DOCUMENT RESUME

ED 187 531 SE 030 729

AUTHOR Robitaille, David F.; And Others
TITLE Mathematics Achievement Test Project. Technical Manual.
INSTITUTION British Columbia Dept. of Education, Victoria.
PUB DATE Feb 80
NOTE 57p.: Contains occasional heavy, blurred print.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Tests; Criterion Referenced Tests; *Diagnostic Tests; Educational Diagnosis; Elementary Secondary Education; *Item Banks; *Mathematics Education; Test Construction; Testing Programs; *Test Items; Tests

ABSTRACT

The Mathematics Achievement Test Project (MATP) was set up by the Learning Assessment Branch of the Ministry of Education of British Columbia to develop testing materials to assist teachers throughout the province in the task of evaluating students' performance in mathematics. The two major objectives of the project were to construct "banks" of test items referenced to the curriculum guide for mathematics (1978 edition) in grades 3, 7, and 10 and use these items to construct tests which could be used either as achievement measures or diagnostic tests. Chapter 1 of this document discusses the strand test concept and the administration and scoring of the strand tests. Chapter 2 describes the development of the items. Chapter 3 describes statistical procedures. The appendices include lists of the MATP strands and the item writers. (Author/MK)

MATHEMATICS ACHIEVEMENT TEST PROJECT

TECHNICAL MANUAL

by

David F. Robitaille
James M. Sherrill
Thomas J. O'Shea

MATP Project Team

David F. Robitaille (Director) — Faculty of Education, U.B.C.
James M. Sherrill — Faculty of Education, U.B.C.
Heather J. Kelleher — Teacher, New Westminster S.D.
John Klassen — Teacher, North Vancouver S.D.

Thomas J. O'Shea — Research Assistant

Learning Assessment Branch
Ministry of Education
Province of British Columbia

February 1980

Mathematics Achievement Test Project

Advisory Committee

Florine Carlson	Helping Teacher Maple Ridge School District
Nancy Greer (Chairman)	Learning Assessment Branch Ministry of Education
Dennis Hamaguchi	Teacher Vernon School District
Harry N. Hufty	Supervisor Prince George School District
Penelope Jesske	Principal Williams Lake School District
Betty-Ann Kuhn	Principal Kamloops School District
Werner Liedtke	Faculty of Education University of Victoria
Alan Taylor	Teacher Coquitlam School District
Brian Tetlow	Teacher Victoria School District

Technical Advisor

Mary Cooper	B.C. Research

Table of Contents

List of Tables

List of Figures

Chapter 1

OVERVIEW

The Mathematics Achievement Test Project (MATP) was set up by the Learning Assessment Branch of the Ministry of Education to develop testing materials to assist teachers throughout the province in the task of evaluating students' performance in mathematics. The two major objectives of the project were to construct "banks" of test items referenced to the curriculum guide for mathematics (1978 edition) in Grades 3, 7, and 10, and to use these items to construct tests which could be used either as achievement measures or as diagnostic tests.

For the classroom teacher, the task of evaluating students' performance in mathematics is a demanding one. If a test is to be used in this process, a serious effort must be made to control for the influence of factors such as the level of difficulty, reliability, and content validity of the items, the length of the test, and the adequacy of "coverage" of the content. Unless factors such as these are controlled, the results obtained may be misleading or meaningless.

In constructing the MATP tests, a serious attempt was made to take into account the influence of factors such as those just mentioned. Moreover, the tests were designed with several attributes which should enhance their usefulness to teachers.

1. <u>All items are curriculum-based.</u> Each item developed for the MAT Project is linked to an objective from the curriculum guide for a particular grade. With the exception of a small number of those objectives which do not lend themselves to paper-and-pencil testing, all objectives in the guide are included.

2. <u>The test items are sound.</u> Each item accepted for inclusion in an item bank went through several stages of review by teachers of mathematics, and pilot testing on at least 300 students selected at random from across the province. Any item which failed to meet stringent psychometric standards (See Chapter 3 for details.) was rejected. This means that the items have content validity and that they meet accepted test development standards.

3. <u>The test items may be used for diagnostic purposes.</u> Virtually all the items included in the item bank are of the multiple-choice type, and the distractors were developed primarily to reflect frequently-occurring errors, wherever appropriate. Moreover, since the items are linked to specific curricular objectives, a student's level of mastery of each such objective may be validly assessed. This means that teachers will have access to the kind of information they need to plan remedial programs when required.

4. The test items for a given grade were calibrated on a common scale. All the items for a given grade were calibrated on a common measurement scale so that item results may be meaningfully compared with each other even if the results are obtained from different samples. The use of this calibration procedure, which is based on the Rasch model (see Chapter 3), will provide two important benefits to users.

a. Tests may be constructed to provide the most accurate information at any ability level desired by the user. Thus, unlike standardized tests which are constructed to be most suitable for the "average" student, tests constructed from the item bank may be designed to be most appropriate for a particular student ability level. For example, as different test forms become available in the future, a teacher could administer one test form to low ability students, a second form to better students, and yet still be able to make direct comparisons of their relative performances.

b. Results obtained from different tests may be compared since all items have been calibrated on a common scale. This means, for example, that different test forms may be used as pre- and post-test measures and that valid estimates of students' gains in achievement may be made. This would be true even if the two test forms had no items in common. Moreover, use of the Rasch model will enable valid comparisons to be made of different groups of students who take different tests.

To summarize, the MAT Project was intended to provide teachers of mathematics with test items and tests that are psychometrically sound, that are referenced to the curriculum guide, and that can be used with confidence. Initially, eleven Strand Tests were developed, and their construction, use, and application are the subject of this manual.

The MATP Strand Test Concept

At each level, the mathematics curriculum may be divided into a number of major topic areas. The 1978 edition of the curriculum guide lists nine such topic areas or strands for mathematics Grades 1 to 12: Sets and Set Operations, Number and Number Operations, Geometry, Measurement, Problem Solving, Graphs and Functions, Applications of Mathematics, Logical Thinking, and Probability and Statistics. At any particular level, certain strands are emphasized while others are given little or no emphasis. For the Grade 7 level, for example, the curriculum guide lists six strands and for Grade 10, three.

For the purposes of the MAT Project, the strands were redefined (See Chapter 2 for details.), their number reduced, and almost all of the objectives listed in the curriculum guide assigned to MATP Strands. Those few objectives which were excluded were judged to be untestable. For example, a Grade 7 objective which was excluded states that, "The student is able to distinguish between the concepts of number and numeral." It was felt that such an objective would be extremely difficult to test in the context of a paper-and-pencil test.

At the Grade 3 level there are three MATP strands: Sets and Numbers, Operations with Whole Numbers, and Geometry and Measurement. For Grade 7 there are five MATP Strands: Sets and Numbers, Operations on Whole Numbers, Operations on Rational Numbers, Geometry and Measurement, and Applications. At Grade 10 there are three: Algebra, Geometry, and Consumer Mathematics.

A forty-item test was constructed for each strand. All items included on these tests are of the multiple-choice variety with four response choices, each of which is a possible answer to the exercise. There are no responses such as "I don't know" or "None of these." A breakdown of each Strand Test by objective is contained in Appendix A.

Each Strand Test may be used at either of two levels and, for that reason, each is identified with two grade levels: Grade 3/4, 7/8, or 10/11. The Strand Tests may be used either as achievement tests or as diagnostic tests. In the former case teachers may wish to use them to assist in the task of assessing students' mastery of a given strand upon completion of that strand or at the end of the school term. Teachers of the subsequent grade can use the tests for diagnostic purposes to identify areas of particular strength or weakness before they begin to teach a major unit.

<u>Administering and Scoring the Strand Tests</u>

Each Strand Test requires forty-five minutes to administer; this is the time limit that was used in all of the pilot testing of the items. Users who wish to use the tests for assigning grades or for comparing their students' performance to the provincial average should adhere closely to this time limit. Although the tests are power tests and not, strictly speaking, speed tests, any significant departure from the time limit may make such comparisons invalid.

Students should not be allowed to use hand-held calculators or any other computational aids during the tests. Similarly, the use of a compass or protractor is not permitted.

The Teacher's Manual which accompanies each test contains a scoring key and a Class Record Form which may be utilized to analyze the results in detail. For example, the Class Record Form can be used to identify a student's strengths and weaknesses. Areas of weakness will be evident if the student frequently or consistently answers items corresponding to a particular objective incorrectly.

Such analyses, however, must be done with caution. Information from a single test item or from a small set of items is subject to a certain amount of measurement error, and should only be used as one piece of data to be confirmed by other performance records. As with any single test score, identification of a possible student weakness on the basis of these tests should be considered tentative until confirmed by further information.

Teachers may obtain diagnostic information by examining the incorrect response choices made by students because the distractors were constructed to include frequently-occurring errors. For example, consider the following Grade 3/4 item:

$$\text{Subtract: } \begin{array}{r} 429 \\ -\underline{346} \end{array}$$

a. 83
b. 183
c. 123
d. 223

Each incorrect response choice (b,c,d) is illustrative of a faulty algorithm or mislearned concept. Students who select response c may have learned to subtract by merely finding the absolute difference of each pair of digits in the exercise. Thus, " 6 from 9 is 3. 2 from 4 is 2. 3 from 4 is 1." Students selecting responses b or d may be having difficulty with place value concepts, and in particular with the concept of regrouping or "borrowing" in subtraction.

For more information concerning the attributes of a particular test, teachers should examine the test itself and the Teacher's Manual which accompanies it. The remainder of this manual is devoted to a description of the developmental process which led to the creation of the Strand Tests which have appeared thus far. Chapter 2 is concerned with the development of the MATP Item Specification Model, the Amplified Objectives, and the test items themselves. Chapter 3 contains a description of the procedures that were followed in pilot testing the items and in putting the tests together. Also included in Chapter 3 is a discussion of latent-trait theory and the Rasch model which was used to calibrate the items in the MATP item banks.

Chapter 2

DEVELOPMENT OF ITEMS

For each grade, the objectives listed in the curriculum guide for mathematics indicate, in general terms, the content that is to be taught. These objectives constituted the basis upon which the MATP items were generated.

The objectives in the curriculum guide are stated in a form that allows teachers a fairly wide degree of latitude in the selection of the actual material that is to be presented, and teachers are encouraged to read and interpret the content of the guide in the light of their experience and knowledge of their students' needs and abilities. Such latitude is clearly incompatible with a test development program where it is essential that all concerned agree upon the scope and intent of the content to be tested. For that reason, it was necessary to rewrite each of the objectives in more detailed form so that homogeneous sets of test items could be written.

Amplified Objectives

For each objective in the curriculum guide which was appropriate for assessment by means of paper and pencil testing, an Amplified Objective was developed. An Amplified Objective is an expanded statement of an objective which provides a detailed description of the topic and behavior implied by the objective. This expanded statement reduces ambiguity about what would constitute a satisfactory test item to assess the degree of mastery the student has attained. The major purpose of the Amplified Objective is to assist item-developers to produce coherent sets of test items.

An Amplified Objective has two major components: Stimulus Elements and Learner Response Options. The Stimulus Elements component consists of a thorough description of the content of a test item which would be valid for that objective. It includes a list of the restrictions that are to be placed on the items for a given objective. The Learner Response Options section is a statement of the nature of the methods the student will be expected to use in responding to the item. For multiple-choice items, this would consist largely of a description of the answer choices, both the correct answer and the distractors.

To illustrate the concept of an Amplified Objective an example is given below. The objective, II.C.7, is taken from section II.C of the curriculum guide for Grade 7.

Specific Objective (II.C.7)

The student is able to:
Write sets of equivalent fractions.

Amplified Objective

Stimulus Elements

A. Students should be able to produce (recognize) sets of two or more (up to 5) fractions equivalent to a given fraction.
B. Given a proportion with one missing term, the student will be able to identify the correct missing term.
C. Numerators and denominators used in the items should be restricted to "reasonable" numbers. For example, denominators should be limited to the following: 2, 3, 4, 5, 6, 8, 10, 12.
D. Items concerning fractions equivalent to zero, one, and other whole numbers should be used.
E. The expression "lowest terms" should be used in exercises asking students to reduce fractions.

Learner Response Options

Distractors should include the reciprocals of the correct response. They may also be formed by adding or subtracting to the numerator and denominator rather than by multiplying or dividing.

Sample Item

Which fraction is equivalent to 6/8 ?

a. 8/10
b. 8/6
c. 3/16
d. 18/24

While not everyone would necessarily agree with all of the restrictions contained in this Amplified Objective, it is evident that the objective is much more specific and less ambiguous in this form than in the original. In studying the example, it is important to bear in mind that these restrictions are to be placed upon the test items to be developed, and are not intended as directives regarding the way in which the topic should be taught. For example, the Amplified Objective restricts the fractions to be used in these items to those with denominators selected from a very small set. A teacher might well decide that this is too restrictive insofar as the teaching of the topic is concerned, and such a decision might well be appropriate in the circumstances. However, to ensure that there is general agreement about what constitutes a suitable test item for such an objective, it is imperative that an operational definition be provided. This is the function of the Amplified Objective.

The existence of an Amplified Objective makes it possible to decide, on grounds of content and face validity, whether or not a given item is acceptable as a measure of a given objective.

Acceptable Item:	Unacceptable Item:
Which fraction is equivalent to 6/8 ?	Which fraction is equivalent to 11/51 ?
a. 8/10	a. 12/52
b. 8/6	b. 77/357
c. 3/16	c. 21/61
d. 18/24	d. 1/41

The unacceptable item violates one restriction of the Stimulus Elements and one restriction of the Learner Response Options. The fraction 11/51 should not be used in a test item for this objective. If the item were to be allowed, the distractors would have to be constructed differently. The inverted form of the correct response is not in the list as required.

Amplified Objectives were developed for every objective from the 1978 edition of the curriculum guide for mathematics which was included in the MAT Project.

Item Development

The MATP Item Specification Model consists of three dimensions: item content, level of difficulty, and level of cognitive behavior. The model for Grade 7/8 is shown in Figure 1, by way of illustration. Only the content dimension varied among grades.

Item Content

Prior to the development of the Amplified Objectives and the actual test items, the objectives from the curriculum guide for each grade were partitioned into strands. Each strand represents a major block of content for a given grade. The Teacher's Manual for each Strand Test contains a list of the actual objectives from the curriculum guide which are included in that strand. These may also be found in Appendix A.

The content of each strand was further specified by the construction of the Amplified Objectives. As was mentioned earlier, an Amplified Objective was written for each objective that was included in the MATP strands.

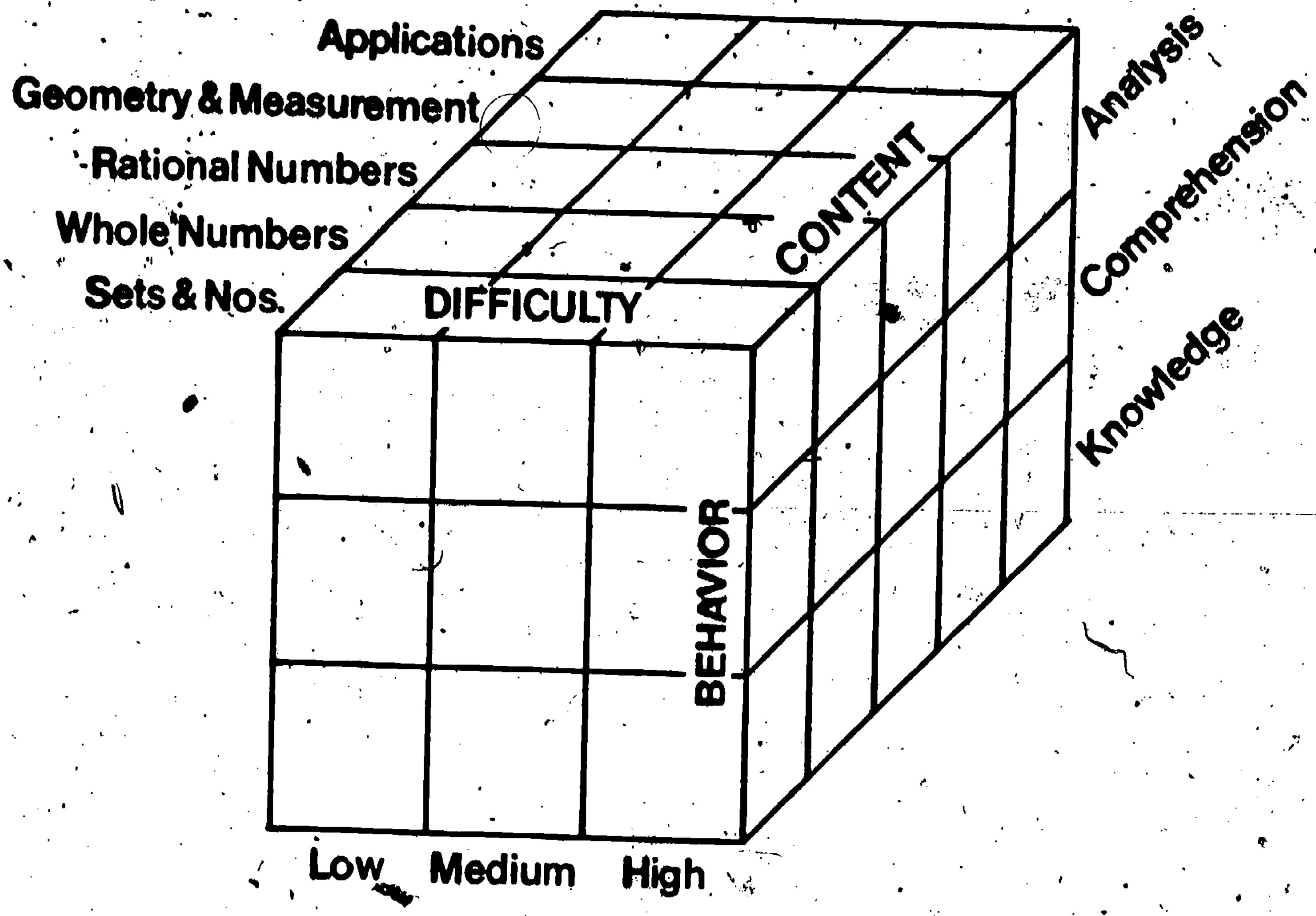

Figure 1. The MATP Item Specification Model - Grade 7/8.

Level of Difficulty

An effort was made to ensure that the items produced would be of varying levels of difficulty. Item developers were asked to produce items that were of low, medium, or high difficulty, basing their judgment of the difficulty of an item on their experience as teachers of mathematics. Items of low difficulty were defined to be those items which had a predicted p-value of approximately 80% to 100%; medium difficulty items were predicted to have a p-value of approximately 30% to 79%; high difficulty items were predicted to have a p-value of approximately 0% to 29%. It was intended that, for each objective, 40% of the items produced would be of low difficulty; 40% of medium difficulty; and 20% of high difficulty.

Level of Cognitive Behavior[30]

The three levels of cognitive behavior used in the MAT Project were Knowledge, Comprehension, and Analysis. Just as when one proceeds from low to high difficulty items the items become "harder", so too as one proceeds from Knowledge to Analysis the items become cognitively more complex. The use of these three categories of cognitive behavior ensured that items would be produced at varying levels of complexity. Such a scheme was necessary because the purpose of the project was to generate items at varying levels, and not merely items to test minimal competencies.

The categorization of each item by cognitive level was done on an a priori basis, as it was for level of difficulty. Frequently, this involved a difficult and somewhat arbitrary decision because the dividing line between the categories is rather ill-defined and vague. The classification of an item can probably only be made accurately for an individual student, taking into account his or her individual background. However, such a categorization, in spite of its weaknesses, does serve to ensure that items will be generated at various cognitive levels.

There would appear to be relatively widespread agreement about what constitutes behavior at the Knowledge level. As the behaviors become more complex, however, the definitions become more difficult to apply to a population of students as opposed to an individual student. In some categorization schemes, there may be five or more cognitive behavior levels used. For the MAT Project there were three. It was intended that, for each objective, 50% of the items produced would be at the Knowledge level, and 25% at each of the other two levels.

The Knowledge level, in which a student is not required to make decisions or use complex memory, consists of knowledge of specific facts, knowledge of terminology, and the ability to use algorithms. The Comprehension level, in which a student is required to make decisions and use complex memory, consists of knowledge of concepts, knowledge of principles, rules, and generalizations, ability to transform problem elements from one mode to another, and ability to read and interpret a mathematics problem. The Analysis level, in which a student is required to give a sequence of responses, consists of the ability to solve routine problems, to analyze data, and to recognize patterns.

A. Knowledge: The Knowledge level includes the least complex behaviors which one expects from students as outcomes of mathematics instruction. The items in this category should require only simple recall and routine manipulations.

A.1. Knowledge of Specific Facts. Items should require students to reproduce or recognize material presented in the same form as it was presented in class. Also included are basic units of knowledge to which the student has been exposed over long periods of time .

Examples

Grade 3: 3 + 5 =

a. 15 b. 35 c. 8 d. 2

Grade 7: Which of the following is NOT a whole number?

a. 0 b. 3 c. 1/2 d. 4

Grade 10: Which of the following is not defined for real numbers?

a. 3+0 b. 3x0 c. 0/3 d. 3/0

A.2. Knowledge of Terminology: Items should require the student to recall the meaning of terms and symbols.

Examples

Grade 3: Which sign says "add"?

a. - b. X c. + d. <

Grade 7: Which of the following is a prime number?

a. 33 b. 35 c. 37 d. 39

Grade 10: 5! equals

a. 5 x 5 x 5 x 5 x 5 b. $5\sqrt{5}$

c. 5 x 4 x 3 x 2 x 1 d. 5 + 4 + 3 + 2 + 1

A.3. Ability to Use Algorithms: Most teachers of mathematics would probably consider this to be the most important subcategory of the Knowledge level. The items should not require a student to select the proper algorithm, i.e., make a decision, but merely to apply the required procedure to the elements in the stimulus.

Examples

Grade 3: 418
+ 231

a. 227 b. 638 c. 418 d. 649

Grade 7: 12) 204

a. 16 b. 17 c. 18 d. 19

Grade 10: 0.02 X 0.13 =

a. 0.26 b. 0.026 c. 0.0026 d. 0.00026

B. Comprehension. The Comprehension level consists of behaviors which are more complex than those at the Knowledge level. The dividing line between categories is somewhat ill-defined and items at this level frequently require behaviors from the Knowledge level as well as from the Comprehension level. Items at this level should require more than simple recall; the student should have to make a decision about which concept or algorithm is appropriate.

B.1. Knowledge of Concepts: Since concepts are abstract and abstractions require some implicit decision-making, Knowledge of Concepts is in the Comprehension level. Items should require the student to make a decision about the applicaton of a concept or whether or not an object is an instance of a specified concept.

Examples

Grade 3: To measure the length of your classroom which of the following units would be used?

a. metre b. gram c. litre d. kilogram

Grade 7: Write the prime factors of 36.

a. 1, 36 b. 2, 3 c. 6, 6 d. 4, 9

Grade 10: If two angles are complementary, then both are

a. acute b. obtuse c. right d. congruent

B.2. Knowledge of Principles, Rules, and Generalizations: For items in this category the student is expected to recognize a principle, rule, or generalization and its application. Whether or not a given item from this sub-category is a Comprehension-level item will depend upon what the student has been taught. If the item requires that the student generate the rule, principle, or generalization, then that item would belong to a higher cognitive level for that student.

Examples

Grade 3: If you add 0 to a number *n*, the sum

a. will be greater than *n*

b. will be less than *n*

c. will be equal to *n*

d. may be more or less than *n*

Grade 7: If the decimal point in a number is moved <u>three places to the right</u>, we are

a. dividing the number by 1000

b. dividing the number by 100

c. multiplying the number by 1000

d. multiplying the number by 100

Grade 10: The sum of the angles of a triangle

a. is between 90° and 180°.

b. is 180°.

c. is between 180° and 360°.

d. depends on the sizes of the angles.

B.3. Ability to Transform Problem Elements: This subcategory is central to the Comprehension level. To solve items from this category the student may be required to change the form of a number or to change a verbal, symbolic, or a pictorial representation to either of the other forms.

Examples

Grade 3: John has 8 marbles. His brother gave him 5 more marbles. How many marbles does John have now?

Which of the following could be used to solve the story problem given above?

a. 13 - 5 = b. 8 - 5 =

c. 8 + 5 = d. 8 X 5 =

Grade 7: Suppose that an operation * on any numbers a and b is defined by $a * b = a + (a \times b)$. Then 5 * 2 equals

a. 10 b. 12 c. 15 d. 20

Grade 10: 1/8 % of N =

a. 0.000125N b. 0.00125N c. 0.8N d. 0.08N

B.4. Ability to Read and Interpret a Mathematics Problem: The ability to actually solve mathematics word problems belongs at the Analysis level. While the skills involved in this subcategory are a necessary step to solving mathematical problems, they are not sufficient. Accordingly, they belong at the Comprehension level.

Items for this category may involve problems where reading and interpreting the question is the difficult part of the solution. That is, the manipulations of the problem elements are not difficult, but the comprehension of what elements to manipulate may be. This category may also include items where the student is required to answer questions about a problem and how to solve it, but is not required to produce the solution.

Examples

Grade 3: It is true our sum is 60
And we differ just by 2
If you want to find our names
Some thinking you must do.
WHO ARE WE?

Which of the following is true about the numbers in the story given above?

a. Both numbers must be less than 30

b. Both numbers must be greater than 30

c. Only one number must be greater than 60

d. Only one number must be less than 60.

Grade 7: Mrs. Brown bought a table at a sale. The price tag stated that the original price was \$60. The tables were selling at a 20% discount. What was the amount of the discount? (Answer the question below; do not solve.)

Which of the following is true about the story problem given above?

a. Mrs. Brown paid \$60 for the table.

b. Mrs. Brown paid more than \$60 for the table

c. Mrs. Brown paid less than \$60 for the table

d. Mrs. Brown received a discount of \$20.

Grade 10; What whole numbers make the following sentence true?

$$5 < N + 3 < 10.$$

a. $N < 10$ b. $N > 5$

c. $5 < N < 10$ d. $2 < N < 8$

C. Analysis. The Analysis level includes items that require a sequence of responses from a student. It is this "sequence of responses" that distinguishes this level from the other two levels. As with the first two levels, Analysis-level behaviors are closely linked to a course of study, and Analysis-level items should be similar, but not identical, to items studied in class.

C.1. Ability to Solve Routine Problems: In its simplest form, an item at this level would require the student to select and utilize an algorithm. The items may be made more complex by involving the student in selecting a principle or rule, in choosing the appropriate algorithm(s), or selecting the proper sequence of steps to be followed.

Examples

Grade 3: Susan had read 62 pages of a book. Then she read 25 more pages. How many pages of the book has she read?

a. 87 b. 37 c. 62 d. 25

Grade 7: A class of 48 pupils decided to impose a book-loss insurance fee based on the average book price of $2.40 and the rate of loss in the previous year of 9 books lost per 24 pupils. What fee should each pupil be charged in order to cover the total expected loss for the current year?

a. $0.45 b. $0.30 c. $0.80 d. $0.90

Grade 10: Mrs. Roland bought a fur coat priced at $500. The sales tax was 4%, and the luxury tax on the coat was 15%. What was her total bill?

a. $519 b. $586.50 c. $595 d. $690

C.2. Ability to Analyze Data: This subcategory involves reading data, interpreting data, organizing data, making decisions about data, and drawing conclusions based on the data. Such abilities are necessary to separate a problem into its component parts, to distinguish relevant information from extraneous information, and to connect the sub-problems which have already been solved in order to solve the stated problem.

Examples

Grade 3:

Race horse	64 km/h
Deer	80 km/h
Homing Pigeon	100 km/h
Cheetah	110 km/h

A homing pigeon was taken 400 km from home. About how many hours would it take the homing pigeon to fly back home?

a. 3 b. 4 c. 5 d. 6

Grade 7: Five spelling tests are to be given in Deanna's class. Each test has a value of 25 points. Deanna's average for the first four tests is 20. What is the lowest score she can get on the fifth test to have a 75% average on all five tests?

a. 13 b. 14 c. 15 d. 20

Grade 10: Harriet wanted 6 pairs of socks. Store X sold them at 2 pairs for $1.25; the same brand sold in store Y at 3 pairs for $1.98. To be economical, what should Harriet do?

a. Buy the socks at Store X

b. Buy the socks at Store Y

c. Buy the socks at either store; it makes no difference.

d. It cannot be determined from the facts given.

C.3. Ability to Recognize Patterns: The student may be required to recall relevant information, transform problem elements, manipulate the problem elements into a sequence, and recognize a relationship. Basically, the student is required to find something familiar in a set of data.

Examples

Grade 3: Mary had four pieces of paper with the numbers 1 to 4 written on them as shown.

1	2	3	4

She made a list of all the 2-digits numbers she could make with the pieces of paper. What number is missing from her list?

12, 13, 14, 21, 23, ____, 31, 32, 34, 41, 42, 43.

a. 22 b. 44 c. 24 d. 42

Grade 7: In an election 356 people vote to choose one of five candidates. The candidate with most votes is the winner. What is the smallest number of votes the winner could receive?

a. 179 b. 178 c. 71 d. 72

Grade 10: The last digit in 4^{10} is

a. 0 b. 2 c. 4 d. 6

Item Validation

Most of the items for a given grade were developed by a team of classroom teachers, mathematics supervisors, and mathematics educators during a week-long writing session. Each item then went through a three-level validation process. At each level the items were scrutinized using the following criteria:

1. Would success on the item be a valid indication of mastery of the objective?

2. Was the item appropriate for the specified grade level?

3. Did the item meet all of the restrictions contained in the Amplified Objective?

4. Was the item properly categorized with respect to cognitive level?

5. Was the item clearly written?

Each pair of item-writers was responsible for a specific strand, and they conducted the first level of validation of the items for that strand. The second validation was done by another pair of item writers at the same grade level.

After the item writing session was completed the third validation was performed. This was done by the members of the project team who were not involved in writing the items. Items that survived this validation procedure were then prepared for field-testing.

References

Popham, W. James. Selecting objectives and generating test items for objectives-based tests. Problems in Criterion Referenced Measurement, CSE Monograph Series in Evaluation, Number 3. Los Angeles: Center for the Study of Evaluation, 1974, pages 13-25.

Wilson, James W. Evaluation of learning in secondary school mathematics. In B.S. Bloom, J.T. Hastings, & G.F. Madaus (Eds.), Handbook on Formative and Summative Evaluation. New York: McGraw-Hill Book Company, 1971.

Chapter 3

STATISTICAL PROCEDURES

In addition to the constraints imposed upon MAT? item development and validation procedures by pedagogical considerations such as those discussed earlier, there were several important psychometric considerations as well. In particular, it was intended that

1. the tests would provide greatest precision for measuring the achievement of students at the pass/fail level, since classification errors at this point are more serious than at other grading locations, and

2. by placing all the items for a given grade on a common scale of difficulty, scores obtained on different tests measuring performance on the same content would be directly comparable.

It was apparent that traditional, sample-dependent approaches to test development were inadequate to meet these conditions. Instead, procedures derived from latent trait theory were proposed as a means of creating tests having the required characteristics.

Latent Trait Models

Two latent trait models were investigated for their potential application in the MAT Project: the three-parameter model, and the one-parameter model. In the former case, the probability that a person of a given ability will correctly answer an item is assumed to be a non-linear function of three parameters: the difficulty of the item, the item discrimination, and the probability of obtaining the correct answer by guessing. In the one-parameter model, the performance of an individual of a given ability is assumed to be a function only of the difficulty of the item.

An extensive review of the literature and discussions with measurement specialists who were active in the area of latent trait theory led to the adoption of the one-parameter model. Although the three-parameter model probably yields a "truer" picture of the testing situation, there are a number of unresolved issues concerning its use. For example, the suitability of the procedures used to estimate the parameters has not been adequately demonstrated. Moreover, the inclusion of a guessing parameter in the model is not defensible for open-ended questions which were originally planned to constitute a portion of the item bank. Hence, although the one-parameter model requires assuming that all of the items have similar discriminating power and that the probability of success on any item tends to zero for students of very low ability, it was concluded to be the more useful of the two models. The successful implementation of an item-bank project in Portland, Oregon based upon the one-parameter model also facilitated the decision to base the MATP item bank upon it as well.

The Rasch Simple Logistic Model

The model used in developing the MATP tests is based upon the conceptualization of the measurement situation put forward in the 1960's by Georg Rasch. Rasch argued that the best that raw scores or percentages on a test can do is rank students in order of achievement. However, if one wishes to compare the performance of two individuals, percentages indicate only which person performed better, not by how much better. For example, if person A obtains a score of 60% on a test and person B obtains a score of 30%, it cannot be concluded that A achieved twice as well as B. If such a conclusion were warranted, what would be the score of person C be who did twice as well as A?

In setting up the model, Rasch attempted to build in several desirable characteristics. First, the relative performances of two individuals should be uniquely determined regardless of the items used on the test. Secondly, quantification of the relative difficulties of two items should not depend upon the abilities of the persons to whom they are administered. Finally, if one person is twice as able as another, the more able person should solve a problem twice as difficult as one given to the other person with the same "expenditure of effort" as the less able person needs to exert in order to solve the less difficult.

Rasch also wished to incorporate in his model a feature relating person-ability and item-difficulty such that, for a very capable person faced with a very easy item, the probability should be very nearly one that the person would obtain the correct answer. Also, if an item were neither too difficult nor too easy for the respondent, the outcome would be uncertain, and the probability for a correct solution should therefore be about 0.5. In other words, there should be about a 50-50 chance of a correct answer.

A probability function meeting the requirements suggested by Rasch is the following:

$$P(si) = \frac{e^{a(s) - d(i)}}{1 + e^{a(s) - d(i)}}$$

where, P(si) is the probability that person *s* correctly answers item *i*,
a(s) is the ability of person *s*,
d(i) is the difficulty of item *i*, and
e is the base of natural logarithms.

Suppose that a test consisting of 40 items is administered to a sample of 300 persons. The computer program, BICAL (Wright and Mead, 1978), may be used to estimate the difficulty, d(i), of each item and the ability, a(s), of each person. All persons obtaining the same raw score, *r*, are assumed to have the same ability. The common scale underlying both a(s) and d(i) is an equal-interval scale ranging, in theory, from positive infinity to negative infinity. The unit of measure for each parameter is the logit and, in practice, the usual range for each parameter is approximately -4 to +4 logits. Negative values for a(s) indicate low ability persons, and negative values for d(i) indicate easy items.

A simple manipulation of the model gives the following odds-on-success, O(si), for a given individual faced with a particular item:

$$O(si) = e^{a(s) - d(i)}$$

The nature of the equal-interval scale may be seen by examining this formula. Suppose persons X, Y, and Z have abilities of 1.30, 1.60, and 1.90 logits respectively. Note that the difference between X and Y, and between Y and Z are equal, namely 0.30 logits. Now suppose that each is faced with an item of difficulty -0.20 logits. The odds-on-success for person X would be $e^{1.30 - (-0.20)} = e^{1.50} = 4.50$. Similarly, the odds-on-success for persons Y and Z would be 6.05 and 8.17 respectively. The odds-on-success for Y compared to those for X are 6.05/4.50 = 1.35 times as great. The odds-on-success for Z compared to those for Y are 8.17/6.05 = 1.35 times greater. Hence, the equal intervals on the ability scale yield equal ratios for odds-on-success. In other words quantifiable meaning has thereby been given to the assertion that one student is "twice as able" as another.

There are several other advantageous properties of the model and these have been investigated and supported through the analysis of real data. First, within reasonable limits the estimates of item difficulty are independent of the sample of persons used to calibrate the items. That is, regardless of whether a sample consists of high ability or low ability subjects, the difficulty value assigned to any given item will be the same. This is in direct contrast to the traditional model which uses the p-value or percentage of correct response in the population tested as an index of the difficulty of an item.

Secondly, estimates of a person's ability may be obtained from any collection of previously calibrated items. Within limits, two persons may be assessed using completely different items, and the estimates of their abilities will be directly comparable. This characteristic is particularly desirable in the construction and utilization of item banks where the objective is to match items to persons. Low difficulty items are best suited to measuring persons of low ability, and high difficulty items give the most precise estimates of the abilities of high achievers. Using Rasch-calibrated items, such tests can be constructed and the two abilities compared on the common scale.

Closely related to the foregoing is the matter of the standard error of the ability estimates. In the Rasch model, the standard error of ability is smallest for measures derived from central scores, becoming larger as the scores tend toward the extremes. Thus, the best estimate of a person's ability will be obtained from a test containing items for each of which the probability of success for that individual is 0.5. On very difficult questions the problem of guessing may confound the issue, and on very easy questions the problem of boredom may arise. Hence, if the intention is to create a test which best measures the abilities of low achievers, items having low levels of difficulty should be selected from the item pool.

The Item Calibration Procedure

A fundamental requirement in the creation of an item bank is the determination of the difficulty of each component item on a common scale. By definition an item bank consists of a large number of items. Since each student in a calibration sample cannot be expected to respond to all items contemplated for the bank, an alternative procedure using the responses of different students to different sets of items must be employed. It is here that the power of the Rasch model is realized. By utilizing blocks of common items as links between different tests, the difficulty values of items may be interrelated. Various linking procedures were used at the three grade levels for which items were developed. The calibration procedure for the Grade 10/11 items is described here in some detail as a comprehensive example.

Three strands were required for Grade 10/11: Algebra, Geometry, and Consumer Mathematics. Initially, the items in each strand were calibrated independently. A common procedure was followed for each strand, and that procedure is represented diagrammatically in Figure 2.

A, B, C, D, and E represent five test booklets each containing 40 items from a single strand. Each booklet consisted of four links or blocks of ten items each. Each of the ten blocks of items - a, b, c, ..., j - linked two booklets. As a result one hundred items were calibrated in ten links of ten items each.

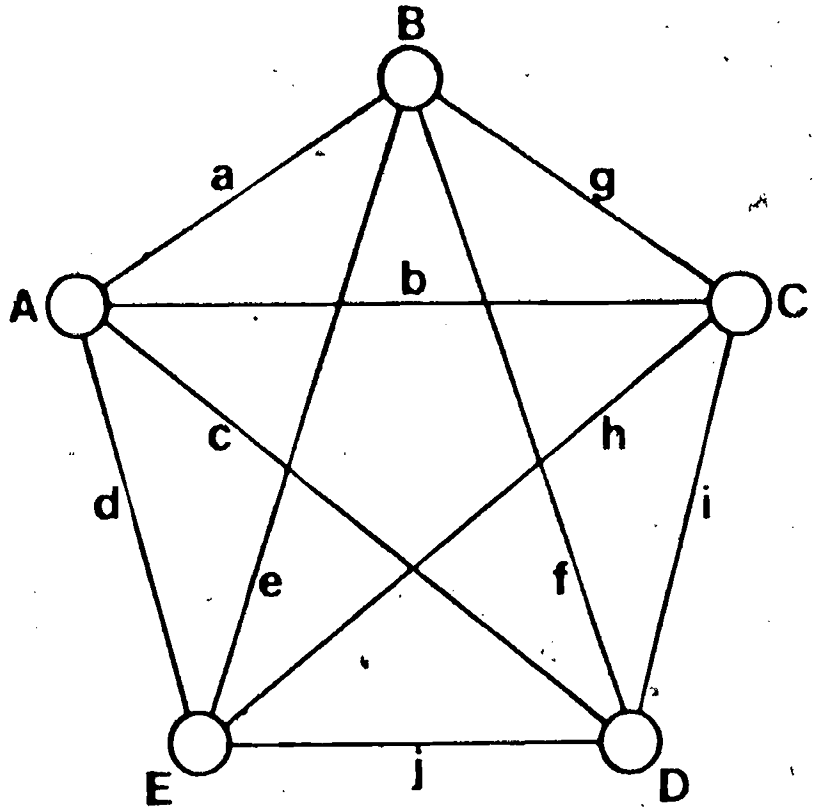

Figure 2. Linking network for items in a Grade 10 strand.

A provincial probability sample of 1500 students was obtained for each strand by selecting an appropriate number of Grade 10 classes, stratified by geographic location and size of school, as follows:

1. All schools in the province which contained classes completing Grade 10 Mathematics in the first semester were sorted into six geographic zones. The sample size chosen in each zone was proportional to the reported Grade 10 enrollment as of September 30, 1978.

2. Within each zone, schools were ranked in order of grade enrollment, and the estimated number of classes in each school was also listed. This estimate was based on the average district class size for each particular grade.

3. Within each zone, grade size strata were defined in such a way that, within each stratum, two or three classes would be selected. This ensured that schools of all sizes were represented in the final sample.

4. Within each stratum the first class was selected randomly while the second (and third, if necessary) was selected systematically. This ensured that the schools selected represented the larger and the smaller schools within each size stratum.

5. Within schools, principals were asked to select classes by a random procedure described in a letter to them. Most schools were asked to select only one class but some of the larger schools were asked to select two or more.

The return rate was approximately 175 per booklet, well below the design level of 300, probably due to inadequate information regarding semestering. The entire procedure was repeated at the end of the second semester with a design sample size of 400 per booklet. As a result, an average of 457 returns per booklet were available for calibration.

The responses to each of the sixteen different Grade 10/11 test booklets were first analyzed by the computer program LERTAP (Nelson, 1974) using traditional item analysis techniques. Hoyt reliabilities were typically around 0.80. Items were identified for closer content analysis if their results showed one or more of the following characteristics:

1. the point-biserial correlation for the correct response was less than 0.20;

2. the point-biserial correlation for an incorrect response was greater than 0.05;

3. the percent correct response was considerably lower than the percent choosing one of the distractors;

4. a distractor was selected by less than one percent of the respondents.

Some such items were omitted entirely from any further consideration, some were set aside to be modified and calibrated at a later date, and some were retained as tested because of their content validity. A summary of the results is given in Table 1.

The remaining items from each booklet were subjected to Rasch calibration using the computer program BICAL (Wright and Mead, 1978). This program generates several statistics which indicate whether or not items fit well with the rest of the collection. Such a lack of fit might occur when students guess at the correct answer or, perhaps, when the wording of an item discriminates in favor of a particular sub-group of persons. The critical statistic used for determining lack of fit was the <u>fit mean square</u>, with expected value of 1.0. An item was rejected if

Table 1

Summary of Items Deleted - by Strand

Number of Items	Algebra	Geometry	Consumer Mathematics
Piloted	100	100	100
Set aside for modification	8	11	6
Deleted (LERTAP)	0	2	4
Deleted (BICAL)	0	1	1
Retained	92	86	89

its fit mean square exceeded 1.0 by three standard errors on each of the two forms on which that item appeared. The final number of retained items is shown in Table 1.

Having decided which items were to be retained in the bank, the next step was to use the links to bring all calibrated difficulties to a common scale. Figure 2 above shows booklets A and B connected by the ten linking items in link a. Since A and B contain items of varying difficulty, and since the booklets are calibrated independently, the difficulty of each item in link a will be assigned two values: one resulting from its inclusion in booklet A and one from booklet B. However, because of the equal-interval scale underlying the model, one set of difficulty values may be taken as the standard, say those for booklet A, and each value for booklet B may then be adjusted by the same amount to bring all the items into line. The amount of adjustment was determined by first averaging the difficulties of the linking items as determined from booklet A, then averaging the difficulties of the linking items as determined from booklet B, and finding the difference between these two means. This difference should be constant for all ten pairs of difficulty values, within a margin of error. If any difference exceeded that margin, the item was omitted from the link and a new constant calculated. In this way the development of stable linking values was ensured. The result was a direct, one-way link from A to B.

As a additional check, three possible two-way linking values were calculated:

A --> C --> B, A --> D --> B, and A --> E --> B.

These three values were averaged and the average of that value and the one-way value was used to determine the final linking constant for each form.

Finally, the appropriate linking constant was added to each item difficulty to obtain the final calibrated difficulty parameters. The result was three independently calibrated sets of items: Algebra, Geometry, and Consumer Mathematics.

An attempt was made to calibrate the Grade 10/11 items on the same scale as the Grade 7/8 items. This possibility was investigated by means of a sixteenth test booklet containing ten items from each of the three Grade 10/11 strands, viz. one item from each of the ten links within a strand, and ten items from the previously calibrated Grade 7/8 items. The feasibility of a tie-in with the Grade 7/8 item bank was determined by examining the calibrated values of the ten common items. If each pair of calibrated values differed by a constant, within a certain margin of error, the inter-connection would be feasible. Using a margin of three standard errors it was found that 6 of the 10 items failed to behave as required. Three were much easier than had been predicted by the Grade 7/8 calibration, and three were more difficult. Since more than half of the items failed to meet the criterion, the result was judged to be sufficiently negative to reject calibrating the Grade 10/11 items on the Grade 7/8 scale.

The sixteenth booklet, which contained ten items from each strand, was also used to test the hypothesis that all three Grade 10/11 strands could be calibrated on the same scale. Using the procedure outlined in the preceding paragraph for comparing item difficulties, the difficulties of the items in each group of ten were matched with the difficulties as determined by the scale of the appropriate strand. The results of this procedure were positive; only one item failed to meet the criterion. This was judged to be acceptable, and all three strands were calibrated to the same scale using the constants derived from the final booklet as linking values.

In all cases, the original logit scale was transformed into a more convenient one. The transformed unit of measure was named the BRIT (BRitish Columbia unIT). The centre of the scale was set at 1000 and each logit was set equal to 100 BRITs. Item difficulties on the new scale ranged from 700 to 1200 BRITs, approximately.

Finally, an attempt was made to establish letter grade boundary values on the BRIT scale. Each teacher in the pilot testing procedure provided a letter grade for each student. In making their decisions as regards these letter grades, teachers were asked to consider only the students' achievement in mathematics and not to take into consideration such factors as motivation, work habits, and the like. In particular, the teachers were asked to consider six dimensions of achievement, and to use them in deciding which letter grade should be assigned to each student. The six dimensions were:

I. Achievement in solving mathematics problems.

II. Achievement in performing routine mathematical tasks (e.g., arithmetic skills).

III. Achievement in grasping new mathematical concepts.

IV. Achievement in making mathematical generalizations.

V. Performance on mathematics tests.

VI. Performance on mathematical quizzes and everyday assignments.

These assigned letter grades were then translated into four variables, dichotomously scored, yielding the A/B, B/C, C/P, and P/F cut-off points, respectively. They were treated as any other item and calibrated onto the BRIT scale.

Test Construction Procedure

In selecting the items from the bank for the Strand Tests, two factors were borne in mind. The first and most important criterion was content validity. That is, it was necessary that the objectives related to each strand as defined by the curriculum guide be fairly represented. In every case, a judgment was made as to a fair weighting of the objectives and this was reflected in the number of items selected for each. Secondly, the tests were designed to yield the most precise measure for students on the borderline between success and failure. To meet this requirement the items selected were generally taken from the lower difficulty end of the scale.

Having selected the specific forty items for each Strand Test, a conversion table of raw scores to Rasch ability scores on the BRIT scale was determined; that is, each score from 1 to 39 was assigned its corresponding BRIT value. As with any measurement process, each measurement is subject to uncertainty; in the Rasch model the standard error of the ability measure is least for scores in the centre. Assuming that each raw score was centrally located in a confidence interval of ±2 standard errors, the previously determined grade cut-off levels were used to determine raw score ranges for each grade level. This information was summarized in pictorial form as shown in Figure 3.

In order to verify the appropriateness of the grading procedure, each of the final Grade 7/8 Strand Tests was administered to a probability sample of 1000 Grade 7 students in June, 1979. It was hoped that this procedure would confirm that the information to be provided to test users was reliable, thus freeing the test developers from the necessity of using a separate norming sample to obtain letter grade information, and resulting in considerable economies of time and money. The

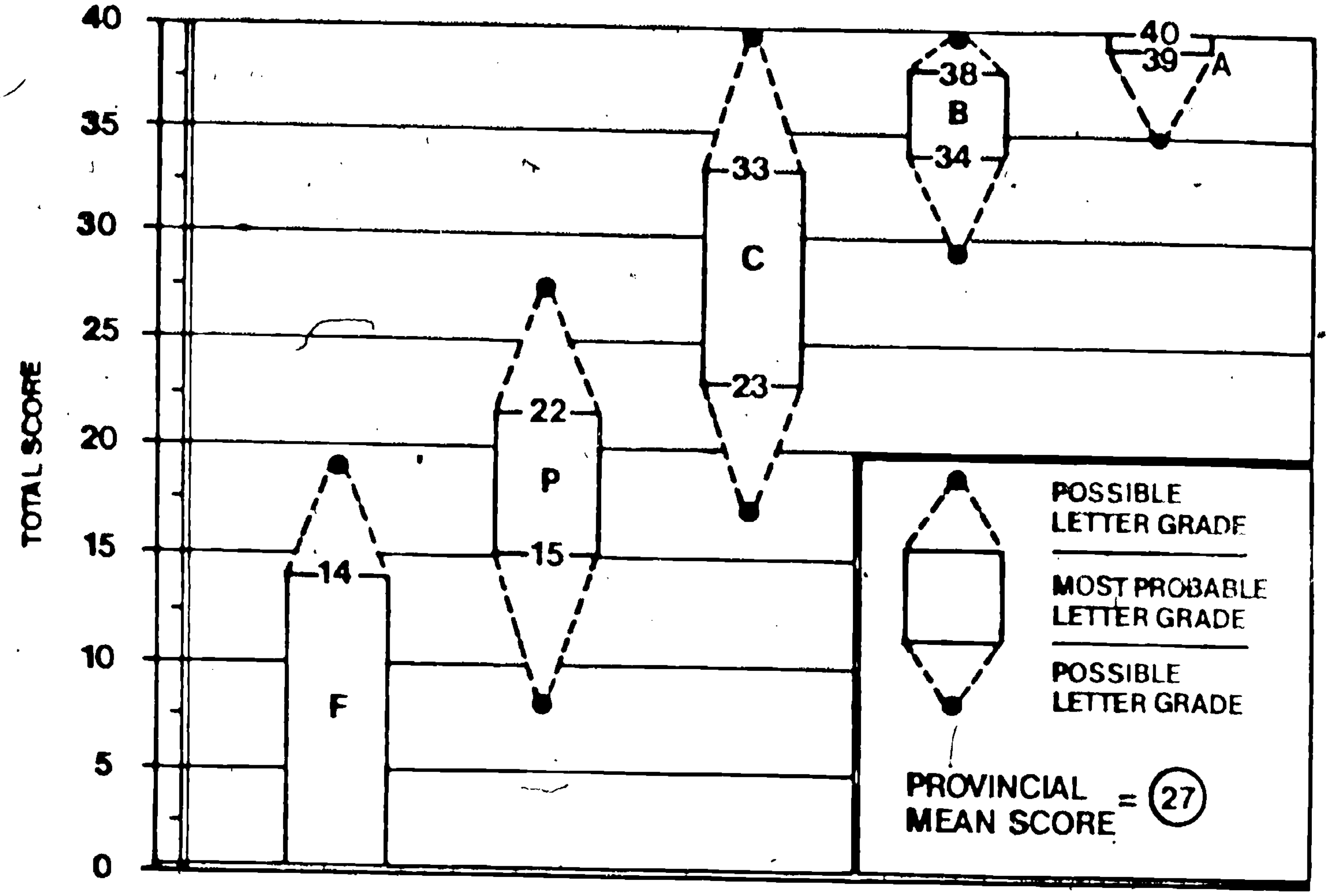

Figure 3. Assignment of letter grades to raw scores. (Grade 10 Algebra Test)

results were satisfactory in terms of the traditional point-biserial and reliability indices, and in predicting the mean scores on the tests. However, a serious discrepancy was evident between the number of respondents scoring in each grade category and the number predicted from the model. Actual performance on the tests was considerably more peaked than had been predicted. Typical findings are displayed in Table 2.

Table 2

Discrepancies between Predicted and Obtained Grade Frequencies

Grade	Suggested Raw Score Range	Predicted % of Respondents	Obtained % of Respondents
A	39-40	10	2
B	35-38	22	11
C	23-34	34	52
P	11-22	21	32
F	0-10	13	3

The problem was resolved through discussions with various persons experienced in the application of the Rasch model, and by a re-analysis of the procedure used to generate the letter grade cut-off scores. It appeared that the difficulty lay in the probabilistic nature of the model, as is illustrated in the following example.

Assume that the calibrated difficulty value at the A/B cut-off point is 1200 BRITs, and that a test of just one item of difficulty 1200 BRITs is administered to a sample of 560 persons with true abilities as shown in Table 3. The probability of success on the item for each person at each ability level is given in the second column. Multiplying the probability of success by the number of students at each level yields the expected number of persons who will be successful on the item. As a result, ninety-seven persons, i.e. 17% of the sample, will likely answer the item correctly or, in the earlier conceptualization, be judged as A-students. However, the true number of A-students in the sample, that is those with true ability levels of 1200 or above, is 60. This is just eleven percent of the sample. It appears that the estimated number will always exceed the actual number for the extreme grade levels, A and F, whenever the samples exhibit normal or normal-like distributions.

The failure of the procedure to yield valid cut-off scores posed a serious problem. For the Grade 7/8 tests, the results of the norming study could be used to establish score ranges yielding the observed percentages in each grade category based on the teachers' judgments. However, for the Grade 3/4 and 10/11 tests, this option was not available since no norming studies were carried out at those grade levels.

An alternative procedure was therefore investigated. For each of the sixteen test forms which were used in May 1979 to calibrate additional items for the three item banks, the section of the BICAL output showing the distribution of person scores was isolated. For each form, the ability scores were adjusted to

Table 3

Probable Proportions of a Normal-Like Sample "Passing" Item A/B of Difficulty 1200 BRITs

True Ability Level (BRITs)	Probability of Success on A/B	Number in Sample	Probable Number "Passing" Item A/B
700	0.01	20	0
800	0.02	40	1
900	0.05	100	5
1000	0.12	240	29
1100	0.27	100	27
1200	0.50	40	20
1300	0.73	20	15
		560	97

a common scale by subtracting the mean ability on that form. Ability scores for the entire sample of approximately 6600 students at each level were then aggregated into a frequency distribution, and the cumulative percent frequency distribution was graphed. Using the teacher estimates of grade distributions on the original calibration sample, ability scores corresponding to the grade boundary points were estimated from the ability ogive. This graph is shown in Figure 4.

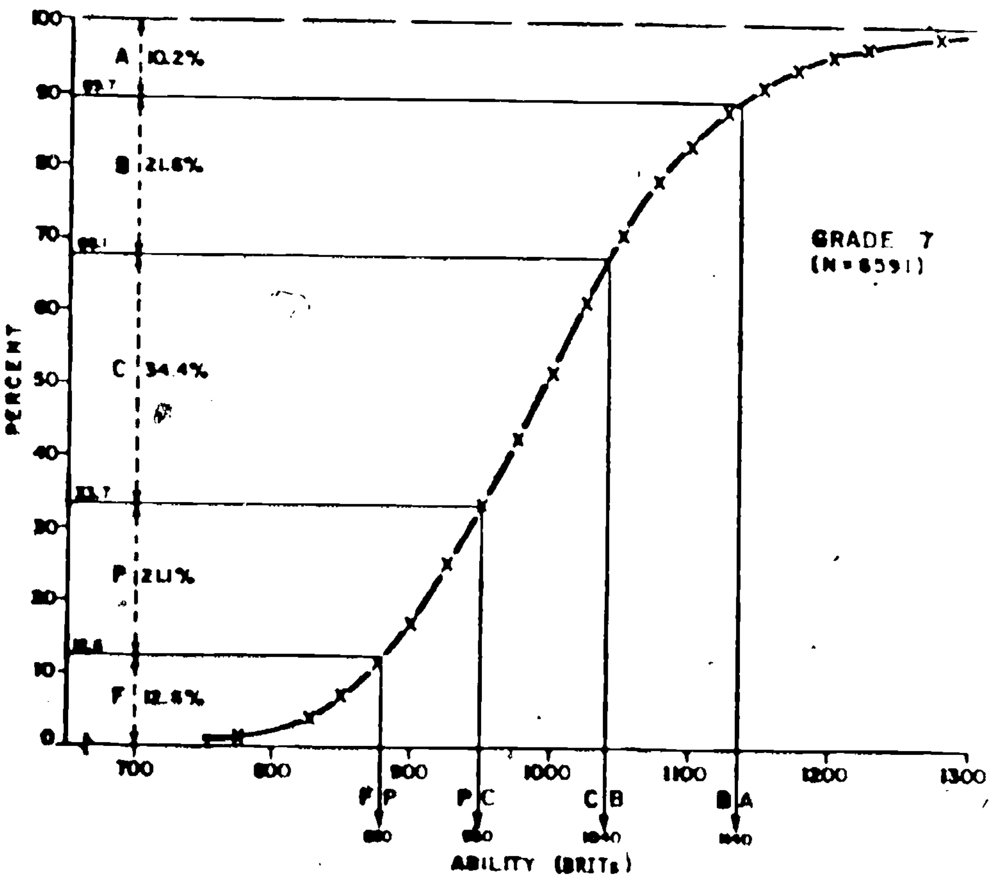

Figure 4. The use of the cumulative frequency distribution to establish grade cut-off scores.

Operating on the assumption that the distribution of abilities on each Strand Test replicated that on the aggregate of items, the ability cut-off scores were used to establish raw score boundaries on each test.

The revised cut-off scores for Grade 10 Algebra are shown in Figure 5 along with the original cut-off scores. The greatest discrepancies occur at the extremes. Grade level categorizations based upon the original values would have underestimated the

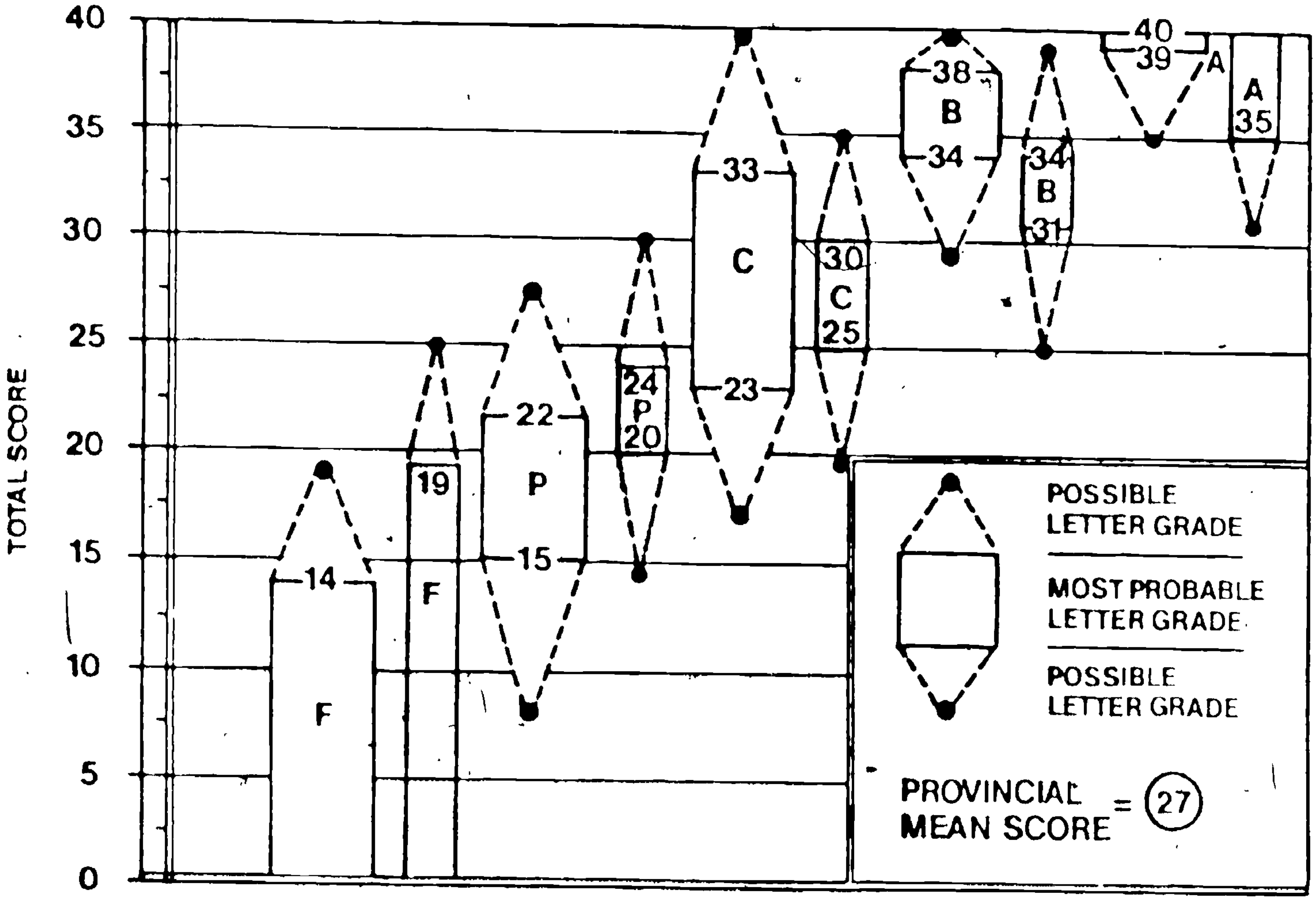

Figure 5. A comparison of the original and revised cut-off scores for Grade 10 Algebra.

number of low ability and high ability students. In general, the lower the school grade level, the less serious the discrepancies became, ranging from 5 raw score points for the P/P cut-off in the example given, to a difference of 1 raw score point for the same cut-off point in the Grade 3 Sets and Numbers test.

The information obtained in the norming study for the Grade 7/8 tests was used to validate the foregoing procedure. The grade distributions from the calibration sample were used to establish cut-off points on the histogram for each of the Strand Tests in the norming sample. The results are shown in Table 4.

Table 4

Comparison of Grading Intervals Suggested by the Calibration Ogive and Those Obtained in the Norming Sample

Grade	Interval from Ogive	n	% of Sample	Target %	Closest Available	New Interval
Sets and Numbers						
A	35-40	118	11.8	10.2	11.8	35-40
B	29-34	260	26.0	21.6	21.6	30-34
C	22-28	343	34.3	34.4	34.5	23-29
D	16-21	193	19.3	21.2	20.0	17-22
F	0-15	85	8.5	12.6	10.2	0-16
Operations with Whole Numbers						
A	37-40	168	16.4	10.2	12.3	38-40
B	32-36	307	30.0	21.6	21.7	34-37
C	26-31	309	30.2	34.4	35.3	28-33
D	21-25	157	15.3	21.1	19.4	22-27
F	0-20	82	8.0	12.6	11.3	0-21
Operations with Rational Numbers						
A	36-40	199	19.3	10.2	10.7	38-40
B	31-35	215	20.9	21.6	22.3	33-37
C	24-30	261	25.3	34.4	35.4	23-32
D	18-23	172	16.7	21.1	19.0	16-22
F	0-17	183	17.8	12.6	12.6	0-15
Geometry and Measurement						
A	36-40	137	13.0	10.2	13.0	36-40
B	31-35	250	23.6	21.6	23.6	31-35
C	24-30	338	32.0	34.4	32.0	24-30
D	18-23	209	19.8	21.1	19.8	18-23
F	0-17	123	11.6	12.6	11.6	0-17
Applications						
A	35-40	142	12.6	0.2	12.6	35-40
B	30-34	225	19.9	1.6	19.9	30-34
C	22-29	416	36.8	4.4	36.8	22-39
D	16-21	203	18.0	1.1	18.0	16-21
F	0-15	144	12.7	2.6	12.7	0-15

On two of the tests - Applications, and Geometry and Measurement - the intervals suggested from the calibration and those obtained from the norming were identical. On Sets and Numbers, the A/B cut-off was the same while the remaining ones differed by one unit. For Operations with Whole Numbers and Operations with Rational Numbers, the cut-off points differed by one or two units. At the critical P/F cut-off point, Sets and Numbers and Operations with Whole Numbers were somewhat conservative. The only area of concern might have been the P/F cut-off for Operations with Rational Numbers, but the region of uncertainty for the P interval extended well below the "true" P/F cut-off.

On balance, the intervals determined from the calibration ogive would appear to be reasonable. A similar procedure, without benefit of the norming study that was employed at the Grade 7/8 level, was used to establish grade cut-off points for the Grade 3/4 and 10/11 tests.

References

Choppin, B.H. Recent developments in item banking: A review. In N.M. De Gruijter & L.J.Th. van der Kamp (Eds.) Advances in Psychological and Educational Measurement. New York: Wiley, 1976.

Nelson, L.R. Guide to LERTAP Use and Interpretation. Dunedin, New Zealand: University of Otago, 1974.

Rasch, G. An item analysis which takes individual differences into account. British Journal of Mathematical and Statistical Psychology, 1966, 19, 49-57.

Rentz, R.R. and Rentz, C.C. Does the Rasch Model Really Work? A Discussion for Practitioners. Princeton: ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, 1978. Reprinted in Measurment in Education, National Council on Measurement in Education, 10, 2, 1979.

Whitely, S.E. & Davis, R.V. The nature of objectivity with the Rasch model. Journal of Educational Measurement, 1974, 11, 163-178.

Wilmott, A.S. & Fowles, D.E. The Objective Interpretation of Test Performance: The Rasch Model Applied. Windsor, Berks.: National Foundation for Educational Research, 1974.

Wright, B.D. Solving measurement problems with the Rasch model. Journal of Educational Measurment, 1977, 14, 97-116.

Wright, B. D. & Mead, R.J. BICAL: Calibrating items and scales with the Rasch model. Research Memorandum No. 23A, Statistical Laboratory, Department of Education, University of Chicago, 1978.

Appendix A

DEFINITION OF MATP STRANDS

Table 1

Definition of MATP Strand:
SETS AND NUMBERS
Grade 3/4

Objective Code	Statement of Objective	Item Nos.
	During Year THREE the student:	
3.2[1]	Reads and writes numerals to 9999	7,11,19,33,36
3.3	Uses expanded notation in renaming to 9999	3,10,18,22, 27
3.4	Understands the place value of numerals to 9999	2,8,9,12,14, 16,21,29,31, 32,38,39,40
3.10	Recognizes and uses correctly the following terms: odd and even	6,20
3.11	Identifies unit fractions of objects and sets. Reads and writes equations using these fraction names	5,13,23,34, 35
3.13	Is able to round off numbers to nearest 10 or 100	4,17,24,37
3.25	Skip counts by 2's, 3's, 5's, 10's	15,25,26,30
2.9	Uses ordinals to twentieth	1,28

[1] 3.2 means objective 2 from the year 3 objectives in the curriculum guide, pages 11-12. The year 2 objective (2.9) was included for completeness.

Table 2

Definition of MATP Strand:
OPERATIONS WITH WHOLE NUMBERS
Grade 3/4

Objective Code	Statement of Objective	Item Nos.
	During YEAR THREE the student:	
3.1[1]	Recognizes and uses the symbols +, -, and x	2
3.5	Solves addition and subtraction examples with 2, 3, and 4-digit numerals	
	a) addition, without regrouping	1,3
	b) addition, with regrouping	7,9,13,16,20, 23,24,27,32
	c) subtraction, without regrouping	22,35
	d) subtraction, with regrouping	4,5,8,10,11, 14,15,17,25, 28,29,33,36
3.6	Solves multiplication examples	12,19
3.8	Recalls with reasonable speed and accuracy the multiplication and division facts to 50	30
3.9	Translates a word problem into mathematical symbols and solves using appropriate operations	18,21,26,37
3.10	Recognizes and uses correctly selected terms	39
3.12	Multiplies by 10 and 100	6,34,38
2.3	Solves addition and subtraction equations including concept of inverse operations	31,40

[1] 3.1 means objective 1 from the Grade 3 objectives in the curriculum guide, pages 11-12. The year 2 objective (2.3) was included for completeness.

Table 3

Definition of MATP Strand:
MEASUREMENT AND GEOMETRY
Grade 3/4

Objective Code	Statement of Objective	Item Nos.
	During YEAR THREE the student:	
3.1[1]	Recognizes and uses the symbols $ and ¢	6,25,26,37
3.14	Constructs simple geometric models of solids and plane shapes	13,16,32
3.15	Recognizes the axis of symmetry	35
3.16	Uses graphs as a means of recording	1,2,9,14,33
3.17	Estimates and measures length in arbitrary and metric units to kilometres	10,12,15,21, 24,27,29,30, 34
3.18	Estimates and measures arbitrary and metric units in mL and litres	19,39
3.19	Determines area by covering 2-D spaces with cm^2	17
3.20	Determines volume by filling 3-D spaces with cm^3	3,22
3.21	Determines mass by balancing in kilograms and grams	18
3.22	Reads the clock to tell time and records in the conventional way	8,20,23,28
3.23	Recognizes the division of units of time into years, months, weeks, days and hours	7,11,31
3.26	Reads the thermometer in °C	5,38,40
2.25	Classifies solids according to faces, edges, vertices	4,36

[1] 3.1 means objective 1 from the Grade 3 objectives in the curriculum guide, pages 11-12. The year 2 objective (2.25) was included for completeness.

Table 4

Definition of MATP Strand: SETS AND NUMBERS Grades 7/8

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
I.1	use the following symbols {a,b,c,...}	11,17,34
I.2	state or demonstrate whether two sets are in one-to-one correspondence	10,26
I.3	specify the elements of a set when given a word description of the set	7,28
I.4	specify, draw, or state	
	a) the union of two or more sets	18
	b) the intersection of two or more sets	5,13,20,36
II.A.2	state the value represented by each digit in a multi-digit decimal numeral	1,3,21,22, 31
II.A.3	write decimal numerals in expanded form using	
	a) powers of ten	12,37
	b) powers of ten in exponent form	16,35
II.A.4	use the common mathematical symbols	2,23,29,30, 33,40
II.B.1	identify the elements belonging to the following sets of numbers:	
	a) natural or counting numbers	6
	b) rational numbers	25
II.B.3	list the factors of a number	8,14,24,39
II.B.4	list the multiples of a number	19

Table 4 (continued)

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
II.B.5	identify the elements belonging to specified sets of numbers:	
	a) odd numbers	4
	b) prime numbers	9,27,38
II.D.1	recognize and use the following properties:	
	a) associative for addition	15
	b) multiplication by zero	32

Table 5

Definition of NATP Strand:
OPERATIONS WITH WHOLE NUMBERS
Grades 7/8

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
II.C.1	perform the operations of addition, subtraction, multiplication, and division with whole numbers	
	a) addition	1,4,25
	b) subtraction	3,10,12,17, 20,29
	c) multiplication	16,19,23,34
	d) division	2,8,13,21, 33,39,40
II.C.4	write a whole number as a product of its prime factors	15,28,30,32
II.C.5	calculate the G.C.F. of two or more whole numbers	7,9,11,14
II.C.6	calculate the L.C.M. of two or more whole numbers	6,18,26,31, 37,38
II.C.13	perform, in conventional order, a calculation involving a series of operations	5,22,24,27, 35,36

Table 6

Definition of MATP Strand: OPERATIONS WITH RATIONAL NUMBERS Grades 7/8

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
II.C.1	perform the operations of addition, subtraction, multiplication, and division with rational numbers in	
	a) fraction form	1,2,6,15,28, 30,32,35,38
	b) decimal form	13,19
	c) mixed numeral form	16,23,29
II.C.2	round a decimal numeral to a specified place value	7,11,17,20, 24
II.C.7	write sets of equivalent fractions	22,39
II.C.8-11	write an improper fraction for a mixed numeral and vice versa	
	a) improper fraction for mixed numeral	14
	b) mixed numeral for improper fraction	4,5
	write a decimal numeral for a fraction numeral and vice versa	
	a) decimal for a fraction	34,36
	b) fraction for a decimal	9,27,37
	write a percent numeral for a fraction numeral and vice versa	
	a) fraction for a percent	25,40
	b) percent for a fraction	3,8,21,31
II.C.14	solve simple open sentences	10,12,18,26, 33

Table 7

Definition of MATP Strand: GEOMETRY AND MEASUREMENT Grades 7/8

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
III.1	describe or give meaning for the concept of congruency for:	
	a) segments	2,12
	b) angles	6,25,33
III.2	use compass and straight-edge to:	
	a) copy an angle	23,39
	b) bisect an angle	
	c) bisect a segment	
III.3	identify from a diagram, or draw a diagram illustrating	
	a) right angles	16,26
	b) parallel lines	24,40
	c) quadrilaterals	4,11,30
	d) parts of a circle	3,5,15,17, 31
IV.2	use a ruler to measure length in metric units	22
IV.3	use a protractor to measure angles	1,18,21
IV.4	state and use the metric units of length, angular measure (degrees), area, and volume	7,8,19,20, 27,29,32,34
IV.5	perform the basic arithmetic operations using units of measure	
	a) length	10,13,28,37
	b) volume	14,36,38
	c) mass	9
	d) angular measure	35

Table 8

Definition of MATP Strand: APPLICATIONS Grades 7/8

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
V.1	analyze a verbal problem and a) distinguish between what is given and what is to be found b) recognize whether sufficient information is given to solve the problem c) determine what operations are required to solve the problem	6,13,23
V.2	construct a "model" such as a diagram as an aid to solving a problem	30,33
V.3	translate verbal problems into open sentences	4,16,24,26,36
V.5	test the appropriateness of an answer to a problem	10,20,25,27,38
VI.1	calculate the output number when given input numbers and a function rule	2,11,34,35
VI.2	graph on the number line a) whole numbers b) rational numbers	 37 3,39,40
VII.1	draw and interpret scale diagrams	1,7,8,21,32
VII.2	solve problems involving percents	12,18,19,22,31
VII.3	calculate perimeters and areas of a) rectangles b) triangles	 5,15 9,14,29
VII.4	calculate volumes of rectangular prisms	17,28

Table 9

Definition of MATP Strand:
ALGEBRA
Grade 10/11

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
I.A.9	use the axioms of real numbers	15, 17, 30
I.A.11	write a given positive integer as the product of its prime factors	24
I.B.1	use the laws of exponents in simplifying expressions with integral exponents	26, 27
I.B.2	evaluate an algebraic expression or formula for given values for the variables	2, 11, 23, 40
I.B.3	add and multiply polynomials and apply these operations to solving open sentences	4, 13, 33
I.B.4	divide a given polynomial by a monomial	28
I.B.6	find the G.C.F. of two or more polynomials	21
I.B.8	given a polynomial containing a common monomial factor, write the polynomial in factored form	8, 12
I.B.10	given a polynomial that is the difference of two squares, write it in factored form	3, 32
I.B.11	write the square of any binomial as a trinomial	37
I.B.13	factor a quadratic trinomial	6, 16
I.C.1	solve open sentences	
	a) simple equations	1, 19, 20, 39
	b) simple inequalities	9, 35

Table 9 (continued)

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
I.D.1	translate an English phrase or sentence into an algebraic expression or open sentence	7,18
I.D.2	solve word or "story" problems algebraically	5,14,22,25, 38
I.E.1	given a table or a list of ordered number pairs, plot a graph	10,36
I.E.2	given an equation in two variables, graph the equation in the coordinate plane	31,34
I.F.1	find the square root of perfect squares of numbers and monomials by factoring	29

Table 10

**Definition of MATP Strand:
GEOMETRY AND MEASUREMENT
Grade 10/11**

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
II.A.1	determine from a given diagram whether two or more lines cut by a transversal are parallel using one of the following properties:	
	a) corresponding angles	23
	b) alternate angles	36
	c) sum of the interior angles	31
II.A.2	determine from a given diagram the measure of specified angles when the diagram embodies parallel lines cut by a transversal	5,23,38
II.B.1	calculate one side of a right-angled triangle given the other two sides, the result being expressed in numerical form	9,35
II.C.1	identify from a diagram or draw a diagram illustrating the parts of a circle	3,37,40
II.C.2	determine from a given diagram the measures of specified segments, arcs or angles using one of the following properties:	
	a) the centre of a circle lies on the perpendicular bisector of a chord	12
	c) two tangents to a circle from an external point are equal and make equal angles with the line joining that point to the centre	1
	f) the central angle of a circle is twice the inscribed angle subtended by the same arc	14
II.C.3	make the following construction:	
	a) locate the centre of a given circle	8
II.D.1	determine whether two triangles are similar	32

Table 10 (continued)

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
II.D.2	calculate a side of one of two similar triangles	6,17
II.D.7	make and read scale drawings	18,24,39
II.E.1	determine whether two triangles are congruent	2,10,13
II.G.1	calculate the perimeter of a polygon and circle	7,33
II.G.2	calculate the area of a	
	a) triangle	21
	d) trapezoid	26
	g) composite figure	28
II.G.3	given the formula, calculate the surface area of a:	
	a) rectangular prism	19
	c) cylinder	11
	e) pyramid	16
II.G.4	given the formula, calculate the volume of a:	
	a) rectangular prism	29
	f) sphere	27
II.G.5	determine from a given diagram the measure of specified angles when the diagram embodies one or more of the following:	
	a) equilateral triangles	25
	b)/c) isosceles triangles/ parallelograms	30
	f) parallel lines	4
II.H.1	distinguish between vector and scalar quantities	34
II.H.2	construct a scaled vector diagram to represent a situation described in words	22
II.H.3	calculate the resultant vector for two given vectors by a scale diagram	15

Table 11

Definition of MATP Strand:
CONSUMER MATHEMATICS
Grade 10/11

Curriculum Guide Objective Code	Statement of Objective	Item Nos.
	The student is able to:	
I.A.1	perform with increasing accuracy the basic operations of arithmetic with the rational numbers	2,3,7,8,18, 27,29
I.A.2	round-off a decimal number to a specified place value	4,6
I.A.3	change a decimal numeral to a percent and conversely	5,13,26
I.A.4	change a fractional numeral to a percent and conversely	17
I.A.5	change a decimal numeral to a fractional numeral and conversely	28,37
I.A.7	convert numerals written in usual notation to scientific notation and conversely	16,33
I.A.8	calculate products and quotients using scientific notation	1,9
III.A.1	use the concept of simple interest in calculations of discounts, commission, profit and loss, etc.	10,15,19, 20,21,22, 23,24,30, 32,34,35
III.A.2	determine the number of days in parts of a year	31,36
III.B.1	explain the meaning of compound interest	11
III.B.2	calculate the number of interest periods in a given length of time	12,38
III.B.3	calculate bank interest (given the formula)	14,39,40
III.C.2	calculate the true rate of interest in installment payments when given the formula	25

Appendix B
ITEM WRITERS

Grade 3/4

Sheila Donnelly	Quilchena Elementary Vancouver
Evelyn Grimston	Gilpin Elementary Burnaby
Kay McKinnon	Greendale Elementary Chilliwack
Cathy Millard	Van Bien Elementary Prince George
Linda O'Reilly	Van Horne Elementary Vancouver
Linda Shortreid	Anniedale Elementary Surrey
Marg Stroyan	District Staff Nanaimo
Doug Super	Mitchell Elementary Richmond
Peggy Williamson	General Gordon Elementary Vancouver

Grade 7/8

Dennis Hamaguchi	W.L. Seaton Junior Secondary Vernon
Bob Holman	James Gilmore Elementary Richmond
Ian Hooper	Gladstone Secondary Vancouver
Lillian Lamb	Charles Dickens Elementary Vancouver
Warren McQuillan	West Whalley Junior Secondary Surrey
Joan Newton	Pitt Meadows Elementary Maple Ridge
Jim Whelan	District Staff Kamloops

Grade 10/11

Dominic Alvaro	Argyle Secondary North Vancouver
Bob Campbell	McRoberts Secondary Richmond
Les Dukowski	D.W. Poppy Junior Secondary Langley
Sue Haberger	Centennial Secondary Coquitlam
Bill Hall	Sentinel Secondary West Vancouver
Joan Madison	Dr. Charles Best Junior Secondary Coquitlam
Gary Mitchell	Mount Baker Secondary Cranbrook
Jake Penner	Prince George Senior Secondary Prince George
Ken Silen	Surrey
Dan Watt	Duchess Park Secondary Prince George