DOCUMENT RESUME

ED 065 275 SE 012 763

AUTHOR	McNaughton, A. E.
TITLE	Evaluation in Secondary School Mathematics.
INSTITUTION	Victoria Education Dept. (Australia).
PUB DATE	70
NOTE	22p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Course Objectives; Grade 9; Item Analysis; Mathematics Education; *Objective Tests; *Secondary School Mathematics; Student Evaluation; *Test Construction; Tests
IDENTIFIERS	Australia

ABSTRACT

This report traces the development of 20 objective test items for secondary mathematics, part of an Evaluation Project designed to compile a reservoir of items to serve as prototypes for teacher-constructed tests. The report first shows how an individual item is written and tested. Next, a classification system of items by subject matter and pupil response is described. The four categories used for pupil response are the abilities to recall standard procedures, to apply them, to choose the correct process, and to devise original methods. This system is illustrated by a test consisting of 20 items covering these four categories in each of five topic areas taken from Form 3 mathematics (approximately grade nine). Lastly, the report describes the standardization of this test, including the preliminary trial, the choice of a representative student sample, the administration of the test, and the item analysis of the results. (MM)

AC.70/43 (C.& R.)

EDUCATION DEPARTMENT - VICTORIA

CURRICULUM AND RESEARCH BRANCH

CIRCULAR OF INFORMATION

SECONDARY MATHEMATICS

NO. MG/5

EVALUATION IN SECONDARY SCHOOL MATHEMATICS

A.E. McNaughton

CURRICULUM AND RESEARCH BRANCH

EVALUATION PROJECT

SECONDARY MATHEMATICS

Much of the current re-appraisal of school practices concerns the role of the class teacher in assessment of his pupils, and a desire to improve on past techniques in this area. The Evaluation Project has been initiated to help teachers construct better tests of mathematics achievement. Its main purpose is to compile a reservoir of suitably-tested evaluation items which can be used as prototypes by teachers if they so desire.

In the field of secondary mathematics, there are additional reasons for the creation of such an item bank. The changing nature of school mathematics means that evaluation techniques must be modified no matter how satisfactory they may have been in the past. It is hoped that teachers will acquire a deeper understanding of the objectives and methods of the new courses through a study of the evaluation items themselves and the manner in which they were developed.

It should be emphasized that the items collected in this Project are merely prototype items. These have been shown to test successfully some levels of cognitive thought relevant to mathematics in schools. If they provide the stimulus for teachers to re-examine their own evaluation techniques then the project will have been worth while.

This report traces the development of twenty items for the secondary mathematics item bank at the form 3 level, during 1969. The time-consuming nature of item construction and validation is illustrated by following the procedure from the beginning. It is shown that items cannot be judged on appearance alone - an item which appears satisfactory on the surface may in fact be worthless or misleading as an evaluation instrument.

Collection of Items.

The initial step in the construction of an item bank is to collect a large number of raw items from various sources. It is most unwise to use items taken from published tests. By so doing the test will be invalidated. Most published tests depend for their effectiveness on the pupils being unfamiliar with the items in the test; familiarity with some items may render the test worthless. Text books and experienced teachers with some knowledge of measurement techniques have been the main sources of items used in this report.

Questions obtained directly from text books are often unsatisfactory evaluation items. This is not a criticism of the questions found in text books, as they are not designed for this purpose. Most modern text book authors design a large number of their exercises so that the pupils learn from doing them - they are teaching items, not evaluation items. However, text book questions can provide the stimulus for the construction of more satisfactory items.

An example of the refinement that is necessary to produce satisfactory evaluation items will illustrate the process involved, and also the frustrating nature of item construction. The following question was obtained from a text book:

"Is 6 a factor of $3n$ if n is (i) an odd integer

(ii) an even integer?"

The question asks for two "yes-or-no" answers, necessitating an awareness by the pupil that even members contain 2 as a factor. It is a satisfactory text-book question in that the pupils must examine whether 3n is odd or even. Some class discussion may even take place on this point and the same principle may be used in more complicated examples as later questions. It is a good teaching item if properly used.

The question was not satisfactory for evaluation purposes for the following reasons:

1. There is a 50% chance of guessing each answer.

2. Pupils are aware from past experience that one answer to such questions is "yes" and the other is "no", and the word "even" in the second part will provide the necessary clue as to which are the correct answers. The correct answers may thus be obtained without investigating the question in the intended manner.

3. There is no way of knowing how the pupils obtained their answers, and thus the teacher will not know any more about the pupils' achievement levels than he did before.

The question was therefore altered to the following form:

"If n is an even number, what is the largest factor that 3n MUST have?

A. 2

B. 3

C. 12

D. 6

E. a number which cannot be determined from the information."

The objections raised above were now overcome. The pupils have a choice of five answers, and four are designed to follow from various incorrect methods of approach. Not only can the teacher tell from the selected answer whether or not a pupil can "do" the question but an incorrect answer will reveal a clear deficiency in the pupil's methods of thought. Each distractor is designed to appeal to a particular line of approach by the pupil.

The construction of suitable incorrect distractors is a task that is quite as demanding as constructing the item itself. No distractor should be so obviously wrong that it will be discarded without full consideration. If suitably designed, incorrect choices can be just as informative as correct choice to the teacher who is evaluating his pupils.

There are many pitfalls in the construction of distractors. These are discussed in most books on test construction. Perhaps the most commonly occurring fault is in making the correct answer either much longer or much shorter than the others, thus making it stand out to the alert pupil. Often the correct answer is couched in more exact terms than the others, which is another clue to the pupil searching for the correct response.

A preliminary trial of the item in this form, along with the other 19 items that had been constructed, revealed that it was an unsatisfactory item. The distractor E was much too appealing, and the correct answer (D) was chosen by only a very small number of pupils, as the following table shows. The upper group consisted of the 30 pupils who obtained the marks on the test as a whole, and the lower group consisted of the 30 pupils who obtained the lowest marks on the test as a whole.

Response	A	B	C	D	E
Upper group (N=30)	2	5	1	4	18
Lower group (N=30)	1	2	6	1	20

The item was much too difficult as it stood. Only 5 of the 60 pupils represented in the table obtained the correct answer. An attempt was made to make distractor E less appealing. Perhaps it was preferred because it was longer than the rest, or because it was so all-embracing. The reasons for its popularity can only be surmised at this stage.

The distractor E was therefore changed to "4". A choice thus had to be made between five numbers, and in this form the question appeared in the final trial of the test, with the following results:

Response	A	B	C	D	E
Upper group (N=100)	8	4	71	11	6
Lower group (N=100)	7	14	60	10	9

For unknown reasons, distractor C had become the popular choice. The item was still unsatisfactory as an evaluation instrument. The two main reasons for its unsuitability are:

1. Nearly as many pupils who performed badly on the test as a whole (i.e. lower group) chose the correct answer as pupils who performed well on the test as a whole (i.e. upper group). The item did not discriminate between high achievers and low achievers, and is thus of little value.

2. The item was still too difficult. Only about one-fifth of the pupils tested obtained the correct answer.

The item must be discarded as an evaluation item. It may have uses in the classroom but its value for measurement purposes is non-existent.

It is interesting to reflect on how many test questions which appear to be suitable on the surface but in fact have little value, have been used in schools. An even more disturbing situation occurs when more of the lower group choose the correct response than of the upper group, and the item is working against the abler pupils. This situation is not uncommon.

Classification of Items

A necessary preliminary to the construction of an item bank is a classification system. This is necessary to ensure that items on a particular topic are readily obtained from the bank, and also to ensure that the bank is comprehensive in that it contains items on a wide range of categories.

There are two classifications necessary for each item:

1. Subject Matter Classification (i.e. topic). In the present case (form 3) the topics were taken from Course of Study, Forms I to IV, V.U.S.E.B. As it is convenient to have each topic of roughly equal weight in terms of teaching time in the classroom, some topics were divided and others combined. In this way each topic heading in the classification system assumed the same importance as far as the need for items is concerned.

2. Pupil response classification (objective). It is part of the aim of teachers of all subjects to produce changes in their pupils. Before suitable evaluation can take place, it is necessary to decide on the nature of desirable changes so that test items can be constructed to measure the extent to which the teacher has been successful. Teachers have objectives in mind when they teach their pupils, and wish to evaluate the effectiveness of their teaching.

Anyone who attempts to list the objectives of even the specialized area of mathematics teaching will soon become convinced of the difficulty of such a task, and it is doubtful if full agreement would be reached with many other teachers. The task of defining educational objectives has been attempted by many writers, perhaps the best known of which is an American committee headed by B.S. Bloom. This committee decided that teachers attempt to produce changes in their pupils in three main areas -

1. The Cognitive domain - intellectual abilities and skills.

2. The Affective domain - interests, attitudes, values etc.

3. The Psycho-motor domain - manipulative or motor-skill activities.

Within each domain, educational objectives were defined and illustrated.

The work of this committee has been criticized for the very reasons that the writers advance as its strength. In attempting to define educational objectives covering all subjects, the particular contributions that each subject can make may have been overlooked. In any case, Bloom's classification system may be more suitable for professional test constructors than it is for informal classroom testing. A simple system should suffice.

This Project is restricted at present to the evaluation of intellectual abilities and skills. In the area of school mathematics, four types of pupil response can be distinguished. These are:

(a) ability to recall;

(b) ability to apply standard processes;

(c) ability to choose the correct response;

(d) ability to devise an original method.

These four catagories are defined fully, with examples in the following table.

CLASSIFICATION OF OBJECTIVES IN MATHEMATICS

1. Ability to Recall - reproducing standard theorems, symbols, definitions, formulae, proofs, etc., which are required to be committed to memory,

e.g. What does a probability of $\frac{1}{2}$ imply?

At what point does the graph of $y = 3x + 4$ intersect the y-axis?

Prove the angle in a semi-circle is a right angle.

Standard geometrical constructions.

2. Ability to Apply Standard Processes - applying standard routine formulae and other processes in which there has been extensive practice. No choice of method is involved; the situations are familiar,

e.g. Factorize $16x^2 - 9y^2$

Find $\{x : x^2 - x - 1 = 0\}$ by the "formula" method.

Simplify $\frac{2}{x+1} + \frac{3}{x-1}$

3. Ability to Choose the Correct Process - applying standard processes but firstly choosing the correct one. The pupil must "type" the question. There is the possibility that an incorrect choice may be made,

e.g. Trigonometric identities.

Geometry riders.

Problems requiring a diagram to be drawn from data given.

4. Ability to Devise an Original Method - no "taught" method is available - the problem cannot be "typed" and requires insight, understanding of basic concepts, etc., in a novel situation,

e.g. Which single digit, when doubled, doubles the fraction $\frac{\frac{2}{3}}{\frac{1}{4}+\frac{5}{6}}$?

Some Commonwealth Secondary Scholarship examination questions.

Items can now be classified in two ways, subject matter and pupil response. For a complete item bank, items are needed covering every topic and every type of pupil response. For this purpose, a grid was constructed along the following lines:

Objectives

		1	2	3	4
Topics	1				
	2				
	3				
	4				
	etc.				

Some cells of this grid cannot be filled with items because not every topic that is studied can produce results in every objective. Many topics are studied in schools because of the changes they can produce towards one or two of the objectives only.

The complete grid for this form 3 study is reproduced on the next page, and the twenty questions that are the concern of this report are placed in their appropriate cells.

FORM 3 TOPIC-CATEGORY GRID

	CATEGORY 1	CATEGORY 2	CATEGORY 3	CATEGORY 4
Sets of Rational Numbers				
Expansion of brackets - Distrib. law				
Factorization	Q.1	Q.2	Q.3	Q.4
Mensuration				
Fractions				
Mult. & Div.				
Add & Subtr.				
Partial Fractions				
Solution sets by factorization				
Sets of Real Numbers				
Square roots				
Field laws & other properties	Q.9	Q.10	Q.11	Q.12
Indices				
Logs and std. form				
Other tables				
Pythagoras				
Trig. applications				
Relation and Function				
Linear graphs		Q.14	Q.15	Q.16
Linear equations (algebraic)				
Simultaneous equns. (graph)				
Travel graphs etc.				
Simultaneous equns. (algebraic)				
Quadratic functions				
Quadratic graphs	Q.13			
Quadratic equations (alg.)				
Graphs of a^x and log functions				
Translation to algebraic form				
Theory of sine, cos, tan functions				
Formulae				
Deductive Proof				
Use of symbols & terms	Q.5	Q.8	Q.6	Q.7
Angle sum of Δ, & applications				
Parallelograms				
Isosceles Δ				
Angles in circles				
Additional				
Similar Δ 's				
Congruent Δ 's	Q.17	Q.19	Q.18	Q.20
Symmetry				
Linear inequations (alg.)				
Miscellaneous				

Items Used in This Test

In order to disrupt the normal timetables in the trial schools as little as possible, twenty items were selected and grouped together as a test. In this way, all that was required from a school was the availability of a form 3 class for one period. Four questions, one from each objective category, were chosen in each of the following five topic areas:-

Graphs,
Congruent triangles,
Factorization,
Use of symbols,
Real number system.

There were thus five questions in each objective category, and the twenty items came from twenty different cells of the form 3 grid. As a result of the preliminary trial of these items, this scheme was disrupted slightly because some items were shown to be unsatisfactory, and alteration of them caused one of them to be displaced from its original cell.

Preliminary Trial

The items were administered to one class in each of the co-educational high schools in the metropolitan area. The total number of pupils tested at this stage was 116. The schools were chosen as representing a reasonably wide range of pupil type, being located in different socio-economic areas, and widely separated geographically. There was no need to sample randomly at this stage as no conclusions were to be drawn from the pupils' responses other than that the items appeared satisfactory or unsatisfactory for the main trial. The purpose of a preliminary trial is to point up weaknesses in the structure of questions (poor wording, typist's errors, two correct answers, etc.), and reveal distractors which are either too appealing or not sufficiently appealing (too obviously incorrect etc.). Obviously if such faults can be rectified before the main trial, much time and effort will be saved.

On the basis of the pupils' responses, 14 of the questions were considered satisfactory for further trial. Of the remainder, minor alterations were made to 5 in an endeavour to make them more suitable. These alterations usually consisted of replacing one or two incorrect distractors which were too appealing, or providing more clues to the correct answer in the stem of the question in an effort to make it less difficult. The remaining question was extensively changed. (This is the question described in the early part of this report, concerning the factors of 3n.)

The questions were then duplicated in their final form, ready for trial. As a result of the alterations described above, the test now covered 6 topics instead of 5 but there were still 5 questions in each objective category. The questions are reproduced at the end of this report.

Sample

It is necessary when choosing a sample to ensure that every member of the population has an equal chance of being included. This is the basic requirement of random sampling. Unless this is done, the various statistical techniques that can be applied to the responses may not give correct results. A better procedure, of course, is to try the items out on the whole population (in this case, all form 3 pupils). The need for sampling procedures is brought about because of the impossibility of trialling the whole population.

Before a sample is taken, the population must be clearly defined. In this case the population was chosen to be all form 3 pupils attending co-educational high schools in the metropolitan area, in 1969. Any results obtained from a sample of this population cannot be applied to pupils outside the population. For example, pupils at single-sex high schools, country high schools, girls secondary schools, etc., may perform quite differently. Results obtained for one population cannot with certainty be extrapolated to a different population, although inferences can be made in many cases with little possibility of error.

The larger a sample can be made, the more accurate will be the results obtained for the complete population. The most common method of achieving a satisfactory random sample is to assign a number to each member of the population, and then use a table of random numbers to select the sample. Other common methods are to include all members of the population with birthdays on a certain date (selected randomly), or every fourth (or fifth or twentieth etc.) person from an alphabetical list.

Many "random" samples are not random at all because the sample is selected from a different population from the intended one. A sample of Melbourne residents selected from a telephone directory is really only a sample of Melbourne residents listed in this book, and conclusions obtained will apply to this select group only and not the overall population. Similarly, if all people whose surnames begin with a certain letter are chosen as the sample, then various national groups will be either over-represented or under-represented, (for example, the "Macs" or the O'Briens are either all in the sample or all out of it).

If the population contains distinct groups which may perform differently on the experimental material, it is permissible to obtain a random sample from each of these groups. These small random samples can then be combined to form a random sample of the population as a whole. The sample so obtained is called a stratified random sample. It is necessary to keep the proportion of subjects chosen for the sample constant in all of the groups.

For example, suppose a random sample of high school teachers is desired, and that one in twenty are to be included in the sample. A simple random sample could be obtained by assigning each teacher a number and selecting the sample from a table of random numbers. Alternatively, the population could be stratified according to their classification, and one in twenty principals, senior teachers, assistants, etc. could be chosen. As most teachers have a number in the classified roll for promotion purposes, this method would be easier than the simple sample method.

Alternatively, all the teachers in randomly-selected schools could be included in the sample. In this way, the proportions of senior teachers, unqualified teachers, maths teachers, men married teachers etc. would be correct. If in addition the schools were stratified before a sample was chosen, the result would probably be even more acceptable. It could be decided to select all the teachers in one-twentieth of country schools and of suburban schools, or grade A, B, C, D high schools. As long as the same proportion is chosen from each stratum, and every teacher has an equal chance of being included in the sample, then the sample will be a random one.

In this study, it was decided to include all the pupils in randomly-selected classes. This decision was made for obvious administrative reasons. The problem was thus reduced to choosing classes, which is a simpler matter than choosing individual pupils. Before selecting the classes, it was decided that two important variables to be considered were the geographical distribution of schools, and the size of schools. A two-stage stratified random sample of schools was thus made, and a third form class was then selected randomly from each chosen school.

The following table summarizes the distribution of large and small high schools in the metropolitan area (1969), which fitted the definition of the population (i.e., co-ed high schools).

	Metropolitan Area			
	North	South	East	West
Large schools (A & B grades)	28	13	32	9
Small schools (C & D grades)	7	4	11	1

It was decided to test one class in each of eleven schools, so approximately one in nine of these schools must be selected randomly.

The required numbers were:

	North	South	East	West
Large schools	3	1	4	1

	East	Other
Small schools	1	1

The schools in each of the six categories were assigned a number, and by using a table of random numbers the eleven schools which were to constitute the sample were obtained.

The class that was to be tested within each chosen school had now to be selected. This was done by contacting the schools and nominating a time that would be suitable. The schools were asked if there was a form 3 mathematics class operating at this time which could be made available. If this was not the case, then another time was nominated. At no stage did a school nominate a class to be tested. Any third-form class which contained both boys and girls, and studied mathematics for 7 or 8 periods per week was considered eligible for selection.

Administration of Items.

It is important that exactly the same instructions are given to each group of subjects which do the test. Most published tests include a very rigorous procedure to be observed by the administrator (for example, the Commonwealth Secondary Scholarships examinations). This is done so that variations from group to group reflect accurately differences between groups, and are not manifestations of differences in administration. This is particularly important when more than one person administers the test.

In this case, the front page of the test consisted of printed instructions to the candidate, together with a sample question and answer to illustrate the type of response required. No verbal instructions were given regarding the test. The class was allowed 3 minutes to read the instructions page, and then told to turn the page and commence. Thirty minutes were then allowed for the test.

All the selected classes were tested in October or November 1969. By this stage of the year, differences between schools caused by varying order of topic arrangements for the year are reduced to a minimum - all schools have finished the main work of the year by this stage.

Results

The following table shows the distribution of total scores obtained by the pupils of each class, on the test as a whole. The highest possible score is 20. As can be seen from the table, 317 pupils were tested.

The concept of Upper and Lower groups has been explained earlier in this circular. In this case, these groups were obtained in the following way:

Lower Group

(a) every pupil with a total score of 5 or less,

(b) half of the pupils from each school with a total score of 6, selected randomly;

Upper Group

(a) every pupil with a total score of 10 or more,

(b) every pupil with a total score of 9, except for 6 pupils who were rejected randomly, no more than one from each school.

This process gave 100 pupils in each group - an easy number to work with. The following table also shows how the groups were constituted.

Results of Testing

and

Selection of Upper and Lower Groups

SCHOOL	FORM	TEST SCORE 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Medain	N.	No. in Upper Group	No. in Lower Group
1	3A	-	1	-	2	6	3	4	6	4	2	3	-	-	-	-	-	-	-	-	-	-	6	31	4	14
2	3A	-	-	-	-	-	1	1	4	3	12	5	3	5	-	-	-	-	-	-	-	-	9	34	24	2
3	3E	-	-	-	-	4	3	7	8	3	6	2	4	1	-	-	-	-	-	-	-	-	7	38	13	10
4	3B	-	-	-	1	2	4	7	5	-	3	1	1	1	-	1	-	-	-	-	-	-	6	26	6	10
5	3C	-	-	-	2	4	-	5	5	5	-	3	6	1	-	1	-	-	-	-	-	-	7½	32	11	8
6	3B	-	-	-	-	1	1	1	3	2	3	3	1	3	-	-	-	-	-	-	-	-	9	18	10	3
7	3D	-	-	-	3	4	4	1	10	2	2	1	2	-	-	-	-	-	-	-	-	-	7	29	4	12
8	3B	-	-	-	3	4	5	3	5	5	3	1	-	-	-	-	-	-	-	-	-	-	6½	29	3	14
9	3B	-	-	1	-	1	2	3	1	5	5	3	-	1	1	-	-	-	-	-	-	-	8	23	9	66
10	3E	-	-	1	1	3	3	2	1	5	3	3	2	1	-	-	-	-	-	-	-	-	8	25	9	9
11	3D	-	1	-	1	3	3	8	3	6	3	3	-	1	-	-	-	-	-	-	-	-	6½	32	7	12
TOTAL		0	2	2	13	32	29	42	51	40	42	28	19	14	1	2	0	0	0	0	0	0	7	317	100	100

Analysis of Results

The following statistics were abstracted: item difficulty, item discrimination, and probability. The meanings of those terms are given below.

(i) Item difficulty

This is a measure representing the relative frequency with which an item is correctly or incorrectly answered. It is usually expressed as a percentage-

e.g., if the upper group 80 obtained the correct answer and 20 obtained the incorrect answer, and those results were reversed in the lower group, then

$$\text{Difficulty\%} = \frac{\text{no. wrong}}{\text{total number}} \times \frac{100}{1}$$

$$= \frac{100}{200} \times \frac{100}{1}$$

$$= 50\%.$$

A difficulty of n% means that $\frac{n}{100}$ of the total upper and lower groups obtained an incorrect answer.

(ii) Item discrimination

This index indicates the usefulness of the item. It is a numerical measure showing how the results for this item relate to the results for the test as a whole. If the item is to fit the pattern of the test as a whole, then obviously more of the upper group should obtain the correct answer than of the lower group. If the numbers from both groups are about the same, then the item does not discriminate between high and low performers. The test would have been just as satisfactory if the item had been omitted. This is not to say that the items with satisfactory discrimination indices are measuring the desired traits better than others but simply that the item "fits in" with the test as a whole. Whether the test actually measures the traits it was intended to measure is another matter.

The index of discrimination thus expresses the extent to which each test item discriminates between pupils, and relates this to the extent to which the test as a whole discriminates between pupils.

(iii) Probability

It is always possible that the results obtained for an item could have been due to a chance occurrence. The probability index of an item is a measure of the likelihood of this. Generally, it is required that the likelihood of the results being due to chance be less than 1 in 20 for test items, although 1 in 100 is often demanded if the test is to be used for particularly important purposes.

The following table summarizes the results of the item analysis.

Item Analysis

(the correct answer is underlined)

Question	Group	A	B	C	D	E	N.A.	Difficulty	Item discrimination	Probability
1	U	77	13	3	5	2	-	41%	0.37	<0.001
	L	41	33	10	11	3	2			
2	U	12	59	19	10	7	3	63%	0.47	<0.001
	L	19	14	17	27	15	8			
3	U	1	1	8	6	82	2	34%	0.34	<0.001
	L	9	5	27	6	49	4			
4	U	3	0	44	47	5	1	63%	0.22	<0.01
	L	10	1	49	26	13	1			
5	U	32	5	45	4	13	1	77%	0.22	<0.01
	L	13	15	32	5	30	5			
6	U	14	38	1	24	22	1	73%	0.28	<0.001
	L	11	13	5	39	29	6			
7	U	2	84	3	6	5	-	39%	0.46	<0.001
	L	9	37	10	25	17	2			
8	U	0	2	95	1	2	-	22%	0.40	<0.001
	L	12	7	62	9	9	1			
9	U	21	2	5	2	70	-	46%	0.31	<0.001
	L	38	9	8	1	39	5			
10	U	3	35	11	31	17	3	78%	0.33	<0.001
	L	6	7	13	36	29	7			

Question	Group	A	B	C	D	E	[illegible]	Difficulty	Discrimination	Probability
1	U	77	13	3	5	2	–	41%	0.37	< 0.001
	L	41	33	10	11	3	2			
2	U	12	59	19	10	7	3	63%	0.47	< 0.001
	L	19	14	17	27	15	8			
3	U	1	1	8	6	82	2	34%	0.34	< 0.001
	L	9	5	27	6	49	4			
4	U	3	0	44	47	5	1	63%	0.22	< 0.01
	L	10	1	49	26	13	1			
5	U	32	5	45	4	13	1	77%	0.22	< 0.01
	L	13	15	32	5	30	5			
6	U	14	38	1	24	22	1	73%	0.28	< 0.001
	L	11	13	5	39	29	6			
7	U	2	84	3	6	5	–	39%	0.46	< 0.001
	L	9	37	10	25	17	2			
8	U	0	2	95	1	2	–	22%	0.40	< 0.001
	L	12	7	62	9	9	1			
9	U	21	2	5	2	70	–	46%	0.31	< 0.001
	L	38	9	8	1	39	5			
10	U	3	35	11	31	17	3	78%	0.33	< 0.001
	L	6	7	13	36	29	7			
11	U	8	4	71	11	4	2	89%	< 0.02	approx. 1
	L	7	14	60	10	7	2			

Question	Group	A	B	C	D	E	N.A.	Difficulty	Item discrimination	Probability
12	U	13	12	36	26	13	0	77%	0.32	< 0.001
	L	18	15	9	39	16	3			
13	U	1	7	2	80	10	0	42%	0.44	< 0.001
	L	7	14	7	36	29	7			
14	U	18	17	22	5	36	2	74%	0,24	< 0.001
	L	27	22	16	8	15	12			
15	U	10	12	12	50	12	4	65%	0.32	< 0.001
	L	24	11	15	19	20	11			
16	U	17	15	15	50	1	2	61%	0.22	< 0.01
	L	14	18	27	27	5	9			
17	U	3	12	13	60	11	1	58%	0.35	< 0.001
	L	16	15	24	25	11	9			
18	U	2	79	4	3	11	1	41%	0.40	< 0.001
	L	11	39	14	11	21	4			
19	U	8	5	39	7	38	3	73%	0.28	< 0.001
	L	4	18	14	23	34	7			
20	U	7	13	21	10	43	6	72%	0.36	< 0.001
	L	16	35	17	14	12	6			

Nineteen of the twenty items proved to be satisfactory in terms of discrimination between upper and lower groups. (Q.11 was the unsatisfactory one, with discrimination under < 0.02 and an even chance that the results obtained were not significant at all).

Many of the items could be classified as very difficult. While the percentage difficulties range from 22% - 78%, eleven were over 60%. This is satisfactory if the items are used with this in mind (for example, in conjunction with other items) but taken together as a test, there were too many difficult items for satisfactory results regarding achievement levels.

The test, of course, was not intended to be an entity but merely a collection of items that were being trialled. Thus we have nineteen items for inclusion in the item bank at the form three level, and to act as prototype items.

****

EDUCATION DEPARTMENT - VICTORIA

CURRICULUM AND RESEARCH BRANCH

SECONDARY MATHEMATICS EVALUATION PROJECT

DRAFT TEST NO. 1

This test consists of 20 items. Each item is followed by five suggested answers.

You are to decide which answer you think is the appropriate one, and then circle the letter of this answer on the separate answer sheet. Do not mark the question sheets in anyway.

If you change your mind about an answer, cross out your first circle and draw another one around your final choice.

Example

The sum of two and two is:

A. 3
B. $4\frac{1}{2}$
C. 5
D. 0
E. none of A, B, C, D

Choice E is the correct one, so the letter E is circled on the answer sheet:-

A B C D E

In many questions, you will find several choices which are partly correct, so you will have to decide which one is the best answer.

Work carefully, but do not spend too long on any one item. If one item seems too difficult, leave it out and come back to it if you have time.

Do not mark the question sheets in any way. Rough working can be done on the sheet of foolscap provided, which need not be handed in.

You have 30 minutes to do this test.

DO NOT TURN OVER UNTIL TOLD TO DO SO.

1. The area measure of a parallelogram can be found from:-

   A. base measure x perpendicular height measure.

   B. $\frac{1}{2}$ x base measure x perpendicular height measure.

   C. base measure x sloping height measure.

   D. $\frac{1}{2}$ x base measure x (sum of measures of parallel sides).

   E. $\frac{1}{2}$ x base measure x sloping height measure.

2. What is the magnitude of the area of this circular region?

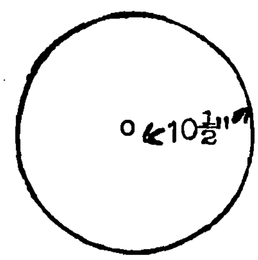
   A. 66 sq. ins.

   B. 346.5 sq. ins.

   C. 693 sq. ins.

   D. 33 sq. ins.

   E. 173.5 sq. ins.



3. Two regions have areas of the same magnitude. One region is bounded by a rectangle 16 ft. x 4 ft, and the other by a square. What is the magnitude of a side of the square?

   A. 12 ft.

   B. 5 ft.

   C. 16 ft.

   D. 10 ft.

   E. 8 ft.

4. In the diagram, part of the whole triangular region has been shaded. What <u>fraction</u> of the whole has been shaded?

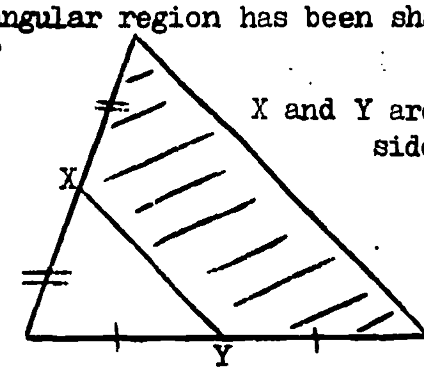
   A. 5/8

   B. 5/6

   C. 2/3

   D. 3/4

   E. 3/5



X and Y are midpoints of sides.

5. If x and y are real numbers, which of the statements below is the correct one?

   A. $x = y \Rightarrow x^2 = y^2$

   B. $x = y \Leftarrow x^2 = y^2$

   C. $x = y \Leftrightarrow x^2 = y^2$

   D. $x = y \equiv x^2 = y^2$

   E. $x = y = x^2 = y^2$

6. On the number line below, $\{x : x^2 \leq 4, \ x \in R\}$ is represented by a closed interval. Which of the following is the correct interval?

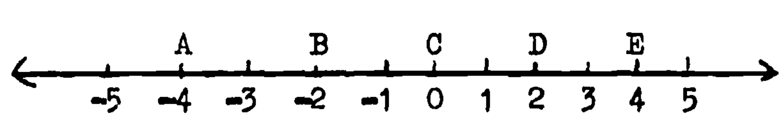
   A. $\overline{CD}$

   B. $\overline{BD}$

   C. $\overline{BC}$

   D. $\overline{AE}$

   E. $\overline{CE}$



7. x and y are two negative numbers. **Which of the expressions** below <u>must</u> be positive?

A. $y - x$

B. xy

C. $\sqrt{x + y}$

D. $x + y$

E. none of the above expressions **can be positive.**

8. Which of the following statements is true?

A. $-7 > -4$

B. $0 < -4$

C. $-5 < -3$

D. $-2 > 1\frac{1}{2}$

E. none of the above.

9. Which one of the following sets of numbers should be inserted in this sentence to make it correct?

$\{\text{rational numbers}\} \cup \{\ldots\ldots\ldots\ldots\ldots\} = \{\text{real numbers}\}$.

A. {integers}

B. {negative numbers}

C. {fractions}

D. {unreal numbers}

E. {irrational numbers}

10. Which one of the following expressions is <u>not</u> equal to $(a+b)(c-d)$ if $a+b \neq 0$ and $c-d \neq 0$?

A. $(b+a)(c-d)$

B. $(b+a)(d-c)$

C. $(c-d)(a+b)$

D. $-(d-c)(a+b)$

E. $(-a-b)(d-c)$

11. <u>Unsatisfactory Item</u>

12. If x and y are whole numbers, each divisible by 7, then which one of the following statements is <u>not necessarily</u> true?

A. xy is divisible by 49

B. $x^2 + y^2$ is divisible by 7

C. $x + y$ is divisible by 14

D. $x - y$ is divisible by 7

E. $7x$ is divisible by 49

13. For the graph shown, an appropriate equation is:

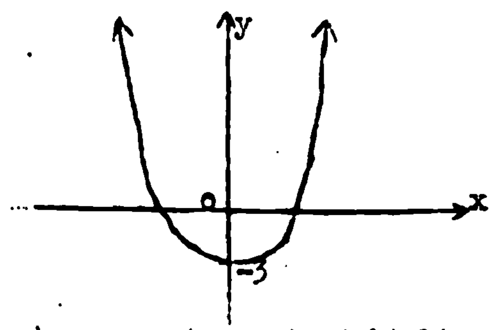
A. $y = -x^2 + 3$

B. $y = -x^2 - 3$

C. $y = x^2 + 3$

D. $y = x^2 - 3$

E. $y \quad 3x - 3$



14. Which of the following equations represents a straight line with gradient 2 and y- intercept 1?

A. $y = x + 2$

B. $2y = x + 1$

C. $y + 2x = 1$

D. $2y - x = 2$

E. $y - 2x = 1$

15. For the graph of $\{(x,y): 3y = x-9;\ x,y \in R\}$ the intersections with the axes occur where.

A. $x = -3$ and $y = -9$

B. $x = 1/3$ and $y = -3$

C. $x = 1$ and $y = -9$

D. $x = 9$ and $y = -3$

E. $x = 1$ and $y = 3$

16. The graph of $\{(x,y): y = 2x + 1,$ where x and $y \in R\}$, is a straight line. Which of the statements suggested below is the best reason for justifying this?

A. The intercepts on each axis can be found.

B. The equation contains a constant term, 1.

C. One-to-one correspondence exists between elements of the domain and range.

D. For every increase of x value by 1, there is an increase of y value by 2.

E. The equation is a simple one, with small numbers.

17. Which one of the following sets of conditions is not sufficient to show that triangles ABC and LMN are congruent?

A. $m(\angle A) = m(\angle L)$, $m(\angle B) = m(\angle M)$, $m(\overline{AB}) = m(\overline{LM})$

B. $m(\angle A) = m(\angle L)$, $m(\overline{AB}) = m(\overline{LM})$, $m(\overline{AC}) = m(\overline{LN})$

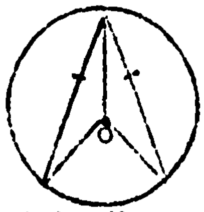
C. $m(\overline{AB}) = M(\overline{LM})$, $m(\overline{AC}) = m(\overline{LN})$, $m(\overline{BC}) = m(\overline{MN})$

D. $m(\angle A) + m(\angle L)$, $m(\angle B) = m(\angle M)$, $m(\angle C) = m(\angle N)$

E. $m(\angle A) = m(\angle L)$, $m(\angle B) = m(\angle M)$, $m(\overline{BC}) = m(\overline{MN})$

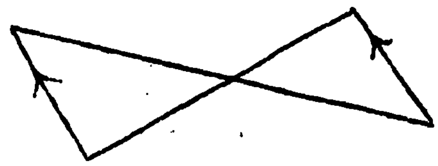
18. Which one of the following diagrams contains triangles which may not be congruent?

A.

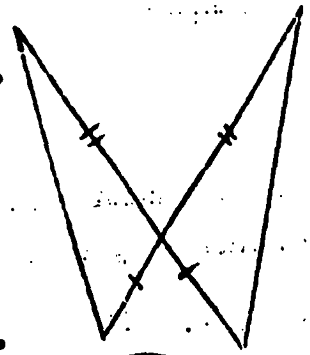


O is the centre

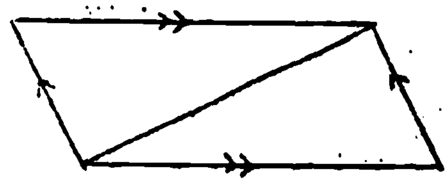
B.



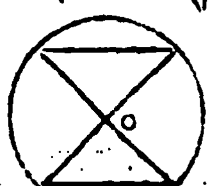
C.



D.



E.



O is the centre

19. Which of the following transformations, when performed on a triangle ABC, will produce an image A'B'C' which may not be congruent to the original triangle?

A. translation parallel to the x-axis

B. rotation about the origin

C. dilation with origin as centre

D. reflection in the y-axis

E. more than one of the above.

20. One of the following sets of figures has all its members congruent. Which one is it?

A. = {parallelograms, each with side length magnitudes 6", 6", 3", 3"}

B. = {parallelograms, each with side length magnitudes 6", 6", 6", 6"}

C. = {parallelograms, each with four right angles and one side length magnitude 6"}

D. = {parallelograms, each with angles of magnitudes 120°, 120°, 60°, 60°}

E. = {parallelograms, each with one right angle and side length magnitudes 6", 6", 6", 6".}

ANSWER SHEET

NAME..............................FORM.................SCHOOL..................

Circle the letter corresponding to the answer of your choice. The questions are grouped according to the page on which they occur.

PAGE 1	QUESTION	1.	A	B	C	D	E
		2.	A	B	C	D	E
		3.	A	B	C	D	E
		4.	A	B	C	D	E
		5.	A	B	C	D	E
		6.	A	B	C	D	E
PAGE 2	QUESTION	7.	A	B	C	D	E
		8.	A	B	C	D	E
		9.	A	B	C	D	E
		10.	A	B	C	D	E
		11.	A	B	C	D	E
		12.	A	B	C	D	E
PAGE 3	QUESTION	13.	A	B	C	D	E
		14.	A	B	C	D	E
		15.	A	B	C	D	E
		16.	A	B	C	D	E
		17.	A	B	C	D	E
PAGE 4	QUESTION	18.	A	B	C	D	E
		19.	A	B	C	D	E
		20.	A	B	C	D	E