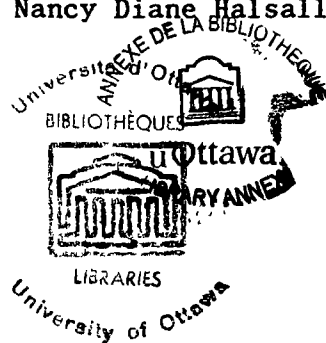


AN INVESTIGATION OF THE EFFECTIVENESS
OF THREE JUDGMENTAL TECHNIQUES
UNDER TWO DEGREES OF DOMAIN ELABORATION
IN ESTABLISHING THE ITEM-DOMAIN CONGRUENCE
OF CRITERION-REFERENCED TEST ITEMS

by

Nancy Diane Halsall



Thesis presented to the School of Graduate
Studies of the University of Ottawa
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy



Nancy Halsall, Ottawa, Canada, 1989

UMI Number: DC53930

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform DC53930
Copyright 2011 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgments

The preparation of a doctoral thesis requires the commitment and time of many people over a long period. Appreciation is extended to my personal friends, colleagues at the Carleton Board of Education, the staff and students at the University of Ottawa and Carleton University who participated in this research, and the personnel of the Ontario Ministry of Education and the British Columbia Ministry of Education for their assistance and encouragement.

Special gratitude is extended to Dr. Marc Gessaroli whose knowledge and dedication have been greatly valued throughout the preparation of this thesis.

Profound thanks are due Dr. Marvin Boss for the assistance he has provided over the years as exhorter, advisor, and teacher about research.

To my husband, Tom, who propitiously became a computer programmer in time to write the many programs to analyze the data for this research, as always, my deep appreciation.

Curriculum Studiorum

Nancy Diane Halsall was born on March 4, 1942, in Massachusetts. She received her B.A. and Teaching Certificate from Jackson College, Tufts University, in 1964. An M.Ed. and Special Education Certificate were received from McGill University in 1973.

Abstract

Because the widespread use of criterion-referenced tests is relatively new, there is a recognized need to study certain aspects of the technology of criterion-referenced test development (Hambleton, 1980; Popham, 1980).

One aspect of test development needing investigation is that of establishing the match between an item and the domain of behavior the item is intended to measure. The match between an item and a domain is termed item-domain congruence.

In the existing literature regarding item-domain congruence, it is apparent that the degree to which a domain is elaborated or described is believed to be an important factor in establishing congruence. This belief has not been tested, nor has there been extensive investigation of the techniques available for collecting ratings of item-domain congruence by knowledgeable individuals.

The purpose of the study was to investigate the effectiveness of three techniques for establishing the item-domain congruence of criterion-referenced test items. The hypothesis that high domain elaboration would result in higher item-domain congruence ratings was tested. As well, three judgmental techniques were investigated and the opinions of the judges regarding the judgmental techniques were studied. A set of geography test items and a set of physical education test items were administered to students. Judges knowledgeable in the fields of geography and of physical education were

asked to rate item-domain congruence for the items and the domains intended to be measured by the items. The judges used three different techniques: (1) the Item-Objective Congruence Index, (2) the Adjective Rating Scales, and (3) the Cross Matching Procedure. Half the judges used high domain elaboration and half used low domain elaboration.

No consistent difference was found among the ratings on each of the three judgmental techniques given by judges using high domain elaboration and judges using low elaboration. In addition, the three judgmental techniques were found to be highly correlated with one another. The examinee response data did not show the test items to be correlated with one another. Hence, the response data were not used to confirm the domain structure suggested by the judges. It was found that assigning items to the domain on which they received the highest rating, then using cut-off values, resulted in most items being assigned to their intended domains and few items assigned to nonintended domains. The Item-Objective Congruence Index emerged as the preferred judgmental technique because items tended not to be assigned to multiple domains. Judges found the Cross Matching Procedure easiest to use but felt the Adjective Rating Scales allowed the best judgment to be made regarding item-domain congruence. Judges were satisfied that the type of elaboration that they used, whether high or low elaboration, was sufficient for judging item-domain congruence.

In future studies of item-domain congruence, control of the effect of instruction on the responses of examinees might provide useful information about the relationship between the domains indicated by judges and the domains resulting from an analysis of examinee response data. Understanding of the functioning of the judgmental techniques would be extended by studying the use of the techniques in establishing item-domain congruence in additional curriculum areas.

Table of Contents

Chapter	Page
1. INTRODUCTION	1
2. REVIEW OF THE LITERATURE	6
Evolving Definition of Criterion-Referenced Testing	6
Methods of Establishing Item-Domain Congruence	9
The development of domain descriptions.	10
Judgmental techniques.	14
Examinee response data.	22
Studies Regarding the Establishment of Item-Domain Congruence of Criterion-Referenced Tests	23
Domain elaboration.	23
Judgments of knowledgeable individuals.	24
Summary of the Review of the Literature	33
The Research Purpose, Hypothesis, and Questions	34
Significance of the Study	35
3. RESEARCH METHODOLOGY	36
Research Design	36
Curriculum areas.	40
Domains.	40
Domain elaboration.	41
Judgmental techniques.	42
Judges.	42
Items.	43

Administration of the Questionnaires to Judges	46
Personal and Attitudinal Data	49
Pilot Study	50
Examinees	51
Data Analysis	52
Limitations of the Study	55
4. RESULTS OF THE ANALYSIS OF THE DATA	56
The Validity of Judgmental Ratings	56
The Research Hypothesis	74
The First Research Question	84
The Second Research Question	107
5. SUMMARY AND DISCUSSION	115
The Establishment of the Validity of the Judge Group Ratings	115
The Research Hypothesis and Questions	117
6. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH	123
REFERENCES	127
APPENDICES	132
A. Domain Descriptions	132
B. Geography and Physical Education Items	141
C. Presentation to Judges	159
D. Forms for the Three Judgmental Techniques	170
E. Personal and Attitudinal Data Forms	175

List of Tables

		Page
Table 1	Organization of Administration of Geography Items to Judges	38
Table 2	Organization of Administration of Physical Education Items to Judges	39
Table 3	Correlations between Judges' Ratings on Scale 1 and Scale 2 of the ARS	59
Table 4	Judge Group Ratings for Geography Fifth Domain Items	61
Table 5	Judge Group Ratings for Physical Education Fifth Domain Items	63
Table 6	Variances of Ratings for Geography Fifth Domain Items by Judge Group	65
Table 7	Variances of Ratings for Physical Education Fifth Domain Items by Judge Group	67
Table 8	Years of Teaching Experience of Judge Groups	71
Table 9	First Language of Judges by Groups	73
Table 10	Judge Group Ratings for Intended Domain Geography Items on Each Judgmental Technique	75
Table 11	Judge Group Ratings for Intended Domain Physical Education Items on Each Judgmental Technique	78
Table 12	Mean, Median, and Standard Deviation for the Total Test Using Ratings on Intended Domains	81
Table 13	Correlations among Judgmental Techniques with High and with Low Domain Elaboration using Ratings on Intended and Nonintended Domains	83
Table 14	Correlation Matrix for Geography Intended Domain Items	86
Table 15	Correlation Matrix for Physical Education Intended Domain Items	89

Table 16	Geography Item Difficulties and Item-Total Domain Correlation Coefficients	92
Table 17	Physical Education Item Difficulties and Item-Total Domain Correlation Coefficients	93
Table 18	Difficulties and Standard Deviations of Total Test and Intended Domains for Geography and Physical Education	94
Table 19	Combined Elaboration Judge Group Ratings for Intended Domain Geography Items on Each Judgmental Technique	97
Table 20	Combined Elaboration Judge Group Ratings for Intended Domain Physical Education Items on Each Judgmental Technique	98
Table 21	Correlations among Combined High and Low Elaboration Ratings on Intended and Nonintended Domains for the Three Judgmental Techniques	100
Table 22	Correlations among the Three Judgmental Techniques by Domain Using Combined High and Low Elaboration Ratings on Intended and Nonintended Domains	101
Table 23	Number of Items Assigned to their Intended Domain and to Nonintended Domains using Judge Group Ratings with Various Cut-Off Values	103
Table 24	Number of Items Assigned to their Intended Domain and to Nonintended Domains using the Highest Judge Group Ratings in Combination with Cut-Off Values	105
Table 25	Judgmental Technique Found Easiest to Use by Judges	108
Table 26	Judgmental Technique Felt to Allow Best Judgment	109
Table 27	Judges' Feeling Regarding Sufficiency of Type of Elaboration Used	110
Table 28	Judges' Feeling Regarding Type of Elaboration Not Used	111

Chapter 1

Introduction

In both regular and special education, there has been a move in the last two decades toward the individualization of instruction in order to enhance the learning of all students. This interest in the individual learning needs of students may be traced to the establishment of compensatory education programs in the United States in the 1960s and in Canada in the development of a network of special education services in the 1970s. The trend toward individualization of curriculum was entrenched in practice with the passage of legislation such as The Education Amendment Act (1980) in Ontario. What was desired and what individualization has come to mean is that each student's current knowledge in an area is evaluated and a program of instruction is planned to ensure that the objectives of the program which have not been mastered will be taught. Individualization required that accurate information be gathered regarding appropriate tasks to include in student programs. Also, frequently, the terms of educational grants have required that evidence of student progress through the curriculum be provided.

During the same period of time from the 1960s to the 1980s, psychologists introduced the notion of programmed instruction to educators. Programmed instruction, which was purported to be able to be carried out by techniques such as programmed texts and computer-aided instruction, provided a way for student programs to be

individualized. When students worked at their own pace, their progress could no longer be monitored by reference to the progress of other members of their group. Therefore, a means of evaluating students in terms of curriculum goals was needed.

Traditional or norm-referenced tests are composed of items which are chosen because they contribute variability to test scores. It is difficult to use an item on such a test to evaluate a specific program or an individual's performance in that program, since the domain covered by the item is often not clearly indicated (Popham, 1978). Without precise information about what each item measures, it is difficult to plan a specific program for an individual based on his test performance. Also, there is a penalty for good teaching and learning. As Hambleton, Swaminathan, Algina, and Coulson (1978) noted, items tapping concepts taught successfully by a great number of teachers will contribute little to test score variability as most students will answer the items correctly. Hence, items reflecting learning may not be included in a test.

Both Traub (1972) and Millman (1974) recognized that it is possible to interpret an examinee's performance on a norm-referenced test item in terms of a domain if the item can be matched with a specific domain. Over the years, educational diagnosticians using norm-referenced tests have developed clinical insights and ways of linking examinee response data to instructional goals in order to arrive at understanding an individual's achievement in a program and

to make recommendations about the individual's future program needs. However, experience and intuition are not constant, and testing to clarify what learning an individual has mastered would be better served by methodical, empirical procedures which emerge directly from the test itself. Criterion-referenced testing offers a solution to the problem of establishing an individual's achievement and future program needs by couching the examinee's responses in the context of a clearly stated domain of behavior. A behavioral domain has been defined as a broad set of skills (Hambleton, 1980). The skills included in a domain of behavior may, in the terminology of Bloom (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956), be cognitive skills, affective skills, or psycho-motor skills. Specific skills which are subsumed by the broader domain would tend to be similar in that they could be categorized as belonging to the same domain.

The aims of individualizing curriculum and evaluating students in terms of their progress through the designated curriculum converge in criterion-referenced testing. Haladyna and Roid (1983b) have described this convergence of educational interests regarding individualization and assessment as an effort to integrate teaching and testing. Essential to the integration of teaching and testing are clear statements of instructional intent and tests which tap the knowledge and skills reflected in those statements (Haladyna & Roid, 1983b). Criterion-referenced tests have been developed to answer this need for items referenced to a specific skill and tests which

are interpreted by referring to a specified set of tasks.

As this new emphasis on the use of criterion-referenced tests in educational assessment has progressed, it has become apparent that there is a need for a review of the technology of test development to ascertain what procedures might best serve the purposes of criterion-referenced tests (Berk, 1978, 1980b; Hambleton, 1980; Hambleton & Eignor, 1978; Hambleton, Swaminathan, Algina, & Coulson, 1978; Popham, 1978, 1980).

One area which has been of particular concern is that of establishing the validity of criterion-referenced test items. Since each item is an indicator of what the student can or cannot do, test developers have devoted increasing emphasis to demonstrating that each item on a test is a measure of the intended skill. This concern with the match between item and skill is a part of the test validation process. The entire process of gathering evidence concerning the inferences that might be made regarding a test, called validation, enhances the understanding of test results. Validity has been said to be the most important consideration in evaluating a test (American Psychological Association, 1985). The establishment of the validity of a test involves a continuing examination of the soundness of the interpretation of a test used in varying situations (American Psychological Association, 1985; Cronbach, 1971). Item validity in criterion-referenced tests is established when it can be shown that there is item-domain congruence, that is, a match between

an item and the domain it was intended to measure.

Further investigation is needed of the techniques available to rate the match between an item and the domain of behavior the item is intended to measure. In the literature on item-domain congruence in criterion-referenced tests, it is contended that the degree to which a domain is elaborated or described is an important factor in establishing congruence (Popham, 1978).

The purpose of this research was threefold. The first purpose was to investigate the notion that the extent to which a domain is elaborated is a major factor in establishing item-domain congruence for criterion-referenced test items. The second purpose was to investigate and compare the effectiveness of three judgmental techniques for establishing item-domain congruence. The third purpose was to determine the opinions of the judges of item-domain congruence regarding the three judgmental techniques.

This research report consists of six chapters. In chapter 1, the research study is introduced. In chapter 2, the literature regarding techniques for establishing item-domain congruence is reviewed and the research questions are discussed. In chapter 3, the research methodology is explained. In chapter 4, the results of the analysis of the data are described. In chapter 5, the results are summarized and discussed. In chapter 6, the conclusions are drawn and suggestions for further research are given. As well, references and appendices are included.

Chapter 2

Review of the Literature

Chapter 2 begins with a discussion of the evolving definition of criterion-referenced measurement. Next, the methods of establishing item-domain congruence are reviewed. Then the literature regarding establishing item-domain congruence is reviewed. The chapter is concluded with a summary of the review of the literature, the purpose of the research, the statement of the research questions, and a discussion of the significance of the study.

Evolving Definition of Criterion-Referenced Testing

Interest in criterion-referenced test development has resulted from attempts in education to make instructional goals more explicit in order to improve and expedite learning. Instructional programs based on clearly defined curricula and outcomes required new ways of measuring individual student performance. By contrast with norm-referenced testing, which provides an indication of an individual's standing in relation to other examinees, criterion-referenced testing is designed to define an individual's status with respect to a domain of behavior (Hambleton et al., 1978; Millman, 1974; Popham, 1978). The examinee's performance is referenced to a behavioral domain, which is a collection of skills organized around a broader skill (Hambleton, 1980). The set of skills or tasks must be well-defined so that it is possible to determine whether a given task

belongs to the set or not (Zwarts, 1982). Thus, criterion-referenced testing is a means of providing information regarding the skills an individual has mastered and, consequently, a basis for individualizing student programs. In addition, criterion-referenced testing enables the monitoring of the progress of students as well as judging the appropriateness of educational instruction and placement. The use of the term criterion-referenced testing is not static and is still evolving as the field is explored and developed.

Ebel (1962) and Glaser (1963) are credited with the first conceptualization of criterion-referenced testing and with distinguishing between norm-referenced and criterion-referenced testing. Popham and Husek (1969) then elaborated the advantages and disadvantages of the two approaches to testing for individual instruction and program decisions. Popham (1978) subsequently made explicit that the criterion to which an individual's status was referenced was a well-defined behavioral domain. Hambleton et al. (1978) found three terms where confusion and overlap existed. These terms were: criterion-referenced tests, domain-referenced tests, and objectives-referenced tests. In attempting to clarify the matter of definition, it was concluded that Popham's definition of a criterion-referenced test is essentially the same as Millman's (1974) definition of a domain-referenced test. Hambleton et al. (1978) further believed that objectives-referenced tests in which the test items are matched to instructional objectives do not define a domain of behavior. In a criterion-referenced test, the items are

representative of a clearly defined domain of behavior. By contrast, no domain of behavior is specified with an objectives-referenced test, and items are not considered to be representative of any specific behavioral domain.

Further refining and classifying of the definition of criterion-referenced tests has taken place. Berk (1980a) delineated two types of criterion-referenced measurement: domain-referenced tests and mastery tests. Berk's definition of a domain-referenced test is in keeping with Hambleton's and Millman's. Mastery tests are used to classify students as masters and nonmasters of an objective. According to Berk, domain-referenced tests and mastery tests fit the definition of criterion-referenced tests as defining an individual's status with respect to a behavioral domain. Both types of tests require rigorous and precise domain specifications to heighten the interpretability of a domain score.

Nitko (1980) provided a framework for integrating the many concepts of criterion-referenced testing. The major distinction elucidated by Nitko was that between sequentially ordered and unordered domains. Ordered domains are those in which a hierarchical learning sequence is specified. Unordered domains, although they can be defined well, cannot be arranged in a hierarchy of skills.

There will undoubtedly be additional classification schemes which emerge from insights gained from the development and use of criterion-referenced tests. In summary, there does, as noted by

Berk (1980a), appear to be general agreement that a criterion-referenced test is intended to reference an individual's score to a well-defined domain (Berk, 1980a; Hambleton et al., 1978; Livingston, 1977; Millman, 1974; Nitko, 1980; Popham, 1978). The methods by which the relationship between a domain and a test item may be established are next discussed.

Methods of Establishing Item-Domain Congruence

The relationship between an item and the domain of which it is purported to be a measure has been said to be the most important consideration in the construction and use of criterion-referenced tests (Berk, 1980b; Dahl, 1971; Popham, 1978, 1981). The relationship between an item and the objective it is intended to measure is termed item-objective congruence. By extension, the relationship between an item and the domain it is intended to measure is termed item-domain congruence.

There are three junctures in the development of a criterion-referenced test at which item-domain congruence can be studied and enhanced. The first point is during test construction, where item-domain congruence can be enhanced by the careful description of domains and the careful development of items to measure those domains. The second point is after the items and domains have been developed, when judgmental techniques may be used to evaluate item-domain congruence. The third point is after a field test of the items, when examinee response data may be examined. Each of these

opportunities for establishing item-domain congruence will now be examined.

The development of domain descriptions. Considerable thought has been given to ways of developing domain descriptions in order to enhance item-domain congruence.

Hambleton and Eignor (1978) pointed out that statements of the specific skills and accomplishments expected of a student are not sufficient in and of themselves for establishing item-domain congruence. Without a description of the broad domain which encompasses the specific skills it would then not be possible to determine whether a representative sample of test items has been selected to measure the domain. Since the purpose of criterion-referenced testing is to infer an examinee's performance in the total domain by administering items which sample the domain, statements of specific skills alone may not be sufficient to indicate what is being measured. Without a domain by which to judge the appropriateness of an item, the validity of the test items could not be established for item-domain congruence is a necessary condition for test validity (Berk, 1980b).

Because of the difficulty in establishing domains from behavioral objectives, Popham (1978, 1980) and Roid and Haladyna (1980), have recommended the use of amplified objectives. An amplified objective has been defined (Millman, 1974) as an expanded statement of an educational goal which provides boundary specifications regarding testing situations, response alternatives, and criteria of

correctness.

Popham (1980) refined the amplified objective to an elaborate test specification which consisted of (1) a general description of the behavior being assessed, (2) an illustrative item, (3) attributes of the class of stimuli the examinee would encounter, (4) attributes of the selected or constructed responses, and (5) a supplement providing extra information such as a detailed listing or explanation of eligible content. Popham (1978) stated that domain specifications should sufficiently delineate the class of behavior under consideration so that independent judges would register high agreement regarding whether particular test items are measures of the behavior described in that domain.

Another approach to domain description, the item form scheme, was demonstrated by Hively, Maxwell, Rabehl, Sension, and Lundin (1973). The approach was even more elaborate than Popham's in that it required a very complete elaboration of the domain. It has been pointed out (Hambleton & Eignor, 1978) that while the Hively et al. approach serves highly structured curriculum areas such as mathematics, it is not as applicable to less highly structured areas such as language arts. Popham (1980) found the approach too constraining and overly specific, and Hively himself is reported (Popham, 1980) to have doubted that the precision of domain specifications in a mathematical field could be duplicated in other fields.

Yet another approach, termed facet design or analysis, has been developed to define a domain and construct items (Guttman & Schlesinger, 1966, 1967). In facet design, the domain is defined by a mapping sentence which links the facets or the dimensions on which the items may differ from one another. One facet, or part of the item, may be substituted for a facet of another item. The mapping sentence is used to generate descriptions of the domain. This approach is said to be useful for hierarchically ordered objectives (Zwarts, 1982).

Popham (1980) attempted to systematize the many techniques for domain description. Popham envisaged a continuum based on specificity of the domain. Specificity for Popham (1980) meant the degree of detail or elaboration of a domain. The degree of elaboration could range from a low range of elaboration to the high degree of elaboration of amplified objectives. The degree of elaboration of a domain was said to determine the extent to which generalizations can be made from the test performance to the broader behavior which the test is intended to measure (Popham, 1978). Accordingly, Hambleton et al. (1978) pointed out that without the domain being specified, although one may be implied or assumed, the generalizations that can be made regarding test scores are limited.

There is a lack of clarity in the writings regarding the theoretical relationship between domains and elaboration and the manner in which these may be described by amplified objectives. Hambleton

and Eignor (1978) stated that instructional objectives do not define a domain of behavior. Yet Popham (1980) wrote of amplified objectives as a means of describing a domain. Hambleton and Eignor (1978) said that domain specifications are a logical extension of amplified objectives. Finally, Popham (1980) stated that behavioral objectives are domain descriptions with a low degree of elaboration. The clarification of terminology in this area would clearly be useful in establishing generally accepted usage of terms which in turn would facilitate understanding of the theoretical issues regarding domains and elaboration.

The specifications for a test serve as instructions for writing test items which must be in keeping with the guidelines contained in the domain descriptions. A well-developed literature exists regarding the generation of test items (Cronbach, 1970; Gronlund, 1976; Hambleton & Eignor, 1978; Thorndike & Hagen, 1977). When item generation rules are used (Hambleton & Eignor, 1978), the process of item writing itself should ensure a relationship between an item and the domain. In summary, when domain descriptions are well-delineated and items are written using those guidelines, there is reason to believe that there will be congruence between items and the domain they are intended to measure. Thus, test construction provides an opportunity to strengthen item-domain congruence. The second opportunity to enhance item-domain congruence arises through the use of judgmental techniques.

Judgmental techniques. In addition to the careful description of domains and the development of items, evidence for item-domain congruence may also be provided by soliciting the opinions of individuals who are knowledgeable about the domains measured by the test. Individuals knowledgeable in the curriculum area covered by a test may be test developers, curriculum supervisors, classroom teachers, university professors, advanced students, or practitioners in a specific field. Such individuals, referred to as judges, may be asked whether a test item is congruent with, or matches, a domain which has been described.

Three approaches to gathering judgmental data were identified by Rovinelli (1976). These approaches were categorizing, rating, and matching of items and objectives. Rovinelli located and/or developed a specific technique for each approach to gathering judgments. For the categorizing approach, Rovinelli located the index of item homogeneity (Hemphill & Westie, 1950) which he extended and named the item-objective congruence index (IOCI). For the rating approach, Rovinelli located the semantic differential technique (Osgoode, Suci, & Tannenbaum, 1957) which he developed into an adjective rating scale (ARS). For the matching approach, Rovinelli developed a cross-matching procedure (CMP). These techniques were originally put forth in the context of objectives-referenced tests, but the techniques are equally applicable in the context of domain-referenced tests. Each technique will now be discussed.

The item-objective congruence index (IOCI) was based on Hemphill and Westie's (1950) categorizing procedure to assess the validity of test items. They developed a series of scales to describe the characteristics of social groups and illustrated a method of using judgments in the assignment of items to the categories comprising the descriptive system. This procedure was called the index of item homogeneity. Judges were asked to rate items by assigning a value of +1, 0, or -1 to indicate the extent to which items were felt to be a measure of each objective intended to be measured by the test. A rating of +1 indicated that the judge felt an item was a measure of an objective; 0 indicated that the judge was undecided about whether or not the item was a measure of the objective; and -1 indicated that the item was not a measure of an objective. The computational formula for the index of item homogeneity (Hemphill & Westie, 1950) was:

$$I_{it} = \frac{(N \sum_{j=1}^n X_{ijt} - \sum_{i=1}^N \sum_{j=1}^n X_{ijt})}{(2 * 2 * n * (N-1) + \sum_{i=1}^N \sum_{j=1}^n X_{ijt} - \sum_{j=1}^n X_{ijt})}$$

where:

I_{it} = index of homogeneity for item congruence on objective i

N = number of objectives ($i = 1, 2, 3, \dots, N$)

n = number of content specialists ($j = 1, 2, 3, \dots, n$)

X_{ijt} = the value of the rating (-1, 0, or +1) assigned to item congruence on objective i by content specialist j

Rovinelli (1976) found two difficulties in using the index of item homogeneity. The first difficulty was felt to be that there were not predefined limits to the possible values resulting from the calculation of the index. The second difficulty was felt to be that the values for different items could not be compared if different numbers of judges rated the items. Hence, Rovinelli developed the item-objective congruence index (IOCI) to overcome the difficulties with the Hemphill-Westie index of item homogeneity.

In the IOCI, item-domain congruence was represented by a value from -1 to +1. A value of -1 would indicate that all the judges rating an item assigned a -1 to the item for the appropriate domain and a +1 to the item for all the other domains measured on the test. Conversely, a value of +1 would indicate that all the judges assigned a +1 to the item for the appropriate domain and a -1 to the item for all the other objectives. The value of the IOCI would not depend on the number of judges or the number of domains. The formula for the IOCI (Rovinelli, 1976) is:

$$I_{ik} = ((N-1) * \sum_{j=1}^i X_{1jk} - \sum_{i=1}^i \sum_{j=1}^i X_{ijk} + \sum_{j=1}^i X_{ijk}) / (2 * (N-1) * n)$$

where:

I_{ik} = index of item-objective congruence for item k
on objective i

N = number of objectives ($i = 1, 2, 3, \dots, N$)

n = number of judges ($j = 1, 2, 3, \dots, n$)

X_{ijk} = the rating (-1, 0, or +1) of item k as a measure
of objective i by judge j

The value obtained by using the IOCI indicates the extent to which judges assigned a perfect rating to the match of the item with the domain. The value of the IOCI is affected by the domain the item is intended to measure. As well, the IOCI is calculated in such a way that the value for items judged to measure more than one domain will be lowered. The format of the answer sheet was such that it was possible for the judges to be cognizant of their responses regarding all the objectives. A scale from 3 to 1 and the following format were used by Rovinelli (1976):

Remember that you should use the following rating scale:

- 3 - If you feel the item is definitely a measure of the objective.
- 2 - If you cannot make a decision whether the item is a measure of the objective.
- 1 - If you feel the item is definitely not a measure of the objective.

Item	Objective					
	1	2	3	4	5	6
1	—	—	—	—	—	—
2 (p. 186)	—	—	—	—	—	—

The adjective rating scale (ARS) was based on the work of Osgoode, Suci, and Tannenbaum (1957) who presented a semantic differential technique in which adjectives which were opposites defined the limits of the possible judgments. Judges were asked to indicate the meaning of a concept on a series of bipolar adjective scales.

Drawing on the semantic differential technique, Rovinelli (1976) developed a 5-point ARS. Judges were asked to indicate the extent to which they felt an item was a relevant measure of each objective. Judges were provided a list of objectives and a list of test items. Judges were asked to assign a rating of 1 to 5 (1 = irrelevant, 2 = slightly relevant, 3 = somewhat relevant, 4 = relevant, 5 = highly relevant) to indicate the extent to which an item was a relevant measure of each objective. Thus, the ARS, while based on the semantic differential technique, did not strictly consist of the bipolar adjectives of relevant and irrelevant, but consisted of 5 points along a relevance continuum. The format of

the ARS answer sheet (Rovinelli, 1976) was:

Remember that you should assign a rating of "1" to "5" (1 = irrelevant, 2 = slightly relevant, 3 = somewhat relevant, 4 = relevant, 5 = highly relevant) to indicate the extent to which you feel each item is a relevant measure of each objective.

Item	Objective					
	1	2	3	4	5	6
1	—	—	—	—	—	—
2	—	—	—	—	—	—
(p. 179)						

Hambleton and Eignor (1978) described the ARS as a semantic differential technique. The ARS was suggested for use in establishing the item-domain congruence of criterion-referenced test items by Rovinelli (1976) and Rovinelli and Hambleton (1977). Judges are presented with an objective and all items to be rated. Judges are then asked to indicate on a 5-point scale, the term which best describes the relationship between the item and objective. The judges' ratings can be used to determine the average scale score for each item on each objective.

The following example of an objective, an item and two possible adjective rating scales was provided by Hambleton and Eignor (1978):

Beside each objective, indicate the item numbers corresponding to the items that you feel measure it.

<u>Objective</u>	<u>Items Measuring the Objective</u>
------------------	--------------------------------------

1	
---	--

2 (p. 191)	
---------------	--

Rovinelli and Hambleton (1977) suggested that a contingency table might be constructed by calculating the number of judges matching each item with a particular domain. A visual analysis of the table would show the extent of agreement among the judges as well as the nature of any disagreements among judges.

Hambleton and Eignor (1978) suggested that the accuracy of a judge's rating could be studied by including a number of items which were not a measure of any of the domains of interest in the item list. A judge's effectiveness as a rater could be measured by the extent to which he was able to determine which items were not measures of the domains of interest. Hambleton and Eignor (1978) commented that this method of evaluating judges must be used cautiously, as it would tend to favor overly critical judges.

In summary, the IOCI, ARS, and CMP are three judgmental techniques which may be employed to evaluate item-domain congruence. The use of judgmental techniques and the development of well-defined domain descriptions are the first two opportunities to enhance item-domain congruence. The third opportunity to review item-domain congruence arises during the study of examinee response data.

Examinee response data. Once a test has been administered during a trial administration or field test, examinee response data may be studied to gather information about item-domain congruence. There is an extensive, well-developed literature regarding the use of statistical indices with criterion-referenced tests (e.g., Berk, 1978). Hambleton (1980) stated that although examinee response data could not be used to answer directly the question of whether the content of an item is appropriate to measure a domain, examinee response data could be examined in reviewing the decisions of judges. Rovinelli and Hambleton (1977) explained that the analysis of examinee response data was a way of gathering information about test items and about the homogeneity of test domains. Anastasi (1976) stated that establishing content validity could include the study of test scores. Thus, items may be reviewed using examinee response data for evidence that the items are homogeneous. Using this approach, the domain would be treated as an achievement construct and the question would be whether the items functioned in a unified way, that is, were homogeneous (Anastasi, 1986). If the items functioned homogeneously when administered to examinees, then evidence is accumulated that there is unity to the characteristic which is being measured. Logically, this characteristic is taken to be the achievement construct described by the domain description. Anastasi (1986) further commented that the constructs employed in test development are derived from observed behavioral consistencies and defined through a network of observed interrelationships.

Thus, item-domain congruence can be studied and enhanced when domains are described, when judgments are gathered, and when examinee response data is analyzed. These three stages in test development are interrelated in that they offer opportunities to examine the consistency of the domains under study. The first two stages have not been examined extensively. Therefore, in the next section, the literature regarding domain description and the use of judges in establishing item-domain congruence is examined to determine what issues are in need of clarification.

Studies of the Establishment of Item-Domain Congruence of Criterion-Referenced Tests

From the evolving understanding of criterion-referenced measurement and the accompanying concern with establishing item-domain congruence, two major areas in which investigation is needed have emerged. These areas are domain elaboration and the use of judges in establishing item-domain congruence. A discussion of these areas and the unresolved issues emerging from them follows.

Domain elaboration. One issue which has received attention is that of the amount of description or elaboration necessary to define a domain sufficiently to establish item-domain congruence (Berk, 1980b; Hambleton, 1980; Popham, 1978). Domain elaboration refers to the degree of amplification of information regarding the domain to which test items are referenced. It is posited that the more elaborated the domain the greater the congruence between item and domain

will be. This relationship is believed to exist because the more explicitly a domain is described the less chance the item writer will produce an item which is not congruent with its domain (Popham, 1978).

There has been recognition that studies of domain elaboration are needed (Hambleton et al., 1978; Popham, 1980). Popham (1980) stressed the need for research regarding the efficacy of using various procedures of domain description in order to begin the development of an experimentally tested technology of criterion-referenced test construction.

In summary, although the need for research regarding domain description has been noted, there has not been a study purposely designed to investigate how domain elaboration functions in relation to item-domain congruence. The literature regarding the use of judges in establishing item-domain congruence, which will now be reviewed, is somewhat more extensive.

Judgments of knowledgeable individuals. A second issue in the evolving understanding of the establishment of item-domain congruence involves the use of individuals knowledgeable in a curriculum area to evaluate test items. Three studies bearing on the establishment of item-domain congruence will now be reviewed. In the major study (Rovinelli, 1976), behavioral objectives rather than domains were employed.

Rovinelli (1976), in reviewing methods for validating criterion-referenced tests, studied the use of judgmental techniques

using two data sets. The first data set was used only with the IOCI. The second data set was used with the IOCI, the ARS, and the CMP.

The first data set resulted from the administration of the IOCI to 21 science teachers. The teachers were asked to judge 40 science items on 11 objectives from a grade nine individualized science curriculum. The objectives used with this study were short, consisting of one sentence each. The objectives were statements of what the students were expected to do and hence took the form of very specific behavioral objectives. An item with a score of .70 and above on the IOCI was considered to belong to the intended objective. Using this procedure, 13 of the 40 items were identified as not being measures of their intended objective. Ratings ranged from .38 to .95 on the intended objectives. The ratings for the items on nonintended objectives were not presented but Rovinelli commented that the ratings were always considerably higher when the items were assessed on the intended objectives than when assessed on nonintended objectives. This finding was expected since in the calculation of the IOCI, the value for an item is lowered if the item is judged to measure more than one objective.

The second data set resulted from the administration of the three judgmental techniques to 30 science teachers using 48 items and 12 objectives. The 12 objectives and their four intended items were divided into three sets. Thus there were three sets of items

with four objectives and their four corresponding items. Each set had a total of 16 test items. Then, two additional objectives without items were assigned to each set. Each set of 6 objectives and 16 items was then assigned to a group of ten judges in such a way that each judge group used a different set of items and objectives with each judgmental technique. Using this procedure for any one set of objectives and items, data were then gathered from three different judge groups using three different judgmental techniques.

For the IOCI, a cut-off score of .70 was used. Eleven of the 48 items were identified as not being measures of their intended objectives. The ratings of the items on nonintended objectives were not presented but were reported to be lower than on intended objectives.

Rovinelli (1976) also used the second data set with an ARS. The terms ranged from irrelevant to highly relevant. It was found that the judges tended to rate items higher on the objectives they were intended to measure. This was cited as an indication that the technique provided meaningful evidence for the validity of items. Rovinelli also compared the ratings for the items on their intended and nonintended domains in order to determine whether an item was a measure of more than one objective.

When the CMP was used with the second data set, discrepancies between the expected matches of an item to an objective and the actual matches indicated invalid items. A criterion of a minimum of 70 percent agreement among the judges was used to establish whether

an item matched an objective. Six of the 48 items were found to be in need of revision or deletion using this procedure.

On the basis of the study of the two data sets, Rovinelli asserted that teachers could make meaningful judgments regarding the congruence of items and objectives. Rovinelli recommended that the IOCI be used when items were expected to be measures of only one objective because of the IOCI's inherent bias in favor of items which measure only one domain. Where it is unclear whether items measure more than one objective, Rovinelli recommended the use of either the ARS or CMP. Because the ARS and CMP are simpler to calculate, Rovinelli further recommended that one of these techniques be used when the number of items and objectives was large.

Rovinelli's (1976) study was done with two data sets from tests under development and was designed to demonstrate the feasibility of using various approaches and indices in validating criterion-referenced test items. The comparison of the approaches to establishing validity of criterion-referenced test items used by Rovinelli was limited to a listing of which items would be identified as aberrant using the various approaches. The effectiveness of the judgmental techniques was compared only on the basis of which items would be considered as flawed using each technique. Similarly, the examinee response data were used to determine which items were aberrant. The examinee response data was not used to determine which of the judgmental techniques was most effective. Thus, there seems not to be

empirical evidence to support Rovinelli's suggestions regarding which technique to use in varying situations. These suggestions appear to be based on Rovinelli's experience using the techniques and on logical understanding of the functioning of the techniques. There were also several difficulties with the study which contributed to this lack of a comparison of the effectiveness of the judgmental techniques.

Rovinelli's study is limited with regard to the use of the ratings on the judgmental techniques on nonintended objectives. The intended objective is the objective an item has been written to measure. A nonintended objective is any other objective than the one which the item has been written to measure. With the IOCI and the ARS, information regarding the ratings on some nonintended objectives was gathered but not presented. It was reported, however, that the ratings on nonintended objectives were lower than on intended objectives. The CMP was administered in such a way that judges were not required to judge each item on each objective. Thus, it was not possible to determine in this study whether an item was a measure of an objective other than the intended objective. This is understandable given Rovinelli's assumption that an item should be a measure of only one objective. However, if one does not wish to make this assumption, judgmental techniques are obviously useful in assessing the extent of congruence with all objectives or domains under consideration.

Another problem is that the judges for data set two did not

rate all the items on all the objectives. As the judgmental techniques were used, judges were exposed to objectives which had no matches with items. These objectives were then used with another technique and another set of items which did match them. Thus, the judges' exposure to the objectives varied with the judgmental technique used.

Further, although for any given technique there were objectives without matching items, there were no items without objectives they were intended to measure. Judges might thus be inclined to believe that every item matched an objective and would search for the most suitable.

Since the objectives for intended items presented to the judges changed with each technique, it was not possible to compare the functioning of the three techniques on all the objectives.

Only one cut-off value was used with each technique in deciding whether an item was aberrant. The effect of various cut-off values was not studied and guidelines for the use of the techniques have not been established.

Most importantly, the science objectives used in Rovinelli's study were highly specific and were more similar to behavioral objectives than to domains. In some cases there were only a few items to measure an objective. In fact, it seems possible that the objectives could be incorporated into one large domain of the structure of matter.

Thus, understanding of the establishment of item-domain congruence is limited in Rovinelli's (1976) work by a lack of comparison of the effectiveness of the judgmental techniques and by attendant methodological limitations. In addition, the attitudes and opinions of the judges toward the three techniques were not investigated. Since a judge's perception of the judgmental techniques might influence the use of the techniques, this type of information would have been interesting and useful in understanding the effectiveness of the three judgmental techniques.

Two further studies regarding the use of a judgmental technique with criterion-referenced tests have been conducted.

Secolsky (1983), in illustrating a procedure for detecting items which examinees found ambiguous, examined the divergence of examinee and instructor opinions regarding the domain associated with an item. A test of 16 objectively scored items covering 14 domains of electricity was administered to 81 architecture students. The students then classified the items according to the 14 content categories using the IOCI. A content category was a section heading of the electricity unit of the course manual. Although these content categories are referred to as domains, Secolsky (1983) pointed out that they were not domain specifications and that the domains were actually unspecified. The instructor (who did not write the test items) also classified the items which could belong to more than one category. Agreement between students and instructor ranged from .095 to .816 where +1 indicated that all the students agreed

with the instructor's judgment and -1 indicated that no student agreed with the instructor. The study illustrated that some students could judge the item-domain congruence of some items using the IOCI. Secolsky concluded that researchers need to know more about how such indices used to judge item-domain congruence function with various ability levels and with different tests. As well, Secolsky indicated that a more well-defined domain-referenced classification scheme would have enhanced the value of the student judgment data. Finally, Secolsky asserted that the practice of developing domain categories after items have been generated should be avoided.

Zwarts (1982) reported the construction of domains of mental arithmetic and beginning reading skills. Item-domain congruence was studied by asking judges to rate the representativeness of items from the domains. No statistical analysis was presented, but it was reported that the content validity of the tests comprised of the items was judged favorably by curriculum specialists. Although the term favorable was not defined, the thesis was that favorable judgments were rendered because of the careful construction of the domains and the explicitness of the domain descriptions. While the reading tests were directly referenced to a remedial curriculum, the arithmetic tests were not associated with instructional objectives. A classification system was constructed to specify the domains from the existing items using an item bank, facet design, item forms, and computer-based methods. Thus, Zwarts concluded that domains could

be constructed for items after the test was completed and that those domains could serve as the basis for establishing item-domain congruence. This finding contrasts with Secolsky's (1983) admonition against constructing domains after items have been generated.

Thus, in these three studies (Rovinelli, 1976; Secolsky, 1983; Zwarts, 1982), an interest is demonstrated in establishing item-domain congruence for criterion-referenced tests using the judgments of knowledgeable individuals. However, the notion of a domain has evolved over time and it is now recognized that behavioral objectives do not constitute domain descriptions. Only Zwarts (1982) could be said to have used actual domains, yet the study is limited because the techniques used to establish item-domain congruence are not reported.

Thus, three issues have emerged in the attempt to understand and clarify the idea of item-domain congruence in relation to criterion-referenced tests. The three interrelated areas needing investigation are: (1) the effect of the degree of domain elaboration on the ratings of item-domain congruence, (2) the relative effectiveness of the available judgmental techniques for analyzing item-domain congruence, and (3) the attitude of judges toward using the techniques.

The review of the literature concerning item-domain congruence will now be summarized.

Summary of the Review of the Literature

For a criterion-referenced test to be considered valid, test developers and users must be confident that each item comprising the test is congruent with the domain it is intended to measure. The congruence between an item and a domain is especially important with criterion-referenced tests because of their focus on defining an individual's status with respect to a domain of behavior.

In order to demonstrate item-domain congruence, it is necessary to have carefully defined domains. Descriptions of domains can range from low elaboration to high elaboration. The degree of elaboration of a domain is believed to be an important variable in establishing item-domain congruence. Item-domain congruence can also be explored by having judges who are knowledgeable of content rate the correspondence between items and domains. Three techniques which have been developed to obtain and analyze judgmental data are the IOCI, the ARS, and the CMP. Examinee response data may also be reviewed in establishing item-domain congruence. When the responses of examinees show that a set of items is homogeneous, the inference can be made that the items reflect the domain for which the items were written.

In reviewing studies of item-domain congruence, first, it was determined that the belief that the higher the degree of domain elaboration the greater the item-domain congruence is virtually untested empirically. Second, it was determined that, though limited, there has been some research on the use of judges knowl-

edgeable in the curriculum of interest in establishing item-domain congruence. The major work in the field has been done by Rovinelli (1976), but that study was limited by assumptions and methodology previously discussed. Third, it was determined that the attitudes and opinions of judges toward the three techniques have not been investigated.

It would be useful to have an investigation of the effectiveness of the judgmental techniques for establishing item-domain congruence under different degrees of domain elaboration. It is therefore proposed to conduct such a research project. The purpose of the research is discussed next.

The Research Purpose, Hypothesis, and Questions

The purpose of this research was to investigate three techniques for establishing the item-domain congruence of criterion-referenced test items in three interrelated ways.

First, this research was designed to test the contention that by describing a domain in extensive detail, judges would be better able to evaluate item-domain congruence. To investigate this belief, the following hypothesis was formulated: When high domain elaboration is used, item-domain congruence ratings will be higher than when low domain elaboration is used.

The second purpose of this research was to investigate the effectiveness of three judgmental techniques, the IOCI, the ARS, and the CMP, by using examinee response data. Thus, the following

research question was posed: What is the effectiveness of each of the judgmental techniques?

The third purpose of the research was to gather attitudinal data from the judges regarding the three techniques. The research question addressed to this purpose was: What were the opinions of the judges regarding the IOCI, the ARS, and the CMP?

Significance of the Study

Knowledge regarding the establishment of item-domain congruence in criterion-referenced tests would be advanced by an examination of the belief that the ratings of items given by judges using the three judgmental techniques are higher when high domain elaboration was used as compared to low domain elaboration.

In addition, an understanding of the effectiveness of each of the three judgmental techniques for establishing item-domain congruence under high and low domain elaboration would be of assistance to curriculum and test developers in the selection of appropriate techniques to measure item-domain congruence. Educational diagnosticians and classroom teachers would benefit from increased confidence that the skills designed to be measured by a test item are indeed being measured.

In chapter 3, the research design is described.

Chapter 3

Research Methodology

In chapter 3, the research design is explained. The administration of the questionnaires to the judges and the collection of personal and attitudinal data are then described. The pilot study is explained and the examinees are described. Finally, the plan for the analysis of the data is given.

Research Design

The establishment of item-domain congruence is critical for criterion-referenced tests. Since there has not been extensive reporting of research findings regarding the establishment of item-domain congruence for criterion-referenced tests, it was decided to undertake a study of item-domain congruence. The purpose of the study was: (1) to test the belief that by describing a domain in high detail, judges would be better able to evaluate item-domain congruence; (2) to investigate the effectiveness of the judgmental techniques; and (3) to gather attitudinal data from the judges regarding judgmental techniques.

Two curriculum areas, geography and physical education, were used in the study. There were four domains in each curriculum area. These domains are referred to as intended domains. The intended domains were each described using a high degree of elaboration and a low degree of elaboration. Three judgmental techniques were used in

the study, the IOCI, the ARS, and the CMP. In each curriculum area, 84 judges rated the match between items and domains. The judges were divided into groups of 14 judges. Each judge group assessed one-third of the items from each domain using the IOCI, another third of the items using the ARS, and another third of the items using the CMP. The judge groups in each curriculum area who assessed the items using high domain elaboration were designated as groups 1.1, 1.2, and 1.3. The judge groups in each curriculum area who assessed the items using low domain elaboration were designated as 2.1, 2.2, and 2.3. Thus, in geography, judge groups 1.1 and 2.1 using technique IOCI assessed items 1, 2, and 3 in each of four domains using high and low domain elaboration respectively. The same judge groups using technique ARS assessed items 4, 5, and 6 in each of four domains using high and low domain elaboration respectively. The same judge groups using technique CMP assessed items 7, 8, and 9 in each of four domains using high and low domain elaboration respectively. In geography, there were nine items for each of four intended domains. As well, a fifth domain consisting of nine items from nonintended domains and having no description was included. In physical education, there were eight items for each of three intended domains and nine items for one intended domain. There were also nine items from other nonintended domains comprising a fifth domain. The research design is summarized in Table 1 for geography and Table 2 for physical education. Each aspect of the study will now be discussed in detail.

Organization of Administration of Geography Items to Judges

Degree of elaboration	Judge group	IOCI					Techniques ARS					CMP				
		Domains					Domains					Domains				
		CC	CK	CF	CI	5	CC	CK	CF	CI	5	CC	CK	CF	CI	5
		Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items
High	1.1	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	4,5,6	4,5,6	4,5,6	4,5,6	4,5,6	7,8,9	7,8,9	7,8,9	7,8,9	7,8,9
High	1.2	4,5,6	4,5,6	4,5,6	4,5,6	1,2,3	7,8,9	7,8,9	7,8,9	7,8,9	4,5,6	1,2,3	1,2,3	1,2,3	1,2,3	7,8,9
High	1.3	7,8,9	7,8,9	7,8,9	7,8,9	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	4,5,6	4,5,6	4,5,6	4,5,6	4,5,6	7,8,9
Low	2.1	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	4,5,6	4,5,6	4,5,6	4,5,6	4,5,6	7,8,9	7,8,9	7,8,9	7,8,9	7,8,9
Low	2.2	4,5,6	4,5,6	4,5,6	4,5,6	1,2,3	7,8,9	7,8,9	7,8,9	7,8,9	4,5,6	1,2,3	1,2,3	1,2,3	1,2,3	7,8,9
Low	2.3	7,8,9	7,8,9	7,8,9	7,8,9	1,2,3	1,2,3	1,2,3	1,2,3	1,2,3	4,5,6	4,5,6	4,5,6	4,5,6	4,5,6	7,8,9

Note. Domain 5 includes all items not intended to measure domains 1 to 4.

Note. For actual items, refer to Appendix B using the following key:

Domain CC items	Domain CK items	Domain CF items	Domain CI items	Domain 5 items
1=CC02	1=CK03	1=CF02	1=CI03	1=CB08
2=CC07	2=CK16	2=CF05	2=CI08	2=CJ01
3=CC63	3=CK19	3=CF06	3=CI05	3=CL01
4=CC03	4=CK14	4=CF03	4=CI12	4=CB09
5=CC05	5=CK17	5=CF19	5=CI01	5=CJ02
6=CC58	6=CK20	6=CF18	6=CI10	6=CL09
7=CC06	7=CK15	7=CF04	7=CI07	7=CB10
8=CC61	8=CK18	8=CF13	8=CI02	8=CJ12
9=CC62	9=CK21	9=CF17	9=CI11	9=CL13

Table 2

Organization of Administration of Physical Education Items to Judges

Degree of elabor- ation	Judge group	IOCI					Techniques ARS					CMP				
		Domains					Domains					Domains				
		RU	PC	BS	FN	5	RU	PC	BS	FN	5	RU	PC	BS	FN	5
		Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items	Items
High	1.1	1,2,3	1,2,3	1,2	1,2,3	1,2,3	4,5,6	4,5	3,4,5	4,5,6	4,5,6	7,8	6,7,8	6,7,8	7,8,9	7,8,9
High	1.2	4,5,6	4,5	3,4,5	4,5,6	1,2,3	7,8	6,7,8	6,7,8	7,8,9	4,5,6	1,2,3	1,2,3	1,2	1,2,3	7,8,9
High	1.3	7,8	6,7,8	6,7,8	7,8,9	1,2,3	1,2,3	1,2,3	1,2	1,2,3	4,5,6	4,5,6	4,5	3,4,5	4,5,6	7,8,9
Low	2.1	1,2,3	1,2,3	1,2	1,2,3	1,2,3	4,5,6	4,5	3,4,5	4,5,6	4,5,6	7,8	6,7,8	6,7,8	7,8,9	7,8,9
Low	2.2	4,5,6	4,5	3,4,5	4,5,6	1,2,3	7,8	6,7,8	6,7,8	7,8,9	4,5,6	1,2,3	1,2,3	1,2	1,2,3	7,8,9
Low	2.3	7,8	6,7,8	6,7,8	7,8,9	1,2,3	1,2,3	1,2,3	1,2	1,2,3	4,5,6	4,5,6	4,5	3,4,5	4,5,6	7,8,9

Note. Domain 5 includes all items not intended to measure domains 1 to 4.

Note. For actual items, refer to Appendix B using the following key:

Domain RU items	Domain PC items	Domain BS items	Domain FN items	Domain 5 items
1=RU01	1=PC03	1=BS04	1=FN05	1=BM02
2=RU08	2=PC06	2=BS09	2=FN10	2=BM11
3=RU14	3=PC18	3=BS20	3=FN12	3=PS01
4=RU17	4=PC22	4=BS24	4=FN13	4=BM33
5=RU21	5=PC23	5=BS28	5=FN16	5=PS02
6=RU30	6=PC29	6=BS34	6=FN19	6=PS04
7=RU35	7=PC32	7=BS39	7=FN26	7=BM38
8=RU25	8=PC36	8=BS15	8=FN31	8=PS03
			9=FN37	9=PS05

Curriculum areas. Two curriculum areas, geography and physical education, were used in the study so that the findings in the two areas might be compared. Two provincial ministries of education had developed databases which were used in this study. The Ontario Ministry of Education initiated the development of pools of test items in a variety of curriculum areas. Through the Ontario Assessment Instrument Pool (OAIP), test items were made available to school boards to enable the assessment of student achievement. The Intermediate Geography document (Ministry of Education, Province of Ontario, 1982) was used as a source of domains and items for the present study. The British Columbia Ministry of Education developed items to measure knowledge in physical education as part of the Physical Education Learning Assessment (PELA). The Grade 11 PELA document (Ministry of Education, Province of British Columbia, 1980) was also used as a source of domains and items for this study.

Domains. Broad domain descriptions were used in the study rather than specific behavioral objectives. Four domains were selected in geography and physical education.

The geography domains tapped knowledge of (1) physical environment terms (CC), (2) urbanization concepts (CK), (3) Canadian resources (CF), and (4) Canadian population patterns (CI). The researcher worked with a geography curriculum specialist who was the head of a high school geography department and who had previously worked on the document. The researcher explained the concept of a

domain to the geography specialist. Drawing on the geography specialist's knowledge of the curriculum, the researcher and the geography specialist developed amplified domain descriptions for the four geography domains. In physical education, the intended domains were (1) physical education rules (RU), (2) physiological concepts (PC), (3) human anatomical systems (BS), and (4) fitness and nutrition (FN). The researcher worked with a university professor of measurement and physical education to develop domains from the brief objectives provided in the PELA document. The professor had previously worked on the PELA research project. The researcher attended to the requirements of describing domains while the professor provided expertise regarding the physical education curriculum.

Domain elaboration. The domains for each curriculum area were described in two ways: (1) with a high degree of description or elaboration and (2) with a low degree of elaboration. For high domain elaboration, Popham's (1980) categories for test specifications were used. These categories are (1) general description, which is a statement of the behavior being assessed; (2) sample item, which is an illustrative item; (3) stimulus attributes, which are a description of the class of stimuli the examinee will encounter; (4) response attributes, which are the attributes of the selected responses; and (5) specification supplement, which provides extra information regarding eligible content for items. For low domain elaboration, only the domain description and sample item were used. Thus, the domain description and sample item were identical

for the high and low domain elaboration, but the high elaborations included the three additional categories of stimulus attributes, response attributes, and specification supplement. A complete description of the high elaboration domains for both geography and physical education is contained in Appendix A.

Judgmental techniques. Three judgmental techniques were selected for study. The three techniques were the IOCI, the ARS, and the CMP. These techniques were selected because they represented the three approaches to establishing item-domain congruence of categorizing, rating, and matching items and domains. In addition, since Rovinelli (1976) had employed the three techniques, it was felt that if these techniques were used, there would be a basis of comparison to draw on with the previous research.

Judges. Although it was felt that practicing teachers would be good judges, for practical reasons, it was not possible to assemble the required number of teachers with knowledge of the appropriate levels of geography and physical education curriculum. Therefore, university students majoring in geography and physical health sciences acted as judges of whether an item was a measure of a particular domain.

The geography majors were drawn from one large senior undergraduate class at each of the two universities in the Ottawa area. The students responded to questionnaires aimed at eliciting their opinions regarding the match between the geography items and

the domains. The questionnaires were administered in April during regular class sessions. Participation in the research was encouraged by the professors, but not mandated. In one class, approximately half of the class chose to participate. In the other class, there was almost total participation. Students who did not wish to participate were free to leave the classroom. The geography professors did not stay with their classes, although one introduced the researcher and encouraged the students to participate. A total of 113 geography majors handed in questionnaires.

The physical education judges were advanced students majoring in physical health sciences such as physical education, kinanthropology, and nursing. Two introductory measurement classes for physical health science students and one physical education theory class at one university in Ottawa participated in the research during February and March. The administration of the questionnaires was during regular class time and students were strongly encouraged by their professors to participate. No students left the classroom and most students complied by completing the questionnaires. All the physical education professors remained in the classroom and assisted the researcher. A total of 106 physical health science majors handed in questionnaires.

Items. In geography, nine different items were selected from each of four domains for a total of 36 items. Nine additional items were selected from domains other than the four intended domains created for this research. These items were included in order to

avoid possible attempts by judges to indicate a match between every item and a domain. Within the context of this research, these nine items in effect create a fifth domain of items which is referred to as the fifth domain. In fact, the nine items comprising the fifth domain were actually selected from three additional domains in the OAIP geography document. Three items were selected from each of the three domains in the OAIP document. The three domains from which the fifth domain items were drawn were: (1) Canada in relation to other parts of the world, (2) the effects of industrialization and technological change, and (3) the significance of transportation systems on Canadian society.

In physical education, only eight items were available for three domains, so these were used without selection to measure each of three intended domains. Nine items were selected to measure the fourth intended domain. As well, nine items were selected for a fifth domain. The nine items in the fifth domain in physical education were selected from two sets of items for two domains in the PELA document. The fifth domain items consisted of items covering: (1) basic movement principles and (2) psychological concepts. There were four items regarding basic movement principles and five items regarding psychological concepts. The geography and physical education items are given in Appendix B.

In order to vary the sequence of presentation of items to judges, each judge's items were presented in a different random

order. However, the same order was used in the presentation to judges in the high and low elaboration groups.

The rotating of sets of one-third of the items across the three judgmental techniques and the randomization of the presentation of items to judges were intended to serve a two-fold purpose. First, the procedure was meant to lessen any possible judge effect for any particular technique, and second, to lessen any effect that the order of presentation of items might have on the ratings assigned by the judges. Items from the fifth domain were held constant within each judgmental technique. This approach enabled a comparison of the decisions made by each judge group and allowed the use of the fifth domain items as a validity check on the decisions made by the judges.

Each geography item selected as a measure of one of the four intended domains was checked with every aspect of the appropriate high elaboration domain description to ensure that it could be considered to meet the requirements of its intended domain. The same procedure was used with the physical education items and the four intended domains. Using Bloom's system of taxonomy (Bloom et al., 1956), all items would be considered to be at a knowledge level where memory of facts and recall are required to answer the item correctly. One taxonomic level of item was used in the study in order not to complicate the interpretation of findings by using items requiring the different types of skills referenced in other levels of the taxonomy. As well, the knowledge level is considered

by some to be theoretically close to Popham's amplified objectives which were used for high domain elaboration.

Both the geography and the physical education items and domain descriptions were reviewed with the five professors in charge of the classes where the students acted as judges. In addition, two research officers of a public school board familiar with the development of items for curriculum domains reviewed the items and domains in order to locate any difficulties with the match between items and domains.

Administration of the Questionnaires to Judges

Each judge received a package of materials consisting of (1) an outline of the task, (2) questionnaires including instructions and response sheets for each of three judgmental techniques, (3) domain descriptions, (4) sets of test items in either geography or physical education, and (5) a personal and attitudinal data questionnaire.

In the package of materials given to the judges, a written outline was provided for the judges to follow during the introductory remarks. At the beginning of each administration of the questionnaires to each class, a brief explanation of terms, the purpose of the research, and examples of high and low domain elaboration were given verbally to the judges. This presentation was made to facilitate the judges' task of matching items with domains and providing attitudinal data at the conclusion of the task. Instructions were then given for the completion of the questionnaires.

The protocol for the introductory remarks and presentation to the judges is contained in Appendix C along with examples of the charts used in the presentation. The introductory remarks took approximately 20 minutes. Judges worked independently. A check was made that each judge understood the task and could proceed. Questions by the judges regarding the procedures to be followed were answered when they arose.

The package given to each judge also contained four domain descriptions. Approximately half the participating geography judges and approximately half the participating physical education judges worked with high domain elaboration and the second half of the geography and physical education judges worked with low domain elaboration. In the domain descriptions, the heading 'domain description' was used in place of 'general description' to emphasize the notion of domain with the judges.

Also included in the package given to judges were instructions and response sheets for each of the three judgmental techniques. For purposes of simplifying terminology for the judges, the IOCI was referred to as categorizing; the ARS, as rating; and the CMP, as matching. Examples of the questionnaires are included in Appendix D. Each judge used each of the three judgmental techniques. Each judgmental technique was used with 15 geography items or 14 physical education items. The order in which the questionnaires for the three judgmental techniques were placed in the package was rotated

so that there would be variation in the sequence in which the judges applied the techniques in considering the match between an item and a domain. Thus, the order in which the first judge used the techniques was IOCI, ARS, CMP. The second judge used the techniques ARS, CMP, IOCI. The third judge used the techniques CMP, IOCI, ARS. This rotation was carried on to the end of the judges.

The judges worked with items, domain descriptions, and judgmental technique response sheets at one time. At the top of each response sheet, the researcher had indicated the item set and domain which were to be used.

Using the IOCI, judges were asked to rate whether an item measured a domain by circling +1, 0, or -1 to indicate that the judge felt the item was a measure of the domain, felt undecided about whether the item was a measure of the domain, or felt the item was not a measure of the domain, respectively.

Using the ARS, judges were asked to circle the number of the term which was found to describe the relationship between the domain and the item for the two scales. In the first scale, 5 indicated very relevant; 4, somewhat relevant; 3, no feeling; 2, somewhat irrelevant; and 1, very irrelevant. Similarly, in the second scale, the 5 points indicated very suitable to very unsuitable. The items for the two scales were presented to the judges in the same order. Half the judges used scale 1 first and half used scale 2 first.

Using the CMP, judges were asked to indicate whether the domain was measured by the item by circling yes or no. Each item

was to be assessed.

The materials were packaged so that all items in a set were rated on one intended domain before the items were matched with the next intended domain. The researcher distributed and collected the questionnaires. The questionnaires were completed by the judges in 45 to 90 minutes.

Personal and Attitudinal Data

After completing the three judgmental technique questionnaires, the judges completed one sheet of personal background data. The forms for the collection of the personal and attitudinal data are contained in Appendix E. It was anticipated that some of the university students who acted as judges might have teaching experience which might affect their judgments of the match between items and domains. Therefore, the judges were asked the number of years they had taught. As well, judges were asked the number of years of teaching experience they had in the area of geography or physical education.

The first language of the judges was believed to be another factor which might affect the decisions of the judges. Therefore, the judges were asked to report their first language.

The judges were also asked their opinions regarding the ease of use of the judgmental techniques and which technique they felt allowed the best judgment to be made. This information was gathered in order to compare the judges' opinions regarding the three techni-

ques.

Also, judges were asked whether they felt the type of elaboration they used was sufficient and whether it would have been more helpful (or sufficient) to have used the other type of elaboration. This question was asked in order to ascertain whether judges had formed impressions of the usefulness of types of elaboration.

Finally, judges were asked to make any further comments they wished.

Pilot Study

To ensure that university level students were able to do the task, a pilot study was conducted using younger students. Two grade 8 students and four high school students identified as academically capable by their teachers were administered the geography questionnaires. Six academically capable high school students were administered the physical education questionnaires. Apart from clarification of terminology, these students had no difficulty understanding the task, following instructions, and making decisions regarding the match between items and domains. Therefore, it was concluded that university students would be appropriate subjects for this research and that the questionnaires were designed in a way which would reflect the decisions of the subjects regarding item-domain congruence.

Examinees

High school students studying geography and physical education responded to the geography and physical education items, thus providing the examinee response data used in the study.

The high school students studying geography (N=360) consisted of grade 9 males and females in two high schools in a large school board in the Ottawa area. One high school was in a rural community and the second was in a large suburban setting. The curriculum content covered by the geography items was part of the material intended by the school board to be covered in the grade 8 studies. These geography students were administered a test consisting of the 45 items selected for the intended and fifth domains in geography. The test was administered to the students in October of grade 9 by their geography teachers. Three forms of the test were used in order to vary the sequence in which the items were presented to the students. The three forms of the geography test were created by dividing the 45 randomly ordered items into 3 sets of 15 items each. The sets of 15 items were identified as A, B, and C. Test form 1 was presented with items in the order of set A, B, and C. Test form 2 was presented in the order B, C, A. Test form 3 was presented in the order C, A, B. The tests were given out to students in the following order: form 1, 2, 3, 1, 2, 3, etc.

The high school students studying physical education (N=858) were grade 11 males and females who responded to the physical education items as part of the British Columbia Learning Assessment

(Ministry of Education, Province of British Columbia, 1980). One aspect of the learning assessment was to investigate the state of knowledge of the students in the province. Therefore, it cannot be assumed that the curriculum content was taught in any methodical way prior to the administration of the items to the students. The physical education students were a sample of grade 11 students in physical education throughout the province. All sizes of schools in various geographic regions of British Columbia were represented in the original sample of 500 males and 500 females. It was felt by the administrators of the Learning Assessment that the students absent (N=142) when the items were given to their classmates undoubtedly caused the sample to be somewhat less than representative but that the outcome of any analysis would not be unduly affected because of the large number in the sample.

Data Analysis

The plan for the analysis of the data is briefly outlined in this section. The plan is discussed in detail in chapter 4.

The judges' ratings for each item using the three techniques under the two degrees of elaboration were calculated. Then the data were examined to ensure that the decisions of the judges were valid. The analysis was conducted by comparing the ratings given to fifth domain geography and physical education items by the judge groups. In addition, the variances of the ratings of the judge groups on fifth domain items were examined to determine whether the ratings

provided by any judge group were consistently different from the ratings of other judge groups. The teaching experience of the judges was also investigated because it was thought that if any judge group contained a large number of very experienced teachers, the ratings of the judge group would be affected. As well, the first language of the judges was investigated in order to determine if any judge group might be unduly affected by working in a second language.

The data were then analyzed in order to address the research hypothesis and questions.

The hypothesis that when high domain elaboration is used, item-domain congruence will be higher than when low domain elaboration is used was first investigated. The mean, median, and standard deviation of the judge group ratings for all items on their intended domain in geography and in physical education were calculated. The mean, median, and standard deviation for items rated by judges using high elaboration were compared with the same statistics for the same items rated by judges using low elaboration for each judgmental technique. In addition, the correlations between judge group ratings for items rated with high elaboration and with low elaboration were studied in order to determine the strength of any relationship between high elaboration ratings and low elaboration ratings regardless of the magnitude of those ratings.

Next, the research question of the effectiveness of each of the

judgmental techniques was to be investigated through confirmatory factor analyses of the examinee response data for various domain structures. The domain structures to be examined were: (1) the intended domain structure, (2) the structures suggested by the judge group ratings using various cut-off values, and (3) the structure suggested by the highest judge group ratings on the four domains. In preparation for the confirmatory factor analyses of these domain structures, tetrachoric correlations were calculated among the geography and among the physical education items using the examinee response data. The matrix of tetrachoric correlations was to be analyzed by an unweighted least-squares estimation procedure and the adequacy of model fit was to be assessed by examining the residual correlations. The domain structures having the lowest residuals would be considered to have been produced by effective judgmental techniques. In addition, the effectiveness of the judgmental techniques was to be investigated by calculating the internal consistency of the examinee response data for the various domain structures. High coefficient alphas would be an indication of an effective judgmental technique.

The opinions of the judges regarding the three judgmental techniques were studied by examining the judges' comments regarding which method they found easiest to use and which technique they felt allowed them to make the best judgment of item-domain congruence. Judges' opinions regarding the sufficiency of the type of elaboration used and the sufficiency of the type of elaboration not used

were also gathered. The comments of the judges were also studied.

Limitations of the Study

In this research, item-domain congruence was studied in only two curriculum areas: geography and physical education. Both these curriculum areas can be considered skill application areas. It may be that the findings would be different in basic skills areas such as mathematics and language arts. It may also be that the findings would be different had the items used in the study been at taxonomic levels other than the knowledge level.

In addition, in order to obtain a sufficient number of judges for the research design, it was necessary to use university students to provide the item-domain congruence ratings. While the university students were knowledgeable in the two curriculum areas used in the study, they could not be assumed to have expertise in writing test items or in test development.

As well, the conditions created by administering the judgmental techniques to large groups of judges and by working within the time limits imposed by class sessions are not typical of the conditions under which such judgments would typically be made.

In chapter 4, the analysis of the data is discussed.

Chapter 4

Results of the Analysis of the Data

In chapter 4, the results of the analysis of the ratings by judges using the three judgmental techniques are presented. The analysis of the examinee response data is also presented.

The research hypothesis and questions to be addressed in the analysis of the data were:

(1) When high domain elaboration is used, item-domain congruence ratings will be higher than when low domain elaboration is used.

(2) What is the effectiveness of each of the judgmental techniques?

(3) What were the opinions of the judges regarding the IOCI, the ARS, and the CMP?

The Validity of Judgmental Ratings

The data were first examined to investigate the validity of judgmental decisions.

Preliminary steps were taken to ensure the validity of the judgmental ratings in the preparation of the judgmental questionnaires for data processing. Each judge group was oversampled because it was anticipated that some judges might have to be eliminated from the analysis because they were unable to use the judgmental techniques. The total number of geography judges was 113.

In physical education, the total number of judges was 106. A random sample of questionnaires was drawn from each of the six geography and six physical education judge groups. When a judge had omitted a technique or had used an obvious response pattern, all the questionnaires supplied by that judge were discarded and another questionnaire was drawn. An example of an obvious response pattern would be the selection of the same response for each item rated on a particular domain. When a judge occasionally omitted a response regarding an item, the judge was excluded from rating that particular item on the four domains. The random sampling proceeded until there were 14 judges for each judge group for a total of 42 geography and 42 physical education judges who used high elaboration and the same numbers using low elaboration who met the criteria described.

Next, the ratings given by each judge to each item on each domain were compared in order to determine whether any judge was consistently different from the other judges in assigning ratings to items. Judges who rated items higher or lower than ten of the other judges in a judge group were noted. If a judge who rated five of the intended domain items consistently different from the other judges using one of the judgmental techniques had been found, the appropriate ratings would have been removed from further analysis. Using this criterion, it was not necessary to remove any judges.

At this point, the judge group ratings for all items on all domains for all judgmental techniques under the two degrees of

elaboration were calculated. The IOCI was calculated using the formula provided by Rovinelli and Hambleton (1977). The ARS item values were calculated by combining the values on scale 1 (relevance) and scale 2 (suitability) and taking the average. The average of the two scales is referred to as the combined adjective rating scales (CARS). The two scales were combined because they were found to be highly correlated with one another. The correlations between judges' ratings on scale 1 and 2 are shown in Table 3. The CMP ratings were calculated by taking the percentage of judges who chose yes, indicating that there was a match between the item and the domain.

Next, the nine items in the geography fifth domain and the nine items in the physical education fifth domain which were held constant for the six judge groups using each judgmental technique were studied. The fifth domain items were held constant in order to determine whether there were consistent differences in the ratings assigned by the six judge groups in each curriculum area. In geography and in physical education, there were 42 judges who reviewed three fifth domain items using the IOCI, three items using the two ARS, and three items using the CMP under each degree of elaboration. In contrast, for items in the intended domains, only 14 judges reviewed a particular item using a specific technique and degree of elaboration. The use of the fifth domain items with all the 42 high elaboration judges and 42 low elaboration judges in a

Table 3

Correlations between Judges' Ratings on Scale 1 and Scale 2
of the ARS

Curriculum Area/Elaboration	Correlation between Scales 1 and 2
Geography/High	.984
Geography/Low	.984
Physical Education/High	.991
Physical Education/Low	.969

curriculum area enabled the comparison of the ratings given by the six judge groups. As well, a comparison of the ratings of the three high elaboration judge groups and three low elaboration judge groups on each item was possible. The ratings for fifth domain items by geography and physical education judge groups are presented in Tables 4 and 5.

Overall, for the three fifth domain items used with the IOCI, the ratings of the six judge groups were similar. There was no discernible difference among elaboration groups or any pattern of differences among judge groups in geography and physical education. The physical education items BM02 and BM11 rated on the IOCI, items BM33 and PS04 rated on the CARS, and items BM38 and PS07 rated on the CMP tended to be viewed as belonging to domain RU. An inspection of the physical education fifth domain items led to the conclusion that these judgments were plausible. Therefore, this phenomenon was interpreted as a valid assignment of items to domains rather than as reluctance of the physical education judges to rate an item as not belonging to any intended domain.

Next, the variances of the ratings of the 6 geography judge groups and the 6 physical education judge groups on their respective fifth domain items were examined. The variances for the judge groups on the 9 fifth domain geography items and the 9 fifth domain physical education items are given in Tables 6 and 7. The variances of these items were examined in order to determine whether the judges in any group had consistently higher variation in their

Judge Group Ratings for Geography Fifth Domain Items

Technique	Item	Elabora- tion	Judge group	Domains CC	CK	CF	CI
IOCI	CB08	High	1.1	0.80	-0.25	-0.35	-0.20
			1.2	0.54	-0.27	-0.18	-0.08
			1.3	0.46	-0.39	0.13	-0.20
		Low	2.1	0.57	-0.29	-0.24	-0.05
			2.2	0.59	-0.18	-0.18	-0.23
			2.3	0.32	-0.11	-0.11	0.11
	CJ01	High	1.1	-0.21	0.02	-0.21	0.40
			1.2	-0.13	0.20	-0.42	0.35
			1.3	-0.51	0.15	0.01	0.35
		Low	2.1	-0.07	0.21	-0.36	0.21
			2.2	-0.35	0.37	-0.20	0.18
			2.3	-0.14	0.14	-0.33	0.33
	CL01	High	1.1	0.14	0.19	-0.38	0.05
			1.2	0.13	0.18	-0.30	-0.01
			1.3	0.02	0.21	0.07	-0.31
Low		2.1	0.40	0.07	-0.36	-0.12	
		2.2	0.21	0.17	-0.31	-0.07	
		2.3	0.14	0.00	-0.14	0.00	
CARS	CB09	High	1.1	3.46	2.60	1.50	2.43
			1.2	2.64	1.82	1.73	1.78
			1.3	2.72	1.72	1.33	2.14
		Low	2.1	3.68	2.54	1.65	2.75
			2.2	3.89	2.12	1.46	2.23
			2.3	2.85	2.62	2.08	3.20
	CJ02	High	1.1	3.50	2.68	1.93	1.90
			1.2	3.50	2.35	2.00	2.57
			1.3	3.47	2.79	1.86	2.93
		Low	2.1	3.27	3.04	2.19	3.00
			2.2	3.54	2.40	1.79	2.57
			2.3	2.50	2.40	2.14	2.64
	CL09	High	1.1	2.82	1.64	3.68	1.68
			1.2	3.04	1.40	4.00	1.33
			1.3	2.43	1.07	4.07	1.07
Low		2.1	3.04	1.22	4.11	1.72	
		2.2	2.10	1.43	4.29	1.21	
		2.3	2.57	1.43	4.11	1.29	

Table 4 continues.

Table 4 continued

Technique	Item	Elabora- tion	Judge group	Domains CC	CK	CF	CI	
CMP	CB10	High	1.1	42.86	14.29	14.29	14.29	
			1.2	57.14	35.71	14.29	35.71	
			1.3	64.29	14.29	0.00	28.57	
		Low	2.1	57.14	42.86	14.29	42.86	
			2.2	57.14	28.57	21.43	50.00	
			2.3	57.14	7.14	0.00	28.57	
		CJ12	High	1.1	57.14	21.43	64.29	35.71
				1.2	14.29	28.57	28.57	7.14
				1.3	21.43	28.57	64.29	21.43
	Low		2.1	28.57	21.43	57.14	14.29	
			2.2	14.29	42.86	71.43	35.71	
			2.3	35.71	14.29	50.00	14.29	
	CL13	High	1.1	0.00	35.71	14.29	28.57	
			1.2	7.14	28.57	7.14	35.71	
			1.3	0.00	28.57	7.14	35.71	
Low		2.1	7.14	42.86	14.29	42.86		
		2.2	7.14	50.00	28.57	50.00		
		2.3	0.00	50.00	7.14	35.71		

Table 5

Judge Group Ratings for Physical Education Fifth Domain Items 63

Technique	Item	Elabora- tion	Judge group	Domains RU	PC	BS	FN
IOCI	BM02	High	1.1	0.94	-0.35	-0.24	-0.35
			1.2	0.85	-0.23	-0.33	-0.28
			1.3	0.87	-0.27	-0.32	-0.27
		Low	2.1	0.83	-0.26	-0.17	-0.40
			2.2	0.89	-0.30	-0.30	-0.30
			2.3	0.88	-0.17	-0.36	-0.36
	BM11	High	1.1	0.96	-0.32	-0.32	-0.32
			1.2	0.85	-0.20	-0.30	-0.35
			1.3	1.00	-0.33	-0.33	-0.33
		Low	2.1	0.96	-0.32	-0.32	-0.32
			2.2	1.00	-0.33	-0.33	-0.33
			2.3	0.93	-0.31	-0.31	0.31
	PS01	High	1.1	0.51	-0.15	-0.25	-0.11
			1.2	0.25	0.01	-0.18	-0.08
			1.3	0.54	-0.23	-0.18	-0.13
		Low	2.1	0.69	-0.26	-0.26	-0.17
			2.2	0.55	-0.21	-0.21	-0.12
			2.3	0.43	-0.05	-0.19	-0.19
CARS	BM33	High	1.1	3.85	1.62	1.96	1.23
			1.2	4.46	1.18	1.50	1.33
			1.3	4.18	1.72	1.40	1.32
		Low	2.1	4.75	1.75	1.54	1.50
			2.2	4.61	1.57	1.97	1.33
			2.3	4.21	1.50	1.93	1.93
	PS02	High	1.1	2.57	1.36	1.36	2.14
			1.2	2.07	1.18	1.39	1.97
			1.3	2.54	1.50	1.57	2.86
		Low	2.1	2.40	1.47	1.14	1.78
			2.2	3.04	1.47	1.54	2.54
			2.3	2.83	1.28	1.18	3.22
	PS04	High	1.1	3.68	1.75	1.21	1.57
			1.2	3.50	1.07	1.14	1.43
			1.3	3.29	1.71	1.32	1.28
		Low	2.1	3.93	1.83	1.35	1.64
			2.2	4.11	1.61	1.32	1.50
			2.3	3.57	1.89	1.22	1.97

Table 5 continues.

Table 5 continued

Technique	Item	Elabora- tion	Judge group	Domains RU	PC	BS	FN	
CMP	BM38	High	1.1	92.86	21.43	21.43	7.14	
			1.2	92.86	7.14	42.86	0.00	
			1.3	100.00	0.00	21.43	0.00	
		Low	2.1	92.86	21.43	14.29	7.14	
			2.2	100.00	28.57	35.71	7.14	
			2.3	100.00	14.29	7.14	0.00	
		PS03	High	1.1	78.57	7.14	0.00	0.00
				1.2	85.71	0.00	7.14	0.00
				1.3	71.43	0.00	0.00	0.00
	Low		2.1	50.00	7.14	0.00	0.00	
			2.2	64.29	0.00	0.00	0.00	
			2.3	71.43	7.14	7.14	0.00	
	PS07	High	1.1	85.71	7.14	0.00	14.29	
			1.2	92.86	7.14	7.14	7.14	
			1.3	71.43	0.00	0.00	0.00	
Low		2.1	71.43	14.29	0.00	0.00		
		2.2	71.43	7.14	0.00	0.00		
		2.3	100.00	0.00	0.00	0.00		

Table 6

Variations of Ratings for Geography Fifth Domain Items by Judge 65
Group

Technique	Item	Elabora- tion	Judge group	Domains CC	CK	CF	CI
IOCI	CB08	High	1.1	0.2041	0.1224	0.0000	0.3112
			1.2	0.8827	0.0000	0.2653	0.2041
			1.3	0.6982	0.1302	0.1775	0.5444
		Low	2.1	0.5306	0.1224	0.3112	0.6786
			2.2	0.6982	0.3314	0.3314	0.1302
			2.3	0.6378	0.5306	0.5306	0.3878
	CJ01	High	1.1	0.6582	0.5714	0.8010	0.5306
			1.2	0.7143	0.3929	0.6735	0.3469
			1.3	0.2840	0.8402	0.3314	0.6982
		Low	2.1	0.7143	0.5306	0.6735	0.5306
			2.2	0.6735	0.5357	0.6582	0.8827
			2.3	0.6582	0.6378	0.3724	0.8010
	CL01	High	1.1	0.7806	0.5510	0.3469	0.7806
			1.2	0.7806	0.5714	0.3469	0.7755
			1.3	0.5325	0.3905	0.3905	0.5562
		Low	2.1	0.8010	0.6939	0.3112	0.6735
			2.2	0.5969	0.5510	0.5306	0.7398
			2.3	0.5714	0.4541	0.5306	0.7398
CARS	CB09	High	1.1	6.6378	7.7398	2.0000	7.1224
			1.2	10.1983	6.4132	3.8843	3.8843
			1.3	10.5306	5.8163	1.0867	7.4898
		Low	2.1	4.9439	7.4949	2.4898	8.9643
			2.2	4.3314	6.3314	2.8402	6.2485
			2.3	7.9053	3.4083	5.2071	5.4675
	CJ02	High	1.1	6.1429	7.3724	4.4082	4.0255
			1.2	4.8571	3.6327	3.4286	6.2653
			1.3	6.3520	7.6735	4.0612	6.8367
		Low	2.1	6.8639	5.6095	4.0828	5.5385
			2.2	4.3520	6.0255	3.9592	5.4082
			2.3	6.0000	4.8827	6.7755	8.3469
	CL09	High	1.1	7.9439	4.3469	7.0867	4.2296
			1.2	6.6378	3.1684	4.5714	1.3724
			1.3	4.5510	0.2653	5.4082	0.2653
		Low	2.1	4.3520	1.2449	5.4541	5.9592
			2.2	6.0255	3.2653	4.8163	1.2449
			2.3	6.6939	3.2653	5.1684	2.5306

Table 6 continues.

Table 6 continued

Technique	Item	Elabora- tion	Judge group	Domains CC	CK	CF	CI	
CMP	CB10	High	1.1	0.2449	0.1224	0.1224	0.1224	
			1.2	0.2449	0.2296	0.1224	0.2296	
			1.3	0.2296	0.1224	0.0000	0.2041	
		Low	2.1	0.2449	0.2449	0.1224	0.2449	
			2.2	0.2449	0.2041	0.1684	0.2500	
			2.3	0.2449	0.0663	0.0000	0.2041	
		CJ12	High	1.1	0.2449	0.1684	0.2296	0.2296
				1.2	0.1224	0.2041	0.2041	0.0663
				1.3	0.1684	0.2041	0.2296	0.1684
	Low		2.1	0.2041	0.1684	0.2449	0.1224	
			2.2	0.1224	0.2449	0.2041	0.2296	
			2.3	0.2296	0.1224	0.2500	0.1224	
	CL13	High	1.1	0.0000	0.2296	0.1224	0.2041	
			1.2	0.0663	0.2041	0.0663	0.2296	
			1.3	0.0000	0.2041	0.0663	0.2296	
		Low	2.1	0.0663	0.2449	0.1224	0.2449	
			2.2	0.0663	0.2500	0.2041	0.2500	
			2.3	0.0000	0.2500	0.0663	0.2296	

Table 7

Variances of Ratings for Physical Education Fifth Domain Items 67
by Judge Group

Technique	Item	Elabora- tion	Judge group	Domains RU	PC	BS	FN
IOCI	BM02	High	1.1	0.0710	0.0000	0.2840	0.0000
			1.2	0.2840	0.3314	0.0710	0.2840
			1.3	0.3112	0.0663	0.0000	0.0663
		Low	2.1	0.0000	0.5153	0.5357	0.2653
			2.2	0.2653	0.0663	0.0663	0.0663
			2.3	0.0000	0.5306	0.2653	0.2653
	BM11	High	1.1	0.0663	0.0000	0.0000	0.0000
			1.2	0.2653	0.4898	0.2653	0.0663
			1.3	0.0000	0.0000	0.0000	0.0000
		Low	2.1	0.0710	0.0000	0.0000	0.0000
			2.2	0.0000	0.0000	0.0000	0.0000
			2.3	0.2653	0.0000	0.0000	0.0000
	PS01	High	1.1	0.5510	0.1224	0.0000	0.3112
			1.2	0.7755	0.5153	0.0663	0.1684
			1.3	0.6939	0.0000	0.0663	0.1224
		Low	2.1	0.5306	0.0000	0.0000	0.2653
			2.2	0.7755	0.2653	0.2653	0.4898
			2.3	0.6378	0.1684	0.0000	0.0000
CARS	BM33	High	1.1	6.6746	5.2544	6.2249	0.7101
			1.2	1.0663	0.5153	2.7143	1.3724
			1.3	4.3724	4.9592	1.1684	1.6582
		Low	2.1	0.6786	4.8214	3.4949	3.1429
			2.2	0.8827	2.1224	4.4949	0.8010
			2.3	3.9592	2.0000	4.6939	1.9796
	PS02	High	1.1	6.6939	2.0612	1.4898	9.0612
			1.2	4.6939	0.5153	1.7398	7.2092
			1.3	5.9235	4.0000	3.1224	10.2041
		Low	2.1	5.0255	2.0663	0.6327	4.5306
			2.2	5.9235	3.0663	2.3520	8.0663
			2.3	6.3724	0.8163	0.5153	10.1020
	PS04	High	1.1	6.2296	5.5357	0.6735	3.8367
			1.2	5.5714	0.2653	0.3469	2.8367
			1.3	7.5306	3.8163	1.6582	1.3878
		Low	2.1	6.9796	6.9439	1.3469	2.6327
			2.2	2.1684	2.0255	0.8010	2.8571
			2.3	3.1224	4.8827	0.6735	4.7806

Table 7 continues.

Table 7 continued

Technique	Item	Elabora- tion	Judge group	Domains RU	PC	BS	FN
CMP	BM38	High	1.1	0.0663	0.1684	0.1684	0.0663
			1.2	0.0663	0.0663	0.2449	0.0000
			1.3	0.0000	0.0000	0.1684	0.0000
		Low	2.1	0.0663	0.1684	0.1224	0.0663
			2.2	0.0000	0.2041	0.2296	0.0663
			2.3	0.0000	0.1224	0.0663	0.0000
	PS03	High	1.1	0.1684	0.0663	0.0000	0.0000
			1.2	0.1224	0.0000	0.0663	0.0000
			1.3	0.2041	0.0000	0.0000	0.0000
		Low	2.1	0.2500	0.0663	0.0000	0.0000
			2.2	0.2296	0.0000	0.0000	0.0000
			2.3	0.2041	0.0663	0.0663	0.0000
	PS07	High	1.1	0.1224	0.0663	0.0000	0.1224
			1.2	0.0663	0.0663	0.0663	0.0663
			1.3	0.2041	0.0000	0.0000	0.0000
Low		2.1	0.2041	0.1224	0.0000	0.0000	
		2.2	0.2041	0.1224	0.0000	0.0000	
		2.3	0.0000	0.0000	0.0000	0.0000	

ratings than other judge groups. Higher variation was defined as a pattern for a judge group of variances on an item which were higher across the four domains. Consistently higher variation meant that this pattern occurred for the same judge group on at least two of the three fifth domain items rated using any one of the judgmental techniques. If a judge group with consistently higher variation was found, it would be necessary to locate the aberrant judge(s) in order to eliminate the ratings of that judge(s) from the analysis of the data. The search for a judge group with consistently higher variance than other groups was done for each judgmental technique for the three high elaboration groups and for the three low elaboration groups. An example of one item with the highest variances of the three elaboration judge groups across the four domains is item CB10 as rated by judge group 1.2 using the CMP. However, since judge group 1.2 did not have a second item with the highest variances of the three elaboration judge groups across the four domains, this group did not meet the criterion which would indicate a need to locate the aberrant judge(s). Similarly, the high variances of judge group 1.1 on item CL09 used with the CARS was not repeated with other items. Therefore, the criterion for searching for aberrant judges was not met for any geography judge group.

The same analysis was repeated with physical education fifth domain items. Judge group 1.2 using the IOCI with item BM11 had the highest variances of the three high elaboration judge groups

across the four domains. Judge group 2.2 using the IOCI had the highest variances of the three judge groups for item PS01 on the four domains. However, these patterns were not repeated for other items, so the criterion for searching for aberrant judges was not met for any physical education judge group.

Although some judge groups had higher variability than other groups on their ratings on an item across the four domains, no consistent pattern of high variability was found for any judge group on any judgmental technique.

The next step in establishing the validity of judgmental ratings was the examination of two characteristics of the judges. First the teaching experience of the judges was studied. Then the first language of the judges was studied.

It was felt that teaching experience might affect the judgments made regarding item-domain congruence because of experience in writing test items to measure a curriculum. Although the judges were university students, it was believed possible that they might also be experienced teachers. Therefore, the judges were queried regarding their teaching experience. The teaching experience reported by the judges is shown in Table 8. Only a small number of judges reported having some teaching experience. Among the 84 geography judges, 19 reported having some teaching experience. The teaching experience of the six judge groups was evenly distributed across the groups with the exception of judge group 1.3 in which no teaching experience was reported. The teaching experience of the judges

Table 8

Years of Teaching Experience of Judge Groups

		Any teaching experience (years)				Experience teaching geography or physical education (years)			
		Geography judges		Physical education judges		Geography judges		Physical education judges	
Elaboration	Judge groups of 14 judges	N	Range	N	Range	N	Range	N	Range
High	1.1	3	.5 - 5	4	1 - 5	1	5	3	1 - 3
	1.2	4	1 - 7	3	2 - 6	0		2	2 - 4
	1.3	0	0	2	1 - 11	0		2	1 - 11
Low	2.1	4	.5 - 15	1	12	1	3	1	3
	2.2	4	1 - 3	0	0	1	1	0	0
	2.3	4	1 - 5	5	3 - 7	1	1	4	3 - 5

directly related to geography was negligible, amounting to one or no persons per judge group. Among the 84 physical education judges, there were from zero to five judges in the groups who had teaching experience, most of it related to physical education. Discussion with the students and an instructor led to the conclusion that this teaching experience was mainly team supervision and did not include designing and testing curriculum. The conclusion was drawn that teaching experience was probably not an influence on the overall decisions of judge groups.

All the judges were in courses being instructed in English. However, it was felt that if there were a large number of judges for whom English was a second language in any judge group, the judgments of the match between items and domains might be affected. This did not appear to be the case. The first language of the judges is reported in Table 9. The first language of both the geography and physical education judges was predominantly English. The judges doing the questionnaires in a second language were reasonably dispersed across the judge groups. Therefore, it did not seem that the language of the judges in any judge group would affect one group differently from another.

Thus, it was determined by the examination of the judgmental ratings of all the items on all domains, a comparison of the judge group ratings and variances on the fifth domain items, and the background information provided by the judges that the judges were

Table 9

First Language of Judges by Groups

		Geography judges				Physical education judges			
Elaboration	Judge group	N	English	French	Other	N	English	French	Other
High	1.1	11	9	1	1	14	13	1	0
	1.2	14	11	1	2	14	13	0	1
	1.3	14	13	1	0	14	13	0	1
Low	2.1	14	14	0	0	14	14	0	0
	2.2	14	10	4	0	14	13	1	0
	2.3	14	13	1	0	14	12	2	0

able to use the techniques adequately and that there was no apparent bias caused by teaching experience and language of the judges.

The analysis of the data proceeded with an examination of the research hypothesis.

The Research Hypothesis

The research hypothesis was: When high domain elaboration is used, item-domain congruence ratings will be higher than when low domain elaboration is used.

The ratings of the judge groups for each item on each judgmental technique under high and low elaboration are shown in Tables 10 and 11.

In order to compare the effect of high and low elaboration on the three judgmental techniques, the mean, median, and standard deviation of the judge group ratings for items were calculated. These statistics were calculated for judge group ratings for items on their intended domains. The results of these analyses for the geography and physical education tests comprised of all intended domain items are shown in Table 12. The mean for the ratings of all items on their intended domains is higher for the IOCI with both geography and physical education tests when high elaboration was used. The mean was higher for the CARS when low elaboration was used for both geography and physical education tests. The mean for the CMP was higher with high elaboration for the geography test and with low elaboration for the physical education test. The differ-

Table 10

Judge Group Ratings for Intended Domain Geography Items on Each Judgmental Technique

IOCI

Item	High Elaboration					Low Elaboration				
	Judges	Domains				Judges	Domains			
	N	CC	CK	CF	CI	N	CC	CK	CF	CI
CC02	14	0.98	-0.36	-0.36	-0.26	14	0.74	-0.21	-0.17	-0.36
CC03	14	0.82	-0.37	-0.08	-0.37	14	0.58	-0.23	-0.08	-0.27
CC05	14	0.74	-0.31	-0.12	-0.31	14	0.75	-0.30	-0.15	-0.30
CC06	13	0.54	-0.44	0.38	-0.49	14	0.51	-0.44	0.32	-0.39
CC07	14	0.73	-0.56	0.20	-0.37	14	0.75	-0.49	0.13	-0.39
CC58	14	0.57	-0.48	0.29	-0.38	14	0.58	-0.37	0.15	-0.37
CC61	13	0.76	-0.42	0.09	-0.42	14	0.48	-0.38	0.38	-0.48
CC62	13	0.71	-0.42	0.14	-0.42	14	0.40	-0.36	0.40	-0.45
CC63	14	0.86	-0.43	-0.10	-0.33	14	0.68	-0.42	0.06	-0.32
CK03	14	-0.54	0.46	-0.39	0.46	13	-0.45	0.47	-0.60	0.58
CK14	14	-0.48	0.43	-0.43	0.48	13	-0.54	0.49	-0.54	0.59
CK15	13	-0.38	0.44	-0.54	0.49	14	-0.35	0.37	-0.44	0.42
CK16	14	-0.24	0.48	-0.52	0.29	14	-0.49	0.46	-0.44	0.46
CK17	14	-0.40	0.45	-0.36	0.31	13	-0.55	0.42	-0.35	0.47
CK18	13	0.09	0.29	-0.27	-0.12	14	-0.02	0.07	0.17	-0.21
CK19	14	-0.05	0.57	-0.57	0.05	14	-0.25	0.42	-0.30	0.13
CK20	14	-0.26	0.21	0.21	-0.17	14	-0.43	0.38	0.29	-0.24
CK21	13	-0.29	0.73	-0.29	-0.14	14	-0.54	0.46	-0.11	0.18
CF02	14	0.43	-0.57	0.71	-0.57	13	0.28	-0.44	0.59	-0.44
CF03	14	0.10	-0.24	0.71	-0.57	14	0.21	-0.36	0.55	-0.40
CF04	13	0.42	-0.40	0.47	-0.50	13	0.32	-0.55	0.58	-0.35
CF05	14	0.17	-0.26	0.69	-0.60	13	-0.18	-0.13	0.69	-0.38
CF06	14	-0.08	-0.32	0.77	-0.37	14	-0.11	-0.44	0.85	-0.30
CF13	13	0.15	-0.41	0.72	-0.46	14	0.10	-0.24	0.57	-0.43
CF17	14	0.38	-0.48	0.43	-0.33	14	0.24	-0.24	0.33	-0.33
CF18	13	0.36	-0.41	0.46	-0.41	13	0.26	-0.26	0.21	-0.21
CF19	14	0.20	-0.42	0.44	-0.23	14	0.35	-0.37	0.20	-0.18
CI01	14	-0.38	0.38	-0.52	0.52	14	-0.44	0.23	-0.39	0.61
CI02	13	-0.46	0.21	-0.46	0.72	14	-0.35	0.32	-0.49	0.51
CI03	14	-0.43	0.29	-0.48	0.62	14	-0.37	0.25	-0.46	0.58
CI05	14	-0.24	0.10	-0.52	0.67	14	-0.36	0.36	-0.60	0.60
CI07	13	-0.26	0.26	-0.51	0.51	14	-0.30	0.42	-0.44	0.32
CI08	14	-0.46	0.49	-0.56	0.54	14	-0.57	0.52	-0.48	0.52
CI10	13	-0.46	0.51	-0.62	0.56	14	-0.29	0.19	-0.33	0.43
CI11	13	-0.51	0.46	-0.36	0.41	14	-0.58	0.46	-0.30	0.42
CI12	14	-0.18	0.20	-0.37	0.35	14	-0.38	0.29	-0.33	0.43

Table 10 continues.

Table 10 continued

CARS

Item	High Elaboration					Low Elaboration				
	Judges		Domains			Judges		Domains		
	N	CC	CK	CF	CI	N	CC	CK	CF	CI
CC02	14	4.11	1.54	1.65	1.43	14	3.93	1.71	1.72	1.86
CC03	14	4.54	1.43	2.04	1.54	14	4.58	1.35	2.04	1.29
CC05	14	4.83	1.57	2.07	1.72	14	4.61	1.64	1.60	1.14
CC06	14	4.36	1.40	3.25	1.33	14	4.64	1.43	3.29	1.71
CC07	14	4.61	1.54	2.04	1.68	14	4.18	1.54	3.22	1.57
CC58	14	4.33	1.50	2.82	1.97	14	4.29	1.78	2.46	1.64
CC61	14	4.39	1.33	2.90	1.10	14	4.71	1.21	2.90	1.43
CC62	14	4.00	1.29	3.04	1.29	14	4.61	1.14	2.93	1.36
CC63	14	4.32	1.79	1.75	1.43	14	4.21	1.22	3.57	1.75
CK03	14	1.57	3.72	1.43	4.54	14	1.83	4.57	1.82	4.43
CK14	14	2.00	4.86	1.72	4.04	14	2.14	4.68	1.83	4.64
CK15	14	2.10	4.86	1.54	4.86	14	1.75	4.54	1.78	4.86
CK16	14	1.75	3.75	2.04	3.93	14	1.97	3.90	2.35	3.93
CK17	14	2.68	4.80	2.07	4.29	14	2.47	4.50	2.18	4.25
CK18	14	2.57	3.58	2.65	2.79	14	3.14	3.68	2.43	3.29
CK19	14	1.83	4.50	1.39	2.89	14	2.50	4.29	1.85	3.10
CK20	14	2.54	2.64	3.14	2.65	14	2.07	3.40	3.36	2.68
CK21	14	1.40	3.85	2.15	3.60	14	1.29	3.75	2.33	2.68
CF02	14	3.61	1.82	4.04	1.93	14	3.86	2.25	4.42	2.40
CF03	14	3.29	2.00	3.97	1.68	14	3.14	1.85	4.17	1.54
CF04	14	4.07	1.64	4.50	1.72	14	3.79	1.57	4.36	1.57
CF05	14	2.29	2.29	4.04	1.79	14	2.46	2.54	3.93	1.75
CF06	14	2.57	2.07	4.07	1.71	14	2.18	2.25	4.47	2.57
CF13	14	3.79	2.04	4.33	1.85	14	3.64	1.97	4.67	1.64
CF17	14	3.46	1.15	3.32	1.68	14	3.71	1.36	3.00	1.65
CF18	14	3.75	1.83	3.14	2.04	14	3.35	1.64	3.14	1.43
CF19	14	3.50	2.22	2.83	2.43	14	2.93	1.64	3.90	2.00
CI01	14	1.65	3.90	1.79	4.68	14	1.60	3.08	2.14	4.07
CI02	14	1.90	4.36	1.57	4.33	14	1.29	3.47	1.75	4.57
CI03	14	1.36	2.97	1.43	3.54	14	2.04	3.32	1.47	3.71
CI05	14	1.50	2.82	1.47	4.04	14	1.71	2.93	1.64	3.86
CI07	14	2.07	4.50	1.64	4.75	14	1.50	4.00	1.50	4.29
CI08	14	1.57	3.50	1.57	4.57	13	1.79	4.32	1.43	4.25
CI10	14	2.04	4.29	1.43	4.32	14	1.86	4.40	2.00	4.57
CI11	14	1.36	4.15	2.29	4.40	14	1.36	3.86	2.71	3.97
CI12	14	1.83	4.00	1.40	3.68	14	1.47	3.86	1.64	3.90

Table 10 continues.

Table 10 continued

CMP										
High Elaboration						Low Elaboration				
Item	Judges Domains					Judges Domains				
	N	CC	CK	CF	CI	N	CC	CK	CF	CI
CC02	14	92.86	7.14	7.14	14.29	14	100.00	7.14	7.14	7.14
CC03	14	92.86	0.00	0.00	28.57	14	92.86	14.29	28.57	7.14
CC05	14	92.86	0.00	14.29	14.29	14	92.86	21.43	28.57	0.00
CC06	14	100.00	0.00	42.86	0.00	14	100.00	7.14	57.14	7.14
CC07	14	85.71	0.00	42.86	14.29	14	92.86	7.14	35.71	7.14
CC58	14	85.71	7.14	42.86	7.14	14	92.86	14.29	64.29	14.29
CC61	14	100.00	0.00	35.71	0.00	14	85.71	7.14	42.86	7.14
CC62	14	92.86	0.00	42.86	0.00	14	100.00	7.14	50.00	7.14
CC63	14	78.57	0.00	14.29	0.00	14	100.00	0.00	28.57	7.14
CK03	14	21.43	85.71	21.43	85.71	14	7.14	85.71	28.57	92.86
CK14	14	0.00	85.71	7.14	92.86	14	0.00	71.43	14.29	85.71
CK15	14	0.00	92.86	0.00	100.00	14	0.00	100.00	14.29	100.00
CK16	14	7.14	100.00	42.86	92.86	14	7.14	85.71	28.57	85.71
CK17	14	7.14	92.86	14.29	71.43	14	0.00	64.29	28.57	85.71
CK18	14	64.29	64.29	42.86	50.00	13	61.54	61.54	69.23	53.85
CK19	14	21.43	92.86	21.43	71.43	14	14.29	85.71	14.29	50.00
CK20	14	14.29	64.29	35.71	21.43	14	7.14	57.14	50.00	42.86
CK21	14	0.00	57.14	21.43	42.86	14	7.14	85.71	7.14	64.29
CF02	14	85.71	0.00	85.71	21.43	14	71.43	7.14	92.86	14.29
CF03	14	42.86	14.29	78.57	14.29	14	21.43	21.43	71.43	14.29
CF04	14	85.71	0.00	92.86	0.00	14	92.86	0.00	92.86	7.14
CF05	14	64.29	28.57	85.71	21.43	14	42.86	14.29	85.71	7.14
CF06	14	21.43	21.43	92.86	7.14	14	7.14	7.14	85.71	14.29
CF13	14	78.57	0.00	92.86	7.14	14	71.43	14.29	78.57	7.14
CF17	14	64.29	0.00	50.00	0.00	13	69.23	7.69	38.46	7.69
CF18	14	85.71	0.00	64.29	7.14	14	71.43	21.43	57.14	14.29
CF19	14	71.43	14.29	78.57	14.29	14	64.29	35.71	50.00	21.43
CI01	14	0.00	35.71	21.43	85.71	14	7.14	64.29	21.43	85.71
CI02	14	0.00	85.71	0.00	100.00	14	0.00	64.29	14.29	100.00
CI03	14	7.14	57.14	14.29	92.86	14	0.00	42.86	7.14	100.00
CI05	14	7.14	50.00	7.14	85.71	14	7.14	57.14	14.29	92.86
CI07	14	0.00	78.57	0.00	100.00	14	7.14	71.43	14.29	92.86
CI08	14	14.29	92.86	28.57	85.71	14	7.14	85.71	7.14	92.86
CI10	14	7.14	71.43	0.00	100.00	14	7.14	71.43	21.43	78.57
CI11	14	0.00	64.29	0.00	92.86	14	0.00	85.71	21.43	78.57
CI12	14	7.14	57.14	7.14	64.29	14	7.14	64.29	21.43	78.57

Table 11

Judge Group Ratings for Intended Domain Physical Education Items
on Each Judgmental Technique

IOCI

Item	High Elaboration					Low Elaboration				
	Judges N	Domains				Judges N	Domains			
		RU	PC	BS	FN		RU	PC	BS	FN
RU01	14	0.89	-0.35	-0.20	-0.35	14	0.77	-0.27	-0.13	-0.37
RU08	14	0.95	-0.33	-0.33	-0.29	14	0.98	-0.26	-0.36	-0.36
RU14	14	0.82	-0.37	-0.18	-0.27	14	0.89	-0.35	-0.20	-0.35
RU17	14	0.87	-0.27	-0.27	-0.32	14	0.94	-0.30	-0.35	-0.30
RU21	13	0.79	-0.18	-0.33	-0.28	14	0.92	-0.27	-0.37	-0.27
RU25	14	0.94	-0.20	-0.39	-0.35	14	0.98	-0.31	-0.36	-0.31
RU30	14	1.00	-0.33	-0.33	-0.33	14	1.00	-0.33	-0.33	-0.33
RU35	14	0.89	-0.44	-0.01	-0.44	14	0.94	-0.25	-0.30	-0.39
PC03	14	-0.56	0.63	-0.27	0.20	14	-0.67	0.52	0.00	0.14
PC06	14	-0.25	0.04	0.27	-0.06	14	-0.38	0.19	0.29	-0.10
PC18	14	-0.48	0.33	-0.05	0.19	14	-0.46	0.30	-0.08	0.25
PC22	14	-0.40	0.12	-0.07	0.36	14	-0.43	0.43	-0.29	0.29
PC23	14	-0.25	-0.01	0.18	0.08	14	-0.37	0.01	0.20	0.15
PC29	14	-0.45	0.21	-0.12	0.36	14	-0.57	0.52	-0.05	0.10
PC32	14	-0.39	0.37	-0.15	0.18	14	-0.33	0.33	-0.14	0.14
PC36	14	-0.46	0.54	-0.18	0.11	14	-0.56	0.30	0.15	0.11
BS04	14	-0.49	0.46	0.42	-0.39	14	-0.60	0.31	0.45	-0.17
BS09	14	-0.69	0.36	0.40	-0.07	14	-0.73	0.27	0.46	-0.01
BS15	14	-0.33	0.14	0.62	-0.43	14	-0.14	0.14	0.43	-0.43
BS20	14	-0.38	0.00	0.33	0.05	14	-0.33	0.10	0.24	0.00
BS24	14	-0.65	0.35	0.15	0.15	14	-0.74	0.31	0.26	0.17
BS28	14	-0.52	0.38	0.38	-0.24	14	-0.54	0.27	0.61	-0.35
BS34	14	-0.57	0.29	0.62	-0.33	14	-0.43	0.10	0.71	-0.38
BS39	14	-0.54	0.13	0.46	-0.06	14	-0.68	0.13	0.46	0.08
FN05	14	-0.76	0.33	0.00	0.43	14	-0.63	0.18	0.13	0.32
FN10	14	-0.38	0.19	0.10	0.10	14	-0.12	0.45	-0.26	-0.07
FN12	14	0.31	-0.07	-0.12	-0.12	14	-0.05	0.10	0.19	-0.24
FN13	14	-0.17	-0.02	0.12	0.07	14	-0.26	0.07	0.07	0.12
FN16	14	-0.60	0.26	0.07	0.26	14	-0.64	0.26	-0.07	0.45
FN19	13	-0.50	0.22	-0.04	0.32	14	-0.36	0.21	-0.07	0.21
FN26	14	-0.70	0.44	0.06	0.20	14	-0.21	0.17	0.26	-0.21
FN31	14	-0.57	0.33	0.29	-0.05	14	-0.30	0.27	0.18	-0.15
FN37	14	-0.24	0.10	-0.24	0.38	14	-0.39	0.18	0.23	-0.01

Table 11 continues.

Table 11 continued

CARS

Item	High Elaboration					Low Elaboration				
	Judges		Domains			Judges		Domains		
	N	RU	PC	BS	FN	N	RU	PC	BS	FN
RU01	14	4.86	2.00	2.10	2.18	14	4.40	2.43	2.25	2.14
RU08	14	4.79	1.21	1.07	1.22	14	4.96	1.32	1.33	1.33
RU14	14	4.93	2.07	1.75	1.79	14	4.79	2.14	2.22	1.75
RU17	14	4.83	1.07	1.29	1.22	14	4.86	1.36	1.53	1.50
RU21	14	4.54	1.07	1.25	1.18	14	4.72	1.43	1.18	1.32
RU25	14	3.97	1.40	1.29	1.07	14	4.68	1.18	1.07	1.07
RU30	14	4.43	1.36	1.29	1.29	14	4.83	1.18	1.15	1.10
RU35	14	4.83	1.22	1.86	1.32	14	4.47	1.60	1.71	1.36
PC03	14	2.47	3.29	2.18	4.00	14	1.86	3.64	2.64	3.72
PC06	14	2.57	3.18	3.64	3.10	14	2.60	2.97	4.00	2.86
PC18	14	1.57	3.22	2.25	3.54	14	1.64	3.79	2.54	3.79
PC22	14	2.00	4.50	3.40	3.86	14	2.43	4.40	3.79	2.85
PC23	14	2.29	4.00	3.46	3.72	14	2.65	4.07	3.75	3.00
PC29	14	1.14	3.58	2.96	3.54	14	1.36	3.90	3.32	3.43
PC32	14	2.36	4.07	3.22	3.18	14	2.39	4.32	2.97	4.04
PC36	14	1.64	4.25	3.57	2.86	14	2.46	4.39	4.11	4.36
BS04	14	1.64	4.21	4.54	3.07	14	1.50	3.82	4.36	2.40
BS09	14	1.68	3.64	4.00	2.90	14	2.22	3.60	4.43	3.18
BS15	14	3.22	3.40	4.54	2.54	14	3.25	3.82	4.75	3.00
BS20	14	1.57	3.43	3.57	2.21	14	1.86	3.39	3.93	2.54
BS24	14	1.29	4.33	4.36	3.35	14	1.47	4.21	4.29	3.14
BS28	14	1.57	4.11	4.75	2.83	14	1.53	3.57	4.67	2.57
BS34	14	1.54	3.04	4.21	2.15	14	1.58	3.10	4.75	2.10
BS39	14	1.47	3.18	4.46	2.50	14	1.47	3.47	4.47	2.93
FN05	14	1.43	4.07	3.40	4.32	14	1.72	3.29	3.18	4.72
FN10	14	2.10	4.36	2.79	3.40	14	2.61	3.93	3.33	3.25
FN12	14	3.86	2.50	2.86	3.57	14	3.92	2.93	3.60	2.93
FN13	14	2.90	4.32	3.64	4.29	14	3.15	4.11	4.11	3.25
FN16	14	1.54	3.90	3.71	4.43	14	1.68	4.21	3.93	4.68
FN19	14	2.25	4.46	3.60	4.40	14	2.68	3.97	3.82	3.54
FN26	14	2.71	4.17	4.14	3.60	14	2.97	4.68	4.42	3.83
FN31	14	2.40	3.00	3.54	3.10	14	2.25	2.97	3.29	3.29
FN37	14	1.85	3.82	3.15	3.79	14	2.36	3.22	2.64	3.82

Table 11 continues.

Table 11 continued

CMP

Item	High Elaboration					Low Elaboration				
	Judges Domains					Judges Domains				
	N	RU	PC	BS	FN	N	RU	PC	BS	FN
RU01	14	100.00	7.14	28.57	14.29	14	100.00	21.43	28.57	14.29
RU08	14	92.86	0.00	14.29	7.14	14	100.00	0.00	0.00	0.00
RU14	14	92.86	7.14	28.57	0.00	14	92.86	21.43	42.86	7.14
RU17	14	100.00	0.00	14.29	7.14	14	100.00	0.00	0.00	0.00
RU21	14	100.00	0.00	0.00	0.00	14	92.86	7.14	0.00	0.00
RU25	14	100.00	0.00	0.00	0.00	14	85.71	7.14	0.00	0.00
RU30	14	100.00	0.00	0.00	0.00	14	100.00	0.00	0.00	0.00
RU35	14	100.00	7.14	21.43	0.00	14	100.00	7.14	21.43	0.00
PC03	14	14.29	85.71	21.43	71.43	14	7.14	71.43	42.86	71.43
PC06	14	21.43	64.29	71.43	71.43	14	35.71	64.29	92.86	78.57
PC18	14	14.29	28.57	28.57	50.00	14	7.14	71.43	14.29	64.29
PC22	14	42.86	100.00	57.14	71.43	14	14.29	92.86	64.29	78.57
PC23	14	28.57	71.43	64.29	85.71	14	28.57	57.14	78.57	57.14
PC29	13	0.00	61.54	38.46	61.54	14	7.14	78.57	50.00	92.86
PC32	14	35.71	92.86	28.57	50.00	14	14.29	100.00	28.57	57.14
PC36	14	7.14	100.00	57.14	35.71	14	7.14	85.71	78.57	78.57
BS04	14	0.00	64.29	71.43	7.14	14	7.14	78.57	85.71	21.43
BS09	14	7.14	71.43	92.86	42.86	14	7.14	85.71	85.71	35.71
BS15	14	35.71	78.57	78.57	7.14	14	57.14	50.00	85.71	21.43
BS20	14	7.14	50.00	57.14	21.43	14	21.43	35.71	42.86	28.57
BS24	14	7.14	92.86	50.00	50.00	14	7.14	85.71	71.43	57.14
BS28	14	7.14	57.14	92.86	14.29	14	7.14	35.71	85.71	21.43
BS34	14	0.00	64.29	85.71	21.43	14	0.00	28.57	100.00	14.29
BS39	14	0.00	64.29	92.86	35.71	14	7.14	50.00	100.00	42.86
FN05	14	21.43	78.57	71.43	85.71	14	7.14	85.71	50.00	92.86
FN10	14	21.43	85.71	28.57	42.86	14	35.71	71.43	42.86	50.00
FN12	14	71.43	28.57	50.00	42.86	14	100.00	42.86	78.57	64.29
FN13	14	57.14	85.71	64.29	85.71	14	71.43	57.14	78.57	78.57
FN16	14	14.29	92.86	42.86	78.57	14	7.14	78.57	71.43	92.86
FN19	14	21.43	92.86	57.14	100.00	14	21.43	71.43	64.29	78.57
FN26	14	21.43	100.00	71.43	71.43	14	14.29	100.00	64.29	57.14
FN31	14	7.14	57.14	57.14	50.00	14	7.14	50.00	64.29	35.71
FN37	14	14.29	85.71	42.86	71.43	14	14.29	78.57	50.00	71.43

Table 12

Mean, Median, and Standard Deviation for the Total Tests Using Ratings on Intended Domains

Geography				
Judgmental Technique	Elaboration	Mean	Median	SD
IOCI	High	.5853	.5650	.1701
IOCI	Low	.4997	.5000	.1632
CARS	High	4.1297	4.1712	.5576
CARS	Low	4.1753	4.2119	.4367
CMP	High	85.7100	86.8300	13.0000
CMP	Low	83.5200	84.9400	15.5900

Physical Education				
Judgmental Technique	Elaboration	Mean	Median	SD
IOCI	High	.4532	.3800	.3237
IOCI	Low	.4261	.4300	.3659
CARS	High	4.1403	4.2900	.5449
CARS	Low	4.1870	4.3900	.5890
CMP	High	80.0000	85.7100	20.4100
CMP	Low	80.9500	85.7100	17.9200

ence between the means for each technique using high and low elaboration was significant ($t(70)=2.1836$, $p <.05$) only for the IOCI used with the geography items. No consistent differences between the mean, median, and standard deviation for the techniques used with high and with low domain elaboration were found in the two curriculum areas.

Correlations between each judgmental technique used with high elaboration and the same judgmental technique used with low elaboration were also studied in order to determine the effect of elaboration on the judgmental techniques. These correlations for geography and physical education judge group item ratings on intended and nonintended domains and on intended domains only are shown in Table 13. In each curriculum area, when ratings on intended and nonintended domains were used, there was a strong correlation between each technique used with high elaboration and the same technique used with low elaboration. There is consistency in the two curriculum areas in the findings of strong correlations between each technique used with high and with low elaboration. This is further evidence that there is little difference between the ratings given using high and low elaboration. However, when ratings on intended domains only were used, the correlations between each technique used with high elaboration and the same technique used with low domain elaboration were moderate to strong. These correlations were not as strong as when the ratings on both the intended and nonintended domains were used. A possible reason for the lower correlations

Table 13

Correlations among Judgmental Techniques with High and
with Low Domain Elaboration

Using Judge Group Ratings on Intended and Nonintended Domains

Geography

Technique/Elaboration	IOCI/High	IOCI/Low	CARS/High	CARS/Low	CMP/High
IOCI/Low	.946				
CARS/High	.912	.913			
CARS/Low	.919	.911	.940		
CMP/High	.917	.906	.926	.938	
CMP/Low	.925	.929	.929	.928	.942

Physical Education

Technique/Elaboration	IOCI/High	IOCI/Low	CARS/High	CARS/Low	CMP/High
IOCI/Low	.925				
CARS/High	.829	.837			
CARS/Low	.823	.835	.942		
CMP/High	.855	.846	.927	.893	
CMP/Low	.849	.843	.918	.920	.914

Using Judge Group Ratings on Intended Domains

Geography

Technique/Elaboration	IOCI/High	IOCI/Low	CARS/High	CARS/Low	CMP/High
IOCI/Low	.706				
CARS/High	.333	.423			
CARS/Low	.362	.365	.718		
CMP/High	.353	.376	.621	.753	
CMP/Low	.582	.591	.571	.566	.642

Physical Education

Technique/Elaboration	IOCI/High	IOCI/Low	CARS/High	CARS/Low	CMP/High
IOCI/Low	.890				
CARS/High	.590	.651			
CARS/Low	.736	.775	.762		
CMP/High	.661	.664	.717	.631	
CMP/Low	.688	.768	.785	.754	.749

might be a restriction of range because ratings might not vary greatly when items were considered only on their intended domains.

Thus, based on the comparison of the mean, median, and standard deviation of judge group ratings of items, domain elaboration does not appear to have an effect on the ratings given by judges on the three judgmental techniques. The finding of a significantly higher mean for the IOCI used with geography items was not also found with the physical education items. The correlations between judge group ratings using each judgmental technique with high and with low elaboration suggested that for each technique, the judge group ratings when high elaboration was used were similar to the judge group ratings when low elaboration was used.

The analysis of the data proceeded with the investigation of the first research question.

The First Research Question

The first research question was: What is the effectiveness of each of the judgmental techniques?

Confirmatory factor analyses of the geography and physical education examinee response data were planned in order to investigate and confirm the domain structure suggested by the judge group ratings. In preparation for factor analyses of the geography and physical education examinee response data, tetrachoric correlations were calculated among the geography items and among the physical education items using the response data. These correlations are

presented in Tables 14 and 15. The correlations among the geography items and among the physical education items were so low that it was not considered worthwhile to do factor analyses with the data, for, as Tabachnick & Fidell (1983) pointed out, if the variables are uncorrelated with one another there is probably nothing to factor analyze.

The low correlations between items are further confirmed by the item-total domain correlations presented in Tables 16 and 17. Therefore, the difficulties of the items, also presented in Tables 16 and 17, were studied for an explanation of the low interitem correlations.

On the whole, the items comprising the geography and physical education domains were not answered consistently by the students. The difficulties for the geography items were all below .65. The geography items with four alternatives would have a chance probability of .25. The difficulties for the individual physical education items were somewhat higher than for the geography items, ranging from .06 to .82. The chance probability for the physical education items is more difficult to establish because there were four choices and a choice of "I don't know." In Table 18, the difficulties and standard deviations of the total test and intended domains are given. The statistics for the two curriculum areas were similar. The difficulty level of the total test for geography was .43 and for physical education, .48. However, physical education domains RU and PC were easier for the students than other domains. The standard deviations were similar for

Table 14

Correlation Matrix for Geography Intended Domain Items

	CC02	CC03	CC05	CC06	CC07	CC58	CC61	CC62	CC63
CC02	1.000								
CC03	0.371	1.000							
CC05	0.067	0.186	1.000						
CC06	0.293	0.227	0.120	1.000					
CC07	0.214	0.086	0.168	0.250	1.000				
CC58	0.207	0.332	0.261	0.345	0.236	1.000			
CC61	0.017	0.091	0.203	0.008	0.033	0.211	1.000		
CC62	0.310	-0.067	-0.005	0.162	0.132	0.201	0.153	1.000	
CC63	0.252	0.040	0.126	0.164	0.138	0.217	0.047	0.055	1.000
CK03	0.242	0.282	0.296	0.165	0.458	0.325	0.147	0.108	0.085
CK14	-0.188	-0.024	0.125	0.011	0.039	-0.075	0.062	-0.156	-0.205
CK15	0.061	0.072	0.028	0.048	0.141	0.042	-0.123	-0.001	0.029
CK16	0.179	0.213	0.186	0.136	0.179	0.195	0.260	0.044	0.208
CK17	0.110	0.189	0.100	0.093	0.125	-0.059	0.005	-0.057	0.030
CK18	0.243	0.227	0.115	0.129	0.175	0.264	0.142	0.094	0.202
CK19	0.358	0.126	0.115	0.055	0.157	0.126	0.183	0.067	0.208
CK20	-0.049	0.061	0.056	-0.078	0.009	0.059	0.119	0.057	0.078
CK21	0.039	0.131	0.042	0.112	0.022	0.122	0.085	-0.082	0.105
CF02	0.166	0.090	0.167	0.187	0.210	0.333	0.038	0.111	0.142
CF03	0.146	-0.076	0.061	0.123	0.088	0.207	0.137	-0.053	0.061
CF04	0.058	0.032	-0.023	0.198	0.211	0.269	-0.145	0.002	0.026
CF05	0.147	0.306	0.176	0.052	0.067	0.175	0.009	0.073	0.044
CF06	0.181	0.025	0.029	0.184	0.107	0.027	-0.045	0.066	-0.056
CF13	0.220	0.140	-0.030	0.065	0.152	0.109	-0.004	-0.013	0.115
CF17	0.143	0.110	0.298	0.059	0.113	0.205	0.119	0.098	0.026
CF18	0.178	0.229	0.124	0.201	0.201	0.340	0.090	0.158	0.263
CF19	0.308	0.239	0.258	0.179	0.212	0.316	0.082	0.278	0.195
CI01	0.217	0.179	-0.024	0.048	0.141	0.181	0.074	-0.094	-0.064
CI02	0.030	0.125	0.159	-0.012	0.186	0.096	0.101	-0.030	0.039
CI03	0.005	-0.004	0.218	0.004	0.169	0.139	0.114	0.014	0.039
CI05	0.276	0.107	0.167	0.157	0.106	0.126	0.134	-0.040	0.182
CI07	0.007	0.157	-0.019	-0.037	0.060	0.008	0.180	0.033	0.035
CI08	0.061	0.179	0.063	0.118	0.141	0.199	0.092	0.111	0.197
CI10	-0.020	0.182	0.019	0.012	0.218	0.266	0.006	0.139	0.014
CI11	0.038	0.071	0.045	0.153	0.109	0.161	-0.009	0.088	0.258
CI12	0.127	0.227	0.097	0.121	0.281	0.121	0.215	0.214	0.162

Table 14 continues.

Table 14 continued

	CK03	CK14	CK15	CK16	CK17	CK18	CK19	CK20	CK21
CK03	1.000								
CK14	-0.001	1.000							
CK15	0.134	0.075	1.000						
CK16	0.259	-0.054	0.243	1.000					
CK17	-0.033	-0.091	0.155	0.240	1.000				
CK18	0.207	-0.130	0.112	0.069	0.126	1.000			
CK19	0.140	0.050	0.320	0.294	0.120	0.241	1.000		
CK20	0.106	-0.064	-0.109	0.032	-0.064	0.167	0.018	1.000	
CK21	0.235	-0.234	0.011	-0.039	0.053	0.181	0.035	0.037	1.000
CF02	0.204	0.075	0.137	0.191	0.193	0.165	0.149	0.047	0.087
CF03	0.273	-0.073	0.096	0.259	0.070	0.112	0.143	-0.034	0.052
CF04	0.047	-0.053	0.002	0.055	-0.089	0.157	0.052	-0.095	0.123
CF05	-0.045	0.060	0.059	0.096	0.178	0.089	0.198	-0.015	0.042
CF06	0.021	-0.192	-0.001	0.061	0.033	0.082	0.171	0.157	-0.060
CF13	0.253	0.043	0.101	0.128	0.114	0.172	0.014	0.011	0.048
CF17	0.099	-0.004	0.112	0.220	0.045	-0.005	0.134	0.054	0.111
CF18	0.239	-0.008	0.159	0.086	-0.036	0.200	0.221	0.235	0.020
CF19	0.223	0.056	0.141	0.280	0.209	0.204	0.229	0.064	-0.020
CI01	0.249	-0.055	0.049	-0.040	0.058	0.182	0.079	-0.057	0.112
CI02	0.176	-0.033	0.351	0.157	0.178	0.092	0.298	-0.006	-0.075
CI03	0.203	-0.054	0.045	0.098	0.121	0.139	0.066	0.162	0.175
CI05	0.366	0.092	0.069	0.208	-0.015	0.138	0.275	0.067	-0.116
CI07	0.094	0.037	-0.006	0.050	0.059	-0.047	-0.148	-0.071	-0.075
CI08	0.204	-0.037	0.240	0.155	0.193	0.217	0.219	0.073	0.162
CI10	0.102	0.014	0.240	0.217	-0.026	0.190	0.038	0.061	-0.027
CI11	0.151	0.084	0.157	0.195	0.007	0.234	0.280	-0.031	0.054
CI12	0.270	0.066	0.143	0.235	-0.017	0.172	0.125	0.149	0.241

Table 14 continues.

Table 14 continued

	CF02	CF03	CF04	CF05	CF06	CF13	CF17	CF18	CF19
CF02	1.000								
CF03	0.043	1.000							
CF04	0.106	0.070	1.000						
CF05	0.078	-0.089	0.009	1.000					
CF06	0.087	0.145	0.072	0.014	1.000				
CF13	0.066	0.248	0.083	0.132	-0.039	1.000			
CF17	0.238	0.216	0.113	0.326	0.103	-0.033	1.000		
CF18	0.159	0.130	0.079	0.141	0.093	0.075	0.118	1.000	
CF19	0.195	0.003	0.104	0.133	0.114	0.196	0.162	0.208	1.000
CI01	-0.021	-0.010	-0.068	0.136	-0.001	0.066	-0.124	0.054	0.069
CI02	0.162	0.183	-0.096	0.095	0.191	0.030	0.216	0.163	0.097
CI03	-0.090	0.108	-0.073	0.249	0.027	0.119	0.098	0.132	0.081
CI05	0.142	0.206	0.124	0.021	0.223	0.140	0.111	0.094	0.160
CI07	-0.042	0.001	0.095	0.022	0.000	0.183	-0.170	0.064	0.112
CI08	0.223	0.061	-0.103	0.136	0.034	0.083	0.112	0.245	0.318
CI10	0.026	0.061	0.197	0.082	0.025	0.062	0.268	0.021	0.041
CI11	0.068	0.170	0.074	0.123	-0.009	0.128	0.162	0.160	0.118
CI12	0.053	0.072	0.140	0.028	0.114	0.123	0.171	0.139	0.325

	CI01	CI02	CI03	CI05	CI07	CI08	CI10	CI11	CI12
CI01	1.000								
CI02	0.020	1.000							
CI03	0.259	0.056	1.000						
CI05	0.069	0.187	0.231	1.000					
CI07	-0.024	-0.037	-0.169	-0.157	1.000				
CI08	0.067	0.401	0.153	0.124	0.172	1.000			
CI10	0.004	0.033	0.064	-0.092	0.045	0.026	1.000		
CI11	0.014	0.119	0.117	0.113	0.053	0.157	0.116	1.000	
CI12	0.017	0.183	0.116	0.150	0.099	0.196	0.215	0.093	1.000

Table 15

Correlation Matrix for Physical Education Intended Domain Items

	RU01	PC03	BS04	FN05	PC06	RU08	BS09	FN10
RU01	1.000							
PC03	0.007	1.000						
BS04	0.055	0.104	1.000					
FN05	-0.005	0.106	0.211	1.000				
PC06	-0.038	0.176	0.081	0.139	1.000			
RU08	0.001	0.163	0.100	0.075	0.028	1.000		
BS09	0.060	0.119	0.101	0.055	0.068	0.053	1.000	
FN10	0.088	0.049	0.122	0.177	0.030	-0.005	0.070	1.000
FN12	-0.021	0.083	-0.016	0.035	0.048	0.009	0.012	0.042
FN13	0.029	0.131	0.079	0.128	0.070	0.145	0.058	0.200
RU14	-0.060	0.130	-0.015	0.136	-0.003	0.050	0.065	0.044
BS15	-0.129	0.081	0.114	-0.006	-0.018	-0.030	-0.007	-0.033
FN16	0.135	0.036	0.092	0.134	0.076	-0.020	0.001	0.094
RU17	0.051	0.249	0.176	0.013	0.026	0.167	0.121	0.021
PC18	-0.089	0.218	-0.026	0.001	0.096	0.021	0.084	-0.073
FN19	-0.099	0.097	0.156	0.117	0.136	0.176	0.146	0.361
BS20	0.014	0.226	0.053	0.131	0.073	0.175	0.155	0.146
RU21	-0.092	0.059	0.109	0.076	-0.002	0.077	0.035	0.039
PC22	-0.067	0.195	0.161	0.101	0.154	0.078	0.100	-0.004
PC23	-0.026	0.028	0.096	0.050	0.042	-0.088	0.127	-0.002
BS24	-0.077	0.177	0.221	0.248	0.146	0.089	0.171	0.097
RU25	0.035	0.182	0.207	0.079	0.056	0.136	0.129	0.062
FN26	0.041	0.143	0.250	0.192	0.165	0.062	0.145	0.060
BS26	-0.090	0.053	0.093	-0.033	0.056	0.016	0.093	0.016
PC29	-0.094	0.092	-0.010	0.061	0.100	-0.112	-0.019	0.032
RU30	-0.023	0.103	0.004	-0.039	-0.021	0.083	0.117	0.145
FN31	0.046	0.039	-0.020	0.074	0.046	0.105	0.220	-0.019
PC32	-0.085	0.099	0.087	0.060	0.101	0.055	0.027	0.086
BS34	0.027	-0.015	0.121	0.044	0.176	-0.083	0.073	0.208
RU35	-0.032	0.206	0.024	-0.023	-0.001	0.209	0.060	-0.026
PC36	-0.073	0.168	0.237	0.228	0.074	0.221	0.221	0.082
FN37	-0.077	0.171	0.232	0.063	0.182	0.087	0.162	0.098
BS39	-0.007	0.197	0.084	0.096	0.158	0.042	0.188	0.174

Table 15 continues.

Table 15 continued

	FN12	FN13	RU14	BS15	FN16	RU17	PC18	FN19
FN12	1.000							
FN13	0.000	1.000						
RU14	0.095	0.076	1.000					
BS15	-0.050	-0.112	-0.078	1.000				
FN16	-0.076	-0.075	0.072	0.074	1.000			
RU17	-0.152	0.175	0.091	0.058	-0.041	1.000		
PC18	-0.027	0.131	0.106	0.010	-0.052	0.337	1.000	
FN19	-0.041	0.261	0.044	-0.005	-0.007	0.230	0.123	1.000
BS20	0.010	0.070	0.078	-0.011	0.026	0.199	0.175	0.160
RU21	0.037	0.084	0.119	0.209	-0.102	0.056	-0.005	0.076
PC22	0.033	0.120	0.039	0.003	-0.073	0.184	0.185	0.045
PC23	0.045	0.024	0.071	-0.011	0.053	0.022	0.094	-0.044
BS24	-0.051	0.174	0.114	0.004	0.035	0.299	0.214	0.280
RU25	-0.000	0.046	0.026	-0.030	-0.033	0.169	0.129	0.166
FN26	0.042	0.089	0.060	0.004	0.107	0.206	0.096	0.232
BS28	-0.047	0.155	0.108	-0.007	-0.025	0.150	0.154	0.069
PC29	-0.027	-0.049	0.054	0.051	0.044	0.041	0.030	0.084
RU30	0.062	0.027	0.122	-0.113	0.029	0.093	0.148	0.117
FN31	0.122	0.080	0.074	0.071	-0.058	0.101	0.098	0.107
PC32	-0.026	0.131	0.054	0.134	0.083	0.140	0.182	0.198
BS34	0.150	-0.052	-0.078	-0.047	0.153	-0.090	0.025	0.057
RU35	-0.010	0.122	0.110	0.003	-0.051	0.291	0.199	0.125
PC36	-0.044	0.196	0.031	0.017	0.012	0.337	0.269	0.264
FN37	0.021	0.107	0.144	-0.060	0.042	0.270	0.125	0.337
BS39	0.055	0.056	-0.002	-0.031	0.096	-0.004	0.104	0.103

Table 15 continues.

Table 15 continued

	BS20	RU21	PC22	PC23	BS24	RU25	FN26	BS28
BS20	1.000							
RU21	0.049	1.000						
PC22	0.062	0.112	1.000					
PC23	-0.002	-0.056	0.044	1.000				
BS24	0.288	0.022	0.192	0.094	1.000			
RU25	0.172	0.097	0.171	-0.024	0.161	1.000		
FN26	0.294	0.015	0.096	0.103	0.271	0.045	1.000	
BS28	0.145	0.070	0.105	0.144	0.044	0.037	0.170	1.000
PC29	0.136	-0.028	0.001	0.041	0.120	-0.096	0.077	0.018
RU30	0.109	0.068	0.120	0.020	0.121	0.174	0.033	0.025
FN31	0.049	0.085	0.220	0.080	0.150	0.087	0.122	0.099
PC32	0.232	0.012	-0.012	0.005	0.141	-0.045	0.181	0.085
BS34	0.058	0.073	0.106	0.043	0.075	0.081	-0.020	0.007
RU35	0.185	0.056	0.184	-0.063	0.143	0.233	0.151	0.076
PC36	0.200	0.011	0.137	0.021	0.329	0.228	0.299	0.113
FN37	0.265	0.065	0.145	0.126	0.397	0.110	0.296	0.205
BS39	0.164	-0.031	0.036	-0.049	0.165	0.032	0.137	-0.012
	PC29	RU30	FN31	PC32	BS34	RU35	PC36	FN37
PC29	1.000							
RU30	-0.055	1.000						
FN31	-0.057	0.038	1.000					
PC32	0.087	-0.003	-0.022	1.000				
BS34	0.184	-0.029	0.059	0.127	1.000			
RU35	0.003	0.129	0.061	0.083	-0.162	1.000		
PC36	0.050	0.130	0.068	0.194	0.020	0.294	1.000	
FN37	0.166	0.053	0.302	0.122	0.244	0.197	0.222	1.000
BS39	0.099	0.037	-0.086	0.081	0.208	-0.022	0.252	0.034

Geography Item Difficulties and Item-Total Domain
Correlation Coefficients

Item	p	Domains			
		CC	CK	CF	CI
CC02	0.5222	0.1796	0.1335	0.2619	0.2986
CC03	0.4000	0.2158	0.2149	0.1809	0.2167
CC05	0.4889	0.1839	0.1216	0.1801	0.1995
CC06	0.5389	0.1149	0.0989	0.2123	0.2703
CC07	0.4889	0.2246	0.2389	0.2315	0.2123
CC58	0.4750	0.1670	0.2187	0.3386	0.3557
CC61	0.4111	0.1468	0.1550	0.0449	0.1233
CC62	0.3583	0.0116	0.0707	0.1180	0.1540
CC63	0.6472	0.1218	0.1477	0.1364	0.1682
CK03	0.2222	0.1785	0.2831	0.2029	0.2998
CK14	0.3611	-.0614	0.2670	-.0153	-.0550
CK15	0.4667	0.1982	0.2250	0.1350	0.0432
CK16	0.5778	0.2087	0.2210	0.2340	0.2543
CK17	0.3167	0.1028	0.0911	0.1144	0.0787
CK18	0.4278	0.1720	0.2244	0.1990	0.2450
CK19	0.4389	0.2525	0.2178	0.2197	0.2153
CK20	0.1861	0.0163	0.0441	0.0624	0.0388
CK21	0.1944	0.0410	0.0646	0.0585	0.0769
CF02	0.4667	0.2167	0.0976	0.1787	0.2245
CF03	0.5583	0.1551	0.1431	0.1430	0.1053
CF04	0.4861	0.0388	0.0529	0.1198	0.1063
CF05	0.3222	0.1188	0.1399	0.1283	0.1568
CF06	0.4750	0.0483	0.1110	0.1104	0.0877
CF13	0.5028	0.1501	0.1582	0.1313	0.1146
CF17	0.6111	0.1315	0.1391	0.2291	0.1864
CF18	0.4472	0.1879	0.1852	0.1841	0.2793
CF19	0.3833	0.2458	0.2322	0.2048	0.3189
CI01	0.4667	0.0944	0.0636	0.0137	0.1017
CI02	0.5722	0.2107	0.1886	0.1740	0.1038
CI03	0.1750	0.1334	0.1331	0.0919	0.0922
CI05	0.6194	0.1893	0.1119	0.2076	0.1929
CI07	0.4139	-.0191	0.0053	0.0440	0.0615
CI08	0.4667	0.2472	0.2559	0.1866	0.1771
CI10	0.2528	0.1425	0.0784	0.1253	0.1311
CI11	0.5861	0.2049	0.1399	0.1648	0.1364
CI12	0.3944	0.2280	0.1976	0.1983	0.2394

Note. Item-total domain correlation coefficients for items with their intended domain were calculated by excluding the item from the domain total.

Physical Education Item Difficulties and Item-Total Domain
Correlation Coefficients

Item	p	Domains			
		RU	PC	BS	FN
RU01	0.7809	0.1548	0.1139	0.1893	-.0232
RU08	0.3357	0.0191	0.3187	-.0116	0.1398
RU14	0.6678	0.1144	0.3258	0.0723	0.0940
RU17	0.7611	0.0828	0.1214	0.0468	0.1730
RU21	0.5361	0.0481	-.0136	0.3540	0.0870
RU25	0.3473	0.0331	0.3129	0.0624	0.1723
RU30	0.5408	0.2249	0.1165	0.1776	0.1332
RU35	0.7086	0.4227	0.0569	0.1209	0.0203
PC03	0.5303	0.1959	0.0980	0.4186	0.0739
PC06	0.3636	0.1466	0.0576	0.0769	-.0499
PC18	0.8205	0.1939	0.0738	0.0524	0.4442
PC22	0.4907	0.1406	0.4001	0.1052	0.1060
PC23	0.6737	0.0479	0.1233	0.1515	0.0374
PC29	0.2611	0.0737	0.0654	0.0960	0.1147
PC32	0.6946	0.1198	0.0958	0.2297	-.0300
PC36	0.4662	0.1760	0.0842	0.0752	0.4777
BS04	0.2885	0.1560	0.4832	0.1475	0.1481
BS09	0.3706	0.1763	0.1656	0.1448	0.1593
BS15	0.6096	0.0233	0.0587	0.0103	0.4040
BS20	0.8030	0.4445	0.1134	0.1545	0.1552
BS24	0.6702	0.3426	0.0748	0.1800	-.0260
BS28	0.5315	0.2519	0.2548	0.0729	0.1578
BS34	0.0629	0.1140	0.0903	0.0598	0.4643
BS39	0.1888	0.2150	0.4852	0.1130	0.1059
FN05	0.3380	-.0724	0.1579	-.0335	0.2463
FN10	0.2028	0.2530	0.1615	0.2196	0.2311
FN12	0.4196	0.4908	0.0251	0.1754	0.2024
FN13	0.1527	0.1365	0.1164	0.4518	0.1137
FN16	0.6538	0.1404	0.0300	0.1280	0.0581
FN19	0.2937	0.4385	0.2343	0.1147	0.0084
FN26	0.4044	0.0742	0.2144	0.0708	0.4341
FN31	0.6818	0.1353	0.1317	0.4701	0.1150
FN37	0.4009	0.0362	0.2356	0.1067	0.0602

Note. Item-total domain correlation coefficients for items with their intended domain were calculated by excluding the item from the domain total.

Table 18

Difficulties and Standard Deviations of Total Test and Intended
Domains for Geography and Physical Education

Geography					
	Total test	Domains CC	CK	CF	CI
p	.4367	.4811	.3546	.4725	.4385
st. deviation	.1446	.2220	.1824	.2043	.1896

Physical Education					
	Total test	Domains RU	PC	BS	FN
p	.4863	.5847	.5375	.4402	.3942
st. deviation	.1250	.1924	.1970	.1771	.1860

the total test and for the domains for the two curriculum areas. Given the probability of guessing the correct answer and the difficulty levels for the geography and physical education items and total tests, there is little indication that the students had acquired knowledge in the domains comprising the two tests.

A possible reason for the lack of an underlying structure to the examinee response data might be that the content of the geography and physical education domains was not known by the students. Therefore, the examinee response data could not be used to investigate and/or confirm the intended or assigned domain structure.

Consideration was given to gathering another set of geography and physical education examinee response data. However, gathering additional response data presented difficulties in terms of locating a suitable size sample of students who had had the same teaching and coverage of content of the domains. Therefore, the idea of confirming a domain structure suggested by the judge group ratings or exploring other structures using examinee response data was put aside. Thus, for comparison purposes the intended domain structure was used.

For this study, the use of the intended domain structure is supported by the placement of the items during the OAIP and PELA test development efforts. As well, support for the use of the intended domain structure was provided by having two geography and two physical education specialists review the appropriateness of the placement of the items in the intended domain. Two public school

board research officers experienced in the development of test items also reviewed the placement of the items in the domains. Among this group there was agreement that the items were appropriately placed in the domains.

In effect, since the domain structure suggested by the judge group ratings could not be investigated or confirmed, the first research question became: How do the judgmental techniques compare with one another? Therefore, the analysis of the data proceeded with a comparison of the strength of the relationships among the judgmental techniques. Since the degree of elaboration had not been found to have an effect on the judge group ratings on the three judgmental techniques, the ratings of the individual judges regardless of the degree of elaboration used were combined and the judge group ratings for each item on each technique were recalculated. Thus, the combined high and low elaboration ratings were the result of the individual ratings of 28 judges.

The combined elaboration judge group ratings for each item on each judgmental technique are shown in Tables 19 and 20. Based on the judge group ratings for geography items, domains CC and CF were rated as discrete domains on all three judgmental techniques. Domains CK and CI were rated as interrelated on the three techniques. Apparently the judges found overlap between the urbanization concepts of domain CK and the population pattern concepts of domain CI. Physical education domain RU was rated as a discrete domain on all

Table 19

Combined Elaboration Judge Group Ratings for Intended Domain
Geography Items on Each Judgmental Technique

Item	IOCI				CARS				CMP			
	CC	CK	CF	CI	CC	CK	CF	CI	CC	CK	CF	CI
CC02	0.86	-0.29	-0.26	-0.31	4.13	1.65	1.70	1.65	96.43	7.14	7.14	10.71
CC03	0.70	-0.30	-0.08	-0.32	4.55	1.39	2.04	1.41	92.86	7.14	14.29	17.86
CC05	0.74	-0.30	-0.14	-0.30	4.74	1.65	1.87	1.44	92.86	10.71	21.43	7.14
CC06	0.52	-0.44	0.35	-0.44	4.50	1.41	3.27	1.52	100.00	3.57	50.00	3.57
CC07	0.74	-0.52	0.17	-0.38	4.46	1.58	2.75	1.58	89.29	3.57	39.29	10.71
CC58	0.58	-0.42	0.22	-0.38	4.30	1.64	2.64	1.80	89.29	10.71	53.57	10.71
CC61	0.61	-0.40	0.24	-0.45	4.55	1.27	2.89	1.27	92.86	3.57	39.29	3.57
CC62	0.55	-0.39	0.28	-0.44	4.30	1.21	2.98	1.32	96.43	3.57	46.43	3.57
CC63	0.77	-0.42	-0.02	-0.33	4.35	1.54	2.72	1.63	89.29	0.00	21.43	3.57
CK03	-0.49	0.47	-0.49	0.52	1.70	4.14	1.62	4.48	14.29	85.71	25.00	89.29
CK14	-0.51	0.46	-0.48	0.53	2.07	4.77	1.77	4.34	0.00	78.57	10.71	89.29
CK15	-0.36	0.40	-0.49	0.45	1.93	4.70	1.66	4.86	0.00	96.43	7.14	100.00
CK16	-0.36	0.47	-0.48	0.38	1.89	3.93	2.17	4.06	7.14	92.86	35.71	89.29
CK17	-0.48	0.44	-0.35	0.39	2.57	4.68	2.12	4.27	3.57	78.57	21.43	78.57
CK18	0.03	0.18	-0.04	-0.16	2.86	3.62	2.54	3.04	62.96	62.96	55.56	51.92
CK19	-0.15	0.49	-0.43	0.09	2.20	4.37	1.54	3.04	18.52	88.89	18.52	62.96
CK20	-0.35	0.30	0.25	-0.20	2.35	3.04	3.30	2.69	10.71	60.71	42.86	32.14
CK21	-0.42	0.59	-0.20	0.02	1.34	3.80	2.23	3.14	3.70	74.07	14.81	55.56
CF02	0.36	-0.51	0.65	-0.51	3.80	2.09	4.30	2.22	78.57	3.57	89.29	17.86
CF03	0.15	-0.30	0.63	-0.49	3.33	1.93	4.22	1.63	32.14	17.86	75.00	14.29
CF04	0.37	-0.47	0.53	-0.42	3.93	1.61	4.43	1.64	89.29	0.00	92.86	3.57
CF05	0.00	-0.20	0.69	-0.49	2.44	2.40	4.12	1.69	53.57	21.43	85.71	14.29
CF06	-0.10	-0.38	0.81	-0.33	2.38	2.16	4.27	2.14	14.29	14.29	89.29	10.71
CF13	0.12	-0.32	0.64	-0.44	3.71	2.00	4.50	1.75	75.00	7.14	85.71	7.14
CF17	0.31	-0.36	0.38	-0.33	3.59	1.25	3.16	1.66	66.67	3.70	44.44	3.70
CF18	0.31	-0.33	0.33	-0.31	3.55	1.73	3.14	1.73	78.57	10.71	60.71	10.71
CF19	0.27	-0.39	0.32	-0.20	3.21	1.93	3.36	2.21	67.86	25.00	64.29	17.86
CI01	-0.41	0.30	-0.46	0.57	1.62	3.48	1.96	4.38	3.57	50.00	21.43	85.71
CI02	-0.40	0.27	-0.48	0.61	1.61	3.87	1.63	4.46	0.00	75.00	7.14	100.00
CI03	-0.40	0.27	-0.47	0.60	1.72	3.20	1.48	3.78	3.57	50.00	10.71	96.43
CI05	-0.30	0.23	-0.56	0.63	1.61	2.88	1.55	3.95	7.14	53.57	10.71	89.29
CI07	-0.28	0.34	-0.48	0.41	1.79	4.25	1.57	4.52	3.57	75.00	7.14	96.43
CI08	-0.52	0.51	-0.52	0.53	1.68	3.91	1.50	4.41	10.71	89.29	17.86	89.29
CI10	-0.37	0.35	-0.47	0.49	1.95	4.34	1.71	4.45	7.14	71.43	10.71	89.29
CI11	-0.55	0.46	-0.33	0.41	1.36	4.00	2.50	4.18	0.00	75.00	10.71	85.71
CI12	-0.28	0.24	-0.35	0.39	1.64	3.93	1.52	3.79	7.14	60.71	14.29	71.43

Table 20

Combined Elaboration Judge Group Ratings for Intended Domain
Physical Education Items on Each Judgmental Technique

Item	IOCI				CARS				CMP			
	Domains				Domains				Domains			
	RU	PC	BS	FN	RU	PC	BS	FN	RU	PC	BS	FN
RU01	0.83	-0.31	-0.17	-0.36	4.62	2.21	2.18	2.16	100.00	14.29	28.57	14.29
RU08	0.96	-0.30	-0.35	-0.32	4.88	1.27	1.20	1.27	96.43	0.00	7.14	3.57
RU14	0.86	-0.36	-0.19	-0.31	4.86	2.11	1.98	1.77	92.86	14.29	35.71	3.57
RU17	0.90	-0.29	-0.31	-0.31	4.84	1.21	1.41	1.36	100.00	0.00	7.14	3.57
RU21	0.86	-0.23	-0.35	-0.28	4.62	1.25	1.21	1.25	96.43	3.57	0.00	0.00
RU25	0.96	-0.26	-0.38	-0.33	4.48	1.26	1.15	1.04	92.86	3.57	0.00	0.00
RU30	1.00	-0.33	-0.33	-0.33	4.62	1.27	1.21	1.20	100.00	0.00	0.00	0.00
RU35	0.92	-0.35	-0.15	-0.42	4.64	1.41	1.79	1.34	100.00	7.14	21.43	0.00
PC03	-0.61	0.58	-0.14	0.17	2.16	3.46	2.41	3.86	10.71	78.57	32.14	71.43
PC06	-0.32	0.11	0.28	-0.08	2.59	3.07	3.82	2.98	28.57	64.29	82.14	75.00
PC18	-0.47	0.32	-0.07	0.22	1.61	3.50	2.39	3.66	10.71	50.00	21.43	57.14
PC22	-0.42	0.27	-0.18	0.32	2.21	4.45	3.59	3.36	28.57	96.43	60.71	75.00
PC23	-0.31	0.00	0.19	0.12	2.46	4.04	3.61	3.36	28.57	64.29	71.43	71.43
PC29	-0.51	0.37	-0.08	0.23	1.25	3.73	3.14	3.48	3.70	70.37	44.44	77.78
PC32	-0.36	0.35	-0.15	0.16	2.38	4.20	3.09	3.61	25.00	96.43	28.57	53.57
PC36	-0.51	0.42	-0.01	0.11	2.06	4.33	3.80	3.74	7.14	92.86	67.86	57.14
BS04	-0.54	0.39	0.43	-0.28	1.57	4.02	4.45	2.73	3.57	71.43	78.57	14.29
BS09	-0.71	0.32	0.43	-0.04	1.91	3.69	4.22	3.11	7.14	78.57	89.29	39.29
BS15	-0.24	0.14	0.52	-0.43	3.23	3.61	4.64	2.77	46.43	64.29	82.14	14.29
BS20	-0.36	0.05	0.29	0.02	1.71	3.41	3.75	2.38	14.81	44.44	48.15	25.93
BS24	-0.70	0.33	0.21	0.16	1.38	4.27	4.32	3.25	7.14	89.29	60.71	53.57
BS28	-0.53	0.33	0.49	-0.29	1.55	3.84	4.71	2.70	7.41	48.15	92.59	14.81
BS34	-0.50	0.19	0.67	-0.36	1.55	3.07	4.48	2.12	0.00	46.43	92.86	17.86
BS39	-0.61	0.13	0.46	0.01	1.46	3.32	4.46	2.71	3.57	57.14	96.43	39.29
FN05	-0.70	0.26	0.07	0.38	1.57	3.68	3.29	4.52	14.29	82.14	60.71	89.29
FN10	-0.25	0.32	-0.08	0.01	2.36	4.14	3.05	3.32	28.57	78.57	35.71	46.43
FN12	0.13	0.01	0.04	-0.18	3.89	2.71	3.23	3.25	85.71	35.71	64.29	53.57
FN13	-0.21	0.02	0.10	0.10	3.02	4.21	3.88	3.77	64.29	71.43	71.43	82.14
FN16	-0.62	0.26	0.00	0.36	1.61	4.05	3.82	4.55	10.71	85.71	57.14	85.71
FN19	-0.43	0.22	-0.06	0.27	2.46	4.21	3.71	3.96	21.43	82.14	60.71	89.29
FN26	-0.46	0.30	0.16	-0.01	2.84	4.43	4.29	3.71	17.86	100.00	67.86	64.29
FN31	-0.43	0.30	0.23	-0.10	2.32	2.98	3.41	3.20	7.14	53.57	60.71	42.86
FN37	-0.32	0.14	-0.01	0.18	2.11	3.52	2.89	3.80	14.29	82.14	46.43	71.43

judgmental techniques. Domains PC, BS, and FN were not clearly defined on the three judgmental techniques. Thus, in both geography and physical education, the judgmental techniques functioned similarly in showing discrete and overlapping domains.

The correlations among the combined high and low elaboration ratings on intended and nonintended domains for the three judgmental techniques are shown in Table 21. Strong correlations were found among the three judgmental techniques in both geography and physical education. In both curriculum areas, the strongest correlation was between the CARS and CMP. On the basis of these correlations, the three judgmental techniques appear to be very similar.

The three judgmental techniques were also compared by domain using combined high and low elaboration ratings for items on intended and nonintended domains. The correlations among the techniques by domains are shown in Table 22. Again, the three judgmental techniques are highly correlated, in this case, on the four geography and four physical education domains.

The next step in the original plan for the analysis of the data was to investigate the effectiveness of the judgmental techniques by assigning each item to a domain on the basis of the judges' ratings using the highest positive domain scores as well as various cut-off values for each of the judgmental techniques under the two degrees of elaboration. The domains created by this procedure were called assigned domains. In contrast with Rovinelli (1976) who used only one cut-off value with each technique, a range of cut-off values was

Table 21

**Correlations among Combined High and Low Elaboration Ratings
On Intended and Nonintended Domains for the Three Judgmental
Techniques**

Geography		
	IOCI	CARS
CARS	.926	
CMP	.942	.955

Physical Education		
	IOCI	CARS
CARS	.861	
CMP	.884	.948

Table 22

Correlations among the Three Judgmental Techniques by Domain
using Combined High and Low Elaboration Ratings on
Intended and Nonintended Domains

Geography		
	IOCI	CARS
Domain CC		
CARS	0.957	
CMP	0.956	0.955
Domain CK		
CARS	0.942	
CMP	0.970	0.964
Domain CF		
CARS	0.955	
CMP	0.915	0.941
Domain CI		
CARS	0.900	
CMP	0.958	0.968

Physical Education		
	IOCI	CARS
Domain RU		
CARS	0.956	
CMP	0.959	0.972
Domain PC		
CARS	0.847	
CMP	0.870	0.951
Domain BS		
CARS	0.881	
CMP	0.899	0.946
Domain FN		
CARS	0.849	
CMP	0.905	0.918

used in the present study. Rovinelli employed a cut-off value of .70 with the IOCI and 70 for the CMP. The ARS values were used to calculate an SD statistic which took the ratings on other domains into account. A cut-off value of .50 on this SD was used. For the present study, the IOCI values were specified as >.30 to >.90; for the CARS, >3.00 to >4.50; for the CMP, >65 to >95. These values were selected because they represented a range of minimum to strong agreement among the judges that an item was congruent with a domain. The use of cut-off values allowed the assignment of an item to more than one domain. As could be anticipated, the higher the cut-off value, the fewer the items that were assigned to domains and the fewer the items assigned to more than one domain.

A comparison of the effect of technique, cut-off score, and domain elaboration was to have been conducted based on the internal consistency of the assigned domains. However, since the correlation of items affects the values for internal consistency and as the interitem correlations were known to be low, it was not considered appropriate to use the examinee response data for such an analysis. Therefore, the investigation of the amended research question of how the judgmental techniques compared with one another proceeded with a comparison of the ratings of the items on their intended domain and on the nonintended domains.

In Table 23, the number of items assigned to their intended and to nonintended domains using the combined elaboration judge group

Number of Items Assigned to their Intended Domain and to Nonintended Domains using Judge Group Ratings with Various Cut-Off Values

Geography (Items=36)

IOCI

Values	>.30	>.40	>.50	>.60	>.70	>.80	>.90
Intended	34	29	21	13	5	2	0
Nonintended	14	5	3	0	0	0	0

CARS

Values	>3.00	>3.25	>3.50	>3.75	>4.00	>4.25	>4.50
Intended	36	33	33	32	26	21	7
Nonintended	25	19	16	13	6	5	1

CMP

Values	>65	>70	>75	>80	>85	>90	>95
Intended	31	31	29	26	26	12	6
Nonintended	16	14	9	7	6	2	1

Physical Education (Items=33)

IOCI

Values	>.30	>.40	>.50	>.60	>.70	>.80	>.90
Intended	21	16	11	9	8	8	4
Nonintended	6	0	0	0	0	0	0

CARS

Values	>3.00	>3.25	>3.50	>3.75	>4.00	>4.25	>4.50
Intended	33	30	27	24	21	18	11
Nonintended	41	32	26	14	8	2	0

CMP

Values	>65	>70	>75	>80	>85	>90	>95
Intended	24	24	22	20	18	14	9
Nonintended	21	19	11	8	5	1	1

ratings with various cut-off values is shown. The IOCI had items assigned to nonintended domains in geography only with cut-off values of .50 and below. In physical education, the IOCI had items assigned to nonintended domains only with a cut-off value of .30. Thus, on the IOCI in both geography and physical education, items tended not to be assigned to multiple domains. The assignment of items to multiple domains occurred only with low cut-off values. Items were more frequently assigned to multiple domains using the CARS than the CMP, although this pattern did not occur with CARS values >4.00 and CMP values >85 .

In Table 24, in the first column, the number of items assigned to their intended and to nonintended domains using the highest of the judge group ratings on the four domains is shown. Without cut-off values, most items were assigned to their intended domains. This pattern was apparent in both curriculum areas and across the three judgmental techniques. If the best judgmental technique is defined as the number of items receiving their highest ratings on their intended domain minus the number of items assigned to nonintended domains, then the IOCI would be considered the best technique in both curriculum areas. In geography, the CARS and CMP were equal. In physical education, the CARS was second best and the CMP was the worst technique.

In the remaining columns of Table 24, the number of items assigned to domains using the highest domain value in combination with cut-off values is shown. Using this approach of assigning

Table 24

Number of Items Assigned to their Intended Domain and to Nonintended Domains using the Highest Judge Group Ratings in Combination with Cut-Off Values

Geography (Items=36)

IOCI

Values	Highest	>.30	>.40	>.50	>.60	>.70	>.80	>.90
Intended	32	30	26	21	13	5	2	0
Nonintended	4	4	2	0	0	0	0	0

CARS

Values	Highest	>3.00	>3.25	>3.50	>3.75	>4.00	>4.25	>4.50
Intended	30	30	30	29	28	24	21	6
Nonintended	6	6	6	5	3	3	2	1

CMP

Values	Highest	>65	>70	>75	>80	>85	>90	>95
Intended	30	28	28	26	24	24	11	5
Nonintended	6	6	4	5	3	3	1	1

Physical Education (Items=33)

IOCI

Values	Highest	>.30	>.40	>.50	>.60	>.70	>.80	>.90
Intended	25	21	16	11	9	8	8	4
Nonintended	8	3	0	0	0	0	0	0

CARS

Values	Highest	>3.00	>3.25	>3.50	>3.75	>4.00	>4.25	>4.50
Intended	24	24	24	24	22	22	18	11
Nonintended	9	9	9	9	7	4	1	0

CMP

Values	Highest	>65	>70	>75	>80	>85	>90	>95
Intended	23	22	22	22	20	18	14	9
Nonintended	10	8	8	5	5	3	1	1

Note. Highest the highest of the ratings on the four domains

items to the domain on which they received the highest value, then using cut-off values, fewer items were assigned to nonintended domains than when only cut-off values were used.

Thus, three approaches were used in assigning items to their intended and to nonintended domains. The first approach (Table 23) was to assign items to domains using cut-off values only. The second approach (Table 24, column 1) was to assign items to domains using the highest domain value. The third approach (Table 24, columns 2 through 7) was to assign items to domains using the highest domain value in combination with cut-off values. These three approaches were then compared with one another. With the first approach when only cut-off values were used, with low cut-off values, many items were assigned to their intended domains, but many items were also assigned to nonintended domains. When high cut-off values were used, fewer items were assigned to nonintended domains, but fewer items were also assigned to their intended domains. With the second approach of assigning items to the domain on which they received the highest ratings, many items were assigned to their intended domain, but a number of items were assigned to nonintended domains. With the third approach of using the highest domain values in combination with cut-off values, the number of items assigned to nonintended domains at all cut-off values tended to decrease. For example, in geography, the CARS used with the first approach and a cut-off value of 3.75, 32 items were assigned to their intended

domain and 13 items were assigned to nonintended domains. When the second approach was used, 30 items were assigned to their intended domain and 6 items were assigned to nonintended domains. When the third approach was used, with a cut-off value of 3.75, 28 items were assigned to their intended domains and 3 items were assigned to nonintended domains.

Next, the opinions of the judges regarding the three techniques were studied.

The Second Research Question

The second research question was: What were the opinions of the judges regarding the IOCI, the ARS, and the CMP?

At the conclusion of the administration of the judgmental questionnaires, the judges were asked to indicate their attitude toward the three judgmental techniques as well as to comment on any aspect of the study. The forms for the collection of the attitudinal data are included in Appendix E. The results of the analysis of the attitudinal data are presented in Tables 25 to 28.

Judges were queried about the judgmental technique which they found easiest to use. The judges' responses are presented in Table 25. For both the geography and physical education judges using either high elaboration or low elaboration, the technique reported as easiest to use was the CMP. When both elaboration groups were considered together, the IOCI was the second choice for easiest to use. The ARS was reported as easiest to use by a small number of

Table 25

Judgmental Technique Found Easiest to Use by Judges

Elabor- ation	Geography judges				Physical education judges			
	N	IOCI	ARS	CMP	N	IOCI	ARS	CMP
High	38	14	4	20	42	6	8	28
Low	42	9	8	25	41	15	4	22
Total	80	23	12	45	83	21	12	50

Table 26

Judgmental Technique Felt to Allow Best Judgment

		Geography judges				Physical education judges			
Elabora- tion	N	IOCI	ARS	CMP	N	IOCI	ARS	CMP	
High	37	14	18	5	42	7	26	9	
Low	42	7	26	9	41	16	22	3	
Total	79	21	44	14	83	23	48	12	

Table 27

Judges' Feeling Regarding Sufficiency of Type of Elaboration Used

		Geography judges		Physical education judges		
Type of elabora- tion used	N	Suffi- cient	In- suffi- cient	N	Suffi- cient	In- suffi- cient
High	36	33	3	42	38	4
Low	42	32	10	38	29	9

Table 28

Judges' Feeling Regarding Type of Elaboration Not Used

	Geography judges			Physical education judges		
--	---------------------	--	--	------------------------------	--	--

Type of elabora- tion used	N	Yes	No	N	Yes	No
	Type of elaboration not used sufficient?					
High	30	10	20	40	22	18
	Type of elaboration not used more helpful?					
Low	42	21	21	38	18	20

judges. The comments of the judges who found the CMP easiest showed that they found it simpler to make a two-choice decision. The judges who found the ARS easiest explained that it was easiest for them to record their decision accurately when the increased choices offered by the five-point scales were available to them. The judges who found the IOCI easiest to use indicated that they found the three-point scale to be a good compromise between the limited choice provided by a two-point scale and the confusion created by having to indicate a choice on a five-point scale.

The judges were then asked which judgmental technique they felt allowed them to make the best judgment of item-domain congruence. The responses of the judges are presented in Table 26. The order in which both geography and physical education judges felt the techniques to allow the best judgments was the ARS, the IOCI, and the CMP. When the elaboration groups were taken into account, the ARS remained the first choice for all groups, but the IOCI and the CMP varied in the order in which they were chosen as allowing the best judgments to be made. The judges who chose the ARS as allowing the best judgment to be made commented that fine differences of judgments could be recorded on the scales as they offered more options. The judges who chose the IOCI stated that it was best because it was not as confusing as the ARS but not as limited as the CMP. The CMP was seen as being too limited to allow good judgments to be made.

The judges' opinions regarding the sufficiency of the type of elaboration which they employed in answering the questionnaires were

also solicited. The results are presented in Table 27. Most judges reported that the type of elaboration they used, whether high or low, was sufficient for judging item-domain congruence.

Judges who used high elaboration were asked whether they felt low elaboration would have been sufficient to judge the match between the item and the domain. Judges who used low elaboration were asked whether they felt it would have been more helpful to have high domain elaboration. The results are presented in Table 28. Judges who used geography items with high elaboration felt that low elaboration would have been insufficient for the task. The other judges, geography with low elaboration and physical education with both high and low elaboration were approximately equally divided in their opinion of whether the type of elaboration not used would have been sufficient or more helpful for judging item-domain congruence.

An opportunity for judges to make any comment on the experience of rating item-domain congruence was provided at the bottom of the personal and attitudinal data forms. Frequent comments included that the length of time needed to complete the questionnaires created tiredness and less painstaking effort toward the end of the task. Judges reported that they found the indication of the correct answer with an asterisk distracting, especially if they took issue with the item or the answer. Judges doubted the worth of some of the test items. Last, many judges commented that they found no difference between the terms relevance and suitable used with the ARS.

After completing the questionnaire, judges commented that they found the task of rating items time consuming. Some judges commented that keeping all the information from the high elaboration domain descriptions in mind was difficult. Some judges commented that they found the task of matching items and domains interesting.

In chapter 5, the results of the analysis of the data are summarized and discussed.

Chapter 5

Summary and Discussion

In Chapter 5, the results of the analysis of the data are summarized and discussed. The establishment of the validity of the judgmental ratings is discussed first. Then the findings regarding the hypothesis and research questions are considered.

The Establishment of the Validity of the Judge Group Ratings

If a judge was unable to use a particular judgmental technique, the ratings by the judge would affect the judge group ratings and ultimately the interpretation of the findings. Therefore, a careful search for aberrant judges was conducted at the outset of the analysis of the data. Questionnaires were not used if largely incomplete or if they were obviously not conscientiously filled out. No aberrant judge was found in the examination of the individual ratings of the judges on each item. Nor were aberrant judges found by comparing the judge group ratings and the variances of the ratings of the judges in the judge groups on fifth domain items. Teaching experience and the first language of the judges were deemed not to have been an influence on the decisions of the judge groups.

Other research was not found in the area of item-domain congruence in which an approach to locating aberrant judges was reported. In this study, the finding of no aberrant judges did not seem unusual for at least two reasons. First, the discarding of

questionnaires at the outset of the analysis appears to have eliminated problem judges. It appears that judges conscientious enough to complete the questionnaires were able to make reasonable decisions. Second, the task of judging item-domain congruence does not seem to be a difficult one. Secolsky (1983) found that university students could judge item-domain congruence. The university students who acted as judges in the present study did not have difficulty using the judgmental techniques as evidenced by the placement of items generally in their intended domains. In the pilot study, academically capable grade 8 and high school students, without identified expertise in the two curriculum areas, were able to match items with domains. The finding of strong correlations between each technique used with high elaboration and the same technique used with low elaboration is also an indication that the judges were using the techniques effectively. However, the judges rated the items of geography domains CK and CI as belonging to the same domain. Physical education domains PC, BS, and FN were also rated as not clearly defined. It is possible that more experienced judges could have made more precise distinctions between domains.

Therefore, there is evidence to suggest that the selection of curriculum experts to act as judges of item-domain congruence is an issue which requires examination and possible revision.

The validity of the judge group ratings having been established, the research hypothesis and questions were then considered.

The Research Hypothesis and Questions

Because of the belief that high domain elaboration would enable judges to determine item-domain congruence better, it was expected that the judge group ratings for items on the three judgmental techniques would be higher when high domain elaboration was used. This expectation was not met because the mean of the judge group ratings under high and low elaboration were not consistently higher for the judgmental techniques when high elaboration was used. As well, each technique used with high elaboration was found to be strongly correlated with the same technique used with low elaboration. Thus, the hypothesis that when high domain elaboration was used, item-domain congruence ratings would be higher than when low domain elaboration was used was not supported.

Since no other studies regarding domain elaboration and item-domain congruence were located, these results cannot be compared with the results of other researchers. The fact that a consistent elaboration effect was not found in the present research calls into question the notion that high elaboration will lead to greater item-domain congruence. However, further investigation of the IOCI will be needed to determine the effect of elaboration on that technique. The judges felt that the type of elaboration which they used, whether high or low elaboration, was sufficient for judging item-domain congruence. Opinions were mixed regarding whether the opposite type of elaboration from the one used would have been suf-

ficient or more helpful for the task of assessing item-domain congruence. Since the judges were only briefly exposed to the notion of the two types of elaboration, it is not clear how informed their decisions were regarding types of elaboration. Nonetheless, it is clear that the judges were satisfied with the type of elaboration they used. It appears that the time and expense of creating highly elaborated domains may not be worthwhile. There still might be value in using highly elaborated domains to constrain item-writers to write better and more homogeneous sets of items. However, it does not appear that the use of highly elaborated domains is needed in judging item-domain congruence. Indeed, judging by the comments of some judges after completing the questionnaire, high elaboration may only serve to try the patience of the judges.

Because an elaboration effect was not found with the three judgmental techniques and because the examinee response data could not be used for its intended purpose of confirming the domain structure suggested by the judge group ratings, the first research question was amended to: How do the judgmental techniques compare with one another? The comparison of the techniques was based on the correlations among the techniques and on the number of items assigned to domains by the ratings on the judgmental techniques.

In both curriculum areas, the three judgmental techniques with combined high and low elaboration were strongly correlated with one another. The CARS and the CMP were the two most strongly correlated

techniques for both geography and physical education. The IOCI, while similar, does function somewhat differently from the CARS and CMP, possibly because the ratings on other domains are taken into account in the calculation of the IOCI. The three techniques were also strongly correlated when compared by domains.

Rovinelli (1976), working in the context of objectives rather than domains, also noted this tendency of the IOCI to function differently and also attributed the difference to the fact that the ratings on other objectives were taken into account. The cut-off value used by Rovinelli with the IOCI was .70. Had this value been used in the present study, only 5 geography items and 8 physical education items would have been assigned to their intended domains using the IOCI. No items would have been assigned to nonintended domains. This would have been the case whether the approach of using only cut-off values or the approach of using cut-off values in combination with the highest value was used. The cut-off value used by Rovinelli with the CMP was 70. Had this value been used in the present study, using the approach of only cut-off values, 31 geography items and 24 physical education items would have been assigned to their intended domains while 14 geography and 19 physical education items would have been assigned to nonintended domains. When the approach of cut-off values in combination with the highest value was used, 28 geography items and 22 physical education items would have been assigned to their intended domains while 4 geography and 8 physical education items would have been assigned to nonintended

domains. Rovinelli's use of the SD statistic with the ARS to declare items aberrant does not allow for a comparison with the present study in which the judge group ratings were used. Thus, based on the cut-off values used by Rovinelli, in the present study the IOCI has functioned differently from the CMP.

Although Rovinelli preferred the IOCI, he felt that it was time-consuming to administer and therefore stated that the test constructor would need to take into consideration the information desired and the resources available before choosing a judgmental technique. In the present study, the tendency of the IOCI to assign items to only one domain was apparent. With the CARS and the CMP, the assignment of items tends to be on multiple domains when only cut-off values are used. Rovinelli also found that the ARS indicated items to be measures of objectives other than the intended ones more often than when the IOCI was used. Although this tendency decreased as higher cut-off values were used, if having items measure only one domain is a requirement, Rovinelli's recommendation of the IOCI is supported in the present study if only cut-off values are used.

When only cut-off values were used with the CARS and CMP a dilemma was created. If it is preferred that the maximum number of items be assigned to intended domains and there is little concern for the number of items assigned to nonintended domains, then low cut-off values should be used. If it is preferred that items not be

assigned to multiple domains, high cut-off values should be used. However, fewer items would then be assigned to their intended domain. To circumvent this dilemma, the approach of assigning an item to the domain on which it received its highest rating in combination with cut-off values appears to be a reasonable compromise. Using this approach, fewer items were assigned to nonintended domains than when only cut-off values were used.

When the best technique was determined by the number of items receiving their highest rating on their intended domain minus the number of items assigned to nonintended domains, then the IOCI would be considered the best technique. The CARS and CMP were equal in one curriculum area and in the other curriculum area the CARS was second best and the CMP the worst technique. These findings are in keeping with Rovinelli's (1976) opinion that the IOCI was the preferred technique but that the ARS and CMP were also useful when items might measure more than one domain.

In comparison with the analysis of the judgmental group ratings, the opinion of the judges was that the order in which the techniques allowed the best judgment was: ARS, IOCI, and the CMP. Here judges commented that the scale with most values would enable the most precise judgments to be made. The order of ease of use of the three techniques was reported to be the CMP, IOCI, and ARS. The judges commented that the decision regarding item-domain congruence was easier on a scale which required fewer choices. Obviously, the judges did not equate ease of use with best judgment. Since judges

did not find the ARS easy to use, it is probably preferable to use one of the other techniques. However, it must be kept in mind that in this study, judges used two ARS and their reaction to the ARS might be based on an objection to using two scales which many judges reported as redundant. The correlations between the two scales of .969 and above support the opinions of the judges. It might be that the use of one 5-point scale would be viewed differently from the use of two 5-point scales. Nonetheless, it does not appear to be necessary to use more than one ARS.

Thus, while the IOCI is the preferred technique, the three judgmental techniques are very similar to one another and therefore the approach used to assign items to domains does as preferred by Rovinelli (1976) depend on the information desired and the resources available. As well, the opinion of the judges is a factor which should also be considered in the selection of a judgmental technique.

In the next chapter, the conclusions are drawn and suggestions are made for further research.

Chapter 6

Conclusions and Suggestions for Further Research

In this research, information has been gathered regarding the use of three judgmental techniques and the effect of domain elaboration on the ratings of judges using those techniques.

The value of criterion-referenced tests in establishing the skills which a student has acquired and the skills in which instruction is required has been immeasurable. That value rests to a large extent on the validity of test items in measuring behavioral domains. Judgmental techniques are an encouraging method for establishing item-domain congruence.

The belief that item-domain congruence will be higher when domains are described in high elaboration was not supported in the present study in which the IOCI, CARS, and CMP were used to rate item-domain congruence. The judges who used the three techniques felt that the type of elaboration used by them was sufficient for judging item-domain congruence. The use of highly elaborated domains does not appear necessary for establishing item-domain congruence. However, further research will be necessary to determine whether the IOCI is affected by domain elaboration. In addition, different types of elaboration, perhaps even more detailed than amplified objectives, may lead to the detection of differences not found in this research. Nonetheless, if the only outcome of the use of high elaboration is higher values, users of the judgmental

techniques could simply lower their cut-off scores and use low elaboration domain descriptions.

The IOCI is the preferred technique because of its tendency not to assign items to multiple domains. However, the three judgmental techniques used in this study are highly correlated with one another. Therefore, one might wish to take the judges' opinions that the CMP was easiest to use and that the ARS enabled the best judgment to be made into account in selecting a technique. No advantage to using more than one ARS was found.

The approach of assigning an item to the domain on which it receives its highest rating then using cut-off values to finalize the assignment of items to domains allows for most items to be assigned to their intended domain while minimizing the number of items assigned to nonintended domains.

The notion that curriculum experts must act as judges may be spurious. Had practicing teachers been used as judges, they might have attempted to predict student performance rather than consider the relationship between items and domains.

In this research, item-domain congruence was studied in only two curriculum areas: geography and physical education. Both these curriculum areas can be considered skill application areas. By studying the use of the techniques with basic skill areas the findings might be confirmed and expanded. The study was also limited to one level of Bloom's taxonomy. It would be useful to have other

studies using items and domains at other levels of the taxonomy.

In this study, the link between the domain structure suggested by the judge group ratings and the domain structure suggested by the analysis of examinee response data could not be established. In any future study of judgmental techniques using examinee response data, the effect of instruction on the responses of examinees should be controlled and studied in a variety of situations. For example, a comparison might be undertaken of the domain structure suggested by examinee response data from uninstructed, possibly instructed, presently instructed, and past instructed students. In addition, in future studies of criterion-referenced test items, it would be well to find students who had been instructed in a manner theoretically in keeping with criterion-referenced measurement.

This research study was conducted within the framework of the belief that items should be measures of only one domain. If one does not wish to accept that assumption, then the IOCI technique would not be the preferred technique as it takes into account the ratings on other domains. In addition, once the belief that items should measure only one domain is set aside, the interesting possibility exists of studying patterns of judge group ratings for items on a number of domains of interest. Such patterns might be very useful to educational diagnosticians and to classroom teachers in diagnosing deficiencies and in planning individualized educational programs.

As well, an issue which needs clarification is whether domains are meant to reflect subskills which cluster in the learning of students or subskills which make a convenient unit of teaching. Clarification of this issue would contribute to understanding the link between domain structures as established by judgmental ratings and as established by analysis of examinee response data.

Further, an investigation of the appropriate characteristics of judges of item-domain congruence is needed. A comparative study of judges from differing backgrounds and levels of expertise would be useful in establishing who can act as judges of item-domain congruence.

A further study of the use that judges of differing backgrounds make of domain elaboration would add to the understanding of the importance of domain elaboration in establishing item-domain congruence.

Information about judgmental techniques has begun to accumulate. It would be helpful if a way could be found to share this information widely. Such useful information should not be sequestered in isolated test development efforts but should be made available to a larger audience.

References

- Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: MacMillan.
- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- American Psychological Association, (1985). *Standards for educational and psychological testing*. Washington, DC: Author.
- Berk, R. (1978) *A consumer's guide to criterion-referenced test item statistics*. Paper presented at the annual meeting of NCME, Toronto, ON.
- Berk, R. (Ed.). (1980a). *Criterion-referenced measurement: The state of the art*. Baltimore, MD: Johns Hopkins University Press.
- Berk, R. (1980b). Item analysis. In R.A. Berk (Ed.), *Criterion-referenced measurement: the state of the art*. Baltimore, MD: Johns Hopkins University Press.
- Berk, R. (1984). *A guide to criterion-referenced test construction*. Baltimore: MD: Johns Hopkins University Press.
- Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: Longmans & Green.
- Cronbach, L. (1970). *Essentials of psychological testing*. (3rd ed.). New York: Harper & Row.
- Cronbach, L. (1971). Test validation. In R.L. Thorndike (Ed.) *Educational measurement*, (2nd ed.). Washington, DC: American Council on Education.
- Dahl, T. (1971). *Toward an evaluative methodology for criterion-referenced measures: Objective-item congruence*. Paper presented at the annual meeting of the California Educational Research Association, San Diego, CA.
- Education Amendment Act, c.61 Ontario Legislative Assembly (1980).
- Ebel, R. (1962). Content standard test scores. *Psychological Measurement*, 22, 15-25.

- Fitzpatrick, A. (1983). The meaning of content validity. *Applied Psychological Measurement*, 7, 3-13.
- Gerritz, K. (1984). *Assessing the dimensionality of NAEP reading test items: Expert judgment approach*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Girard, R., & Cliff, N. (1973). A comparison of methods for judging the similarity of personality inventory items. *Multivariate Behavioral Research*, 8, 71-88.
- Glasser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519-21.
- Gronlund, N. (1976). *Measurement and evaluation in teaching*. (3rd ed.). New York: MacMillan.
- Guttman, L., & Schlesinger, I. (1966). Development of diagnostic and mechanical ability tests through facet design and analysis. *Cooperative Research Project* (N OE-4-21-014). Jerusalem: Israel Institute of Applied Social Research.
- Guttman, L., & Schlesinger, I. (1967) Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement*, 27, 159-170.
- Haladyna, T., & Roid, G. (1981). The role of instructional sensitivity in the empirical review of criterion-referenced test items. *Journal of Educational Measurement*, 18, 39-53.
- Haladyna, T., & Roid, G. (1983a). A comparison of two approaches to criterion-referenced test construction. *Journal of Educational Measurement*, 20, 271-281.
- Haladyna, T., & Roid, G. (1983b). Reviewing criterion-referenced test items. *Educational Technology*, 23, (8), 35-38.
- Hambleton, R. (1980). Test score validity and standard-setting methods. In R.A. Berk (Ed.), *Criterion-referenced measurement: the state of the art*. Baltimore, MD: Johns Hopkins University Press.

- Hambleton, R. (1982). Advances in criterion-referenced testing technology. In C. Reynolds & T. Gutkin (Eds.), *The handbook of school psychology*. New York: Wiley.
- Hambleton, R. (1984). Validating the test scores. In R. Berk (Ed.), *A guide to criterion-referenced test construction*. Baltimore: MD: Johns Hopkins University Press.
- Hambleton, R., & Eignor, D. (1978). *A practitioner's guide to criterion-referenced test development, validation, and test score usage* (Research Report No. 70). Amherst, MA: University of Massachusetts Laboratory of Psychometric and Evaluative Research.
- Hambleton, R., Swaminathan, H., Algina, J., & Coulson, D. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research*, 48, 1-47.
- Harris, D. (1983). *Item selection for mastery tests: A comparison of three procedures*. Unpublished doctoral dissertation, University of Wisconsin-Madison.
- Hemphill, J., & Westie, C. (1950). The measurement of group dimensions. *Journal of Psychology*, 29, 325-342.
- Hively, W., Maxwell, G., Rabehl, G., Sension, D., & Lundin, S. (1973). Domain-referenced curriculum evaluation: A technical handbook and a case study from the Minnemast Project. *Monograph Series in Evaluation, No. 1*. Los Angeles: Center for the Study of Evaluation, University of California.
- Hively, W., Patterson, H., & Page, S. (1968). A universe-defined system of arithmetic achievement tests. *Journal of Educational Measurement*, 5, 275-290.
- Linn, R. (1980). Issues of validity for criterion-referenced measures. *Applied Psychological Measurement*, 4, 547-561.
- Livingston, S. (1977). Psychometric techniques for criterion-referenced testing and behavioral assessment. In J. Cone & R. Hawkins (Eds.), *Behavioral assessment: new directions in clinical psychology*. New York: Bruner/Mazel.
- Millman, J. (1974). Criterion-referenced measurement. In W.J. Popham (Ed.), *Evaluation in education: Current practices*. San Francisco: McCutchan Publishers.

- Ministry of Education, Province of British Columbia. (1980). *British Columbia physical education assessment: General report*. Victoria, BC: Author.
- Ministry of Education, Province of Ontario. (1982). *Ontario assessment instrument pool: Geography (Canada): Intermediate division*. Toronto, ON: Author.
- Nitko, A. (1980). Distinguishing the many varieties of criterion-referenced tests. *Review of Educational Research*, 50, 461-485.
- Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.
- Popham, W. (1971). Indices of adequacy for criterion-referenced test items. In W.J. Popham (Ed.), *Criterion-referenced measurement*. Englewood Cliffs, NJ: Educational Technology Publications.
- Popham, W.J. (Ed.) (1978). *Criterion-referenced measurement: An introduction*. Englewood Cliffs, NJ: Prentice-Hall.
- Popham, W. (1980). Domain specification strategies. In R. Berk (Ed.), *Criterion-referenced measurement: The state of the art*. Baltimore, MD: Johns Hopkins University Press.
- Popham, W. (1981). *Modern educational measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Popham, W. (1987). Two-plus decades of educational objectives. *International Journal of Educational Evaluation*, 11, (1), 31-41.
- Popham, W., & Husek, T. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, 6, 1-9.
- Roid, G., & Haladyna, T. (1980). The emergence of an item-writing technology. *Review of Educational Research*, 50, 293-314.
- Roid, G., & Haladyna, T. (1982). *A technology for test-item writing*. New York: Academic Press.
- Rovinelli, R.. (1976). *Methods of validating criterion-referenced test items*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Rovinelli, R., & Hambleton, R. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal for Educational Research*, 2, 49-60.

- Ryan, J. (1968). Teacher judgments of test item properties. *Journal of Educational Measurement*, 5, 301-306.
- Secolsky, C. (1983). Using examinee judgments for detecting invalid items on teacher-made criterion-referenced tests. *Journal of Educational Measurement*, 20, 51-63.
- Tabachnick, B., & Fidell, L. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Thorndike, R., & Hagen, E. (1977). *Measurement and evaluation in psychology and education* (4th ed.). New York: Wiley.
- Traub, R. (1972). *Criterion-referenced measurement: Something old and something new*. Paper presented at the University of Victoria.
- Zwarts, M.A. (1982). On the construction and validation of domain-referenced measurements. *Evaluation in Education*, 5, 119-139.

Appendix A

Domain Descriptions

GEOGRAPHY DOMAIN 1 (CC)

DOMAIN DESCRIPTION

Students will demonstrate recognition of terms describing the physical environment.

SAMPLE ITEM

Relief is defined as

- A. mountainous areas.
- B. the vertical distance above sea level.
- C. an area of valleys and lowlands.
- * D. the differences in elevation of parts of the earth's surfaces.

STIMULUS ATTRIBUTES

1. Items will involve the recognition of at least one term used to describe the physical environment.
2. Terms describing the physical environment will be similar to those in the list in the Specification Supplement.
3. The stem of the item will consist of 30 words or fewer.

RESPONSE ATTRIBUTES

1. Multiple-choice questions with four alternatives will be used.
2. Distractors will consist of physical environment terms or descriptions of terms.
3. Recognition of the term will be demonstrated by either
 - (a) selecting a correct definition(s) for the term(s), or
 - (b) selecting a correct term(s) given a description(s), or
 - (c) selecting correct descriptive terms for a geographical region.

SPECIFICATION SUPPLEMENT

Some examples of descriptive terms for the physical environment are: altitude, plateau, tree-line.

GEOGRAPHY DOMAIN 2 (CK)

DOMAIN DESCRIPTION

Students will recognize urban centers and their characteristics and the factors accounting for urbanization.

SAMPLE ITEM

Which of these cities is not a provincial capital?

- A. St. John's
- B. Fredericton
- * C. Montreal
- D. Victoria

STIMULUS ATTRIBUTES

1. Each item will involve one of the following:
 - (a) factor(s) accounting for urbanization.
 - (b) characteristic(s) of urban centers.
 - (c) names of urban centers.
 - (d) locations of urban centers.
2. The stem of the item will consist of 30 words or less.
3. If negative terms are used in the stem, they will be underlined.

RESPONSE ATTRIBUTES

1. Multiple-choice items with four alternatives will be used.
2. Distractors will consist of parallel terminology.
3. Negative terms will not be used in the choices, e. g., "not,", "never", "no".
4. Choices may include "all of the above" and/or "none of the above".

SPECIFICATION SUPPLEMENT

Examples of factors which may be considered as affecting urbanization are: urban centres and ports.

GEOGRAPHY DOMAIN 3 (CF)

DOMAIN DESCRIPTION

Students will demonstrate knowledge of the availability, renewability and usability of natural resources.

SAMPLE ITEM

You want to buy a large farm and specialize in growing wheat. In which of the following regions would you most likely find a suitable farm?

- A. the Canadian Shield.
- * B. the Interior Plains.
- C. the Great Lakes - St. Lawrence Lowlands.
- D. the Appalachian Region.

STIMULUS ATTRIBUTES

1. A resource will be defined as something which is used by people.
2. The stem of the item will consist of 30 words or less.
3. The item will require one or more of the following:
 - (a) the location of a resource in a specific area of Canada.
 - (b) a demonstration of the knowledge that the resource is not naturally found in Canada.
 - (c) knowledge of which resources are renewable.
 - (d) knowledge that while a resource might be available, it might not be usable.
4. Items may include more than one resource.

RESPONSE ATTRIBUTES

1. Multiple-choice items with four alternatives will be used.
2. When distractors use locations or resources, they will be actual Canadian locations or resources.
3. There will be only one correct answer per item.

SPECIFICATION SUPPLEMENT

1. Both renewable and non-renewable resources may be used.
2. Examples of resources include: minerals, oil, natural gas.

GEOGRAPHY DOMAIN 4 (CI)

DOMAIN DESCRIPTION

Students will demonstrate knowledge of the population patterns which have emerged in the development of Canada.

SAMPLE ITEM

In which of the following provinces was the original pattern of population distribution largely determined by a river?

- A. Saskatchewan.
- B. Nova Scotia.
- * C. Quebec.
- D. Newfoundland.

STIMULUS ATTRIBUTES

1. Each item will involve knowledge of one or more of the following:
 - (a) the population characteristics of a region(s).
 - (b) the population characteristics of Canada.
 - (c) the location(s) of a population group(s).
 - (d) changes in population for a region(s).
 - (e) factors affecting population distribution.
2. The stem of the item will consist of 25 words or fewer.
3. Only population groups found in Canada are eligible for inclusion in an item.

RESPONSE ATTRIBUTES

1. Multiple-choice items with four alternatives will be used.
2. Each alternative will consist of not more than three provinces or areas or three dates or numbers.
3. No fictitious names will be used.
4. There will be only one correct answer.

SPECIFICATION SUPPLEMENT

Examples of factors accounting for the population patterns in Canada are: employment opportunities and founding groups.

PHYSICAL EDUCATION DOMAIN 1 (RU)

DOMAIN DESCRIPTION

Students will demonstrate knowledge of the rules, techniques, strategies and movement principles which are applicable to sport or physical activities.

SAMPLE ITEM

If you are throwing a ball with your right hand, you should:

- * A. step forward on your left foot.
- B. step backward with your left foot.
- C. step forward on your right foot.
- D. spread the weight evenly on both feet.
- E. I don't know.

STIMULUS ATTRIBUTES

1. Items will involve a physical activity or sport.
2. Items will assess knowledge of rules, techniques, strategies and movement principles as demonstrated by one or more of the following:
 - (a) identifying principles which will accomplish a specific movement.
 - (b) recall of rules governing a particular sport(s).
 - (c) recognition of the proper performance techniques for physical activities or sports.
 - (d) appreciation of appropriate strategies for use in physical activities or sports.
3. The form of the stem may consist of either a question or a partial statement to be completed by one of the choices.
4. The stem of the item will consist of 25 words or fewer.

RESPONSE ATTRIBUTES

1. Multiple-choice items with five alternatives will be used.
2. The fifth choice will be "I don't know."
3. Choices may include "all of the above."
4. Distractors will not use:
 - (a) movements which are not physically possible.
 - (b) equipment which is not part of the specified sport or activity(ies).
 - (c) players and positions which are not part of the specified sport or activity(ies).
 - (d) terminology which is not usually associated with the sport or activity(ies).

SPECIFICATION SUPPLEMENT

Examples of physical activities are: basketball, running, badminton.

PHYSICAL EDUCATION DOMAIN 2 (PC)

DOMAIN DESCRIPTION

Students will demonstrate knowledge of the interrelationships among the factors which affect physiological efficiency.

SAMPLE ITEM

It is important to rest after long hard exercise because the:

- A. muscles need to rebuild tissue.
- B. lungs need to rest.
- C. nerves need to rest.
- * D. body needs to rebuild its supply of energy.
- E. I don't know.

STIMULUS ATTRIBUTES

1. Items will include knowledge of one or more of the following:
 - (a) the effects of substances which influence physiological efficiency.
 - (b) the effect of poor physical fitness.
 - (c) the effect of a physical activity(ies) on physiological functioning.
 - (d) the effect of training on physiological functioning.
2. The stem of the item will consist of a question or an unfinished statement.
3. The stem of the item will consist of 30 words or fewer.
4. The stem may include more than one sentence.

RESPONSE ATTRIBUTES

1. Multiple-choice items with five alternatives will be used.
2. The fifth choice will be "I don't know."
3. Where choices refer to physical activities either individual or group activities may be used.
4. Choices may include terms describing one or more of the following:
 - (a) part(s) of the anatomy and/or anatomical system.
 - (b) an anatomical/physiological path necessary for the correct functioning of an anatomical/physiological system.
 - (c) anatomical/physiological responses to changes in development.
 - (d) factors which effect physiological efficiency.

SPECIFICATION SUPPLEMENT

1. An example of physiological functioning is breathing.
2. Examples of physical activities are swimming and soccer.
3. An example of a factor affecting physical activities is substance use.

PHYSICAL EDUCATION DOMAIN 3 (BS)

DOMAIN DESCRIPTION

Students will demonstrate knowledge of the development, structure, function and influences of human anatomical systems.

SAMPLE ITEM

Messages from your brain to the muscles in your arm are carried by:

- A. veins.
- B. skin.
- C. bones.
- * D. nerves.
- E. I don't know.

STIMULUS ATTRIBUTES

1. Items will involve recognition and/or recall of information regarding human anatomical systems.
2. If negative terms are used they will be underlined.
3. Items may include reference to more than one body system.
4. The stem of the item may be a question or a statement to be completed.

RESPONSE ATTRIBUTES

1. Knowledge of the development, structure, function and influences of human anatomical systems will be demonstrated by selecting one of the following:
 - (a) the correct part of the anatomy and/or anatomical system necessary for a given body function.
 - (b) correct anatomical responses to changes in development.
 - (c) physiological responses to changes in the performance of an anatomical system.
 - (d) the correct path of function of a human anatomical system.
2. Multiple-choice items with five alternatives will be used.
3. The fifth choice will be "I don't know."
4. Choices may include "all of the above" and/or "none of the above."

SPECIFICATION SUPPLEMENT

Examples of body systems include the circulatory, digestive, endocrine, muscular, and nervous systems.

PHYSICAL EDUCATION DOMAIN 4 (FN)

DOMAIN DESCRIPTION

The student will demonstrate knowledge of factors relating to physical fitness and nutrition.

SAMPLE ITEM

Which one of the following is the best way to maintain normal weight?

- * A. Balancing the amount of food eaten and amount of exercise.
- B. Playing actively for two hours a day.
- C. Eating between meals but less at meal time.
- D. Eating a large amount at mealtimes but nothing between meals.
- E. I don't know.

STIMULUS ATTRIBUTES

1. The components of physical fitness will be defined as muscular strength, muscular endurance, cardiorespiratory endurance, flexibility and body structure.
2. Items will consist of questions or unfinished statements.
3. Items will cover knowledge of one or more of the following topics:
 - (a) principles and/or activities used in developing, maintaining and monitoring physical fitness.
 - (b) the effect of dietary or nutritional compounds to the body.
 - (c) identification of activities that are predominantly aerobic or anaerobic.

RESPONSE ATTRIBUTES

1. Multiple-choice items with five alternatives will be used.
2. The fifth choice will be "I don't know".
3. Choices may include terms relating to:
 - (a) physical activities.
 - (b) practices/tests used in fitness training.
 - (c) the effects of practices/tests used in fitness training.
 - (d) body processes.
 - (e) diet/nutrition.

SPECIFICATION SUPPLEMENT

An example of a physical activity is sit-up practice.

An example of a nutritional/dietary compound is protein.

An example of a principle in developing physical fitness is "overload."

Appendix B

Geography and Physical Education Items

GEOGRAPHY ITEMS

Domain 1 (CC)

- CC02 Altitude is defined as
- * A. the vertical distance above sea level.
 - B. degrees north of the equator.
 - C. the elevation of a mountain.
 - D. the horizontal distance above sea level.
- CC03 A continental shelf is
- A. the abrupt edge of a high continental plateau.
 - B. the drainage divide of rivers flowing to oceans on opposite sides of a continent.
 - * C. the extension of a continent covered by oceans to a depth of 100 fathoms or less.
 - D. a legendary continent sometimes called Atlantis.
- CC05 Which one of the following series of words best describes the Appalachian Region?
- * A. worn down, old, rounded mountains.
 - B. high, young, peaked mountains.
 - C. coastal plain, muskeg.
 - D. ice-covered, flat lowlands.
- CC06 The Canadian Shield is composed primarily of
- A. horizontal layers of young sedimentary rock.
 - B. quite young folded sedimentary rocks.
 - * C. very old igneous and metamorphic rocks like granite.
 - D. old worn down sedimentary layers like sandstone.
- CC07 The Interior Plains of Canada are composed primarily of
- A. folded and faulted rocks of all types.
 - B. igneous and metamorphic intrusions.
 - * C. horizontal layers of sedimentary rock.
 - D. volcanic rock.
- CC58 The tree-line refers to the
- A. northern limit of a particular species of tree.
 - * B. approximate northern limit of tree growth.
 - C. line up to which the land has been cleared for agriculture and beyond which it is still in forest.
 - D. blaze marks used by a logging survey crew to identify trees for cutting.

- CC61 Which one of the following phrases correctly defines the term soil profile?
- A. a complete analysis of the chemistry and physical processes of a particular soil.
 - * B. a cross-section of a soil which allows one to observe the various layers or horizons of the soil.
 - C. a kind of "fingerprint" each type of soil has which allows it to be identified.
 - D. a map showing the distribution of different kinds of soils occurring over an area.

- CC62 Humus is best defined as the
- * A. organic part of the soil derived from the decaying plant and animal life.
 - B. inorganic part of soil derived from the breaking down of rock into small particles.
 - C. upper layers of the earth's surface including soil and any loose parent material.
 - D. dead leaves and grass lying on top of the soil.

- CC63 The term muskeg refers to
- * A. swampy areas containing much organic material.
 - B. permanently frozen ground.
 - C. rock covered with a thin layer of soil.
 - D. a soil of low fertility.

Domain 2 (CK)

- CK03 The three cities in Canada with the largest populations are all
- A. provincial capitals.
 - B. steel producing centres.
 - * C. ports.
 - D. auto manufacturing centres.
- CK14 Today, the percentage of Canadians living in urban centres is approximately
- A. 25%.
 - B. 48%.
 - * C. 60%.
 - D. 77%.
- CK15 The province with the greatest percentage of its population living in urban areas today is
- A. Prince Edward Island.
 - B. Quebec.
 - * C. Ontario.
 - D. British Columbia.

- CK16 Which one of the following statements best describes the relationship between population distribution and the concentration of manufacturing?
- A. Where manufacturing concentration is high, urban population is low.
 - B. Where manufacturing concentration is low, urban population is high.
 - * C. Where manufacturing concentration is high, urban population is high.
 - D. There is little correlation between urban population concentration and the concentration of manufacturing.
- CK17 Which one of the following factors best explains why the Maritime region of Canada has not developed a large urban population based on industry?
- A. Many of the young people are leaving the Maritimes.
 - B. The soils are too poor to support a large population.
 - * C. The area is too far from major markets.
 - D. Transportation in the area is difficult.
- CK18 In Canada, the areas that are undergoing the most rapid changes in land use are areas
- A. where soils are poor.
 - * B. on the fringes of urban centres.
 - C. of extensive beef grazing.
 - D. of grain production.
- CK19 The highest proportion of the land in urban areas is used for
- * A. residential purposes.
 - B. industrial purposes.
 - C. commercial purposes.
 - D. institutional purposes.
- CK20 Which one of these cities is not a steel producing centre?
- A. Sydney, Nova Scotia.
 - B. Hamilton, Ontario.
 - C. Sault Ste. Marie, Ontario.
 - * D. Thunder Bay, Ontario.
- CK21 Which one of these cities is not an automobile producing centre?
- A. Oakville.
 - * B. Sudbury.
 - C. Oshawa.
 - D. Windsor.

Domain 3 (CF)

- CF02 If you were exploring for oil in Canada you would most likely be successful if you chose to drill in the
- A. Western Cordillera.
 - * B. Interior Plains.
 - C. Canadian Shield .
 - D. Hudson Bay Lowlands.
- CF03 Most of the natural gas used in Ontario comes from
- A. Nova Scotia.
 - B. Saskatchewan.
 - * C. Alberta.
 - D. United States.
- CF04 A prospector looking for metallic minerals in Canada would most likely be successful if he explored the
- * A. Canadian Shield.
 - B. Interior Plains.
 - C. Great Lakes-St. Lawrence Lowlands.
 - D. Appalachian Region.
- CF05 Which one of the following resources for the steel industry in the Great Lakes - St. Lawrence Lowlands must be imported?
- * A. coal.
 - B. water.
 - C. scrap metal.
 - D. limestone.
- CF06 Which set of manufactured goods listed below is made largely from renewable resources?
- A. gasoline, cod liver oil.
 - B. wool sweater, refrigerator.
 - C. dinner knives, books.
 - * D. paper, cotton, T-shirts.
- CF13 Which one of the following forest zones is noted for the production of lumber rather than pulp and paper?
- A. coniferous forest of Northern Ontario.
 - * B. west coast forest of British Columbia.
 - C. mixed forest of Southern Ontario.
 - D. Acadian forest of the Atlantic Provinces.
- CF17 The term growing season refers to
- A. the days when there is no frost.
 - * B. the period during which plants will grow.
 - C. the period during which the daily temperatures go above 20° C.
 - D. the heat available for plant growth.

- CF18 The growing season can be longer than the frost-free period because
- A. soil fertility determines the length of the growing season.
 - * B. many plants can tolerate some frost and still grow.
 - C. the growing season is determined by precipitation rather than temperatures.
 - D. higher summer temperatures eliminate the effect of frosts.

- CF19 In an arid area which of the following improvements would be most useful in increasing agricultural productivity?
- * A. spending more money on irrigation.
 - B. building a good road network.
 - C. putting in tiles and ditches to improve drainage.
 - D. improving local storage of crops.

Domain 4 (CI)

- CI01 In 1981, the population of Canada was approximately
- A. 22 000 000.
 - * B. 24 000 000.
 - C. 26 000 000.
 - D. 28 000 000.
- CI02 In 1981, the province with the largest population was
- * A. Ontario.
 - B. Quebec.
 - C. British Columbia.
 - D. Alberta.
- CI03 Which one of the following Canadian provinces is approximately 37% French-speaking?
- A. Newfoundland.
 - * B. New Brunswick.
 - C. Manitoba.
 - D. Ontario.
- CI05 Of the following, the province or territory that is home to the largest number of Inuit people is
- A. Quebec.
 - B. Newfoundland.
 - C. Yukon Territory.
 - * D. The Northwest Territories.

- CI07 In which of the following regions of Canada is the population evenly distributed but of low density?
- A. Cape Breton Island.
 - * B. Southern Saskatchewan.
 - C. Fraser Delta.
 - D. Golden Horseshoe.
- CI08 Which of the following areas in Canada has the greatest concentration of population?
- A. between Vancouver and Hope, British Columbia.
 - B. between Calgary and Edmonton, Alberta.
 - * C. between Sarnia and Toronto, Ontario.
 - D. between Montreal and Quebec City, Quebec.
- CI10 From 1975 to 1980, the province which gained the most people through interprovincial migration was
- A. Newfoundland.
 - * B. Alberta.
 - C. Saskatchewan.
 - D. British Columbia.
- CI11 Of the following Canadian provinces, the one which has the highest percentage of its labour force employed in manufacturing is
- A. Saskatchewan.
 - B. Newfoundland.
 - * C. Ontario.
 - D. Nova Scotia.
- CI12 In which region of Canada is unemployment usually higher than in the rest of the country?
- * A. Atlantic Provinces.
 - B. Quebec.
 - C. Prairie Provinces.
 - D. British Columbia.

Domain 5 (Other)

- CB08 The most northerly point of Canada is approximately
- A. 68° N.
 - B. 73° N.
 - C. 78° N.
 - * D. 83° N.

- CB09 The 49th parallel of latitude forms the boundary between the United States and
- A. Atlantic Canada.
 - B. Quebec.
 - * C. the Western Provinces.
 - D. the Yukon Territory.
- CB10 The 45th parallel of latitude forms part of the boundary between
- A. the Eastern Townships of Quebec and Nova Scotia.
 - * B. the Eastern Townships of Quebec and the United States.
 - C. Nova Scotia and the United States.
 - D. the Western Provinces and the United States.
- CJ01 The Canadian Pacific Railway was built in the late 1800's as a transcontinental line. Its main purpose was to
- A. move people and goods north and south in Canada.
 - B. join American rail lines for trade with the United States.
 - C. export manufactured goods from the prairies.
 - * D. join together the pockets of settlement from east to west across the country.
- CJ02 One major difficulty in building the transcontinental railway line in Canada was
- * A. the north-south trend of the mountains in British Columbia.
 - B. crossing the Prairies.
 - C. getting "right-of-way" allowance through highly populated areas.
 - D. competition from other railways.
- CJ12 In Canada, water transportation is used primarily to move
- A. manufactured goods and agricultural commodities.
 - B. petroleum and petroleum products.
 - C. fruit and perishable commodities.
 - * D. bulky goods and raw materials.
- CL01 In geography, the term region is generally used to describe an area having
- A. one central governing body.
 - B. a large concentration of manufacturing activity.
 - * C. enough similarities to be considered as one unit.
 - D. similar types of industries.

CL09 Coke is obtained from

- A. slag.
- B. pig iron.
- C. iron ore.
- * D. coal.

CL13 Which one of the following situations has resulted from the improvements in technology that have occurred over the last fifty years?

- A. People exert more physical effort on their jobs.
- * B. People have more leisure time.
- C. People work longer hours.
- D. People need less education.

PHYSICAL EDUCATION ITEMS

Domain 1 (RU)

- RU01 What should you do when you run?
- A. Lean slightly backwards.
 - * B. Move the arms alternately backwards and forwards.
 - C. Land on the heels of the feet.
 - D. Increase trunk rotation.
 - E. I don't know.
- RU08 If a basketball team is playing a man-to-man system of defense, each member of the team should:
- * A. stay between the designated opponent and the basket when the opponents are attacking.
 - B. use the position of the ball to determine the best defensive position on the court.
 - C. stay with the designated opponent at all times, wherever he is on the court.
 - D. stay with the designated opponent only if he is near the ball.
 - E. I don't know.
- RU14 If a person is attempting to jump as high as possible, which of the following techniques will increase the height of the jump?
- A. Keep the arms beside the body when leaving the floor.
 - B. Lean forward with the trunk after leaving the floor.
 - * C. Bend the knees before leaving the floor.
 - D. Bend the knees after leaving the floor.
 - E. I don't know.
- RU17 If you are passing the ball to a team mate who is running forward, where should the ball be aimed?
- A. At the hands of the receiver.
 - * B. Ahead of the receiver.
 - C. At the eye level of the receiver.
 - D. At the waist of the receiver.
 - E. I don't know.
- RU21 Most team sports involve movement into open spaces. The purpose of moving into an open space is:
- A. to cause an opponent to move with you.
 - B. to make room for the player with the ball.
 - C. to avoid an opponent who remains close to you.
 - * D. all of the above.
 - E. I don't know.

- RU25 Which of the following statements is true when playing both badminton doubles and volleyball?
- A. Overhand serves are illegal.
 - B. Serves that contact the net are legal.
 - C. the serve must land within a designated service area of the court.
 - * D. It is possible for one person to serve an entire game.
 - E. I don't know.
- RU30 Which of the following situations would be most likely to result in a "point" in a badminton game?
- A. The server steps on the serving line.
 - B. The receiving side sends the bird high and deep into the opponent's court.
 - C. The receiving side sends a fast, sharply angled shot to the floor of the opponent's court.
 - * D. The serving side sends a fast, sharply angled shot to the floor of the opponent's court.
 - E. I don't know.
- RU35 When playing volleyball, which body part should be used to receive a ball that is coming at you low and hard?
- A. Palms of your hands.
 - B. Fists.
 - C. Knee.
 - * D. Forearms.
 - E. I don't know.

Domain 2 (PC)

- PC03 The performance of which of the following athletes is least likely to be affected by smoking tobacco?
- A. A swimmer.
 - B. A runner.
 - * C. A high-jumper.
 - D. A tennis player.
 - E. I don't know.
- PC06 Weak abdominal muscles will:
- A. make you less flexible.
 - * B. contribute to poor posture.
 - C. grow stronger with proper diet.
 - D. create digestive problems.
 - E. I don't know.

- PC18 Which of the following questions about a drug should be of most concern to athletes?
- A. Will the drug improve performance?
 - * B. Is the drug safe?
 - C. Will the drug relieve pain?
 - D. Will the drug delay the onset of fatigue?
 - E. I don't know.
- PC22 If the breathing rates of trained and untrained individuals were compared during a long run, what would be the result?
- A. The untrained athlete would breathe shallower and slower.
 - B. The trained athlete would breathe shallower and faster.
 - C. The untrained athlete would breathe deeper and faster.
 - * D. The trained athlete would breathe deeper and slower.
 - E. I don't know.
- PC23 If your class has started a weight training and conditioning program, which of the following results might be expected:
- A. The girls and boys in the class will develop large, bulky muscles.
 - B. The number of muscles in the body will increase.
 - * C. The muscles in the body will become stronger.
 - D. The number of muscle fibres in each muscle will increase.
 - E. I don't know.
- PC29 Which of the following gases is found in tobacco smoke and reduces physical endurance?
- * A. Carbon monoxide.
 - B. Nitrogen.
 - C. Carbon dioxide.
 - D. Sulphur dioxide.
 - E. I don't know.
- PC32 Strenuous activities will place the greatest strain on the body:
- * A. during hot humid weather.
 - B. during cold weather.
 - C. during wet weather.
 - D. at low altitudes.
 - E. I don't know.

- PC36 As a result of intensive physical activity, energy will be used up and waste products will accumulate. The muscles are then in a state of:
- A. Sensitivity.
 - * B. Fatigue.
 - C. Irritability.
 - D. Extension.
 - E. I don't know.

Domain 3 (BS)

- BS04 What is the path followed by the blood in order to get oxygen to the body cells?
- A. Heart, lungs, arteries, capillaries.
 - B. Capillaries, arteries, lungs, heart.
 - * C. Lungs, heart, arteries, capillaries.
 - D. Heart, lungs, capillaries, arteries.
 - E. I don't know.
- BS09 Fuel for muscles is provided by which of the following systems?
- A. Respiratory.
 - B. Nervous.
 - * C. Circulatory.
 - D. Digestive.
 - E. I don't know.
- BS15 Movement is the end result of action by the body's:
- A. bones.
 - B. nerves.
 - C. muscles.
 - * D. all of the above.
 - E. I don't know.
- BS20 Which of the following statements is correct?
- * A. After puberty, boys, on the average, are stronger than girls.
 - B. After puberty, boys, on the average are more flexible than girls.
 - C. Before puberty, boys, on the average, are taller than girls.
 - D. Before puberty, boys, on the average, are heavier than girls.
 - E. I don't know.

- BS24 Which of the following statements describes an efficient circulatory system?
- A. A pulse that increases rapidly during exercise.
 - * B. A pulse that returns quickly to normal after exercise.
 - C. A pulse that returns slowly to normal after exercise.
 - D. A pulse that is high all the time.
 - E. I don't know.
- BS28 Which of the following tissues is not as easily repaired by the body once it is injured?
- A. Muscle.
 - * B. Nerve.
 - C. Bone.
 - D. Skin.
 - E. I don't know.
- BS34 Which of the following is not a type of muscle tissue?
- A. Smooth.
 - B. Skeletal.
 - C. Cardiac.
 - * D. Connective.
 - E. I don't know.
- BS39 Which blood vessels carry nourishment to the heart muscle?
- A. Coronary veins.
 - B. Carotid arteries.
 - * C. Coronary arteries.
 - D. Pulmonary veins.
 - E. I don't know.

Domain 4 (FN)

- FN05 Which of the following food types requires the shortest amount of time to be digested?
- * A. Carbohydrates.
 - B. Proteins.
 - C. Fats.
 - D. There is no difference.
 - E. I don't know.
- FN10 Which of the following activities relies mainly on anaerobic energy?
- A. A 400 metre swim.
 - B. 5000 metre run.
 - C. A marathon run.
 - * D. A 60 metre sprint.
 - E. I don't know.

- FN12 When doing sit-ups, the abdominal muscles work hardest when:
- A. your legs are straight and your hands are locked behind your head.
 - * B. your knees are bent and your hands are locked behind your head.
 - C. your legs are straight and your back is arched.
 - D. your knees are bent and your back is arched.
 - E. I don't know.
- FN13 What is the best method for developing muscle strength?
- A. Exercising with light weights for many repetitions.
 - B. Exercising with light weights for a few repetitions.
 - * C. Exercising with heavy weights for a few repetitions.
 - D. Exercising with a maximum weight for one repetition.
 - E. I don't know.
- FN16 Why is protein essential for a physically active person?
- A. It speeds up digestive processes.
 - B. It helps form simple sugars.
 - C. It absorbs acids.
 - * D. It builds body tissues.
 - E. I don't know.
- FN19 In terms of building long term fitness, the average person should be most concerned with developing their:
- A. muscular system.
 - * B. aerobic energy system.
 - C. anaerobic energy system.
 - D. skilled running patterns.
 - E. I don't know.
- FN26 Why is "overloading" a muscle important in developing muscle strength?
- A. "Overloading" destroys weak muscle fibres and replaces them with stronger ones.
 - * B. Muscles grow larger and stronger only in response to progressively increasing loads.
 - C. "Overloading" stretches muscles beyond their capacity.
 - D. Waste materials do not accumulate during overload.
 - E. I don't know.

- FN31 Which of the following methods is most accurate for locating and recording the pulse?
- A. During activity, keep the thumb of one hand on the wrist of the other arm.
 - * B. After activity, press two fingers against the side of the neck.
 - C. During activity, press two fingers against the side of the neck.
 - D. After activity, keep the thumb of one hand on the wrist of the other arm.
 - E. I don't know.
- FN37 Which of the following tests is not used to measure cardiovascular endurance?
- A. 12 Minute Run.
 - B. Stationary Bicycle Test.
 - * C. The Sit-Up Test.
 - D. The Step Test.
 - E. I don't know.

Domain 5 (Other)

- BM02 What is the main purpose of the approach run in high jumping or pole vaulting?
- A. It allows time for mental preparation.
 - B. It ensures that the jump is taken from the correct foot.
 - * C. It develops momentum.
 - D. It establishes good balance.
 - E. I don't know.
- BM11 What will happen to a ball that is released with topspin?
- A. It will swerve to the right.
 - * B. It will hit the ground sooner than a ball without topspin.
 - C. It will go farther before hitting the ground than a ball without topspin.
 - D. It will go higher before hitting the ground than a ball without topspin.
 - E. I don't know.
- BM33 What is the safest way to fall?
- A. Keep your head up to see where you are going.
 - B. Put your arms out straight ahead of you.
 - C. Land on your hands and knees.
 - * D. Curl up and roll as you fall.
 - E. I don't know.

- BM38 In which position will a skater or a diver rotate, or spin, most quickly?
- A. Arms away from the body; legs together.
 - B. Arms above the head; legs apart.
 - * C. Arms close to the body; legs together.
 - D. Arms close to the body; legs apart.
 - E. I don't know.
- PS01 Which of the following statements suggests a worthwhile benefit of team sports?
- * A. The needs of the team are important as well as the needs of each person.
 - B. People can make the team even though they don't get to play.
 - C. Coaching for team sports is usually of a higher calibre than for individual sports.
 - D. Players on teams are always under a great deal of stress or anxiety.
 - E. I don't know.
- PS02 Which of the following statements best describes an important role of a good Physical Education program?
- A. It prepares students for interschool teams.
 - * B. It teaches students how to make good use of leisure time.
 - C. It prepares students for a profession in athletics.
 - D. It provides a rest from other school subjects.
 - E. I don't know.
- PS03 Which of the following is the least important thing in determining good team play?
- A. Each member appreciates the contribution of every other member.
 - * B. Players on the same team have many different characteristics.
 - C. Each team member has a positive attitude.
 - D. Good communication exists between players and coach.
 - E. I don't know.
- PS04 Your game skills will probably improve most quickly if:
- A. you win all of your games easily.
 - B. you lose all of your games by large scores.
 - * C. all of your games are close.
 - D. all of your games are against rough players.
 - E. I don't know.

PS07 If you want to help your team mates become better players, you should probably:

- A. keep reminding them of all their errors.
- * B. praise them, while pointing out their errors.
- C. be strict with them.
- D. tell them how the professionals do it.
- E. I don't know.

Appendix C
Presentation to Judges

Presentation to Judges

Introduce self. Thank professors and students for their participation.

First, I would like to provide some background about the research you will be assisting with.

[Refer to Chart 1.] In the whole area of knowledge which we know as physical education (geography) there are pockets of knowledge or skills which are called domains. Domains are well-defined sets of skills or knowledge. When we construct a test to measure a domain, we must ask, how do we know each item on a test truly measures the domain?

[Refer to Chart 2.] As you can see from this chart and from the blue sheet at the beginning of your materials, domains can be described with low elaboration, or a low amount of detail such as the example where there is only a domain description and a sample item. Domains can also be described with a high degree of elaboration. As you can see, high domain elaboration includes a domain description which is a general description of the behavior being assessed; an illustrative item; a list of stimulus attributes which are the class of stimuli the examinee will encounter; the response attributes which are the various choices; and perhaps a specification supplement which contains any extra information regarding eligible content. Other definitions which you will need to know [refer to terminology on blue reference sheet] include:

domain = a broad set of behavioral skills
domain description = a description of a set of skills
match = the correspondence between a test item and a domain
multiple choice = items which give a choice among several

answers

alternative = an answer choice
stem = the part of a test item which presents the task but does
not include the alternative choices
* = used to indicate an expected correct answer to an item

[Refer to Chart 3 and to blue sheet.] There are three
techniques for judging the match between test items and domains.

These are:

1. categorizing where you, the judge, rate the match between the item and the domain as +1, 0, or -1.
2. matching where you say 'yes' the item does match the domain or 'no' the item does not match the domain.
3. rating where you will be using two scales of 5,4,3,2,1 to indicate the match between the item and the domain.

[Refer to Chart 4.] You will be working with three sets of materials at once. You will use:

1. Domain Descriptions - which are on white paper and which are used over and over.
2. Answer sheets - which are in yellow and have instructions at the top of each sheet.
3. Items - which are on computer print-out sheets.

The Answer sheets and the items which go together are paper-clipped together and are used once and put aside. Please use the sheets in the order that you find them in your file folder. Fill in all the

yellow sheets. There will be 17 in all [chart 4] :

categorizing = 4 sheets
ratings = 8 sheets (because there are two scales)
matching = 4 sheets
personal data = 1 sheet

There is also one yellow sheet requesting some personal data and your reaction to the techniques you have used. This sheet should be filled in last.

You should be aware that some items may not belong to any domain and some items to more than one domain. Remember to work within one paper-clipped set of papers at a time and re-clip them when you're finished. In red at the top of the yellow sheets is indicated which domain to use and which item set to use. When you're finished, complete the attitudinal and personal data form. Refer to the blue reference sheet if necessary. Please work independently.

You are about to begin to work. I will come around to each person to be sure you understand what to do. If you have questions as you proceed, please give an indication and I will come to where you are to answer.

Thank you again for your cooperation. Please begin.

Chart 1

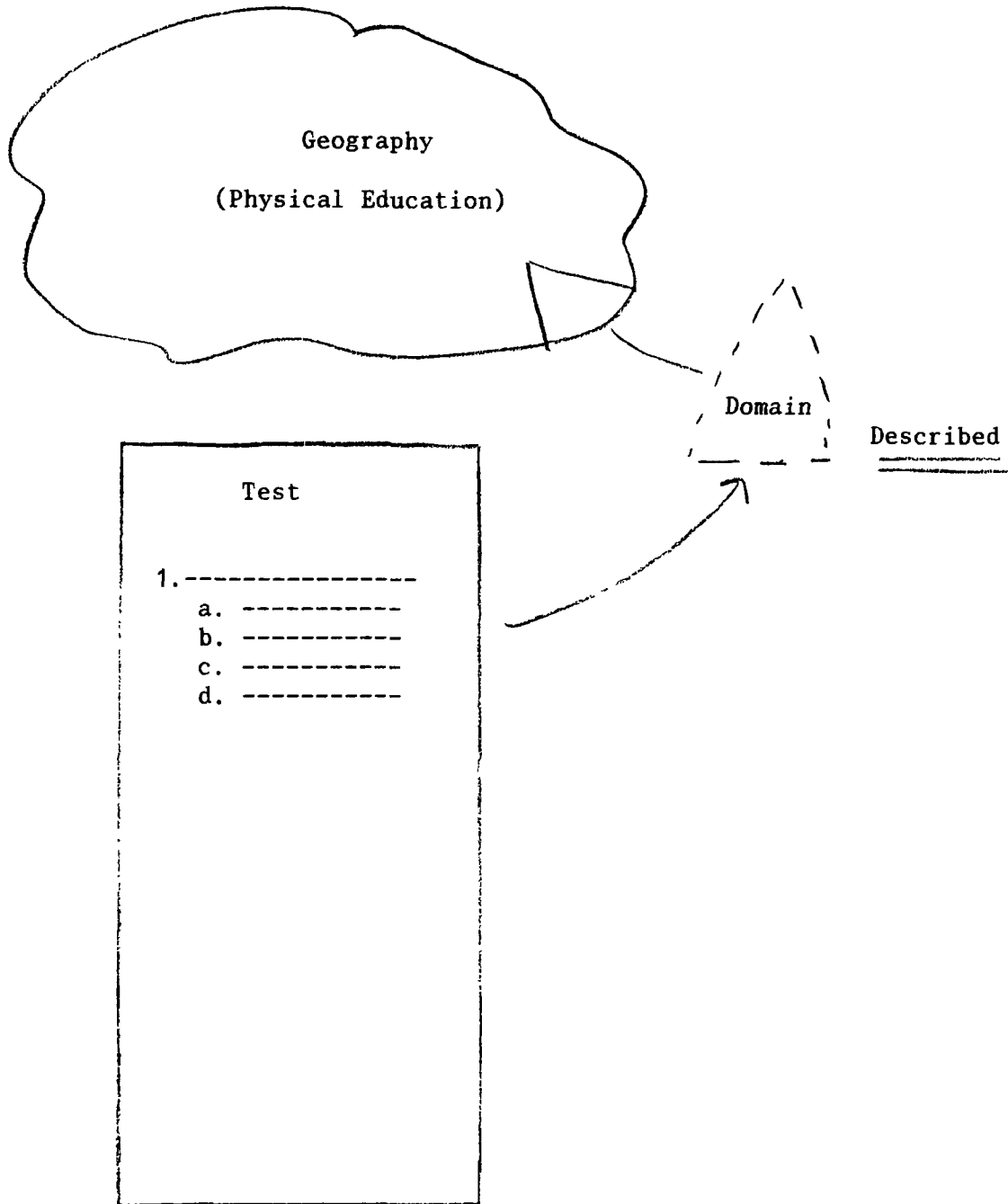


Chart 2
(For Geography Judges)

An Example of Low Domain Elaboration

DOMAIN DESCRIPTION

Students will demonstrate knowledge of the rules, techniques, strategies and movement principles which are applicable to sport or physical activities.

SAMPLE ITEM

If you are throwing a ball with your right hand, you should:

- * A. step forward on your left foot.
- B. step backward with your left foot.
- C. step forward on your right foot.
- D. spread the weight evenly on both feet.
- E. I don't know.

An Example of High Domain Elaboration

DOMAIN DESCRIPTION

Students will demonstrate knowledge of the rules, techniques, strategies and movement principles which are applicable to sport or physical activities.

SAMPLE ITEM

If you are throwing a ball with your right hand, you should:

- * A. step forward on your left foot.
- B. step backward with your left foot.
- C. step forward on your right foot.
- D. spread the weight evenly on both feet.
- E. I don't know.

STIMULUS ATTRIBUTES

1. Items will involve a physical activity or sport.
2. Items will assess knowledge of rules, techniques, strategies and movement principles as demonstrated by one or more of the following:
 - (a) identifying principles which will accomplish a specific movement.
 - (b) recall of rules governing a particular sport(s).
 - (c) recognition of the proper performance techniques for physical activities or sports.
 - (d) appreciation of appropriate strategies for use in physical activities or sports.
3. The form of the stem may consist of either a question or a partial statement to be completed by one of the choices.
4. The stem of the item will consist of 25 words or fewer.

RESPONSE ATTRIBUTES

1. Multiple-choice items with five alternatives will be used.
2. The fifth choice will be "I don't know."
3. Choices may include "all of the above."
4. Distractors will not use:
 - (a) movements which are not physically possible.
 - (b) equipment which is not part of the specified sport or activity(ies).
 - (c) players and positions which are not part of the specified sport or activity(ies).
 - (d) terminology which is not usually associated with the sport or activity(ies).

SPECIFICATION SUPPLEMENT

Examples of physical activities are: basketball, running, badminton.

Chart 2

(For Physical Education Judges)

An Example of Low Domain Elaboration

DOMAIN DESCRIPTION

Students will demonstrate recognition of terms describing the physical environment.

SAMPLE ITEM

Relief is defined as

- A. mountainous areas.
- B. the vertical distance above sea level.
- C. an area of valleys and lowlands.
- * D. the differences in elevation of parts of the earth's surfaces.

An Example of High Domain Elaboration

DOMAIN DESCRIPTION

Students will demonstrate recognition of terms describing the physical environment.

SAMPLE ITEM

Relief is defined as

- A. mountainous areas.
- B. the vertical distance above sea level.
- C. an area of valleys and lowlands.
- * D. the differences in elevation of parts of the earth's surfaces.

STIMULUS ATTRIBUTES

1. Items will involve the recognition of at least one term used to describe the physical environment.
2. Terms describing the physical environment will be similar to those in the list in the Specification Supplement.
3. The stem of the item will consist of 30 words or fewer.

RESPONSE ATTRIBUTES

1. Multiple-choice questions with four alternatives will be used.
2. Distractors will consist of physical environment terms or descriptions of terms.
3. Recognition of the term will be demonstrated by either
 - (a) selecting a correct definition(s) for the term(s), or
 - (b) selecting a correct term(s) given a description(s), or
 - (c) selecting correct descriptive terms for a geographical region.

SPECIFICATION SUPPLEMENT

Some examples of descriptive terms for the physical environment are:
altitude, plateau, tree-line.

Chart 3

THE THREE TECHNIQUES

for judging the match between an item and a domain

MATCHING TECHNIQUE = requires the judge to indicate "yes" or "no"

RATING 1 and RATING 2 = judge indicates one of "5, 4, 3, 2, 1"

CATEGORIZING TECHNIQUE = judge indicates one of: "+1, 0, -1"

Chart 4

1. Domain Descriptions - white - used over and over
2. Answer Sheets - yellow }
3. Items - computer print-out sheets } clipped together

Answer all yellow sheets - 17 pages
categorizing = 4
rating = 8
matching = 4
personal data = 1

Appendix D

Forms for the Three Judgmental Techniques

Judge_____

CATEGORIZING TECHNIQUE

USE ITEM SET_____ and DOMAIN_____

Please indicate how well you feel each item measures the domain by circling +1, 0, or -1.

Items	You feel the item is a measure of the domain.	You are undecided about whether the item is a measure of the domain.	You feel the item is <u>not</u> a measure of the domain.
1	+1	0	-1
2	+1	0	-1
3	+1	0	-1
4	+1	0	-1
5	+1	0	-1
6	+1	0	-1
7	+1	0	-1
8	+1	0	-1
9	+1	0	-1
10	+1	0	-1
11	+1	0	-1
12	+1	0	-1
13	+1	0	-1
14	+1	0	-1
15	+1	0	-1

What item set have you used?_____ What domain have you used?_____

If you realize you have used the wrong items or domain, please let me know before proceeding.

Judge _____

RATING TECHNIQUE 1

USE ITEM SET _____ and DOMAIN _____

Please circle the number of the term that you find describes the relationship between the domain and the item.

Items	very relevant	somewhat relevant	no feeling	somewhat irrelevant	very irrelevant
1	5	4	3	2	1
2	5	4	3	2	1
3	5	4	3	2	1
4	5	4	3	2	1
5	5	4	3	2	1
6	5	4	3	2	1
7	5	4	3	2	1
8	5	4	3	2	1
9	5	4	3	2	1
10	5	4	3	2	1
11	5	4	3	2	1
12	5	4	3	2	1
13	5	4	3	2	1
14	5	4	3	2	1
15	5	4	3	2	1

What item set have you used? _____ What domain have you used? _____

If you realize you have used the wrong items or domain, please let me know before proceeding.

Judge_____

RATING TECHNIQUE 2

USE ITEM SET_____ and DOMAIN_____

Please circle the number of the term that you find describes the relationship between the domain and the item.

Items	very suitable	somewhat suitable	no feeling	somewhat unsuitable	very unsuitable
1	5	4	3	2	1
2	5	4	3	2	1
3	5	4	3	2	1
4	5	4	3	2	1
5	5	4	3	2	1
6	5	4	3	2	1
7	5	4	3	2	1
8	5	4	3	2	1
9	5	4	3	2	1
10	5	4	3	2	1
11	5	4	3	2	1
12	5	4	3	2	1
13	5	4	3	2	1
14	5	4	3	2	1
15	5	4	3	2	1

What item set have you used?_____ What domain have you used?_____

If you realize you have used the wrong items or domain, please let me know before proceeding.

Judge_____

MATCHING TECHNIQUE

USE ITEM SET_____ and DOMAIN_____

Please circle yes or no to indicate whether the domain is measured by the item.

Choose either yes or no; do not leave any items unanswered.

Items		
1	yes	no

2	yes	no

3	yes	no

4	yes	no

5	yes	no

6	yes	no

7	yes	no

8	yes	no

9	yes	no

10	yes	no

11	yes	no

12	yes	no

13	yes	no

14	yes	no

15	yes	no

What item set have you used?_____ What domain have you used?_____

If you realize you have used the wrong items or domain, please let me know before proceeding.

Appendix E

Personal and Attitudinal Data Forms

Judge _____

Your cooperation is appreciated in supplying the following information which will enable better understanding of the three methods for matching items with domains.

1. Number of years teaching? _____
2. Number of years teaching Geography? _____
3. Your first language? _____
4. Which method did you find easiest to use in judging the match between items and domains? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why? _____

5. Which method did you feel allowed you to make the best judgement of the match between an item and a domain? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why? _____

6. a) Did you feel the high domain elaboration provided you with sufficient information to judge the match between the item and the domain? Circle one.

Yes No

- b) Did you feel that a low domain elaboration would have been sufficient to judge the match between the item and the domain? Circle one.

Yes No

Please explain. _____

7. Any comments? _____
-

Judge _____

Your cooperation is appreciated in supplying the following information which will enable better understanding of the three methods for matching items with domains.

1. Number of years teaching? _____
2. Number of years teaching Geography? _____
3. Your first language? _____
4. Which method did you find easiest to use in judging the match between items and domains? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why? _____

5. Which method did you feel allowed you to make the best judgement of the match between an item and a domain? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why? _____

6. a) Did you feel the low domain elaboration provided you with sufficient information to judge the match between the item and the domain? Circle one.

Yes No

- b) Did you feel it would have been more helpful to have high domain elaboration? Circle one.

Yes No

Please explain. _____

7. Any comments? _____

Judge_____

Your cooperation is appreciated in supplying the following information which will enable better understanding of the three methods for matching items with domains.

1. Number of years teaching?_____
2. Number of years teaching Physical Education?_____
3. Your first language?_____
4. Which method did you find easiest to use in judging the match between items and domains? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why?_____

5. Which method did you feel allowed you to make the best judgement of the match between an item and a domain? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why?_____

6. a) Did you feel the high domain elaboration provided you with sufficient information to judge the match between the item and the domain? Circle one.

Yes No

- b) Did you feel that a low domain elaboration would have been sufficient to judge the match between the item and the domain? Circle one.

Yes No

Please explain. _____

7. Any comments?_____
-

Judge _____

Your cooperation is appreciated in supplying the following information which will enable better understanding of the three methods for matching items with domains.

1. Number of years teaching? _____
2. Number of years teaching Physical Education? _____
3. Your first language? _____
4. Which method did you find easiest to use in judging the match between items and domains? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why? _____

5. Which method did you feel allowed you to make the best judgement of the match between an item and a domain? Circle one.

Matching (yes,no) Rating 1&2 (5,4,3,2,1) Categorizing (+1,0,-1)

Why? _____

6. a) Did you feel the low domain elaboration provided you with sufficient information to judge the match between the item and the domain? Circle one.

Yes No

- b) Did you feel it would have been more helpful to have high domain elaboration? Circle one.

Yes No

Please explain. _____

7. Any comments? _____