Figure B is defined by definition B.

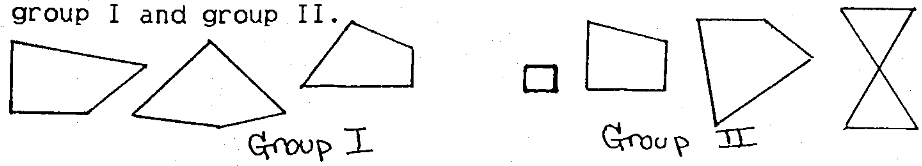
Definition A: A quadrilateral with exactly one pair of parallel sides
Definition B: A quadrilateral with at least one pair of parallel sides

Which of the following statements is true?
(A) The two definitions are the same.
(B) All figures defined by definition A are also defined by definition B. (2.04, 2.07, 2.13)
(C) All figures defined by definition B are also defined by definition A.
(D) No figure defined by definition A is also defined by definition B.
(E) No figure defined by definition B is also defined by definition A.

33. When working with a PARALLELOGRAM, which of (A) to (C) is FALSE?

(A) If told the diagonals are congruent, then you know that they bisect.
(B) If told all four sides are equal then you know that the opposite sides are equal
(C) If told at least one angle is a right angle, then you know all the angles are right angles.
(D) Both (A)and(C) are false
(E) None of (A) - (C) above is false (2.15)

34. A set of shapes was sorted into the two groups shown here, group I and group II.



What characteristic do all figures in group I have which no figure in group II has?

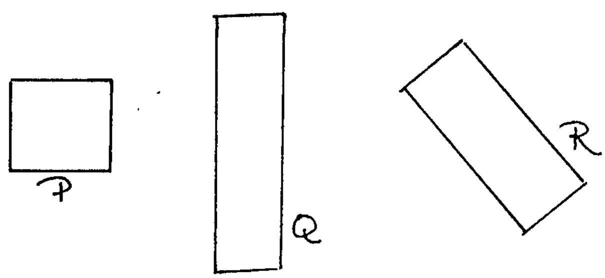
(A) Exactly one right angle. (2.17)
(B) At least one right angle.
(C) At most one right angle.
(D) No right angles.
(E) None of the above.

35. Here are two statements about a quadrilateral.

Statement 1: Quadrilateral QRST has 4 sides of the same length.
Statement 2: The opposite angles in quadrilateral QRST are equal.

Which is correct?
(A) Statments 1 and 2 cannot both be true.
(B) If 1 is true, then 2 is true (2.04, 2.15)
(C) If 2 is true, then 1 is true
(D) If 1 is false, then 2 is true
(E) If 2 is false, then 1 is true

36. Which of these can be called rectangles?



(A) All can (2.01, 2.08)
(B) Q only
(C) R only
(D) P and Q only
(E) Q and R only

37. A certain shape has both sets of opposite sides parallel and diagonals which are equal but not perpendicular. To which class of figures might this shape belong?

(A) Kite
(B) Square
(C) Rhombus
(D) Rectangle (2.16)
(E) Trapezoid

38. Working from the fact that the sum of the angles of a quadrilateral is 360 degrees, what would you say is the sum of the angles of a 6 sided figures? (Some examples are given below)

(A) This cannot be determined
(B) 360 degrees
(C) 540 degrees
(D) 720 degrees (2.18, 2.22)
(E) 1080 degrees

39. Two geometry books define the word rectangle in different ways.
Which is true?

(A) One of the books has an error.
(B) One of the definitions is wrong. There cannot be two different definitions for rectangle.
(C) The rectangles in one of the books must have different properties from those in the other book.
(D) The rectangles in one of the books must have the same properties as those in the other book.
(E) The properties of rectangles in the two books might be different. (2.07)

40. Consider the following suggested definitions for a parallelogram:

Definition 1: A parallelogram is a quadrilateral in which both pairs of opposite sides are parallel.

Definition 2: A parallelogram is a quadrilateral in which both pairs of opposites sides are congruent.

Which statement about these definitions is true?

(A) The definitions are equivalent. (2.07)
(B) Only one definition can be correct.
(C) Definition 1 is a partial definition.
(D) Definition 2 is a partial definition.
(E) Neither is a complete definition.

41. Which of (A) - (D) starts with the same idea statement I ends with and ends with the idea statement I starts with?

Statement I: When two sides of a quadrilateral are parallel to each other and congruent, the figure is a parallelogram.

(A) When two sides of a quarilateral are parallel to each other, the figure is a parallelogram
(B) When two sides of a parallelogram are parallel to each other and congruent, the figure is a quadrilateral.
(C) When a figure is a parallelogram, two sides are parallel.
(D) When a figure is a parallelogram, two sides are parallel and congruent. (2.11)
(E) None of the above

42. Consider these two statements

Statement X: A rectangle is a parallelogram with a right angle
Statement Y: A rectangle with perpendicular diagonals is a square

Which of the following sentences is true?

(A) X and Y are definitions
(B) X and Y are theorems
(C) X and Y are postulates
(D) X is a definition, Y is a theorem (3.07)
(E) X is a postulate, Y is a definition

43. A proof is a list of statements together with a justification for each statment which ends up with the desired conclusion. Which of the following is not a proper type of justification.

(A) Axiom
(B) Given
(C) Theorem
(D) Definition
(E) Measurement (3.07)

44. Which statement is true?

(A) Any statment which seems true should become a postulate.
(B) Theorems are proved only on the basis of definitions and undefined terms.
(C) It is possible to define each geometric term by using simpler geometric terms.
(D) Exact geometric reasoning leads to geometric truths that cannot be deduced with absolute certainty from measurement. (3.07, 3.11)
(E) More than one of the above is true.

45. Here are two statements

I. If a figure is a rectangle, then its diagonals bisect each other.
II. If the diagonals of a quadrilateral bisect each other, the figure is a rectangle.

Which is correct?

(A) To prove I is true, it is enough to prove that II is true.
(B) To prove II is true, it is enough to prove that I is true.
(C) To prove II is true, it is enough to find several rectangles whose diagonals bisect each other.
(D) To prove II is false, it is enough to find one non-rectangle whose diagonals bisect each other. (3.13)
(E) None of (A) - (D) is correct

46. Which of the statements (A) to (C) is an accurate restatement of this fact:

A quadrilateral whose diagonals bisect each other is a parallelogram

(A) If a quadrilateral is a parallelogram, then the diagonals bisect each other.
(B) If the diagonals of a parallelogram bisect each other, then the figure is a quadrilateral
(C) If the diagonals of a quadrilateral bisect each other, then the figure is a parallelogram. (3.04)
(D) Both (A) and (C)
(E) All of the above are accurate statements.

47. What is assumed (given) and what is to be shown (proved) in the following statement: A quadrilateral with supplementary adjacent angles is a parallelogram.

(A) Given: A parallelogram
Prove: the adjacent angles are supplementary

(B) Given: A quadrilateral
Prove: the adjacent supplementary angles are a parallelogram

(C) Given: A parallelogram with supplementary angles
Prove: the angles are adjacent

(D) Given: A quadrilateral with adjacent angles supplementary
Prove: the figure is a parallelogram (3.04)

(E) Given: A quadrilateral with supplementary angles
Prove: the figure is a parallelgram with adjacent angles

48. Consider the following statements

Statement I: If a quadrilateral is convex then condition A holds

Statement II: If condition A holds, then the quadrilateral is convex

Statement III: A quadrilateral is convex if and only if condition A holds.

Which of the following is correct?

(A) Statment I and II say the same thing
(B) Statement I and III say the same thing
(C) All three statements say the same thing
(D) If statement III is true then both statement I and statement II are true (3.08)
(E) There is not enough information to judge

49. Which condition will show that a quadrilateral is a rhombus without first showing that it is a parallelogram.

(A) If it contains a consecutive pair of sides that are equal
(B) If either diagonal bisects two angles
(C) If the diagonals are perpendicular bisectors of each other (3.12, 3.17)
(D) All of the above
(E) None of the above

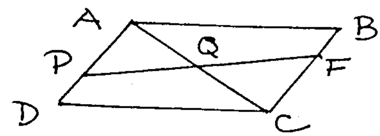
50. Suppose you have proved statements I and II.

I. If p, then q.
II. If s, then not q.

Which statement follows from statements I and II?

(A) If p, then s.
(B) If not p, then not q.
(C) If p or q, then s.
(D) If s, then not p. (3.09, 3.13)
(E) If not s, then p.

51. Figure ABCD is a parallelogram. AP and CF are congruent.



Which of the following strategies can be used to prove or disprove the conclusion that PQ = FQ and AQ = CQ

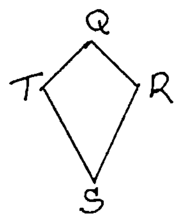
(A) Similar triangles
(B) The midpoint theorem
(C) The diagonals of a parallelogram bisect (3.04, 3.10, 3.12)
(D) Corresponding parts of congruent triangles
(E) If the diagonals of a quadrilalteral are equal, the figure is a parallelogram

52. What conclusions can be drawn from the following true statements?

Statement 1: If P is true, then Q is true .
Statement 2: If R is true, then S is not true.
Statement 3: If Q is true, then S is true.
Statement 4: P is true.

(A) S is true; R is True
(B) S is true; R is False (3.09)
(C) S is false; R is True
(D) S is false; R is True
(E) Only S is true (2.10)

53. Given: Quadrilateral QRST with QR = QT and $\angle R = \angle T$
Prove: SR = ST



To complete the proof, it would be useful to

(A) introduce segment RT (3.01, 3.02)
(B) introduce segment QS
(C) either (A) or (B).
(D) both (A) and (B).
(E) neither (A) or (B).

SAMPLE RESPONSE PAGE

<u>Question 1</u>:

a. Do the question and answer in #1 test the specified level indicators? Yes_____ No______

b. If not, why not?

c. How can this question/answer be clarified, revised or otherwise improved?

<u>Question 2</u>:

a. Do the question and answer in #2 test the specified level indicators? Yes_____ No______

b. If not, why not?

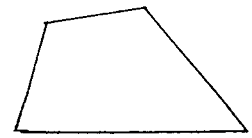
c. How can this question/answer be clarified, revised or otherwise improved?

Appendix C

Quadrilateral Guidelines

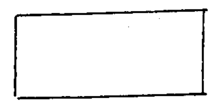
Definitions of Quadrilaterals

A QUADRILATERAL is a four sided polygon

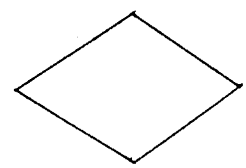


A PARALLELOGRAM is a quadrilateral in which both pairs of opposite sides are parallel

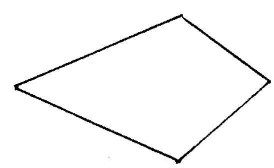
A RECTANGLE is a parallelogram in which at least two consecutive sides are congruent.



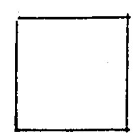
A RHOMBUS is a parallelogram in which at least two consecutive sides are congruent.



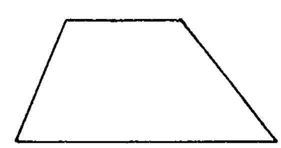
A KITE is a quadrilateral with two distinct pairs of congruent consecutive sides.



A SQUARE is a parallelogram that is both a rectangle and a rhombus.



A TRAPEZOID is a quadrilateral with exactly one pair of parallel sides. The parallel sides are called BASES of the trapezoid. (Sometimes, the TRAPEZOID is defined by "at least" one pair of parallel sides.)

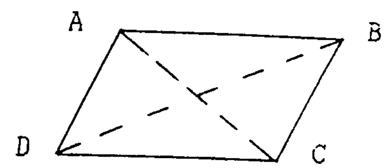


PROPERTIES OF QUADRILATERALS

These properties are derived from the previously listed definitions.

PROPERTIES OF PARALLELOGRAMS:

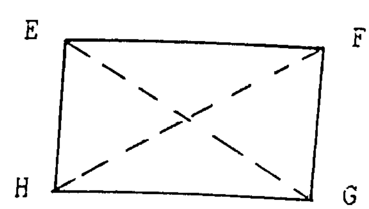
In a parallelogram



1. the opposite sides are parallel (by definition). — $AB \parallel CD$, $BC \parallel DA$
2. the opposite sides are congruent. — $AB = CD$, $BC = DA$
3. the opposite angles are congruent. — $\angle DAB \cong \angle BCD$, $\angle ABC \cong \angle CDA$
4. the diagonals bisect each other. — AC and BD bisect each other
5. any pair of consecutive angles are supplementary. — $\angle DAB$ and $\angle ABC$ are supplementary

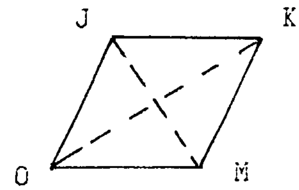
PROPERTIES OF RECTANGLES:

In a rectangle-



1. all the properties of a parallelogram apply (by definition).
2. all angles are right angles. — $\angle E$, $\angle F$, $\angle G$, $\angle H$ are right angles
3. the diagonals are congruent. — $EG = FH$

PROPERTIES OF RHOMBI:
In a rhombus-



1. all the properties of a parallelogram apply (by definition).

2. all sides are congruent (a rhombus is equilateral).

   JK = JO = OM = MK

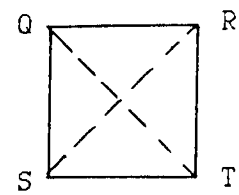
3. the diagonals bisect the angles of the polygon

   JM bisects $\angle$ OMK and $\angle$ OJK: OK bisects $\angle$ JOM and $\angle$ MKJ

4. the diagonals are perpendicular bisectors of each other.

   JM $\perp$ OK, JM bisects OK, and vice versa

PROPERTIES of SQUARES:
In a square--



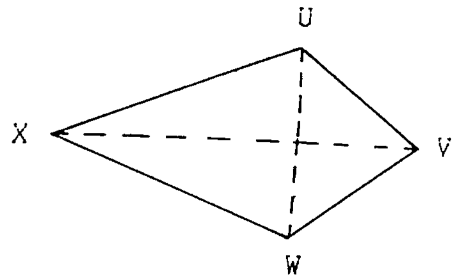
1. all the properties of a rectangle apply (by definition).

2. all the properties of a rhombus apply (by definition).

3. the diagonals form four isosceles right triangles.

   $\triangle$QTS, $\triangle$QRS, $\triangle$TQR, $\triangle$RST are all right, isosceles

PROPERTIES OF KITES:
In a kite-



1. the distinct pairs of consecutive sides are congruent (by definition) — UV = VW, XW = XU
2. one of the diagonals is the perpendicular bisector of the other diagonal — XV $\perp$ bisector of UW
3. if the kite is also a rhombus or a square, it inherits the properties of those figures.

EXAMPLES OF NECESSARY AND SUFFICIENT CONDITIONS: Proving that figures are special quadrilaterals:

Proving that a quadrilateral is a PARALLELOGRAM

1. if both pairs of opposite sides of a quadrilateral are parallel, then the quadrilateral is a parallelogram.

2. if both pairs of the opposite sides of a quadrilateral are congruent, then the quadrilateral is a parallelogram.

3. if two sides of a quadrilateral are both parallel and congruent, then the quadrilateraal is a parallelogram.

4. if the diagonals of a quadrilateral bisect each other, then the quadrilateral is a parallelogram.

5. if both pairs of opposite angles of a quadrilateral are congruent, then the quadrilateral is a parallelogram.

Proving that a quadrilateral is a RECTANGLE

If is can be shown the quadrilateral is a parallelogram then...

1. if a parallelogram contains at least one right angle, then it is a rectangle.

2. if the diagonals of a parallelogram are congruent, then the parallelogram is a rectangle.

Proving that a quadrilateral is a RHOMBUS

If it can be shown that the quadrilateral is a parallelogram then...

1. if a parallelogram contains a consecutive pair of sides that are congruent, then it is a rhombus.

2. if either diagonal of a parallelogram bisects two angles of the polygon, then the parallelogram is a rhombus.

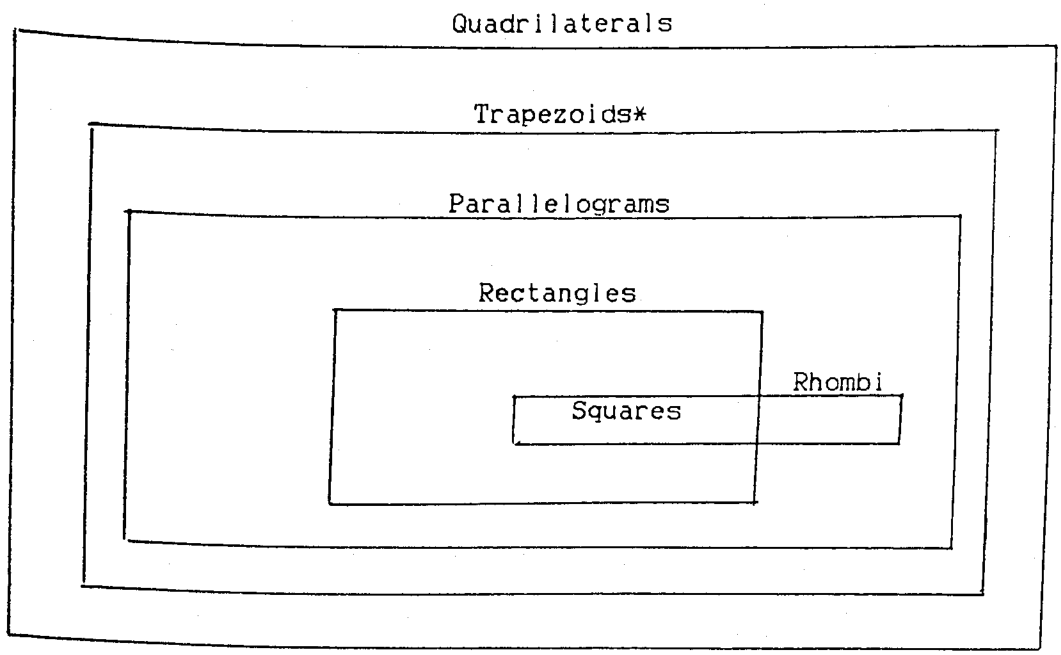
To show that a quadrilateral is a rhombus without first showning that it is a parallelogram:

3. if the diagonals of a quadrilateral are perpendicular bisectors of each other, then the quadrilateral is a rhombus.

Proving that a quadrilateral is a SQUARE

1. If a quadrilateral is both a rectangle and a rhombus, then it is a square.

<u>Figure C.1.</u> Subsets of the regular quadrilaterals



* Trapezoid is defined here as "at least" one set of sides parallel.

Appendix D

Revised Level Indicators

Basic Level (Level 1): Visualization

The student reasons about basic geometric concepts, such as simple shapes, primarily by means of visual considerations of the concept as a whole and without explicit regard to properties of its components. (B&S). He realizes there are differences and similarites among figures. (H) He understands the conservation of the shape of figures in various positions. (H, G)

The student does NOT think of properties as characterizing a concept. (G)

The student:

Verbal

1.01. verbally describes shapes by their appearance as a whole (e.g. a rectangle "looks like a window", a parallelogram "looks like a slanty rectangle", an angle "looks like hands on a clock").(G)

1.02. names or labels shapes and other geometric figures appropriately using standard and/or nonstandard names and labels. (H, G)

1.03. sometimes includes irrelevant attributes when identifying and describing shapes, such a orientation of the figure on the page. (B&S)

Representational

1.04. constructs, draws, or copies a shape (on a geoboard, on dot/graph/grid/plain paper).(G)

1.05. operates on shapes by folding, measuring, coloring, constructing, manipulating (e.g. making patterns with pattern blocks or by coloring a triangular gird; solving a geometric puzzle).(G)

Applied

1.06. identifies shapes and other geometric figures (G/H)
   a. in a simple drawing,
   b. in varying positions/orientations,
   c. in a shape (e.g. angles in a quadrilateral or in two intersecting lines; shapes in a pattern of a triangular grid; edges, faces, vertices of a cube),
   d. in a photograph or physical object (e.g. cutouts).

1.07. compares and sorts shapes
  a. on the basis of their appearance as a whole (e.g. on an "it looks like basis) (G, H),
  b. may be inconsistent (e.g sorting by properties not shared by sorted type). (B&S)

1.08. solves routine problems by operating on shape -- using observation, measuring, counting, overlays, etc. -- rather than by using properties which apply in general (e.g. finds area of a shape by covering it with tiles or counting squares on a grid overlay; trial and error). (G)

Level 2: (Analysis)

The student realizes that geometric concepts have properties and that these properties can be used to distinguish between concepts. (H) He reasons about geometric concepts by means of an informal (empirical) analysis of component parts and attributes. Necessary properties of the concept are established. (B&S)

The student does NOT see how properties are interrelated: does not formulate and use formal definitions; does not explain subclass relationships; does not see need for logical explanations of generalizations discovered empirically. (G)

The student:

Verbal

2.01. recalls and uses appropriate vocabulary for components and relationships (e.g. opposite sides, corresponding angles are congruent, diagonals bisect each other). (G)

2.02. describes a class of figures (e.g.parallelograms) in terms of its properties. (G)

2.03. may describe types of shapes by explicit use of their properties, rather than by type names, even if known. (B&S)

2.04. may reject textbook definitions of shapes in favor of personal characteristics. (B&S)

2.05. explains verbal or symbolic (e.g. a=bh) statements of rules, recognizes when to apply them and does so appropriately. (G)

Representational

2.06. discovers and analyzes relationships among components of a figure (e.g. congruence of opposite sides of a parallelogram; congruence of angles in a tiling pattern) by measuring, drawing, coloring (G); treats geometry as physics. (B&S)

2.07. uses a description of a figure in terms of its properties to draw/construct the figure. (H, G)

Applied

2.08. identifies and test relationships among components of figures (e.g. congruence of opposite sides of a parallelogram) (G)

2.09. based on empirical discoveries, establishes properties for a class of figures (e.g. finds that sum of the angles of a triangle is 180 degrees--by observing several examples). (G)

2.10. given properties, identifies shape(G)

2.11. compares shapes according to their properties (e.g. notes how a square and rectangle are alike and different in terms of sides and angles)

2.12. identifies which properties used to characterize one class of figures also apply to another class of figures (G,H), but prohibits class inclusion. (B&S)

2.13. sorts shapes according to certain properties: when sorting, usually uses a single attribute e.g. properties of sides while neglecting angles, symmetry, etc.; can sort in different ways (B&S)

2.14. when identifying shapes, explaining identifications, and deciding on unidentified shapes, applies a list of necessary properties instead of determining sufficient properties . (G)

2.15. solves geometric problems by using known properties of figures or by insightful approaches. (G)

Level 3: (Abstraction)

The student is able to operate with known relations (vH, 42). He logically orders the properties of concepts; accepts logical partial ordering among types of shapes, including class inclusion. (B&S); uses and forms abstract definitions, can distinguish between the necessity and sufficiency of a set of properties in determining a concept. (B&S) .

The student does NOT grasp the meaning of proof in an axiomatic sense and cannot yet establish interrelationships between networks of theorems. (G)

The student:

Verbal

3.01. makes explicit references to definitions. (B&S)

3.02. formulates sentences showing interrelationships between figures. (H)

3.03. uses language of comparison, quantification and implication: "all", "some", "every", "none" "at least" (G) "if...then", "provided that", "since", "because", "so" (B&S, G)

Representational

3.04. given certain figures, is able to construct other figures related to the given ones. (H)

Applied

3.05. identifies or gives minimum sets of properties which can characterize a concept. (G)

3.06. orders and interrelates properties (G); can deduce one property from another. (U)

3.07. identifies figures which belong to more than one class; uses properties to determine if one class of figures is contained in another class. (H)

3.08. sorts shapes according to a variety of mathematically precise attributes. (B&S)

3.09. Definitions:
   a. applies definitions (G),
   b. modifies definitions, (B&S),
   c. formulates complete definitions (G, H),
   d. immediately accepts and uses definitions of new concepts (B&S),
   e. recognizes equivalence of definitions. (B&S)

3.10. gives informal arguments (using diagrams, cutouts shapes. other materials) (G); discovers new properties by simple deduction (usually based, at least partially, on empirical evidence). (G)

3.11. sometimes gives more than one correct explanation. argument. (G)

3.12. follows a simple deductive argument, perhaps supplying parts of the argument. (G)

3.13. summarizes or give a variation of a simple deductive argument. (G)

3.14. implicitly uses logical forms such as chain rule and modus ponens. (B&S)

3.15. informally recognizes differences between a statement and its converse as opposites (G)

3.16. on the strength of general theorems, can deduce facts. (DvH-G)

3.17. identifies and uses strategies of insightful reasoning to solve problems. (G)

<u>Level 4: (Deduction)</u>

The student reasons formally within the context of a mathematical system, complete with undefined terms. axioms. an underlying logical system, definitions and theorems. (B&S) He recognizes the need for and the structure of undefined terms, definitions, postulates, theorems (G). He implicitly accepts postulates of Euclidean geometry. (B&S) He relies on proof as the final authority in deciding the truth of a mathematical proposition. (B&S)

The student:

<u>Verbal</u>

4.01. gives examples of undefined terms, definitions, postulates, and theorems; can explain interrelationships. (G)

4.02. clarifies ambiguous questions and rephrases problem tasks into precise language. (B&S)

4.03. conjectures frequently and attempts to verify conjectures deductively. (B&S)

<u>Representational</u>

4.04. deduces from given information how to draw or construct a specific figure. (H)

<u>Applied</u>

4.05. identifies what is given in a problem and what is required to find or do (H)

4.06. deduces properties of objects from given or obtained information (H); this includes proving relationships which were explained informally on level II.(G)

4.07. uses proof as the final authority in deciding the truth of a mathematical proposition. (B & S)

4.08. uses rules of logic to develop proof. (H)

4.09. proves relationships between a theorem and related statements (e.g. converse, inverse, contrapositive). (G)

4.10. establishes interrelationships among networks of theorems. (G)

4.11. establishes a general principle that unifies several different theorems (G) or relates objects (H)

4.12. investigates the effects of changing an initial postulate in a logical sequence. (G)

4.13. creates proofs from simple sets of axioms frequently using a model to support arguments. (G)

4.14. generates, compares and contrasts different proofs of theorems. (G)

Appendix E

Pilot Instrument

NAME______________________________

DIRECTIONS

There are 45 written questions in this survey of geometric thinking. You may take as long as you need to answer the questions. No one is expected to answer all of the questions correctly. I am looking for "good" questions and "bad" questions, not trying to find out how smart you are.

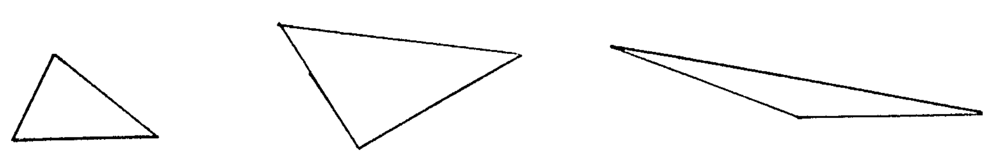
Read each problem carefully. Most questions are multiple choice. Read each choice of answers carefully especially as some examples have combination answer choices such as "All of the above are true", "Some of the above", "(A) and (B) are both true", etc.

* Darken the letter next to your choice of answer (as shown in the examples).
* Erase all incorrectly chosen answers.
* Points are not taken off for incorrectly answered questions.

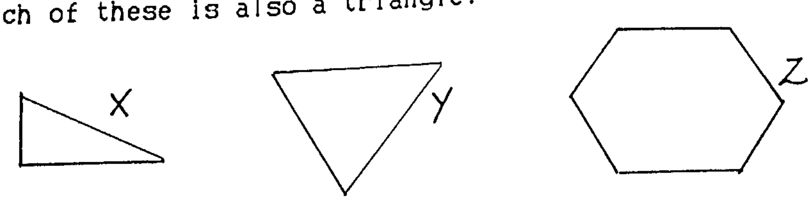
Some example questions are given below.

EXAMPLE #1

These are examples of a figure called a <u>triangle</u>.



Which of these is also a triangle?



note that only one answer choice is darkened

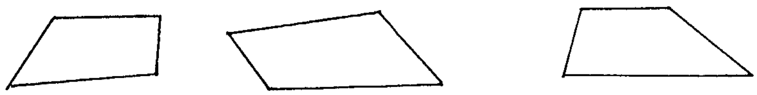
(A) X
(B) Y
(C) Z
(D) X and Y
(E) All of the above are quadrilaterals

note the 'combination' answer choices

A few questions will not have the strict multiple choice format.

EXAMPLE #2 (This problem asks you to explain why you chose your answer. Select an answer and explain your choice.)

These are examples of a figure called a quadrilateral.



Which of these are quadrilaterals?

(A) J
~~(B)~~ K
(C) L
(D) M
(E) N

note there are two parts to answer

EXPLAIN why you chose your answer.

K is the only figure which . . . .

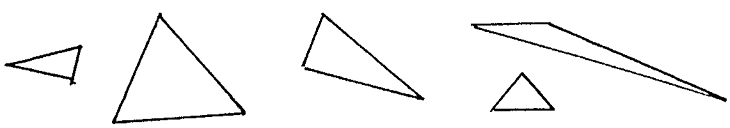
** If you make an educated guess, explain why it was "educated". For example:

I knew it wasn't choice (A) or (B) because ...
or
I know that a rectangle has... but I'm not sure about....

** If you make an uneducated guess, just say so: I guessed!

EXAMPLE #3 (This type just asks you do draw)

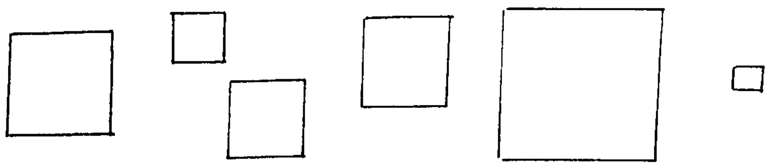
These are examples of figures called a <u>triangle</u>.



Start at point A and draw a triangle.

A •

1. These are examples of a figure called a <u>square</u>.

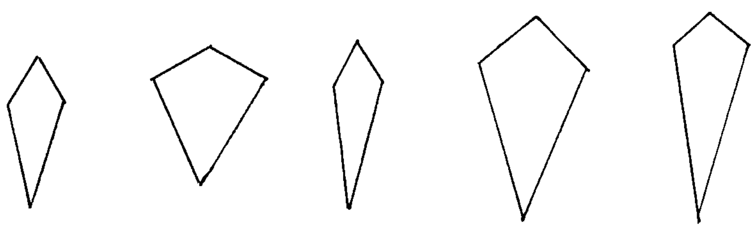


Which of these appear to be a square?

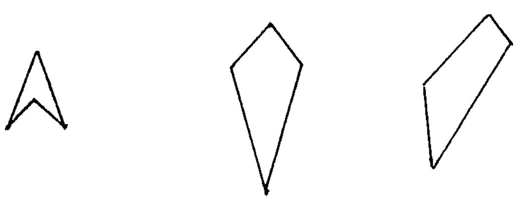
K L M

(A) K only
(B) L only
(C) M only (1.06a, 1.07a)
(D) L and M only
(E) All are squares

2. These are examples of a figure called a <u>quadram</u>.



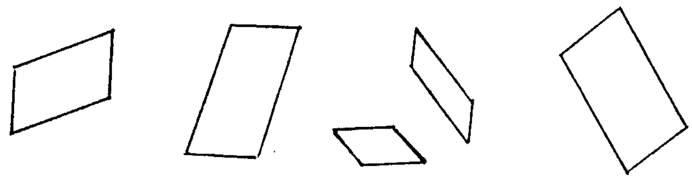
Which of these appear to be a quadram?



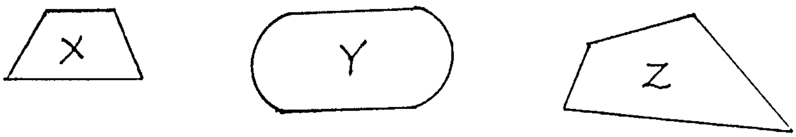
(A) L
(B) M (1.06a, 1.07a)
(C) N
(D) M and N
(E) None of these

EXPLAIN why you chose your answer

3. These are examples of a figure called a <u>parallelogram</u>.



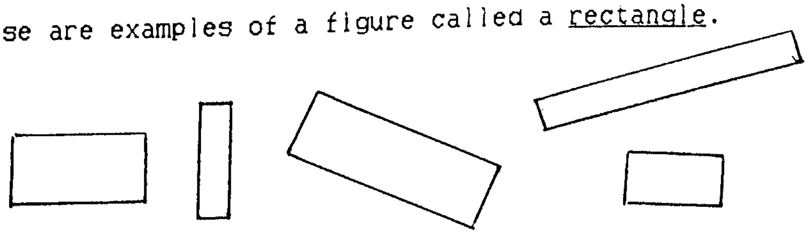
Which of these appear to be parallelograms



(A) X
(B) Y
(C) Z
(D) ALL are parallelograms
(E) NONE are parallelograms (1.07, 1.08)

EXPLAIN why you chose your answer:

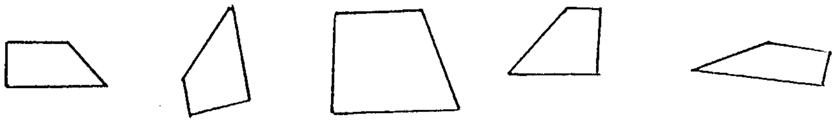
4. These are examples of a figure called a <u>rectangle</u>.



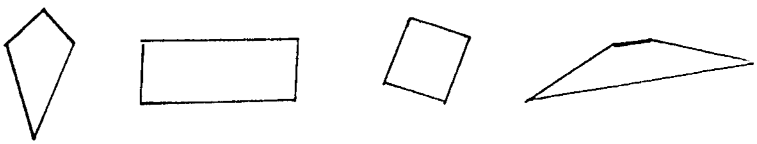
Starting at point A, draw a rectangle on the paper. (1.04)

A •

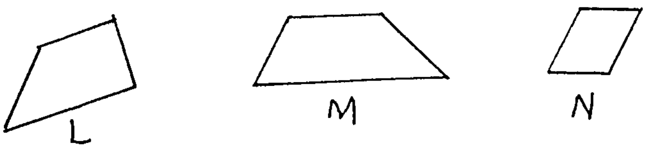
5. These are examples of a figure called a <u>tetragon</u>.



NONE of these figures is a tetragon.



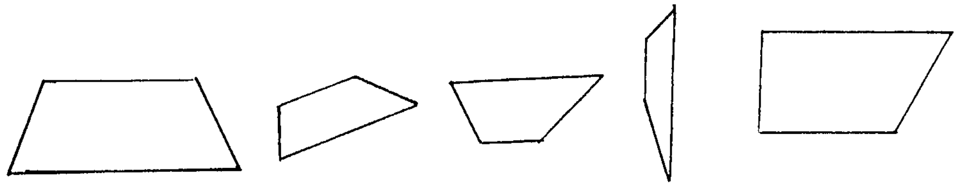
Which of these appear to be a tetragon?



(A) L (1.06b, 1.07a)
(B) M
(C) N
(D) M and N
(E) L, M and N

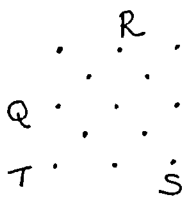
EXPLAIN why you chose your answer:

6. These are examples of <u>trapezoids</u>.

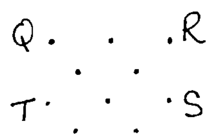


On each segment of dot paper, connect the points QRSTQ. Use straight lines. Connect the points in the order given. (Q to R, R to S, S to T, T to Q)

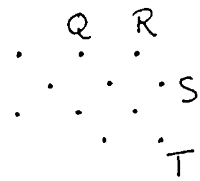
Which choice results in a trapezoid being outlined?



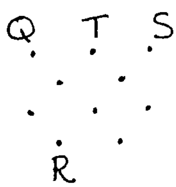
(A)



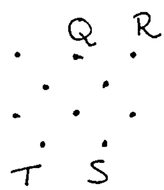
(B)



(C)
(1.06b,c, 1.07a)



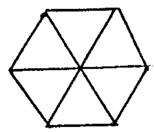
(D)



(E)

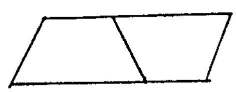
EXPLAIN why you chose your answer:

7. Which shape named in (A) to (D) could be traced on the figure below by following only the lines of the figure. The figure is flat (2-dimensional).



(A) Square
(B) Rectangle
(C) Tetrahedron
(D) Parallelogram (1.06c)
(E) None of the above.

8. Two identical trapezoids are arranged side by side as shown.



Which statement (A) - (C) below would you use as a reason to say that the new figure (outlined) is a parallelogram?

(A) The new figure looks like a parallelogram. (1.08)
(B) You could measure and show that the new figure has all the properties of a parallelogram (2.09)
(C) Using properties of the trapezoid it could be shown that the parallelism is convergent.
(D) Using properties of the trapezoid it could be shown that the new figure has at least one set of opposite sides which are equal and parallel (3.05, 3.17)
(E) It isn't a parallelogram

EXPLAIN why you chose your answer:

9. Consider the following properties of a four sided figure:

1. Opposite sides are equal.
2. Diagonals are equal.
3. Opposite angles are equal.

These properties are ALWAYS true for which type of figure?

(A) Quadrilateral
(B) Parallelogram
(C) Rectangle (2.10)
(D) Kites
(E) Tetrahedron

10. Consider the following properties of a four sided figure:

1. One pair of opposite sides are parallel.
2. No information is available about the other pair of sides.
3. The pair of opposite sides which are known to be parallel are also equal.

These properties are ALWAYS true for which type (or types) of figure?

(A) Square
(B) Parallelogram
(C) Rectangle
(D) All of the above (2.10)
(E) None of the above

11. These are some statements which can be made about four sided figures.

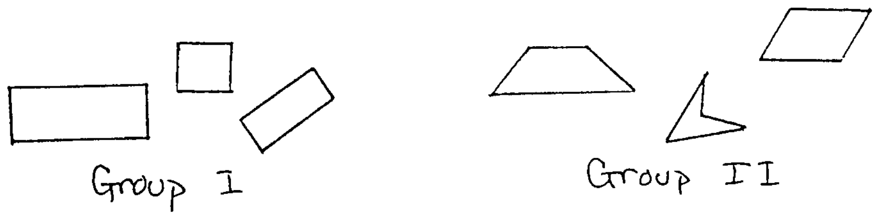
Statement 1: two long sides, two short sides
Statement 2: both pairs of opposite sides are the same length
Statement 3: both pairs of opposite sides are parallel
Statement 4: one angle is a right angle
Statement 5: all 4 angles are right angles.

From the choices below, which selection of these statements is the <u>shortest</u> list needed to GUARANTEE that a four sided closed figure is a RECTANGLE?

(A) 1
(B) 2, 3
(C) 3, 4 (3.05)
(D) 1, 2, 3, 5 (2.14)
(E) None of the lists in (A) to (D) guarantee a rectangle

EXPLAIN why you chose your answer

12. A set of six shapes was sorted into the two different and distinct groups shown here, group I and group II.



What characteristic can be used to describe why figures were put into group I.

(A) They look "balanced" (1.07)
(B) Adjacent sides are equal
(C) The opposite sides are parallel
(D) All the figures are quadrilaterals
(E) No angle is greater than 90 degrees (2.11)

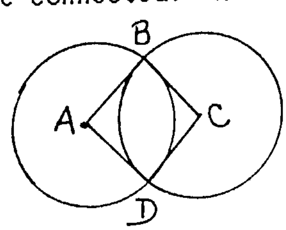
13. What do all squares have that some parallelograms do not have?

(A) Opposite sides equal
(B) Opposite angles equal
(C) Opposite sides parallel
(D) Diagonals bisect each other
(E) Both have all of the above (2.11)

14. What do all rectangles have which some parallelograms do not have?

(A) Opposite sides equal
(B) Opposite angles equal
(C) Diagonals are perpendicular
(D) Diagonals bisect each other
(E) Diagonals are equal (2.11)

15. Two circles intersect in such a way that the figure ABCD is formed when the centers of the circles and the points of intersection are connected. AB=BC=CD=DA.



Which of the following could be used to show that BD is perpendicular to AC?

(A) Properties of a square
(B) Properties of a rhombus (2.10, 2.15)
(C) Properties of a rectangles
(D) Properties of a parallelogram
(E) None of these

EXPLAIN why you chose your answer:

16. In which shape or shapes are 3 sides ALWAYS equal?

(A) A square (2.15)
(B) A kite
(C) A rectangle
(D) Both A and B
(E) None of the above

EXPLAIN why you chose your answer:

17. A rhombus is a four sided figure with all sides the same length. Two or more such figures are called rhombi. The diagonals of a rhombus are straight lines which connect the opposite vertices (corners) of the figure.

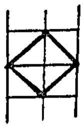
Which of the statements (A) to (E) about diagonals is FALSE for <u>some</u> rhombi?

(A) The diagonals bisect each other.
(B) The diagonals are lines of symmetry.
(C) The two diagonals are perpendicular.
(D) The two diagonals have the same length. (2.08)
(E) Each diagonal bisects two angles of the rhombus.

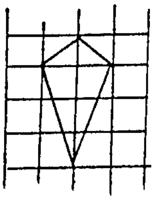
18. A calor is a four sided closed figure. Two adjacent sides are equal ("adjacent" means "next to"). The other two adjacent sides are equal. All four sides are NOT equal.

Which of these shapes is a calor?

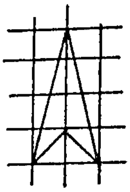
(A)



(B)



(C)



(D) Both B and C are calors. (2.10, 2.15)

(E) All three figures are calors.

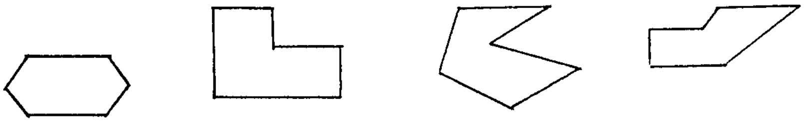
19. Which of (A) to (E) is true for all parallelograms

(A) The sum of the interior angles is 360.
(B) The opposite angles are equal.
(C) The diagonals are lines of symmetry.
(D) Both (A) and (B) are true in all parallelograms.
(E) All of the above are true in all parallelogram.

20. Which of (A) to (E) is FALSE for some rectangles?

(A) There are four sides.
(B) There are four right angles.
(C) The diagonals have the same length.
(D) The opposite sides have the same length.
(E) All of the above are true in every rectangle.(2.08)

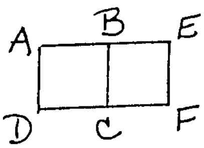
21. Working from the fact that the sum of the angles of a quadrilateral is 360 degrees, what would you say is the sum of the angles of a 6 sided figures? (Some examples are given below)



(A) This cannot be determined
(B) 360 degrees
(C) 540 degrees
(D) 720 degrees (2.09, 2.15)
(E) 1080 degrees

EXPLAIN why you chose your answer:

22. Two identical squares share a common side (BC) as shown.



Which of the following can be used to show that AF = DE

(A) Properties of a quadrilateral
(B) Properties of a rhombus
(C) Properties of a rectangle (2.14)
(D) Properties of a parallelogram
(E) None of these

23. A four-sided closed figure has the following properties

1. Each pair of opposite sides are equal in length.
2. Each pair of opposite sides are parallel.

Based on the above, which of the choices (A) - (D) is sufficient (enough) information to determine that the four sided figure is a parallelogram?

(A) (1) is needed; (2) is not necessarily true.
(B) (2) is needed; (1) is not necessarily true.
(C) Both (1) and (2) are needed (2.12)
(D) Either (1) or (2) (3.05)
(E) Neither (1) or (2) is enough information

EXPLAIN why you choose your answer:

24. What type of a figure can be called both a rhombus and a rectangle?

(A) Square (3.07)
(B) Rhombus
(C) Rectangle
(D) Parallelogram
(E) No figure

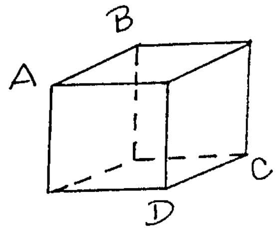
EXPLAIN your choice of answer.

25. Which is true?

(A) All properties of parallelograms are properties of all squares (3.06)
(B) All properties of squares are properties of all parallelograms
(C) All properties of rectangles are properties of all parallelograms
(D) All properties of squares are properties of all rectangles
(E) All properties of rectangles are properties of all quadrilaterals

EXPLAIN why you chose your answer.

26. A cube is a 3-dimensional figure with 6 sides (faces), each of which is a square. The faces are perpendicular to each other. What would be the shape of the plane figure ABCD which results from cutting the cube through the vertices A, B, C, D?



(A) Square
(B) Rectangle (3.17)
(C) Trapezoid
(D) Either A or B
(E) Not enough information

EXPLAIN why you chose your answer

27. In rectangle PQRS, diagonal PR bisects angle SPQ. If PQ = 10, how long is PS?

(A) 5
(B) 10 (3.07, 3.17)
(C) 20
(D) $10\sqrt{2}$
(E) There is not enough information to determine this.

EXPLAIN why you chose your answer.

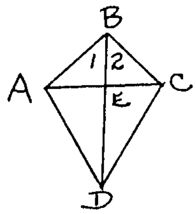
28. Here are three properties of a four sided figure

property A: It has four right angles.
property S: It is a square
property R: It is a rectangle

Which chain of statements is correct? (X "implies" Y means that when X is true, Y must also be true)

(A) A implies S which implies R
(B) A implies R which implies S
(C) S implies R which implies A (3.06)
(D) R implies A which implies S
(E) R implies S which implies A

29. ABCD is a kite with AB = BC and AD = CD. What property or properties of kites is established by the following?



1. We are told that ABCD is a kite, with AB = BC and AD = CD
2. BD = BD (they are the same segment)
3. $\triangle ABD \cong \triangle BCD$ (Side-Side-Side Congruence of triangles)
4. $\angle 1 = \angle 2$ because they are corresponding parts of congruent triangles
5. Since $\triangle$ ABC is isosceles (see step #1), and since BE bisects its vertex angle (see step #4), BE is a altitude of $\triangle$ ABC
6. Furthermore, $\angle AEB \cong \angle CEB$ (from what we know about the properties of altitudes in an isosceles triangle)

Therefore:

(A) A kite is a figure with two sets of adjacent sides congruent
(B) If a quadrilateral is a kite, the diagonals are perpendicular (3.12)
(C) If the diagonals of a quadrilateral are perpendicular, the figure is a kite.
(D) If a figure contains two congruent triangles, the perpendiculars bisect.
(E) All of the above

30. Definition A: A quadrilateral with exactly one pair of parallel sides is called an <u>exacta</u>.

Definition B: A quadrilateral with at least one pair of parallel sides is called a <u>leasta</u>.

Which of the following statements is true?

(A) The two definitions determine the same class of figures.
(B) All exactas are also leastas (3.09)
(C) All leastas are also exactas
(D) No exacta is also a leasta.
(E) No leasta is also an exacta.

31. When working with a PARALLELOGRAM, which of (A) to (C) is FALSE?

(A) If told all four sides are equal then you know that the opposite sides are equal
(B) If told at least one angle is a right angle, then you know all the angles are right angles.
(C) If told the diagonals are congruent, then you know that they bisect the angles too (3.06)
(D) Both (B)and (C) are false
(E) None of (A) - (C) above is false

32. Here are two statements about a quadrilateral.

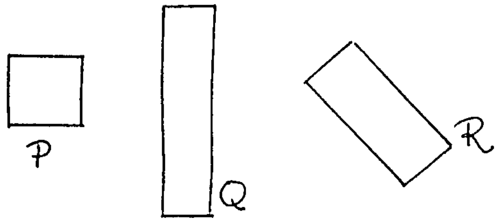
Statement 1: Quadrilateral QRST has 4 sides of the same length.

Statement 2: The opposite angles in quadrilateral QRST are equal.

Which is correct?

(A) Statements 1 and 2 cannot both be true.
(B) If 1 is true, then 2 is true (3.06)
(C) If 2 is true, then 1 is true
(D) If 1 is false, then 2 is true
(E) If 2 is false, then 1 is true

33. Which of these can be called rectangles?



(A) All can (3.07)
(B) Q only
(C) R only
(D) P and Q only
(E) Q and R only

Explain why you chose your answer:

34. Certain quadrilaterals, called Geldof's, have both sets of opposite sides parallel and diagonals which are equal but not perpendicular. To which other class of figures might this shape belong?

(A) Kite
(B) Square
(C) Rhombus
(D) Rectangle (3.07)
(E) None of the above

EXPLAIN why you chose your answer.

35. Consider the following suggested definitions for a parallelogram:

Definition 1: A parallelogram is a quadrilateral in which both pairs of opposite sides are parallel.

Definition 2: A parallelogram is a quadrilateral in which both pairs of opposite sides are congruent.

Which statement about these definitions is true?

(A) The definitions are equivalent (interchangeable). (3.09)
(B) Only one definition can be correct.
(C) Definition 1 is a partial definition.
(D) Definition 2 is a partial definition.
(E) Neither is a complete definition.

36. Which of (A) - (D) starts with the same idea statement 1 ends with and ends with the idea statement 1 starts with?

Statement 1: When two sides of a quadrilateral are parallel to each other and congruent. the figure is a parallelogram.

(A) When two sides of a parallelogram are parallel to each other, the figure is congruent.
(B) When two sides of a parallelogram are parallel to each other and congruent, the figure is a quadrilateral.
(C) When a figure is a parallelogram, two sides are parallel.
(D) When a figure is a parallelogram, two sides are parallel and congruent.(3.15)
(E) None of the above.

37. Which condition will show that a quadrilateral is a rhombus without first showing that it is a parallelogram.
(A) If it contains one adjacaent pair of sides that are equal
(B) If either diagonal bisects two angles
(C) If the diagonals are perpendicular bisectors of each other(3.05,3.09)
(D) All of the above
(E) None of the above

EXPLAIN why you chose your answer:

38. A proof is a list of statements together with a justification for each statment which ends up with the desired conclusion. Which of the following is not a proper type of justification within a proof?

(A) Axiom
(B) Given
(C) Theorem
(D) Definition
(E) Measurement (4.07)

39. Which statement is true?

(A) Any statment which seems true should become a postulate.
(B) Theorems are proved only on the basis of definitions and undefined terms.
(C) It is possible to define each geometric term by using simpler geometric terms.
(D) Exact geometric reasoning leads to geometric truths that cannot be deduced with absolute certainty from measurement. (4.07)
(E) More than one of the above is true. (List which ones here:____________)

40. Consider these to be two unproven statements:

I. If a figure is a square, then its diagonals are perpendicular to each other.
II. If the diagonals of a quadrilateral are perpendicular to each other, the figure is a square.

Which is correct?
(A) To prove I is true, it is enough to prove that II is true.
(B) To prove II is true, it is enough to prove that I is true.
(C) To prove II is true, it is enough to find several squares whose diagonals are perpendicular to each other.
(D) To prove II is false, it is enough to find one non-square whose diagonals are perpendicular to each other. (4.09)
(E) None of (A) - (D) is correct

41. Which of the statements (A) to (C) is an accurate restatement of this fact:

    A quadrilateral whose diagonals bisect each other is a parallelogram

(A) If a quadrilateral is a parallelogram, then the diagonals bisect each other.
(B) If the diagonals of a parallelogram bisect each other, then the figure is a quadrilateral
(C) If the diagonals of a quadrilateral bisect each other. then the figure is a parallelogram. (4.05)
(D) Both (A) and (C) are accurate restatements.
(E) All of the above are accurate restatements.

42. What is assumed (given) and what is to be shown (proved) in the following statement:

    A quadrilateral with supplementary adjacent angles is a parallelogram.

(A) Given: A parallelogram
    Prove: the adjacent angles are supplementary

(B) Given: A quadrilateral
    Prove: the adjacent supplementary angles are a parallelogram

(C) Given: A parallelogram with supplementary angles
    Prove: the angles are adjacent

(D) Given: A quadrilateral with adjacent angles supplementary
    Prove: the figure is a parallelogram (4.05)

(E) Given: A quadrilateral with supplementary angles
    Prove: the figure is a parallelgram with adjacent angles

43. Consider the following statements:

Statement I: If a quadrilateral is convex then condition A holds

Statement II: If condition A holds, then the quadrilateral is convex

Statement III: A quadrilateral is convex if and only if condition A holds.

Which of the following is correct?

(A) Statment I and II say the same thing,
(B) Statement I and III say the same thing,
(C) All three statements say the same thing,
(D) If statement III is true then both statement I and statement II are true (4.09),
(E) There is not enough information to judge.

44. Suppose you have proved statements I and II.

I. If p, then q.
II. If s, then not q.

Which statement follows from statements I and II?

(A) If q, then p.
(B) If not p, then s.
(C) If p, then not s.(4.08)
(D) If not p, then not q.
(E) If not s, then p.

45. Which of the conclusions (A) to (E) can be drawn from the following true statements?

Statement 1: If P is true, then Q is true .
Statement 2: If R is true, then S is not true.
Statement 3: If Q is true, then S is true.
Statement 4: P is true.

(A) S is true; R is True
(B) S is true; R is False (4.07)
(C) S is false; R is True
(D) S is false; R is True
(E) Only S is true (3.12)

Appendix F

Draft Instrument

DIRECTIONS

There are 37 written questions in this survey of geometric thinking. You have all period to answer the questions. No one is expected to answer all of the questions correctly. I am looking for "good" questions and "bad" questions, not trying to find out how smart you are.

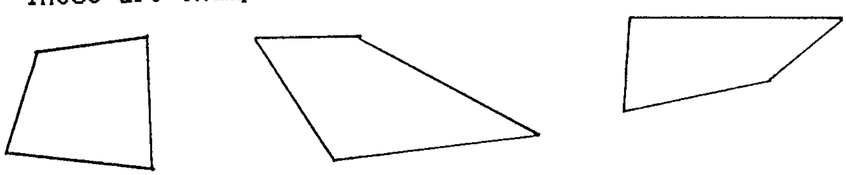
Read each problem carefully. All the questions are multiple choice. Read each choice of answers carefully especially as some examples have combination answer choices such as "All of the above are true", "Some of the above", "(A) and (B) are both true", etc.

* Indicate your answer choice on the answer sheet which is provided. Either put a cross on the letter which corresponds with your choice or darken the letter.
* Erase all incorrectly chosen answers.
* Points are not taken off for incorrectly answered questions.

Some example questions are given below.

EXAMPLE #1

These are examples of a figure called a quadrilateral.



Which of these are quadrilaterals?

J K L M N

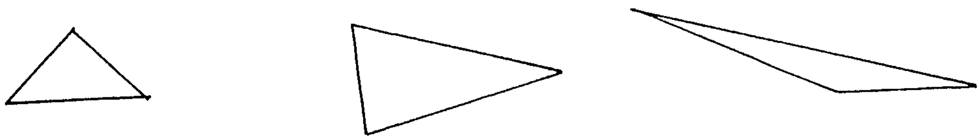
(A) J
correct (B) K
(C) L
(D) M
(E) N

If you choose to cross out the correct answer, your answer sheet would look like this:

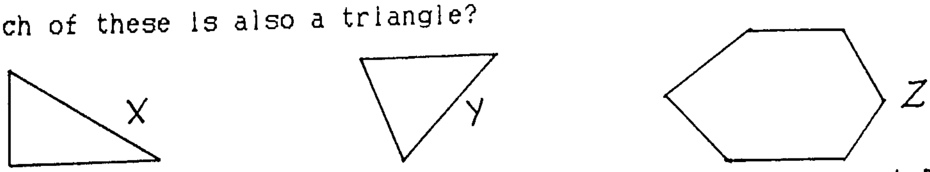
Example #1. A ~~B~~ C D E

EXAMPLE #2

These are examples of a figure called a <u>triangle</u>.



Which of these is also a triangle?



(A) X
(B) Y
(C) Z
(D) X and Y
(E) All of the above are quadrilaterals

correct

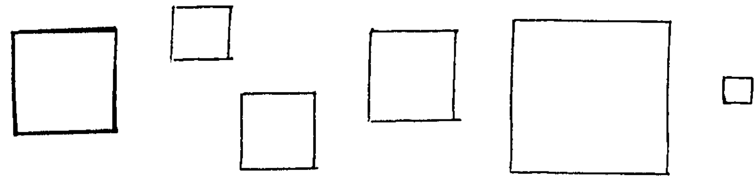
note the 'combination' answer choices

If you choose to darken the correct answer, it would look like this:

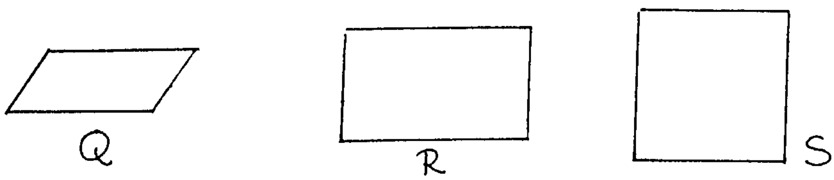
Example #2. A B C D E

YOU MAY BEGIN THIS TEST WHEN THE ADMINISTRATOR SAYS "BEGIN".

1. These are examples of a figure called a <u>square</u>

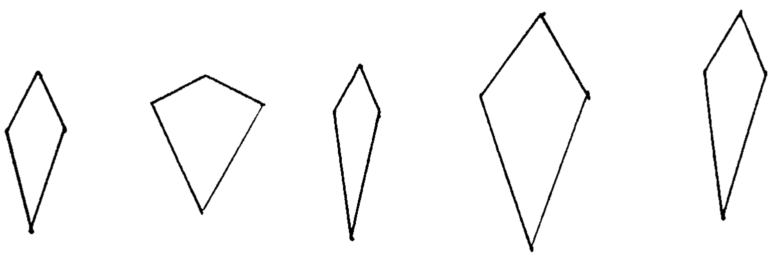


Which of these appear to be a square?

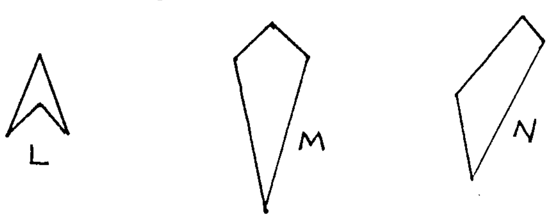


(A) Q only
(B) R only
(C) S only (1.06a, 1.07a)
(D) R and S only
(E) All are squares

2. These are examples of a figure called a <u>quadram</u>.

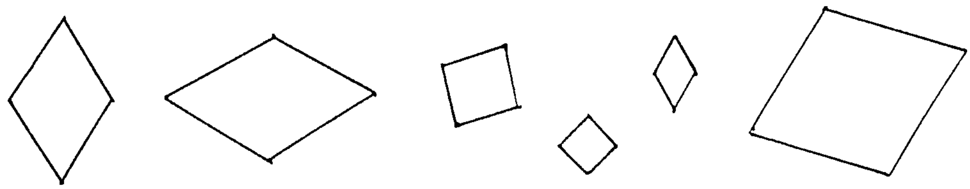


Which of these appear to be a quadram?

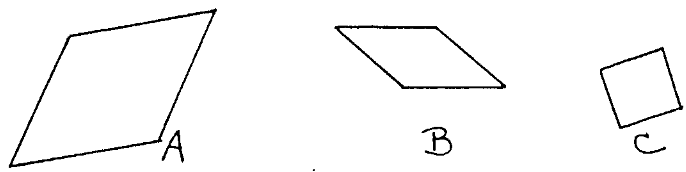


(A) L only
(B) M only (1.06a, 1.07a)
(C) N only
(D) M and N only
(E) None of these

3. These are examples of a figure called a <u>rhombus</u>.

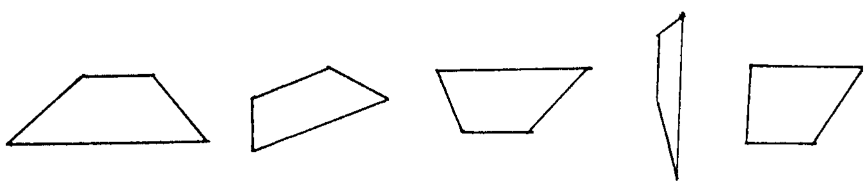


Which of these appear to be a rhombus?

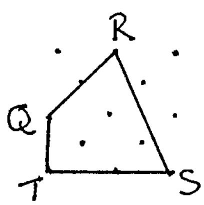


(A) A only
(B) B only
(C) C only
(D) A and C only
(E) A, B and C (1.06, 1.07)

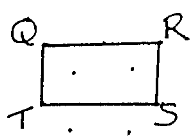
4. These are examples of a figure called a <u>trapezoid</u>.



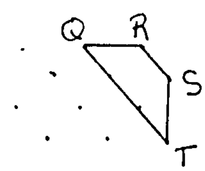
Which of these five figures, QRST, appear to be a trapezoid?



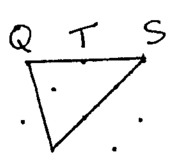
(A)



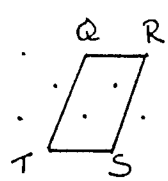
(B)



(C)
(1.06b,c. 1.07a)

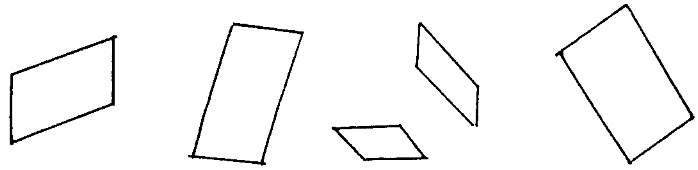


(D)

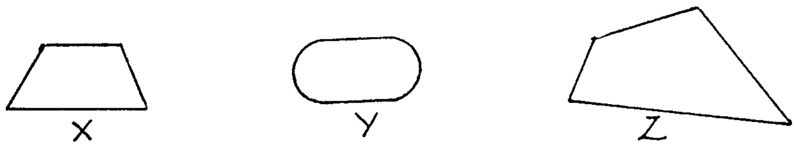


(E)

5. These are examples of a figure called a <u>parallelogram</u>.

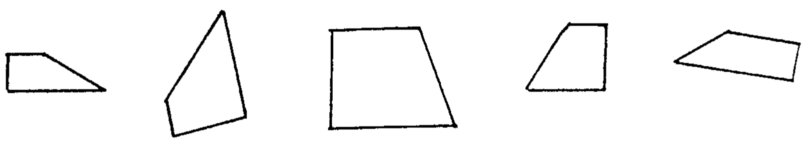


Which of these appear to be parallelograms

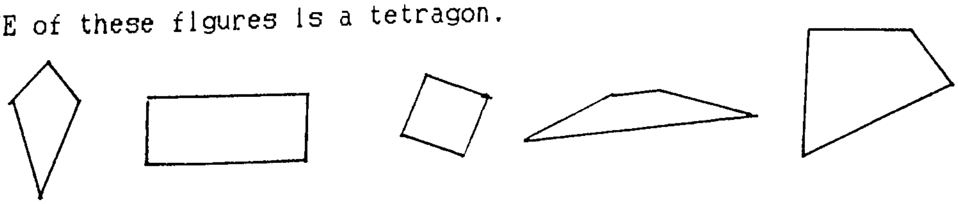


(A) X
(B) Y
(C) Z
(D) ALL are parallelograms
(E) NONE are parallelograms (1.07, 1.08)

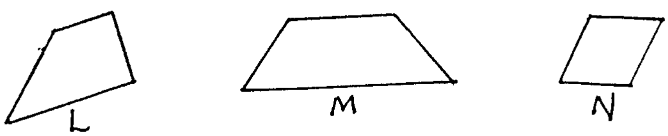
6. These are examples of a figure called a <u>tetragon</u>.



NONE of these figures is a tetragon.

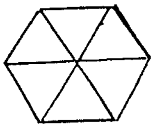


Which of these appear to be a tetragon?



(A) L (1.06b, 1.07a)
(B) M
(C) N
(D) M and N
(E) L, M and N

7. Which shape named in (A) to (D) could be traced on the figure below by following only the lines of the figure. The figure is flat (2-dimensional).

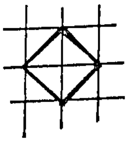


(A) Square
(B) Rectangle
(C) Tetrahedron
(D) Parallelogram (1.06c)
(E) None of the above.

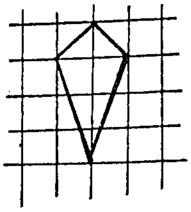
8. A calor is a four sided closed figure. Two adjacent sides are equal ("adjacent" means "next to"). The other two adjacent sides are equal. All four sides are NOT equal.

Which of these shapes is a calor?

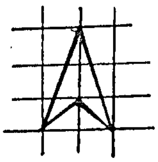
(A)



(B)



(C)



(D) Both B and C are calors.(2.10, 2.15)

(E) All three figures are calors.

9. A rhombus is a four sided figure with all sides the same length. Two or more such figures are called rhombi. The diagonals of a rhombus are straight lines which connect the opposite vertices (corners) of the figure.

Which of the statements (A) to (E) about diagonals is FALSE for <u>some</u> rhombi?
(A) The diagonals bisect each other.
(B) The diagonals are lines of symmetry.
(C) The two diagonals have the same length. (2.08)
(D) Each diagonal bisects two angles of the rhombus.
(E) The two diagonals are perpendicular (meet at right angles).

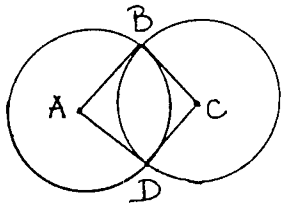
10. Consider the following properties of a four sided figure:

1. Opposite sides are equal.
2. Diagonals are equal.
3. Opposite angles are equal.

These properties are ALWAYS true for which type of figure?

(A) Quadrilateral
(B) Parallelogram
(C) Rectangle (2.10)
(D) Kites
(E) Tetrahedron

11. Two circles intersect in such a way that the figure ABCD is formed when the centers of the circles and the points of intersection are connected. AB=BC=CD=DA.



Which of the following could be used to show that BD is perpendicular to AC?
(A) Properties of a square
(B) Properties of a rhombus (2.10, 2.15)
(C) Properties of a tangent
(D) Properties of a circumference
(E) None of these

12. These are some statements which can be made about four sided figures.

Statement 1: two long sides, two short sides
Statement 2: both pairs of opposite sides are the same length
Statement 3: both pairs of opposite sides are parallel
Statement 4: one angle is a right angle
Statement 5: all 4 angles are right angles.

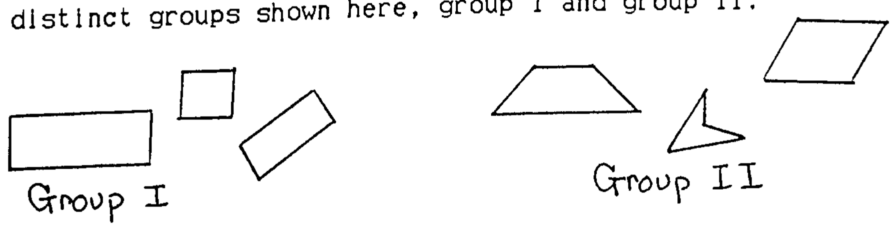
From the choices below, which selection of these statements is the shortest list needed to GUARANTEE that a four sided closed figure is a RECTANGLE?

(A) 1
(B) 2, 3
(C) 3, 4 (3.05)
(D) 1, 2, 3, 5 (2.14)
(E) None of the lists in (A) to (D) guarantee a rectangle.

13. What do ALL squares have that SOME parallelograms do not have?

(A) Opposite sides equal
(B) Opposite angles equal
(C) Opposite sides parallel
(D) Diagonals bisect each other
(E) Both have all of the above (2.11)

14. A set of six shapes was sorted into the two different and distinct groups shown here, group I and group II.



What characteristic can be used to describe why figures were put into group I.

(A) They look "balanced". (1.07)
(B) Adjacent sides are equal.
(C) The opposite sides are parallel.
(D) All the figures are quadrilaterals.
(E) No angle is greater than 90 degrees.(2.11)

15. What do ALL rectangles have which SOME parallelograms do not have?

    (A) Diagonals are equal.(2.11)
    (B) Opposite sides equal.
    (C) Opposite angles equal.
    (D) Diagonals are perpendicular.
    (E) Diagonals bisect each other.

16. In which shape or shapes are 3 sides ALWAYS equal?

    (A) A square (2.15)
    (B) A kite
    (C) A rectangle
    (D) Both A and B
    (E) None of the above.

17. Which of (A) to (D) is FALSE <u>for</u> some rectangles?

    (A) There are four sides.
    (B) There are four right angles.
    (C) The diagonals have the same length.
    (D) The opposite sides have the same length.
    (E) All of the above are true in every rectangle.(2.08)

18. Working from the fact that the sum of the angles of a quadrilateral is 360 degrees, what would you say is the sum of the angles of a 6 sided figures? (Some examples are given below)

    (A) 360 degrees
    (B) 540 degrees
    (C) 720 degrees (3.10, 3.17 OR IS IT 2.09. 2.15)
    (D) 1080 degrees
    (E) This cannot be determined

19. A four-sided closed figure has the following properties

1. Each pair of opposite sides are parallel.
2. Each pair of opposite sides are equal in length.

Based on the above, which of the choices (A) - (D) is sufficient (enough) information to determine that the four sided figure is a parallelogram?

(A) Either (1) or (2). (3.05)
(B) Both (1) and (2) are needed.(2.14)
(C) (1) is needed; (2) is not necessarily true.
(D) (2) is needed; (1) is not necessarily true.
(E) Neither (1) or (2) is enough information.

20. What type of a figure can be called both a rhombus and a rectangle?

(A) Square (3.07)
(B) Rhombus
(C) Rectangle
(D) Parallelogram
(E) No figure

21. Which is true?

(A) All properties of parallelograms are properties of all squares.(3.06)
(B) All properties of squares are properties of all parallelograms.
(C) All properties of rectangles are properties of all parallelograms.
(D) All properties of squares are properties of all rectangles.
(E) All properties of rectangles are properties of all quadrilaterals.

22. In rectangle PQRS, diagonal PR bisects angle SPQ. If PQ = 10, how long is PS?

(A) 5
(B) 10 (3.07, 3.17)
(C) 20
(D) $10\sqrt{2}$
(E) There is not enough information to determine this.

23. Here are three properties of a four sided figure

property A: It has four right angles.
property S: It is a square.
property R: It is a rectangle.

Which chain of statements is correct? (X "implies" Y means that when X is true, Y must also be true)

(A) A implies S which implies R.
(B) A implies R which implies S.
(C) S implies R which implies A.(3.06)
(D) R implies A which implies S.
(E) R implies S which implies A.

24. Definition A: A quadrilateral with exactly one pair of parallel sides is called an exacta.
Definition B: A quadrilateral with at least one pair of parallel sides is called a leasta.

Which of the following statements is true?
(A) All exactas are also leastas. (3.09d)
(B) All leastas are also exactas.
(C) No exacta is also a leasta.
(D) No leasta is also an exacta.
(E) The two definitions determine the same class of figures.

25. When working with a PARALLELOGRAM, which of (A) to (C) is FALSE?

(A) If told that two adjacent sides are equal, then all four sides are equal.
(B) If told at least one angle is a right angle, then you know all the angles are right angles.
(C) If told the diagonals are congruent, then you know that they bisect the angles too (3.06)
(D) Both (B)and (C) are false
(E) (A), (B) and (C) are all true.

26. Here are two statements about a quadrilateral.

Statement 1: Quadrilateral QRST has 4 sides of the same length.

Statement 2: The opposite angles in quadrilateral QRST are equal.

Which is correct?

(A) If 1 is true, then 2 is true. (3.06)
(B) If 2 is true, then 1 is true.
(C) If 1 is false, then 2 is true.
(D) If 2 is false, then 1 is true.
(E) Statements 1 and 2 cannot both be true.

27. Certain quadrilaterals, called Geldof's. have both sets of opposite sides parallel and diagonals which are equal but not perpendicular. To which other class of figures might this shape belong?

(A) Kite
(B) Square
(C) Rhombus
(D) Rectangle (3.07)
(E) None of the above

28. Consider the following suggested definitions for a parallelogram:

Definition 1: A parallelogram is a quadrilateral in which each pair of opposite sides are parallel.

Definition 2: A parallelogram is a quadrilateral in which each pair of opposite sides are congruent.

Which statement about these definitions is the most accurate?

(A) Neither is a complete definition.
(B) Only one definition can be correct.
(C) Definition 1 is a partial definition.
(D) Definition 2 is a partial definition.
(E) The definitions are equivalent (interchangeable). (3.09e)

29. Which of (A) - (D) starts with the same idea statement 1 ends with and ends with the idea statement 1 starts with (in other words, is the converse of statement 1)?

Statement 1: When two sides of a quadrilateral are parallel to each other and congruent, the figure is a parallelogram.

(A) When two sides of a parallelogram are parallel to each other, the figure is congruent.
(B) When two sides of a parallelogram are parallel to each other and congruent, the figure is a quadrilateral.
(C) When a figure is a parallelogram, two sides are parallel.
(D) When a figure is a parallelogram, two sides are parallel and congruent.(3.15)
(E) None of the above.

30. Which condition will show that a quadrilateral is a rhombus without first showing that it is a parallelogram.

(A) If either diagonal bisects two angles.
(B) If it contains one adjacaent pair of sides that are equal.
(C) If the diagonals are perpendicular bisectors of each other.(3.05)
(D) All of the above.
(E) None of the above.

31. Which statement is true?

(A) Any statement which seems true should become a postulate.
(B) Theorems are proved only on the basis of definitions and undefined terms, not with other theorems.
(C) It is possible to define each geometric term by using simpler geometric terms.
(D) Exact geometric reasoning leads to geometric truths that cannot be deduced with absolute certainty from measurement. (4.07)
(E) More than one of the above is true.

32. A proof is a list of statements together with a justification for each statment which ends up with the desired conclusion. Which of the following is not a proper type of justification within a proof?

    (A) Axiom
    (B) Given
    (C) Theorem
    (D) Definition
    (E) Measurement (4.07)

33. Which of the conclusions (A) to (E) can be drawn from the following true statements?

    Statement 1: If P is true, then Q is true .
    Statement 2: If R is true, then S is not true.
    Statement 3: If Q is true, then S is true.
    Statement 4: P is true.

    (A) S is true; R is True.
    (B) S is true; R is False. (4.08)
    (C) S is false; R is True.
    (D) S is false; Q is True.
    (E) Only S is true (3.12).

34. Consider these as two unproven statements:

    I. If a figure is a square, then its diagonals are perpendicular to each other.
    II. If the diagonals of a quadrilateral are perpendicular to each other, the figure is a square.

    Which is correct?
    (A) To prove I is true, it is enough to prove that II is true.
    (B) To prove II is true, it is enough to prove that I is true.
    (C) To prove II is true, it is enough to find several squares whose diagonals are perpendicular to each other.
    (D) To prove II is false, it is enough to find one non-square whose diagonals are perpendicular to each other. (4.08)
    (E) None of (A) - (D) is correct

35. Which of the statements (A) to (C) is the most direct restatement of this fact:

A quadrilateral whose diagonals bisect each other is a trangram.

(A) If a quadrilateral is a trangram, then the diagonals bisect each other.
(B) If the diagonals of a trangram bisect each other, then the figure is a quadrilateral.
(C) If the diagonals of a quadrilateral bisect each other. then the figure is a trangram. (4.05)
(D) Both (A) and (C) are direct restatements.
(E) All of the above are direct restatements.

36. Consider the following statements:

Statement 1: If a quadrilateral is convex then condition A holds.
Statement 2: If condition A holds, then the quadrilateral is convex.
Statement 3: A quadrilateral is convex if and only if condition A holds.

Which of the following is correct?
(A) Statement 1 and 2 say the same thing.
(B) Statement 1 and 3 say the same thing.
(C) All three statements say the same thing.
(D) If statement 3 is true then both statement 1 and statement 2 are true. (4.08)
(E) There is not enough information to judge.

37. Suppose you have proved statements I and II.

I. If p, then q.
II. If s, then not q.

Which statement follows from statements I and II?
(A) If q, then p.
(B) If not p, then s.
(C) If p, then not s.(4.08)
(D) If not p, then not q.
(E) If not s, then p.

Test Number ________

Answer Sheet
Van Hiele Quadrilateral Evaluation

<u>Please print</u>

Name____ Sex: M F
Last First Middle (circle one)

Grade in School: 6 7 8 9 10 11 12 other________

Math Teacher__________________________ Math Class__________

Birth date ______ ______ ______ Test date ______ ______ ______
Day Month Year Day Month Year

Cross out or darken the correct answer

1.	A	B	C	D	E
2.	A	B	C	D	E
3.	A	B	C	D	E
4.	A	B	C	D	E
5.	A	B	C	D	E
6.	A	B	C	D	E
7.	A	B	C	D	E
8.	A	B	C	D	E
9.	A	B	C	D	E
10.	A	B	C	D	E
11.	A	B	C	D	E
12.	A	B	C	D	E
13.	A	B	C	D	E
14.	A	B	C	D	E
15.	A	B	C	D	E
16.	A	B	C	D	E
17.	A	B	C	D	E
18.	A	B	C	D	E
19.	A	B	C	D	E
20.	A	B	C	D	E
21.	A	B	C	D	E
22.	A	B	C	D	E
23.	A	B	C	D	E
24.	A	B	C	D	E
25.	A	B	C	D	E
26.	A	B	C	D	E
27.	A	B	C	D	E
28.	A	B	C	D	E
29.	A	B	C	D	E
30.	A	B	C	D	E
31.	A	B	C	D	E
32.	A	B	C	D	E
33.	A	B	C	D	E
34.	A	B	C	D	E
35.	A	B	C	D	E
36.	A	B	C	D	E
37.	A	B	C	D	E

Appendix G

Field Testing Permission Form

November 7. 1988

Dear Parent,

I am writing to ask your permission to involve your child in a research project. The focus of the research is on the teaching and learning of geometry. As I have previously worked with the adminis- tration and the staff at (insert school name) . I am familiar with the mathematics instruction being offered there. This has led me to request that this school participate in this study. All pertinent school personnel have agreed to the project. subject to parental approval.

I am developing a written test which will assess differences in how individuals think about geometric topics. To validate my instrument, I must administer it to groups of students. The test requires approximately 40 minutes to complete. In order to verify the accuracy of my results, I need to explore verbally. on a one-to-one interview basis, the responses of some of the students to other geometry activities. This interview requires approximately 30 minutes to complete. I am writing, therefore. to ask if your child may participate in both the written test and the interview. Neither activity is a test of intelligence or skill. Rather, they are methods which try to identify how students perceive geometric concepts.

I propose to start my research on (insert date) . The written test will be administered to the students at a time which (insert teacher's name) designates as appropriate. In order to minimally disrupt the students learning, the interviews will also be scheduled through her/him .

On the attached page, you will find a permission slip requesting approval for your child's participation in the two activities. The first request is that your child be allowed to complete the written geometry test. The second request is that. should your child be selected, he/she could participate in the interview activities.

Perhaps some background information about me would also be appropriate. I have been teaching in the School of Education at Dalhousie University since 1975. One of my major areas of responsibility there is working with the secondary school mathematics student teachers. I have also served as a member of the provincial task force for high school mathematics (1977-1983), conducted numerous inservices on the mathematics curriculum and on the use of computers, written for several Canadian textbook publishing houses and published articles in the area of mathematics education. Prior to joining the faculty at Dalhousie, I taught mathematics at Queen Elizabeth High School (1970-1974). Along with the above activies, I have also been pursuing a Doctorate of Philosophy in mathematics education at the University of Maryland. I have completed all of my course work towards that degree and have only the doctoral dissertation to complete. The research I am proposing is the basis of my dissertation.

Please rest assured that the identity of individuals will be kept in strictest confidence. I will be the only person with access to individual results. In any writing or publications which may result from this study, the identity of the school will also be kept in confidence.

If you have any questions about procedures, dates, etc., I would be pleased to answer them. I would also be glad to supply further references and rationale if you so desire. I may be reached at work (424-3369) or home (423-1556) or messages may be left 424-3724.

Thank you for allowing your child to participate in this project. I think research of this type--school based and content specific--will contribute greatly towards improving the learning opportunities we provide children.

Sincerely,

Mary L. Crowley

PERMISSION TO PARTICIPATE IN THE GEOMETRY RESEARCH

PLEASE CHECK THE APPROPRIATE BOXES

I give permission for my child to participate in the following activities (check one or both if the student may participate):

/__/ The written geometry test

/__/ The additional geometry activities (interview times will be selected in consultation with insert teacher's name)

/__/ I do not give permission for my child to participate.

____
Parent's or Guardian's Signature

____
Student's Name

____
Date

PLEASE RETURN THIS SLIP TO insert teacher's name
ON OR BEFORE insert date

Appendix H

Final Test Permission Form

March 28. 1989

Dear Parent,

I am writing to ask your permission to involve your child in a research project, the focus of which is the teaching and learning of geometry at the junior and senior high school levels. All pertinent school personnel have agreed to the project, subject to parental approval.

I am developing a test which assess differences in how individuals think about geometric topics. It is not a test of intelligence or skill. Rather, it is a method which tries to identify how students perceive geometric concepts. To validate my instrument, I must administer it to groups of students. The multiple choice test will require no more than one period to complete. When the scores are interpreted, each students will be identified as one of four "types of thinkers" about geometry.

On the attached page, you will find a permission slip requesting your approval for your child's participation in the testing. I would appreciate having the form returned to your child' mathematics teacher no later than Friday, March 31. 1989. The test will be administered during a regular mathematics class during the week of April 3, 1989.

Perhaps some background information about me would also be appropriate. I have been teaching in the School of Education at Dalhousie University since 1975. One of my major areas of responsibility there is working with the secondary school mathematics student teachers. I have also served as a member of the provincial task force for high school mathematics (1977-1983), conducted numerous inservices on the mathematics curriculum and on the use of computers, written for several Canadian textbook publishing houses and published articles in the area of mathematics education. Prior to joining the faculty at Dalhousie, I taught mathematics in Halifax at Queen Elizabeth High School (1970-1974). Along with the above activies, I have also been pursuing a Doctorate of Philosophy in mathematics education at the University of Maryland. I have completed all of my course work towards that degree and have only the doctoral dissertation to complete. The research I am proposing is the last phase of the data collection for my dissertation.

Please rest assured that the identity of individuals will be kept in strictest confidence. As well, in any writing or publications which may result from this study, the identity of the school will also be kept in confidence.

If you have any questions about procedures, dates, etc., I would be pleased to answer them. I would also be glad to supply further references and rationale if you so desire. I may be reached at 423-1556.

Thank you for allowing your child to participate in this project. I think research of this type--school based and content specific--will contribute greatly towards improving the learning opportunities we provide children.

Sincerely,

Mary L. Crowley

PERMISSION TO PARTICIPATE IN THE
GEOMETRY RESEARCH

PLEASE CHECK THE APPROPRIATE BOX

/__/ I give permission for my child to participate in the research project.

/__/ I do not give permission for my child to participate.

Comments:

____
Parent's or Guardian's Signature

____
Student's Name (Please Print)

____
Date

PLEASE RETURN THIS SLIP TO: (Teacher's Name)

ON OR BEFORE: (Date)

Appendix I

Van Hiele Quadrilateral Test

Test Number ________

Van Hiele Quadrilateral Test
DIRECTIONS

Do NOT open this test booklet until you are told to do so. In addition to this test booklet, you should have an answer sheet and a pencil. If you do not have both of these, please raise your hand NOW and indicate this to the person administering the test.

When you are told to begin:

1. Read each question carefully.

2. Read each choice of answers carefully before selecting which one you think is correct. Some examples have combination answer choices such as "All of the above are true", "Some of the above are true", "(A) and (B) are both true", etc.

3. Indicate your answer choice on the answer sheet by darkening the letter which corresponds to your choice or by crossing it out. Do NOT circle your answer choice.

4. If you wish to change an answer, erase the the first answer completely.

5. If you have NO idea which answer is correct, you may leave the answer blank. Points are not taken off, however, for incorrectly answered questions.

6. Do NOT mark in the test booklet. Use the space provided on your answer sheet, front and back, for scrap paper.

7. You will have 30 minutes to answer the 19 questions on this test. No one is expected to answer all of the questions correctly.

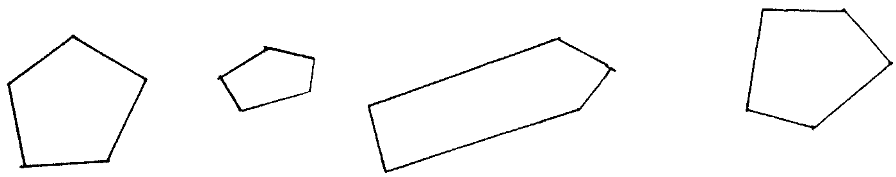
There is a test number in the upper right hand corner of this page. While you wait for the teacher to say you may begin the test, please write this number in the upper right hand corner of your answer sheet. Next, fill in the rest of the information on the top of the answer sheet.

When you have filled in the information on the answer sheet, turn to the next page in this booklet. Wait for the teacher to work through the sample problems before beginning the test.

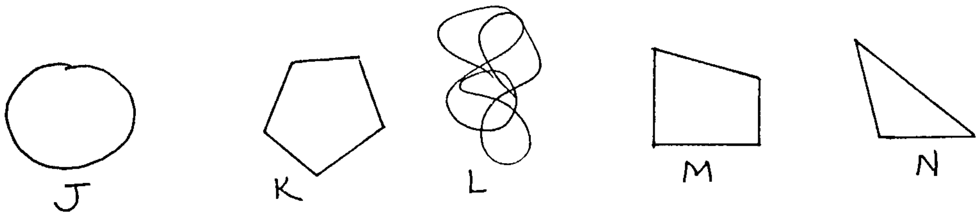
SAMPLE PROBLEMS

EXAMPLE #1:

These are examples of a figure called a pentagon.



Which of these is also a pentagon?

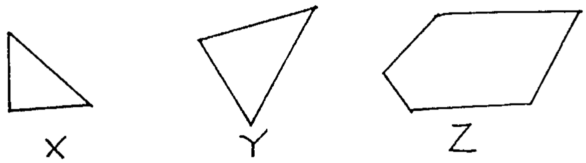


(A) J only
(B) K only
(C) L only
(D) M only
(E) N only

ANSWER: The correct answer is that figure K is the only pentagon. Thus, answer (B) is darken on your answer sheet. (See Example #1 on your answer sheet)

EXAMPLE #2

Which of these figures is a triangle?



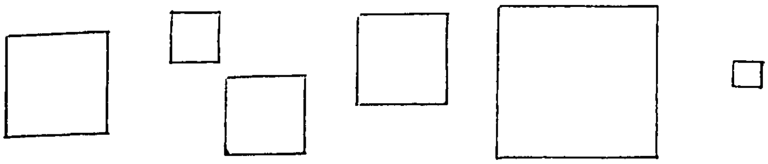
(A) X
(B) Y
(C) Z
(D) X and Y
(E) All of the above are triangles

ANSWER: This is an example of why it is important to read ALL the answer choices before selecting the best answer. Figures X and Y are both triangles, thus the correct answer is (D). It would be incorrect to selected just answer (A) or just answer (B). The correct answer is indicated on your answer sheet next to EXAMPLE #2. This time, the answer is crossed out.

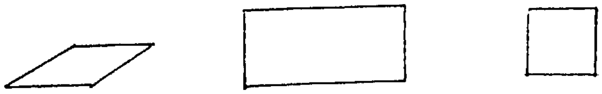
DO NOT START UNTIL THE TEST ADMINISTRATOR SAYS "BEGIN"

Van Hiele Quadrilateral Test

1. These are examples of a figure called a <u>square</u>



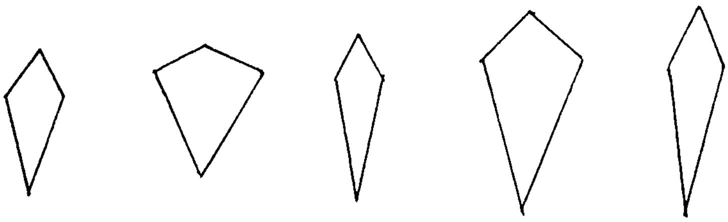
Which of these appear to be a square?



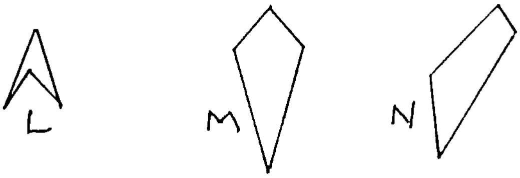
Q R S

(A) Q only
(B) R only
(C) S only (1.06, 1.07)
(D) R and S only
(E) All are squares

2. These are examples of a figure called a <u>quadram</u>.

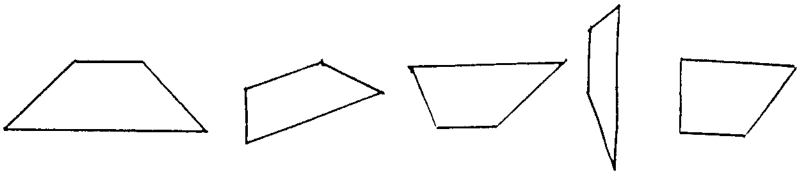


Which of these appear to be a quadram?

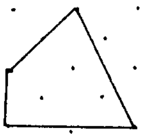


(A) L only
(B) M only (1.06, 1.07)
(C) N only
(D) M and N only
(E) None of these

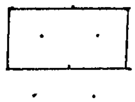
3. These are examples of a figure called a trapezoid.



Which of these appear to be a trapezoid?



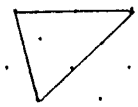
(A)



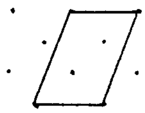
(B)

(C)

(1.06, 1.07)

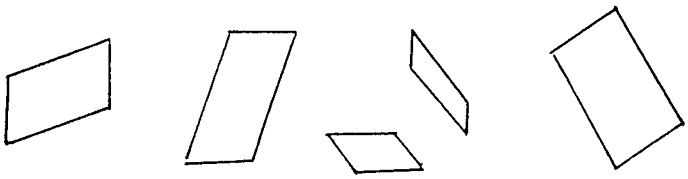


(D)

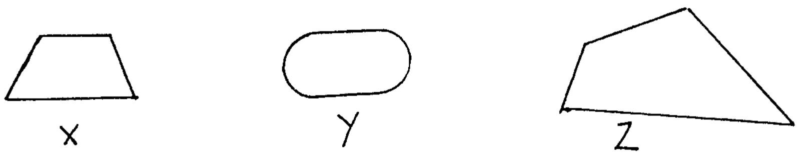


(E)

4. These are examples of a figure called a parallelogram.



Which of these appear to be parallelograms



(A) X
(B) Y
(C) Z
(D) ALL are parallelograms
(E) NONE are parallelograms (1.07, 1.08)

5. Consider the following properties of a four sided figure:

   1. Opposite sides are equal.
   2. Diagonals are equal.
   3. Opposite angles are equal.

   These properties are ALWAYS true for which type of figure?

   (A) Quadrilateral
   (B) Parallelogram
   (C) Rectangle (2.10)
   (D) Kites
   (E) Tetrahedron

6. These are some statements which can be made about four sided figures.

   Statement 1: two long sides, two short sides
   Statement 2: both pairs of opposite sides are the same length
   Statement 3: both pairs of opposite sides are parallel
   Statement 4: one angle is a right angle
   Statement 5: all 4 angles are right angles.

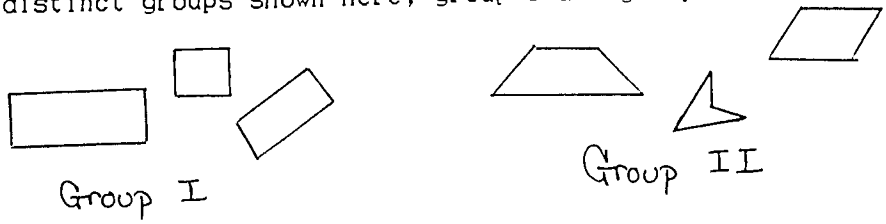
   From the choices below, which selection of these statements is the shortest list needed to GUARANTEE that a four sided closed figure is a RECTANGLE?

   (A) 1
   (B) 2, 3
   (C) 3, 4 (3.05)
   (D) 1, 2, 3, 5 (2.14)
   (E) None of the lists in (A) to (D) guarantee a rectangle

7. What do ALL squares have that SOME parallelograms do not have?

   (A) Opposite sides equal
   (B) Opposite angles equal
   (C) Opposite sides parallel
   (D) Diagonals bisect each other
   (E) Both have all of the above (2.11)

8. A set of six shapes was sorted into the two different and distinct groups shown here, group I and group II.



What characteristic can be used to describe why figures were put into group I.

(A) They look "balanced".
(B) Adjacent sides are equal.
(C) The opposite sides are parallel.
(D) All the figures are quadrilaterals.
(E) No angle is greater than 90 degrees. (2.11)

9. In which shape or shapes are 3 sides ALWAYS equal?

(A) A square (2.15)
(B) A kite
(C) A rectangle
(D) Both A and B
(E) None of the above.

10. What type of a figure can be called both a rhombus and a rectangle?

(A) Square (3.07)
(B) Rhombus
(C) Rectangle
(D) Parallelogram
(E) No figure

11. Which is true?

(A) All properties of parallelograms are properties of all squares. (3.06)
(B) All properties of squares are properties of all parallelograms.
(C) All properties of rectangles are properties of all parallelograms.
(D) All properties of squares are properties of all rectangles.
(E) All properties of rectangles are properties of all quadrilaterals.

12. Definition A: A quadrilateral with exactly one pair of parallel sides is called an <u>exacta</u>.

    Definition B: A quadrilateral with at least one pair of parallel sides is called a <u>leasta</u>.

    Which of the following statements is true?

    (A) All exactas are also leastas. (3.09d)
    (B) All leastas are also exactas.
    (C) No exacta is also a leasta.
    (D) No leasta is also an exacta.
    (E) The two definitions determine the same class of figures.

13. Certain quadrilaterals, called Geldof's, have both sets of opposite sides parallel and diagonals which are equal but not perpendicular. To which other class of figures might this shape belong?

    (A) Kite
    (B) Square
    (C) Rhombus
    (D) Rectangle (3.07)
    (E) None of the above

14. Consider the following suggested definitions for a parallelogram:

    Definition 1: A parallelogram is a quadrilateral in which each pair of opposite sides are parallel.

    Definition 2: A parallelogram is a quadrilateral in which each pair of opposite sides are congruent.

    Which statement about these definitions is the most accurate?

    (A) Neither is a complete definition.
    (B) Only one definition can be correct.
    (C) Definition 1 is a partial definition.
    (D) Definition 2 is a partial definition.
    (E) The definitions are equivalent (interchangeable). (3.09e)

15. Which of (A) - (D) starts with the same idea statement 1 ends with and ends with the idea statement 1 starts with (in other words, is the converse of statement 1)?

Statement 1: When two sides of a quadrilateral are parallel to each other and congruent, the figure is a parallelogram.

(A) When two sides of a parallelogram are parallel to each other, the figure is congruent.

(B) When two sides of a parallelogram are parallel to each other and congruent, the figure is a quadrilateral.

(C) When a figure is a parallelogram, two sides are parallel.

(D) When a figure is a parallelogram, two sides are parallel and congruent. (3.15)

(E) None of the above.

16. A proof is a list of statements together with a justification for each statment which ends up with the desired conclusion. Which of the following is not a proper type of justification within a proof?

(A) Axiom
(B) Given
(C) Theorem
(D) Definition
(E) Measurement (4.07)

17. Consider these as two unproven statements:

I. If a figure is a square, then its diagonals are perpendicular to each other.

II. If the diagonals of a quadrilateral are perpendicular to each other, the figure is a square.

Which of the following is correct?

(A) To prove I is true, it is enough to prove that II is true.
(B) To prove II is true, it is enough to prove that I is true.
(C) To prove II is true, it is enough to find several squares whose diagonals are perpendicular to each other.
(D) To prove II is false, it is enough to find one non-square whose diagonals are perpendicular to each other.(4.08)
(E) None of (A) - (D) is correct.

18. Consider the following statements:

Statement 1: If a quadrilateral is convex then condition A holds.
Statement 2: If condition A holds, then the quadrilateral is convex.
Statement 3: A quadrilateral is convex if and only if condition A holds.

Which of the following is correct?

(A) Statement 1 and 2 say the same thing.
(B) Statement 1 and 3 say the same thing.
(C) All three statements say the same thing.
(D) If statement 3 is true then both statement 1 and statement 2 are true. (4.08)
(E) There is not enough information to judge.

19. Suppose you have proved statements I and II.

I. If p, then q.
II. If s, then not q.

Which statement follows from statements I and II?

(A) If q, then p.
(B) If not p, then s.
(C) If p, then not s. (4.08)
(D) If not p, then not q.
(E) If not s, then p.

Test Number ______

Answer Sheet
Van Hiele Quadrilateral Test

Please print

Name ____ Sex: M F
Last First Middle (circle one)

Grade in School: 6 7 8 9 10 11 12 other ________

Math Teacher ______________________________ Math Class ________

Birth date ______ ______ ______ Test date ______ ______ ______
Day Month Year Day Month Year

Example #1: A ~~B~~ C D E Example #2: A B C ~~D~~ E

Cross out or darken the correct answer

1.	A	B	C	D	E	11.	A	B	C	D	E
2.	A	B	C	D	E	12.	A	B	C	D	E
3.	A	B	C	D	E	13.	A	B	C	D	E
4.	A	B	C	D	E	14.	A	B	C	D	E
5.	A	B	C	D	E	15.	A	B	C	D	E
6.	A	B	C	D	E	16.	A	B	C	D	E
7.	A	B	C	D	E	17.	A	B	C	D	E
8.	A	B	C	D	E	18.	A	B	C	D	E
9.	A	B	C	D	E	19.	A	B	C	D	E
10.	A	B	C	D	E						

Space for drawing or figuring. (You may also use the back)

INSTRUCTIONS FOR PERSON ADMINISTERING THE TEST

1. Before students arrive, check to see that there are enough test booklets, answer sheets and pencils for each individual who will write the test.

2. Write the date on the board in a location visible to all students.

(Note: text in capital letters is to be read aloud, verbatim, to students)

3. After students are seated, say

TODAY YOU WILL BE TAKING A GEOMETRY TEST. THE PURPOSE OF THIS TEST IS TO DETERMINE HOW YOU THINK ABOUT GEOMETRY, NOT TO SEE HOW MUCH YOU KNOW ABOUT THE SUBJECT. THE NUMBER OF CORRECT ANSWERS YOU GET IS NOT IMPORTANT. WHAT IS OF INTEREST IS WHICH QUESTIONS YOU ANSWER.

I WILL NOW DISTRIBUTE THE TEST BOOKLET, AN ANSWER SHEET AND A PENCIL . DO <u>NOT</u> OPEN THE BOOKLET UNTIL INSTRUCTED TO DO SO.

4. Distribute the booklets and answer sheets.

5. Say:

FOLLOW THE DIRECTIONS ON THE FIRST PAGE AS I READ THEM.

(Read the first page of directions out loud)

6. When the students have completed the information section of their answer sheet, say:

PLEASE TURN TO THE SECOND PAGE OF THE DIRECTIONS, THE SAMPLE PROBLEMS. FOLLOW ALONG AS I READ THE PROBLEMS AND THE ANSWERS..

(Read through the examples and the answers. At the appropriate times, have students refer to their answer sheet to see demonstrations of the two methods which can be used to correctly indicate an answer choice.)

7. Say: ARE THERE ANY QUESTIONS?

8. After answering any questions students may have, say

TESTS AND ANSWER SHEETS WILL BE COLLECTED AT THE END OF THE 30 MINUTE TESTING PERIOD. YOU MAY BEGIN THE TEST NOW.

9. You may wish to write on the board the time the test began and the time the test ends. You may also wish to indicate when 5 minutes are remaining.

10. After 30 minutes has elapsed, say

STOP. TIME IS UP. PUT YOUR PENCILS DOWN.

PASS YOUR ANSWER SHEETS FORWARD (or to the left, etc.)
(Wait for those to reach the front)

LOOK CAREFULLY THROUGH YOUR TEST BOOKLET AND ERASE ANY MARKS WHICH YOU FIND IN IT. (pause)

PASS YOUR TEST BOOKLETS FORWARD. (pause)

PASS YOUR PENCILS FORWARD.

Appendix J

Selected Binomial Expansions and Probabilities of Success

Binomial Expansion, n = 20
(master's) probability of success, p = .66

critical value (m)	p	mth term	sum of first m terms. α
0	.66	.0000000	.0000000
1	.66	.0000000	.0000000
2	.66	.0000003	.0000003
3	.66	.0000036	.0000039
4	.66	.0000293	.0000332
5	.66	.0001821	.0002153
6	.66	.0008838	.0010991
7	.66	.0034312	.0045303
8	.66	.0108234	.0153538
9	.66	.0280136	.0433674
10	.66	.0598173	.1031847
11	.66	.1055600	.2087447
12	.66	.1536830	.3624277
13	.66	.1835851	.5460128
14	.66	.1781855	.7241983
15	.66	.1383558	.8625541
16	.66	.0839291	.9464831
17	.66	.0383344	.9848175
18	.66	.0124023	.9972198
19	.66	.0025342	.9997540
20	.66	.0002460	1.0000000

Binomial Expansion, n = 20
(nonmaster's) probability of success, r = .41

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.41	.0000261	.0000261
1	.41	.0003631	.0003892
2	.41	.0023969	.0027861
3	.41	.0099940	.0127802
4	.41	.0295163	.0422964
5	.41	.0656362	.1079326
6	.41	.1140289	.2219615
7	.41	.1584809	.3804424
8	.41	.1789625	.5594050
9	.41	.1658184	.7252234
10	.41	.1267527	.8519761
11	.41	.0800749	.9320510
12	.41	.0417340	.9737849
13	.41	.0178471	.9916320
14	.41	.0062011	.9978331
15	.41	.0017237	.9995568
16	.41	.0003743	.9999312
17	.41	.0000612	.9999924
18	.41	.0000071	.9999995
19	.41	.0000005	1.0000000
20	.41	.0000000	1.0000000

Binomial Expansion, n = 20
(master's) probability of success, p = .7

critical value (m)	p	mth term	sum of first m terms. α
0	.7	.0000000	.0000000
1	.7	.0000000	.0000000
2	.7	.0000000	.0000000
3	.7	.0000005	.0000005
4	.7	.0000050	.0000056
5	.7	.0000374	.0000429
6	.7	.0002181	.0002610
7	.7	.0010178	.0012789
8	.7	.0038593	.0051382
9	.7	.0120067	.0171448
10	.7	.0308171	.0479619
11	.7	.0653696	.1133315
12	.7	.1143967	.2277282
13	.7	.1642620	.3919902
14	.7	.1916390	.5836292
15	.7	.1788631	.7624922
16	.7	.1304210	.8929132
17	.7	.0716037	.9645169
18	.7	.0278459	.9923627
19	.7	.0068393	.9992021
20	.7	.0007979	1.0000000

Binomial Expansion, n = 20
(nonmaster's) probability of success, r = .45

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.45	.0000064	.0000064
1	.45	.0001050	.0001114
2	.45	.0008160	.0009274
3	.45	.0040060	.0049334
4	.45	.0139299	.0188633
5	.45	.0364709	.0553342
6	.45	.0745996	.1299338
7	.45	.1220721	.2520059
8	.45	.1623004	.4143062
9	.45	.1770550	.5913612
10	.45	.1593495	.7507106
11	.45	.1185244	.8692350
12	.45	.0727309	.9419659
13	.45	.0366197	.9785856
14	.45	.0149808	.9935664
15	.45	.0049028	.9984693
16	.45	.0012536	.9997228
17	.45	.0002413	.9999641
18	.45	.0000329	.9999970
19	.45	.0000028	.9999999
20	.45	.0000001	1.0000000

Binomial Expansion, n = 20
(master's) probability of success, p = .75

critical value (m)	p	mth term	sum of first m terms, α
0	.75	.0000000	.0000000
1	.75	.0000000	.0000000
2	.75	.0000000	.0000000
3	.75	.0000000	.0000000
4	.75	.0000004	.0000004
5	.75	.0000034	.0000038
6	.75	.0000257	.0000295
7	.75	.0001542	.0001837
8	.75	.0007517	.0009354
9	.75	.0030068	.0039421
10	.75	.0099223	.0138644
11	.75	.0270608	.0409252
12	.75	.0608867	.1018119
13	.75	.1124062	.2142181
14	.75	.1686093	.3828273
15	.75	.2023312	.5851585
16	.75	.1896855	.7748440
17	.75	.1338956	.9087396
18	.75	.0669478	.9756874
19	.75	.0211414	.9968288
20	.75	.0031712	1.0000000

Binomial Expansion, n = 20
(nonmaster's) probability of success. r = .50

critical value (m)	r	mth term	sum of first m terms. $1-\beta$
0	.5	.0000010	.0000010
1	.5	.0000191	.0000200
2	.5	.0001812	.0002012
3	.5	.0010872	.0012884
4	.5	.0046206	.0059090
5	.5	.0147858	.0206947
6	.5	.0369644	.0576591
7	.5	.0739288	.1315880
8	.5	.1201344	.2517223
9	.5	.1601791	.4119015
10	.5	.1761971	.5880985
11	.5	.1601791	.7482777
12	.5	.1201344	.8684120
13	.5	.0739288	.9423409
14	.5	.0369644	.9793053
15	.5	.0147858	.9940910
16	.5	.0046206	.9987116
17	.5	.0010872	.9997988
18	.5	.0001812	.9999800
19	.5	.0000191	.9999990
20	.5	.0000010	1.0000000

Binomial Expansion, n = 20
(master's) probability of success, p = .80

critical value (m)	p	mth term	sum of first m terms, α
0	.8	.0000000	.0000000
1	.8	.0000000	.0000000
2	.8	.0000000	.0000000
3	.8	.0000000	.0000000
4	.8	.0000000	.0000000
5	.8	.0000002	.0000002
6	.8	.0000017	.0000018
7	.8	.0000133	.0000152
8	.8	.0000866	.0001017
9	.8	.0004617	.0005634
10	.8	.0020314	.0025948
11	.8	.0073870	.0099818
12	.8	.0221609	.0321427
13	.8	.0545499	.0866925
14	.8	.1090997	.1957922
15	.8	.1745595	.3703517
16	.8	.2181994	.5885511
17	.8	.2053641	.7939153
18	.8	.1369094	.9308247
19	.8	.0576461	.9884708
20	.8	.0115292	1.0000000

Binomial Expansion, n = 20
(nonmaster's) probablity of success, r = .55

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.55	.0000001	.0000001
1	.55	.0000028	.0000030
2	.55	.0000329	.0000359
3	.55	.0002413	.0002772
4	.55	.0012536	.0015307
5	.55	.0049028	.0064336
6	.55	.0149808	.0214144
7	.55	.0366197	.0580341
8	.55	.0727309	.1307650
9	.55	.1185244	.2492894
10	.55	.1593495	.4086388
11	.55	.1770550	.5856938
12	.55	.1623004	.7479941
13	.55	.1220721	.8700662
14	.55	.0745996	.9446658
15	.55	.0364709	.9811367
16	.55	.0139299	.9950666
17	.55	.0040060	.9990726
18	.55	.0008160	.9998886
19	.55	.0001050	.9999936
20	.55	.0000064	1.0000000

Binomial Expansion, n = 21
(master's) probability of success, p = .66

critical value (m)	p	mth term	sum of first m terms, α
0	.66	.0000000	.0000000
1	.66	.0000000	.0000000
2	.66	.0000001	.0000001
3	.66	.0000014	.0000015
4	.66	.0000123	.0000138
5	.66	.0000813	.0000951
6	.66	.0004207	.0005158
7	.66	.0017499	.0022657
8	.66	.0059446	.0082103
9	.66	.0166681	.0248784
10	.66	.0388269	.0637053
11	.66	.0753698	.1390751
12	.66	.1219218	.2609969
13	.66	.1638497	.4248466
14	.66	.1817492	.6065958
15	.66	.1646434	.7712392
16	.66	.1198507	.8910899
17	.66	.0684269	.9595168
18	.66	.0295175	.9890343
19	.66	.0090472	.9980814
20	.66	.0017562	.9998377
21	.66	.0001623	1.0000000

Binomial Expansion, n = 21
(nonmaster's) probablity of success, r = .41

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.41	.0000154	.0000154
1	.41	.0002249	.0002403
2	.41	.0015631	.0018034
3	.41	.0068792	.0086826
4	.41	.0215121	.0301948
5	.41	.0508270	.0810218
6	.41	.0941879	.1752097
7	.41	.1402556	.3154653
8	.41	.1705651	.4860303
9	.41	.1712075	.6572378
10	.41	.1427696	.8000075
11	.41	.0992128	.8992203
12	.41	.0574537	.9566740
13	.41	.0276407	.9843147
14	.41	.0109760	.9952907
15	.41	.0035594	.9988501
16	.41	.0009276	.9997777
17	.41	.0001896	.9999673
18	.41	.0000293	.9999966
19	.41	.0000032	.9999998
20	.41	.0000002	1.0000000
21	.41	.0000000	1.0000000

Binomial Expansion, n = 21
(master's) probability of success, p = .70

critical value (m)	p	mth term	sum of first m terms, α
0	.7	.0000000	.0000000
1	.7	.0000000	.0000000
2	.7	.0000000	.0000000
3	.7	.0000002	.0000002
4	.7	.0000019	.0000020
5	.7	.0000147	.0000168
6	.7	.0000916	.0001084
7	.7	.0004580	.0005664
8	.7	.0018703	.0024367
9	.7	.0063035	.0087402
10	.7	.0176498	.0263899
11	.7	.0411828	.0675728
12	.7	.0800777	.1476505
13	.7	.1293563	.2770068
14	.7	.1724751	.4494819
15	.7	.1878062	.6372881
16	.7	.1643304	.8016185
17	.7	.1127758	.9143943
18	.7	.0584763	.9728706
19	.7	.0215439	.9944145
20	.7	.0050269	.9994415
21	.7	.0005585	1.0000000

Binomial Expansion, n = 21
(nonmaster's) probablity of success, r = .45

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.45	.0000035	.0000035
1	.45	.0000606	.0000642
2	.45	.0004961	.0005602
3	.45	.0025705	.0031307
4	.45	.0094641	.0125948
5	.45	.0263274	.0389223
6	.45	.0574417	.0963640
7	.45	.1007095	.1970734
8	.45	.1441976	.3412711
9	.45	.1704154	.5116865
10	.45	.1673169	.6790034
11	.45	.1368957	.8158991
12	.45	.0933380	.9092370
13	.45	.0528698	.9621068
14	.45	.0247183	.9868251
15	.45	.0094379	.9962630
16	.45	.0028957	.9991587
17	.45	.0006968	.9998555
18	.45	.0001267	.9999822
19	.45	.0000164	.9999986
20	.45	.0000013	.9999999
21	.45	.0000001	1.0000000

Binomial Expansion, n = 21
(master's) probability of success, p = .75

critical value (m)	p	mth term	sum of first m terms, α
0	.75	.0000000	.0000000
1	.75	.0000000	.0000000
2	.75	.0000000	.0000000
3	.75	.0000000	.0000000
4	.75	.0000001	.0000001
5	.75	.0000011	.0000012
6	.75	.0000090	.0000102
7	.75	.0000578	.0000681
8	.75	.0003036	.0003716
9	.75	.0013155	.0016871
10	.75	.0047356	.0064227
11	.75	.0142069	.0206296
12	.75	.0355172	.0561468
13	.75	.0737666	.1299134
14	.75	.1264570	.2563704
15	.75	.1770398	.4334101
16	.75	.1991697	.6325799
17	.75	.1757380	.8083179
18	.75	.1171587	.9254765
19	.75	.0554962	.9809727
20	.75	.0166489	.9976216
21	.75	.0023784	1.0000000

Binomial Expansion, n = 21
(nonmaster's) probablity of success, r = .50

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.5	.0000005	.0000005
1	.5	.0000100	.0000105
2	.5	.0001001	.0001106
3	.5	.0006342	.0007448
4	.5	.0028539	.0035987
5	.5	.0097032	.0133018
6	.5	.0258751	.0391769
7	.5	.0554466	.0946236
8	.5	.0970316	.1916552
9	.5	.1401567	.3318119
10	.5	.1681881	.5000000
11	.5	.1681881	.6681881
12	.5	.1401567	.8083448
13	.5	.0970316	.9053764
14	.5	.0554466	.9608231
15	.5	.0258751	.9866982
16	.5	.0097032	.9964013
17	.5	.0028539	.9992552
18	.5	.0006342	.9998894
19	.5	.0001001	.9999895
20	.5	.0000100	.9999995
21	.5	.0000005	1.0000000

Binomial Expansion, n = 21
(master's) probability of success, p = .80

critical value (m)	p	mth term	sum of first m terms, α
0	.8	.0000000	.0000000
1	.8	.0000000	.0000000
2	.8	.0000000	.0000000
3	.8	.0000000	.0000000
4	.8	.0000000	.0000000
5	.8	.0000000	.0000000
6	.8	.0000005	.0000005
7	.8	.0000040	.0000045
8	.8	.0000280	.0000325
9	.8	.0001616	.0001941
10	.8	.0007756	.0009697
11	.8	.0031025	.0040722
12	.8	.0103417	.0144140
13	.8	.0286387	.0430526
14	.8	.0654598	.1085125
15	.8	.1221917	.2307041
16	.8	.1832875	.4139916
17	.8	.2156324	.6296240
18	.8	.1916732	.8212972
19	.8	.1210568	.9423539
20	.8	.0484227	.9907766
21	.8	.0092234	1.0000000

Binomial Expansion, n = 21
(nonmaster's) probablity of success, r = .55

critical value (m)	r	mth term	sum of first m terms. $1-\beta$
0	.55	.0000001	.0000001
1	.55	.0000013	.0000014
2	.55	.0000164	.0000178
3	.55	.0001267	.0001445
4	.55	.0006968	.0008413
5	.55	.0028957	.0037370
6	.55	.0094379	.0131749
7	.55	.0247183	.0378932
8	.55	.0528698	.0907630
9	.55	.0933380	.1841009
10	.55	.1368957	.3209966
11	.55	.1673169	.4883135
12	.55	.1704154	.6587289
13	.55	.1441976	.8029266
14	.55	.1007095	.9036360
15	.55	.0574417	.9610777
16	.55	.0263274	.9874052
17	.55	.0094641	.9968693
18	.55	.0025705	.9994398
19	.55	.0004961	.9999358
20	.55	.0000606	.9999965
21	.55	.0000035	1.0000000

Binomial Expansion, n = 21
(master's) probability of success, p = .85

critical value (m)	p	mth term	sum of first m terms, α
0	.85	.0000000	.0000000
1	.85	.0000000	.0000000
2	.85	.0000000	.0000000
3	.85	.0000000	.0000000
4	.85	.0000000	.0000000
5	.85	.0000000	.0000000
6	.85	.0000000	.0000000
7	.85	.0000001	.0000001
8	.85	.0000011	.0000012
9	.85	.0000088	.0000100
10	.85	.0000601	.0000701
11	.85	.0003404	.0004105
12	.85	.0016073	.0020177
13	.85	.0063055	.0083232
14	.85	.0204178	.0287410
15	.85	.0539937	.0827348
16	.85	.1147367	.1974714
17	.85	.1912278	.3886992
18	.85	.2408053	.6295045
19	.85	.2154574	.8449619
20	.85	.1220925	.9670544
21	.85	.0329456	1.0000000

Binomial Expansion, n = 21
(nonmaster's) probablity of success, r = .60

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.6	.0000000	.0000000
1	.6	.0000001	.0000001
2	.6	.0000021	.0000022
3	.6	.0000197	.0000220
4	.6	.0001333	.0001552
5	.6	.0006796	.0008348
6	.6	.0027184	.0035533
7	.6	.0087378	.0122911
8	.6	.0229368	.0352279
9	.6	.0496964	.0849243
10	.6	.0894535	.1743779
11	.6	.1341803	.3085582
12	.6	.1677254	.4762836
13	.6	.1741764	.6504600
14	.6	.1492940	.7997540
15	.6	.1045058	.9042598
16	.6	.0587845	.9630444
17	.6	.0259344	.9889787
18	.6	.0086448	.9976235
19	.6	.0020474	.9996709
20	.6	.0003071	.9999781
21	.6	.0000219	1.0000000

Binomial Expansion, n = 21
(master's) probability of success, p = .90

critical value (m)	p	mth term	sum of first m terms, α
0	.9	.0000000	.0000000
1	.9	.0000000	.0000000
2	.9	.0000000	.0000000
3	.9	.0000000	.0000000
4	.9	.0000000	.0000000
5	.9	.0000000	.0000000
6	.9	.0000000	.0000000
7	.9	.0000000	.0000000
8	.9	.0000000	.0000000
9	.9	.0000001	.0000001
10	.9	.0000012	.0000014
11	.9	.0000111	.0000124
12	.9	.0000830	.0000954
13	.9	.0005172	.0006127
14	.9	.0026601	.0032728
15	.9	.0111725	.0144453
16	.9	.0377071	.0521524
17	.9	.0998129	.1519653
18	.9	.1996259	.3515912
19	.9	.2836789	.6352700
20	.9	.255311	.8905810
21	.9	.109419	1.0000000

Binomial Expansion, n = 21
(nonmaster's) probablity of success, r = .65

critical value (m)	r	mth term	sum of first m terms. $1-\beta$
0	.65	.0000000	.0000000
1	.65	.0000000	.0000000
2	.65	.0000002	.0000002
3	.65	.0000023	.0000025
4	.65	.0000190	.0000214
5	.65	.0001197	.0001412
6	.65	.0005929	.0007341
7	.65	.0023597	.0030938
8	.65	.0076689	.0107627
9	.65	.0205722	.0313349
10	.65	.0458466	.0771815
11	.65	.0851437	.1623252
12	.65	.1317700	.2940952
13	.65	.1694186	.4635138
14	.65	.1797912	.6433050
15	.65	.1558190	.7991240
16	.65	.1085168	.9076408
17	.65	.0592739	.9669147
18	.65	.0244622	.9913769
19	.65	.0071731	.9985500
20	.65	.0013322	.9998822
21	.65	.0001178	1.0000000

Binomial Expansion, n = 22
(master's) probability of success, p = .66

critical value (m)	p	mth term	sum of first m terms, α
0	.66	.0000000	.0000000
1	.66	.0000000	.0000000
2	.66	.0000000	.0000000
3	.66	.0000006	.0000006
4	.66	.0000051	.0000057
5	.66	.0000358	.0000415
6	.66	.0001967	.0002381
7	.66	.0008726	.0011108
8	.66	.0031761	.0042869
9	.66	.0095906	.0138774
10	.66	.0242021	.0380795
11	.66	.0512515	.089331
12	.66	.0911975	.1805285
13	.66	.1361773	.3167058
14	.66	.1699355	.4866413
15	.66	.1759332	.6625746
16	.66	.1494139	.8119885
17	.66	.1023666	.9143551
18	.66	.0551977	.9695528
19	.66	.0225576	.9921103
20	.66	.0065682	.9986786
21	.66	.0012143	.9998929
22	.66	.0001071	1.0000000

Binomial Expansion, n = 22
(nonmaster's) probablity of success, r = .41

critical value (m)	r	mth term	sum of first m terms. $1-\beta$
0	.41	.0000091	.0000091
1	.41	.0001390	.0001481
2	.41	.0010144	.0011625
3	.41	.0046996	.0058621
4	.41	.0155126	.0213748
5	.41	.0388079	.0601827
6	.41	.0764099	.1365926
7	.41	.1213678	.2579605
8	.41	.1581382	.4160987
9	.41	.1709441	.5870428
10	.41	.1544292	.7414719
11	.41	.1170711	.8585430
12	.41	.0745750	.9331180
13	.41	.0398641	.9729820
14	.41	.0178085	.9907905
15	.41	.0066002	.9973908
16	.41	.0020066	.9993974
17	.41	.0004922	.9998896
18	.41	.0000950	.9999846
19	.41	.0000139	.9999985
20	.41	.0000014	.9999999
21	.41	.0000001	1.0000000
22	.41	.0000000	1.0000000

Binomial Expansion, n = 22
(master's) probability of success, p = .70

critical value (m)	p	mth term	sum of first m terms, α
0	.7	.0000000	.0000000
1	.7	.0000000	.0000000
2	.7	.0000000	.0000000
3	.7	.0000001	.0000001
4	.7	.0000007	.0000007
5	.7	.0000057	.0000065
6	.7	.0000378	.0000442
7	.7	.0002015	.0002458
8	.7	.0008817	.0011275
9	.7	.0032002	.0043277
10	.7	.0097074	.0140351
11	.7	.0247097	.0387448
12	.7	.0528513	.0915961
13	.7	.0948613	.1864574
14	.7	.1422919	.3287493
15	.7	.1770744	.5058237
16	.7	.1807635	.6865872
17	.7	.1488640	.8354512
18	.7	.0964859	.9319372
19	.7	.0473966	.9793338
20	.7	.0165888	.9959226
21	.7	.0036864	.9996090
22	.7	0003910	1.0000000

Binomial Expansion, n = 22
(nonmaster's) probablity of success, r = .45

critical value (m)	r	mth term	sum of first m terms. $1-\beta$
0	.45	.0000019	.0000019
1	.45	.0000349	.0000369
2	.45	.0003001	.0003370
3	.45	.0016370	.0019740
4	.45	.0063620	.0083360
5	.45	.0187389	.0270749
6	.45	.0434403	.0705152
7	.45	.0812390	.1517542
8	.45	.1246280	.2763821
9	.45	.1586174	.4349995
10	.45	.1687112	.6037108
11	.45	.1505852	.7542960
12	.45	.1129389	.8672349
13	.45	.0710804	.9383154
14	.45	.0373865	.9757018
15	.45	.0163141	.9920159
16	.45	.0058397	.9978556
17	.45	.0016863	.9995420
18	.45	.0003833	.9999252
19	.45	.0000660	.9999912
20	.45	.0000081	.9999993
21	.45	.0000006	1.0000000
22	.45	.0000000	1.0000000

Binomial Expansion, n = 22
(master's) probability of success, p = .75

critical value (m)	p	mth term	sum of first m terms. α
0	.75	.0000000	.0000000
1	.75	.0000000	.0000000
2	.75	.0000000	.0000000
3	.75	.0000000	.0000000
4	.75	.0000000	.0000000
5	.75	.0000004	.0000004
6	.75	.0000031	.0000035
7	.75	.0000212	.0000247
8	.75	.0001193	.0001440
9	.75	.0005565	.0007005
10	.75	.0021705	.0028710
11	.75	.0071034	.0099744
12	.75	.0195345	.0295089
13	.75	.0450796	.0745885
14	.75	.0869392	.1615276
15	.75	.1391027	.3006303
16	.75	.1825723	.4832026
17	.75	.1933118	.6765144
18	.75	.1610932	.8376075
19	.75	.1017430	.9393506
20	.75	.0457844	.9851349
21	.75	.0130812	.9982162
22	.75	.0017838	1.0000000

Binomial Expansion, n = 22
(nonmaster's) probablity of success, r = .50

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.5	.0000002	.0000002
1	.5	.0000052	.0000055
2	.5	.0000551	.0000606
3	.5	.0003672	.0004277
4	.5	.0017440	.0021718
5	.5	.0062785	.0084503
6	.5	.0177891	.0262394
7	.5	.0406609	.0669003
8	.5	.0762391	.1431394
9	.5	.1185942	.2617335
10	.5	.1541724	.4159060
11	.5	.1681881	.5840940
12	.5	.1541724	.7382665
13	.5	.1185942	.8568606
14	.5	.0762391	.9330997
15	.5	.0406609	.9737606
16	.5	.0177891	.9915497
17	.5	.0062785	.9978282
18	.5	.0017440	.9995723
19	.5	.0003672	.9999394
20	.5	.0000551	.9999945
21	.5	.0000052	.9999998
22	.5	.0000002	1.0000000

Binomial Expansion, n = 22
(master's) probability of success, p = .80

critical value (m)	p	mth term	sum of first m terms, α
0	.8	.0000000	.0000000
1	.8	.0000000	.0000000
2	.8	.0000000	.0000000
3	.8	.0000000	.0000000
4	.8	.0000000	.0000000
5	.8	.0000000	.0000000
6	.8	.0000001	.0000001
7	.8	.0000012	.0000013
8	.8	.0000088	.0000101
9	.8	.0000547	.0000648
10	.8	.0002844	.0003492
11	.8	.0012410	.0015902
12	.8	.0045504	.0061406
13	.8	.0140011	.0201417
14	.8	.0360029	.0561446
15	.8	.0768062	.1329508
16	.8	.1344108	.2673616
17	.8	.1897565	.4571181
18	.8	.2108405	.6679586
19	.8	.1775499	.8455085
20	.8	.1065299	.9520385
21	.8	.0405828	.9926213
22	.8	.0073787	1.0000000

Binomial Expansion, n = 22
(nonmaster's) probablity of success, r = .55

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.55	.0000000	.0000000
1	.55	.0000006	.0000007
2	.55	.0000081	.0000088
3	.55	.0000660	.0000748
4	.55	.0003833	.0004580
5	.55	.0016863	.0021444
6	.55	.0058397	.0079841
7	.55	.0163141	.0242982
8	.55	.0373865	.0616846
9	.55	.0710804	.1327651
10	.55	.1129389	.2457040
11	.55	.1505852	.3962892
12	.55	.1687112	.5650005
13	.55	.1586174	.7236179
14	.55	.1246280	.8482458
15	.55	.0812390	.9294848
16	.55	.0434403	.9729251
17	.55	.0187389	.9916640
18	.55	.0063620	.9980260
19	.55	.0016370	.9996630
20	.55	.0003001	.9999631
21	.55	.0000349	.9999981
22	.55	.0000019	1.0000000

Binomial Expansion, n = 23
(master's) probability of success, p = .66

critical value (m)	p	mth term	sum of first m terms $\propto$
0	.66	.0000000	.0000000
1	.66	.0000000	.0000000
2	.66	.0000000	.0000000
3	.66	.0000002	.0000002
4	.66	.0000021	.0000023
5	.66	.0000155	.0000179
6	.66	.0000905	.0001083
7	.66	.0004265	.0005348
8	.66	.0016558	.0021906
9	.66	.0053570	.0075477
10	.66	.0145585	.0221062
11	.66	.0333989	.0555050
12	.66	.0648331	.1203382
13	.66	.1064906	.2268288
14	.66	.1476551	.3744839
15	.66	.1719747	.5464586
16	.66	.1669167	.7133753
17	.66	.1334178	.8467931
18	.66	.0863292	.9331223
19	.66	.0441000	.9772223
20	.66	.0171212	.9943435
21	.66	.0047479	.9990914
22	.66	.0008379	.9999293
23	.66	.0000707	1.0000000

Binomial Expansion, n = 23
(nonmaster's) probablity of success. r = .41

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.41	.0000054	.0000054
1	.41	.0000858	.0000911
2	.41	.0006555	.0007466
3	.41	.0031887	.0039353
4	.41	.0110793	.0150146
5	.41	.0292569	.0442715
6	.41	.0609931	.1052646
7	.41	.1029351	.2081997
8	.41	.1430623	.3512620
9	.41	.1656937	.5169557
10	.41	.1612003	.6781560
11	.41	.1323879	.8105439
12	.41	.0919984	.9025422
13	.41	.0540955	.9566378
14	.41	.0268513	.9834890
15	.41	.0111956	.9946847
16	.41	.0038900	.9985747
17	.41	.0011131	.9996878
18	.41	.0002578	.9999456
19	.41	.0000472	.9999928
20	.41	.0000066	.9999993
21	.41	.0000007	1.0000000
22	.41	.0000000	1.0000000
23	.41	.0000000	1.0000000

Binomial Expansion, n = 23
(master's) probability of success, p = .70

critical value (m)	p	mth term	sum of first m terms, α
0	.7	.0000000	.0000000
1	.7	.0000000	.0000000
2	.7	.0000000	.0000000
3	.7	.0000000	.0000000
4	.7	.0000002	.0000003
5	.7	.0000022	.0000025
6	.7	.0000153	.0000178
7	.7	.0000869	.0001047
8	.7	.0004056	.0005103
9	.7	.0015773	.0020875
10	.7	.0051524	.0072399
11	.7	.0142081	.0214480
12	.7	.0331522	.0546002
13	.7	.0654543	.1200545
14	.7	.1090905	.2291450
15	.7	.1527267	.3818716
16	.7	.1781811	.5600528
17	.7	.1711936	.7312464
18	.7	.1331506	.8643970
19	.7	.0817591	.9461562
20	.7	.0381543	.9843104
21	.7	.0127181	.9970285
22	.7	.0026978	.9997263
23	.7	.0002737	1.0000000

Binomial Expansion, n = 23
(nonmaster's) probablity of success. r = .45

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.45	.0000011	.0000011
1	.45	.0000201	.0000212
2	.45	.0001808	.0002019
3	.45	.0010354	.0012373
4	.45	.0042357	.0054731
5	.45	.0131693	.0186424
6	.45	.0323247	.0509671
7	.45	.0642296	.1151966
8	.45	.1051029	.2202995
9	.45	.1433222	.3636217
10	.45	.1641690	.5277907
11	.45	.1587419	.6865326
12	.45	.1298798	.8164124
13	.45	.0899168	.9063292
14	.45	.0525488	.9588779
15	.45	.0257967	.9846746
16	.45	.0105532	.9952278
17	.45	.0035553	.9987831
18	.45	.0009696	.9997528
19	.45	.0002088	.9999615
20	.45	.0000342	.9999957
21	.45	.0000040	.9999997
22	.45	.0000003	1.0000000
23	.45	.0000000	1.0000000

Binomial Expansion, n = 23
(master's) probability of success, p = .75

critical value (m)	p	mth term	sum of first m terms, α
0	.75	.0000000	.0000000
1	.75	.0000000	.0000000
2	.75	.0000000	.0000000
3	.75	.0000000	.0000000
4	.75	.0000000	.0000000
5	.75	.0000001	.0000001
6	.75	.0000010	.0000012
7	.75	.0000076	.0000088
8	.75	.0000457	.0000545
9	.75	.0002286	.0002831
10	.75	.0009600	.0012431
11	.75	.0034037	.0046468
12	.75	.0102112	.0148581
13	.75	.0259208	.0407788
14	.75	.0555445	.0963233
15	.75	.0999800	.1963033
16	.75	.1499701	.3462734
17	.75	.1852571	.5315305
18	.75	.1852571	.7167877
19	.75	.1462556	.8630433
20	.75	.0877534	.9507967
21	.75	.0376086	.9884053
22	.75	.0102569	.9986621
23	.75	.0013379	1.0000000

Binomial Expansion, n = 23
(nonmaster's) probablity of success, r = .5

critical value (m)	r	mth term	sum of first m terms, $1-\beta$
0	.5	.0000001	.0000001
1	.5	.0000027	.0000029
2	.5	.0000302	.0000330
3	.5	.0002111	.0002441
4	.5	.0010556	.0012997
5	.5	.0040113	.0053110
6	.5	.0120338	.0173448
7	.5	.0292250	.0465698
8	.5	.0584500	.1050198
9	.5	.0974166	.2024364
10	.5	.1363833	.3388197
11	.5	.1611803	.5000000
12	.5	.1611803	.6611803
13	.5	.1363833	.7975636
14	.5	.0974166	.8949802
15	.5	.0584500	.9534302
16	.5	.0292250	.9826552
17	.5	.0120338	.9946890
18	.5	.0040113	.9987003
19	.5	.0010556	.9997559
20	.5	.0002111	.9999670
21	.5	.0000302	.9999971
22	.5	.0000027	.9999999
23	.5	.0000001	1.0000000

Binomial Expansion, n = 23
(master's) probability of success, p = .80

critical value (m)	p	mth term	sum of first m terms, α
0	.8	.0000000	.0000000
1	.8	.0000000	.0000000
2	.8	.0000000	.0000000
3	.8	.0000000	.0000000
4	.8	.0000000	.0000000
5	.8	.0000000	.0000000
6	.8	.0000000	.0000000
7	.8	.0000003	.0000004
8	.8	.0000027	.0000031
9	.8	.0000180	.0000210
10	.8	.0001006	.0001217
11	.8	.0004757	.0005974
12	.8	.0019029	.0025003
13	.8	.0064405	.0089408
14	.8	.0184015	.0273423
15	.8	.0441636	.0715058
16	.8	.0883271	.1598330
17	.8	.1454800	.3053129
18	.8	.1939733	.4992862
19	.8	.2041824	.7034686
20	.8	.1633459	.8668145
21	.8	.0933405	.960155
22	.8	.0339420	.994097
23	.8	.0059030	1.000000

Binomial Expansion, n = 23
(nonmaster's) probablity of success, r = .55

critical value (m)	r	mth term	sum of first m terms, 1 − β
0	.55	.0000000	.0000000
1	.55	.0000003	.0000003
2	.55	.0000040	.0000043
3	.55	.0000342	.0000385
4	.55	.0002088	.0002472
5	.55	.0009696	.0012169
6	.55	.0035553	.0047722
7	.55	.0105532	.0153254
8	.55	.0257967	.0411221
9	.55	.0525488	.0936708
10	.55	.0899168	.1835876
11	.55	.1298798	.3134674
12	.55	.1587419	.4722093
13	.55	.1641690	.6363783
14	.55	.1433222	.7797005
15	.55	.1051029	.8848034
16	.55	.0642296	.9490329
17	.55	.0323247	.9813576
18	.55	.0131693	.9945269
19	.55	.0042357	.9987627
20	.55	.0010354	.9997981
21	.55	.0001808	.9999788
22	.55	.0000201	.9999989
23	.55	.0000011	1.0000000

Appendix K

Approximation Tables for Agreement Coefficient and Cohen's Kappa Coefficient

Table K.1

Approximate Values of the Agreement Coefficient Based on the Standardized Cutoff Score, |z|, and a Reliability Coefficient, r

	r								
|z|	.10	.20	.30	.40	.50	.60	.70	.80	.90
.00	.53	.56	.60	.63	.67	.70	.75	.80	.86
.10	.53	.57	.60	.63	.67	.71	.75	.80	.86
.20	.54	.57	.61	.64	.67	.71	.75	.80	.86
.30	.56	.59	.62	.65	.68	.72	.76	.80	.86
.40	.58	.60	.63	.66	.69	.73	.77	.81	.87
.50	.60	.62	.65	.68	.71	.74	.78	.82	.87
.60	.62	.65	.67	.70	.73	.76	.79	.83	.88
.70	.65	.67	.70	.72	.75	.77	.80	.84	.89
.80	.68	.70	.72	.74	.77	.79	.82	.85	.90
.90	.71	.73	.75	.77	.79	.81	.84	.87	.90
1.00	.75	.76	.77	.77	.81	.83	.85	.88	.91
1.10	.78	.79	.80	.81	.83	.85	.87	.89	.92
1.20	.80	.81	.82	.84	.85	.86	.88	.90	.93
1.30	.83	.84	.85	.86	.87	.88	.90	.91	.94
1.40	.86	.86	.87	.88	.89	.90	.91	.93	.95
1.50	.88	.88	.89	.90	.90	.91	.92	.94	.95
1.60	.90	.90	.91	.91	.92	.93	.93	.95	.96
1.70	.92	.92	.92	.93	.93	.94	.95	.95	.97
1.80	.93	.93	.94	.94	.94	.95	.95	.96	.97
1.90	.95	.95	.95	.95	.95	.96	.96	.97	.98
2.00	.96	.96	.96	.96	.96	.97	.97	.97	.98

Table K.2

Approximate Values of the Kappa Coefficient Based on the Standardized Cutoff Score, |z|, and a Reliability Coefficient, r

	r								
|z|	.10	.20	.30	.40	.50	.60	.70	.80	.90
.00	.06	.13	.19	.26	.33	.41	.49	.59	.71
.10	.06	.13	.19	.26	.33	.41	.49	.59	.71
.20	.06	.13	.19	.26	.33	.41	.49	.59	.71
.30	.06	.12	.19	.26	.33	.40	.49	.59	.71
.40	.06	.12	.19	.25	.32	.40	.48	.58	.71
.50	.06	.12	.18	.25	.32	.40	.48	.58	.70
.60	.06	.12	.18	.24	.31	.39	.47	.57	.70
.70	.05	.11	.17	.24	.31	.38	.47	.57	.70
.80	.05	.11	.17	.23	.30	.37	.46	.56	.69
.90	.05	.10	.16	.22	.29	.36	.45	.55	.68
1.00	.05	.10	.15	.21	.28	.35	.44	.54	.68
1.10	.04	.09	.14	.20	.27	.34	.43	.53	.67
1.20	.04	.08	.14	.19	.26	.33	.42	.52	.66
1.30	.04	.08	.13	.18	.25	.32	.41	.51	.65
1.40	.03	.07	.12	.17	.23	.31	.39	.50	.64
1.50	.03	.07	.11	.16	.22	.29	.38	.49	.63
1.60	.03	.06	.10	.15	.21	.28	.37	.47	.62
1.70	.02	.05	.09	.14	.20	.27	.35	.46	.61
1.80	.02	.05	.08	.13	.18	.25	.34	.45	.60
1.90	.02	.04	.08	.12	.17	.24	.32	.43	.59
2.00	.02	.04	.07	.11	.16	.22	.31	.42	.58

Appendix L

Data from Final Testing Stage

Table L.1

Grade 9 van Hiele Quadrilateral Test Data and Nova Scotia Achievement Basic Concepts Test Data

Subject	Raw score on level subtest				Assigned mastery level for each interpretation scheme		Standard score on the Basic Concepts Test
	1	2	3	4	3,3,3,3	3,3,4.3	
N1	3	3	1	2	2	2	60
N2	4	4	1	1	2	2	55
N3	4	0	1	2	1	1	50
N4	4	3	1	0	2	2	57
N5	3	3	2	0	2	2	50
N6	3	2	3	1	1	1	57
N7	4	2	2	0	1	1	59
N8	4	1	2	0	1	1	63
N9	2	0	3	1	0	0	34
N10	3	4	4	0	3	3	60
N11	3	4	2	0	2	2	64
N12	4	1	1	2	1	1	34
N13	4	2	0	0	1	1	59
N14	4	3	0	0	2	2	53
N15	4	4	1	0	2	2	54
N16	3	4	1	1	2	2	50
N17	3	4	0	0	2	2	56
N18	3	4	2	0	2	2	68
N19	3	2	2	0	1	1	50
N20	3	2	1	0	1	1	55
N21	2	1	1	1	0	0	54
N22	3	4	4	1	3	3	66
N23	4	3	0	0	2	2	34
N24	4	3	0	0	2	2	49
N25	3	2	0	1	1	1	64
N26	3	2	1	0	1	1	53
N27	4	3	2	0	2	2	61
N28	4	1	0	2	1	1	60
N29	4	3	1	0	2	2	61
N30	3	3	2	1	2	2	60

(table continues)

Subject	Raw score on level subtest 1	2	3	4	Assigned mastery level for each interpretation scheme 3,3,3,3	3,3,4,3	Standard score on the Basic Concepts Test
N31	4	1	1	0	1	1	59
N32	3	4	3	1	3	2	68
N33	4	2	2	0	1	1	69
N34	4	3	1	2	2	2	67
N35	4	5	1	1	2	2	57
N36	4	4	1	1	2	2	57
N37	3	3	0	2	2	2	57
N38	2	4	2	0	0	0	49
N39	2	2	2	1	0	0	53
N40	3	2	3	0	1	1	63
N41	3	2	0	0	1	1	57
N42	4	4	1	0	2	2	54
N43	4	2	2	0	1	1	60
N44	3	3	3	0	3	2	68
N45	4	3	2	2	2	2	71
N46	4	5	2	4	2	2	71
N47	4	2	2	1	1	1	54
N48	4	2	1	0	1	1	50
N49	4	4	0	0	2	2	59
N50	3	4	2	0	2	2	59

Table L.2

Grade 12 van Hiele Quadrilateral Test Data and Nova Scotia Achievement Basic Concepts Test Data

Subject	Raw score on level subtest				Assigned mastery level for each interpretation scheme		Standard score on the Basic Concepts Test
	1	2	3	4	3,3,3,3	3,3,4,3	
T1	4	5	1	1	2	2	57
T2	3	2	1	0	1	1	54
T3	3	2	0	1	1	1	56
T4	4	3	6	2	3	3	56
T5	3	2	2	2	1	1	56
T6	4	3	4	1	3	3	69
T7	4	5	5	2	3	3	58
T8	4	4	5	2	3	3	60
T9	4	4	4	4	4	4	66
T10	4	5	1	2	2	2	50
T11	4	4	5	3	4	4	68
T12	4	5	6	3	4	4	73
T13	3	1	1	2	1	1	54
T14	4	4	5	3	4	4	68
T15	3	3	1	3	2	2	56
T16	4	4	2	3	2	2	58
T17	3	4	1	2	2	2	56
T18	4	1	2	0	1	1	56
T19	4	1	3	1	1	1	62
T20	3	4	4	3	4	4	63
T21	3	3	3	4	4	2	60
T22	3	3	3	3	4	2	63
T23	4	3	3	2	3	2	66
T24	3	2	1	0	1	1	52
T25	4	1	2	1	1	1	54
T26	4	3	3	1	3	2	60
T27	3	5	1	2	2	2	68
T28	4	4	4	2	3	3	58
T29	4	4	2	2	2	2	56
T30	4	4	6	4	4	4	71

(table continues)

Subject	Raw score on level subtest 1	2	3	4	Assigned mastery level for each interpretation scheme 3,3,3,3	3,3,4,3	Standard score on the Basic Concepts Test
T31	4	3	2	2	2	2	63
T32	4	4	4	3	4	4	63
T33	4	4	3	2	3	2	64
T34	4	5	3	1	3	2	53
T35	4	5	5	3	4	4	73
T36	4	3	1	3	2	2	54
T37	4	5	5	3	4	4	73
T38	4	3	3	3	4	2	56
T39	4	3	4	4	4	4	54
T40	3	3	4	2	3	3	66
T41	4	3	2	2	2	2	62
T42	4	4	2	2	2	2	53
T43	4	3	1	1	2	2	56
T44	4	2	2	0	1	1	57
T45	4	4	1	2	2	2	56
T46	3	3	5	2	3	3	58
T47	3	2	3	1	1	1	60
T48	4	1	4	0	1	1	53
T49	3	3	1	1	2	2	47
T50	4	5	5	4	4	4	68

REFERENCES

Assaf, S. A. (1985). The effects of using Logo turtle graphics in teaching geometry on eighth grade students' level of thought, attitudes toward geometry and knowledge of geometry. Dissertation Abstracts International, 46, 2952A. (University Microfilms No. 8512288).

Berk, R. A. (1980). A consumers' guide to criterion-referenced test reliability. Journal of Educational Measurement, 17(4). 323 - 349.

Berk, R. A. (1984). Selecting the index of reliability. In R. Berk (Ed.) A guide to criterion-referenced test construction (pp. 231 -266). Baltimore: The Johns Hopkins University Press.

Bishop, A. J. (1983). Space and geometry. In R. Lesh & M. Landau (Eds.) Acquisition of mathematics concepts and processes (pp. 175-203). New York: Academic Press.

Bobango, J.C. (1987). Van Hiele levels of geometric thought and student achievenment in standard content and proof writing: The effect of phase-based instruction. Dissertation Abstracts International, 48, 2556A, (University Microfilms No. 8727983).

Burger, W. F. and Shaughnessy, J. M. (1986). Characterizing the van Hiele levels of development in geometry. Journal for Research in Mathematics Education, 17(1), 31-48.

Carpenter, T. P. (1980). Research in Cognitive Development. In R. J. Shumway (Ed.), Research in Mathematics Education (pp. 146-206). Reston, VA: The National Council of Teachers of Mathematics.

Carpenter, T. P., Lindquist, M.M., Matthews, W. & Silver, E.A. (1983). Results of the third NAEP mathematics assessment: Secondary school. Mathematics Teacher, 76, 652 - 659.

Craine, T. V. (1985). Integrating geometry into the secondary mathematics curriculum. In C.R. Hirsch & M. J. Zweng (Eds.), The secondary school mathematics curriculum, 1985 Yearbook of National Council of Teachers of Mathematics (pp. 119-133). Reston,Va: The National Council of Teachers of Mathematics.

Cockroft, W. H. (1982). Report of the committee of inquiry into the teaching of mathematics in schools. London: Her Majesty's Stationery Office.

Cox, P. L. (1985). Informal geometry--more is needed. Mathematics Teacher, 78, 404-405.

Coxford, A. (1978). Research directions in geometry. In R. Lesh & D.Mierkiewicz (Eds.), Recent research concerning the development of spatial and geometric concepts. Columbus, Ohio: ERIC/SMEAC.

Crowley, M. L. (1984). The role of van Hiele levels of geometric thought in Canadian secondary school mathematics textbooks. (Unpublished manuscript).

Crowley, M. L. (in press). Criterion-referenced reliability indices associated with the VAN HIELE GEOMETRY TEST. Journal for Research in Mathematics Education.

Denis, L. P. (1987). Relationships between stage of cognitive development and van Hiele level of geometric thought among Puerto Rican adolescents. Dissertation Abstracts International, 48, 859A. (University Microfilms No. 8715795).

Edwards, A. L. (1973). Statistical methods. (3rd ed.). New York: Holt, Rinehart and Winston.

Fey, J. T. & Good, R. A. (1985). Rethinking the sequence and priorities of high school mathematics curricula. In C.R. Hirsch & M. J. Zweng (Eds.), The secondary school mathematics curriculum, 1985 Yearbook of the National Council of Teachers of Mathematics (pp. 43 - 52). Reston, Va: National Council of Teachers of Mathematics.

Freudenthal, H. (1973). Mathematics as an educational task. Dordrecht, Holland: D. Reidel.

Fuys, D., Geddes, D. & Tischler, R. (1984). English translation of selected writings of Dina van Hiele-Geldof and Pierre M. van Hiele (Grant no. SED 7920640). Washington, D.C.: National Science Foundation.

Fuys, D., Geddes, D. & Tischler, R. (1985). An investigation of the van Hiele model of thinking in geometry among adolescents (Grant no. SED 7920640). Washington, D.C.: National Science Foundation.

Gearhart, G. (1975). What do geometry teachers think about the high school geometry controvery?. Mathematics Teacher, 75, 486-493.

Glass, G. V. & Stanley, J. C. (1970). Statisical Methods in Education and Psychology. Englewood Cliffs, NJ: Prentice-Hall.

Guttman, L. (1944). A basis for scaling qualitative data. American Sociological Review, (9), 139 - 150.

Hoffer, A. (1979). Geometry, A model of the universe. Menlo Park, CA: Addison-Wesley.

Hoffer, A. (1981). Geometry is more than proof. Mathematics Teacher, (74), 11-18.

Hoffer, A. (1983). Van Hiele-based research. In R. Lesh and M. Landau (Eds.) Acquisition of Mathematics Concepts and Processes, (205-227). New York: Academic Press.

Huynh, H. (1976). On the reliability of decisions in domain-referenced testing. Journal of Educational Measurement, (13), 253-264.

Kay, C. S. (1986). Is a square a rectangle? The development of first-grade students' understanding of quadrilaterals with implications for the van Hiele theory of the development of geometric thought. Dissertation Abstracts International. 47. 2934A. (University Microfilms No. 8628890).

Kerr, D. R. (1981). A geometry lesson from national assessment. Mathematics Teacher, (74), 27-32.

Lindquist, M. M., & Shulte, A. P. (Eds.). (1987). Learning and teaching geometry, K - 12: 1987 Yearbook. Reston, VA: National Council of Teachers of Mathematics.

Livingston, S. A. & Wingersky, M. S. (1979). Assessing the reliability of tests used to make pass/fail decisions. Journal of Educational Measurement, 16(4), 247-260.

Lowry, J. A. (1987). An investigation of nine-year-olds' geometric concepts of area and perimeter. Dissertation Abstracts International, 48, 1971A, (University Microfilms No. 8725526).

MacPherson, E. D. (1985). The themes of geometry: Design of the nonformal geometry curriculum. In C.R. Hirsch & M. J. Zweng (Eds.), The secondary school mathematics curriculum, 1985 Yearbook, (pp. 65 - 80). Reston, VA: National Council of Teachers of Mathematics.

Mayberry, J. W. (1981) An investigation in the van Hiele levels of geometric thought in undergraduate preservice teachers. Dissertation Abstracts International, 42, 2008A. (University Microfilms No. 80-23078).

Mehrens, W.A., & Lehmann, I. R. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York: Holt, Rinehart and Winston.

Minium, E. W. (1978). Statistical reasoning in psychology and education (2nd ed.). New York: John Wiley & Sons.

National Council of Supervisors of Mathematics. (1978). Position statement on basic skills. Mathematics Teacher, (71), 147-152.

National Council of Teachers of Mathematics. (1980). An agenda for action. Reston, Va: National Council of Teachers of Mathematics.

Nitko, A. J. (1983). Educational tests and measurement: An introduction. New York: Harcourt Brace Jovanovich.

Nitko, A. J. (1984). Defining "criterion-referenced test". In R. Berk (Ed.) A guide to criterion-referenced test construction (pp. 8 - 28). Baltimore: The Johns Hopkins University Press.

Norris, D. O. (1981). Let's put computers into the mathematics curriculum . Mathematics Teacher, (74), 24-26.

Nova Scotia Department of Education. (1989). Guide to the Nova Scotia Achievement Tests, Level 9. Halifax, N.S.

Nova Scotia Department of Education. (1989). Guide to the Nova Scotia Achievement Tests, Level 12. Halifax, N.S.

Peng, C-Y. J., & Subkoviak, M. J. (1980), A note on Huynh's normal approximation procedure for estimating criterion-referenced reliability. Journal of Educational Measurement. (17), 359 - 368.

Scally, S. P. (1987, July). The effects of learning logo on ninth grade students' understanding of geometric relations. Paper presented at the XI Psychology of Mathematics Education meeting, Montreal, Quebec.

Senk, S. (1985). How well do students write geometry proofs? Mathematics Teacher, (78), 448-456.

Severin, S. T. (1987). Analysis of a grade nine intended, implemented and attained geometry curriculum in terms of the van Hiele levels of thinking. Unpublished master's of education paper, University of Western Ontario, London, Ontario.

Shaughnessy, J. M. and Burger, W. F. (1985). Spadework prior to deduction in Geometry, Mathematics Teacher, (78), 419-428.

Suydam, M. N. (1985). The shape of instruction in geometry: Some highlights from research, Mathematics Teacher, (78), 481-486.

Subkoviak, M. J. (1976). Estimating reliability from a single administration of a criterion-referenced test, Journal of Educational Measurement, 13(4), 265-276.

Subkoviak, M. J. (1984). Estimating the reliability of mastery-nonmastery classifications. In R. Berk (Ed.) A guide to criterion-referenced test construction (pp. 267 -291). Baltimore: The Johns Hopkins University Press.

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests, Journal of Educational Measurement, 25(1), 47-55.

Usiskin, Z. (1982). Van Hiele levels and achievement in secondary school geometry. Chicago: University of Chicago, Cognitive Development and Achievement in Secondary School Geometry Project.

Usiskin, Z. (1987). Resolving the continuing dilemmas in school geometry. In M. M. Lindquist & A. P. Shulte (Eds.). Learning and Teaching Geometry, K-12 (pp. 17-31). Reston. Va: National Council of Teachers of Mathematics.

Van Hiele, P. M. (1984a). English summary by Pierre Marie van Hiele of the problem of insight in connection with school children's insight into the subject matter of geometry. In D. Fuys, D. Geddes and R. Tischler (Eds.), English Translation of selected writings of Dina van Hiele-Geldof and P. M. Van Hiele. (pp. 237-241). Washington, D. C: National Science Foundation. (Original work published in 1957)

Van Hiele, P. M. (1984b). A child's thought and geometry. In D.Geddes, D. Fuys and R. Tischler (Eds.), English Translation of selected writings of Dina van Hiele-Geldof and P. M. Van Hiele. (pp. 243 - 252). Washington, D. C: National Science Foundation. (Original work published 1959)

Van Hiele, P. M. (1986). Structure and insight: A theory of mathematics education. New York: Academic Press.

Van Hiele-Geldof, D. (1984). Dissertion of Dina van Hiele-Geldof entitled: Didactics of geometry in the lowest class of secondary school. In D. Fuys, D. Geddes & R. Tischler. An investigation of the van Hiele model of thinking in geometry among adolescents (Grant no. SED 7920640). Washington, D.C.: National Science Foundation. (Original work published in 1957)

Williams, E. (1980). An investigation of senior high school students' understanding of the nature of mathematical proof. Journal for Research in Mathematics Education, 11, 165-166.

Willson, W.W. (1977). Geometry. London: Blackie and Son.

Wirzup, I. (1976). Breakthroughs in the psychology of learning and teaching geometry. In J. L. Martin (Ed.) Space and geometry: Papers from a research workshop. Columbus, Ohio: ERIC/SMEAC.