The Irish Journal of Education, 1990, *xxiv*, 1, pp. 27-39.

MATHEMATICS ACHIEVEMENT IN SIXTH CLASS IN IRISH PRIMARY SCHOOLS*

Michael O. Martin
Educational Research Centre
St Patrick's College, Dublin

A criterion-referenced mathematics test was developed by the Curriculum Unit of the Department of Education to assess the level of performance of sixth class pupils across the mathematics curriculum. It was administered to pupils in a national sample of sixth classes in May, 1984. The present study investigates the utility of the test as an index of individual differences in pupil performance. Analyses show that the test exhibits the statistical characteristics of a good norm-referenced test and could be used as a satisfactory measure of individual pupil achievement in mathematics. Use of the test in this way revealed large differences in the levels of achievement of the most and least mathematically able pupils. On a test assessing mastery of 41 objectives, pupils scoring in the top 10% on the test as a whole achieved mastery on 30 objectives on average, while those in the bottom 10% achieved mastery on just 8 objectives.

In 1977 the Curriculum Unit of the Department of Education began a programme of assessment of mathematics achievement in primary schools. The purpose of the programme was to assess the level of achievement of pupils in the areas of mathematics specified in the primary-school curriculum. As part of this programme, mathematics tests were developed and administered to national samples of pupils in second and fourth standards in May and June, 1977. A similar test developed for sixth standard was administered to a national sample of pupils in May and June, 1979. An equivalent version of the 1979 sixth-standard test was administered to a national sample of pupils in 1984. A report of the results of that study has already been published (Eire. An Roinn Oideachais, Aonad Curaclam, Brainse an Bhunoideachais, 1985).

Adequate coverage of the curriculum was the principal design criterion in constructing tests for the assessment programme. The assessment of the relative performance of individual pupils on the test as a whole was very much a secondary consideration. However, since it is possible to interpret the performance of an individual pupil on a test of mathematics in terms of the

*This report is based on work funded by the Curriculum Unit of the Department of Education.

'mathematical ability' of the pupil, it is legitimate to investigate the extent to which the criterion-referenced mathematics test devised by the Curriculum Unit could be used in the study of individual differences in mathematics achievement

The present study was commissioned by the Curriculum Unit to address this issue using the data from the 1984 survey The purpose of the study was to investigate the utility of the test used in the survey as an index of mathematical ability and to illustrate its potential use in the study of the nature of individual pupil achievement in mathematics

THE 1984 TEST AS AN INDEX OF MATHEMATICAL ABILITY

Design of the Test

An analysis of the content of the mathematics programme for fifth and sixth standards by members of the Inspectorate of the Department of Education yielded 83 objectives in nine curricular domains Because of practical constraints, only 37 'key' objectives were selected for testing The omitted objectives were either unsuitable for testing in written form or were such that mastery of them could be implied from mastery of the objectives included in the test The curricular domains and the tested objectives are listed in Table 1 An additional domain from the third- and fourth-standard programme, consisting of four objectives concerning Operations with Whole Numbers, was added to the test, giving a total of 41 objectives in ten curricular areas

TABLE 1

SUMMARY STATEMENT OF OBJECTIVES TESTED IN THE 1984 MATHEMATICS SURVEY

Category Y Operations with Whole Numbers
Addition of Column of Numbers
Subtraction of Numbers of not more than four digits
Multiplication by two digit Numbers
Long Division
Category A Whole Number Structure
Identifying Prime and Composite Numbers
Renaming a Composite Number as the Product of Prime Factors using Exponents
Identifying Common Factors
Identifying the HCF of two Numbers
Identifying the LCM of two Numbers
Addition of Directed Numbers

TABLE 1 - Continued

Category B: Fractional Number Structure.
Identifying Fractions on the Number Line.
Ordering Fractions.
Commutative Property of Addition of Fractions.
Commutative Property of Multiplication of Fractions.
Associative Property of Addition and Multiplication of Fractions.
Using the Distributive Property of Multiplication over Addition of Fractions.
Category C: Operations with Fractions.
Addition and Subtraction of two Fractions.
Subtraction of a Fraction from the Sum or Difference of two Fractions.
Multiplication of Fractions.
Division of Fractions.
Category D: Decimals and Percentages.
Converting Fractions to Decimals and vice versa.
Ordering of Decimals.
Expressing a Number as a Percentage of another Number.
Category E: Metric Measure.
Converting Metric Measure from one Unit to another.
Category F: Algebra.
Solving one-step Algebraic Equations.
Solving two-step Algebraic Equations.
Category G: Geometry.
Identifying two-dimensional Shapes.
Using a Protractor.
Recognizing relationships between the Sides, Diagonals, and Angles of a Polygon.
Identifying three-dimensional Shapes.
Recognizing relationships between the number of Vertices, Faces, and Edges of Polyhedra.
Recognizing the relation between the Radius and Circumference of a Circle.
Recognizing the relation between Square Metre and Square Centimetre.
Calculating the Area of Rectangles.
Calculating the Area of Triangles.
Category H: Charts and Graphs.
Interpretation of Charts and Graphs.
Identifying the Co-ordinates of a Point on a Grid.
Category I: Problems.
Problems involving Simple Interest.
Problems involving Averages.
Problems involving Percentages.
Problems involving Percentage Profit and Loss.

Three items were written to test mastery of each selected objective, resulting in a test of 123 items. A pupil who answered correctly at least two out of the three test items was judged to have mastered an objective. All items were open-response rather than multiple-choice format.

The test can be considered on four levels as an index of mathematics achievement At the lowest level, there are 123 test questions or items At the next level, there are 41 objectives, each consisting of three items The 41 objectives can be grouped into ten curricular domains, with the number of objectives tested per domain ranging from one in Metric Measure to nine in Geometry Finally, the ten domains can be grouped to form a single test of mathematics As the focus shifts from items through objectives and domains to a single score, the information contained in the test results changes from detailed performance data on specific tasks to a more general indication of the position of the respondent on a continuum of general mathematical achievement

This approach makes the assumption that performance on the test can to a large extent be explained in terms of a general trait of mathematical ability If there is a single dimension or latent trait of mathematical ability underlying performance on the test, then a positive relationship should be evident between performance on different items, between performance on different objectives, and between performance in different domains Thus, the assumption of a single latent trait of mathematical ability underlying performance on the test can be verified to some extent by empirical methods

Sample

A sample of 142 sixth-class classes was selected from the population of sixth classes in ordinary national schools in May, 1984 using a random sampling procedure The test was administered by members of the inspectorate A total of 2,377 pupils were tested

Assessing the Dimensionality of the Test

The approach adopted to analyzing the dimensionality of the test should be seen as empirical and data-analytic, rather than theoretical The purpose was to extract from the 41 mastery scores (or perhaps the 123 test items) a small number of indices (preferably one) which convey as much information as possible about the mathematical ability of the pupils in the most concise possible form The emphasis was on reducing the data to manageable form rather than producing a theoretical defence of the existence of just one or two mathematical abilities or traits

In the present study, principal-components analyses were carried out at three levels of the data, the item level, the objective level, and the domain level Although principal-components analysis is a good general purpose method for reducing data to more manageable size, difficulties are sometimes found when it is used with item data where the items have only two values, right or wrong

In this case, spurious factors are sometimes found which are more related to the difficulty levels of the items than to a real affinity between them. The possible presence of such spurious factors should not cause problems in the present study since the aim was to find the minimum number of components necessary to account for variation in the data rather than to justify the existence of each observed component. At the objective level, the problem of possible spurious factors was reduced by analyzing scores for objectives rather than mastery levels. The score for an objective is simply the number of correct items for the objective.

Table 2, which summarizes the results of these analyses, shows the size of each of the first five components and the percentage of the total variance at that level which is accounted for by each component. A similar pattern emerges at each of the three levels. In each instance, there is a large first component, followed by a number of smaller components. This implies that performance on the test may be explained by one powerful latent trait and a number of less important ones. At the item level, the first component accounts for 23.9% of the total item variance. Given that there are 123 items contributing to the total variance, this is a respectable figure. At the objective level, the first component

TABLE 2

SIZE OF COMPONENTS AND PERCENTAGE OF VARIANCE ACCOUNTED FOR BY COMPONENTS FOR ANALYSES OF ITEMS, OBJECTIVES, AND DOMAINS

Level	Component	% Variance	Cumulative %
Items	29.4	23.9	23.9
	5.7	4.7	28.6
	3.5	2.9	31.5
	2.8	2.3	33.7
	2.3	1.9	35.6
Objectives	14.4	35.0	35.0
	2.3	5.5	40.6
	1.7	4.1	44.7
	1.3	3.1	47.8
	1.1	2.7	50.4
Domains	5.4	54.3	54.3
	0.8	8.3	62.6
	0.8	7.9	70.5
	0.6	5.5	76.0
	0.5	5.2	81.2

accounts for 35 0% of the variance among scores The increase in percentage of variance explained is caused by the consolidation of item groups into scores for objectives and a consequent reduction in the amount of item-specific variance in the total At the domain level, the percentage of variance explained by the first component rises to 54 3 percent Again, the consolidation of 41 scores for objectives into ten domain scores reduces the amount of objective-specific variance and increases the importance of the common component

The existence of a relatively large first component at each level means that it is not unreasonable to assume the existence of a single trait of mathematical ability underlying performance on the test as a whole At the same time, the fact that the percentage of variance explained by the first component falls well short of 100% at each level implies that the use of a single score as an index of mathematics ability will result in some of the information contained in the test items being discarded A useful compromize in this situation may be to use an overall score for general analyses, and the domain scores when more detailed content information is required

Psychometric Properties of the Test

If all 123 items are to be combined into a single mathematics test for assessing differences in individual pupil achievement, it would be useful to have some information concerning the statistical or psychometric properties of the test The statistical criteria for a good test of individual differences are quite different from those for a good criterion-referenced test. The usual properties of interest for tests of individual differences are the difficulty levels of the items, the discrimination levels of the items, and the reliability of the test as a whole In contrast, the highest priority when selecting items for a domain or criterion-referenced test is that the items form representative samples from each of the domains in question Statistical properties such as discrimination and reliability levels are usually considered to be of secondary importance The psychometric properties of the test used in the present study are summarized in Table 3 In general, the test had good psychometric properties as a measure of individual differences in mathematics achievement.

The most efficient tests (i e , those yielding the most information per item) have average item-difficulty levels (the percentage of pupils correctly answering the item) of approximately 50 percent At 54 7%, the average item difficulty level in the present test is close to this ideal There are quite a few easy items (17 1% of items have difficulty levels above 80%, i e , 17 1% of items are answered correctly by at least 80% of the pupils in the sample), but this is

TABLE 3

DIFFICULTY AND DISCRIMINATION LEVELS OF THE COMPLETE MATHEMATICS TEST

Difficulty Levels

Range	No. of Items	Percentage
80.01-100	21	17.1
60.01- 80	33	26.8
40.01- 60	36	29.3
20.01- 40	26	21.1
0.00- 20	7	5.7

Total No. of Items:	123
Mean Difficulty:	54.7
Mean Raw Score:	67.2
SD Raw Score:	26.6

Discrimination Levels

Range	No. of Items	Percentage	Comment
0.4-1.0	94	76.4	Very Good
0.3-0.39	19	15.5	Reasonably Good
0.2-0.29	9	7.3	Marginal
0.0-0.19	1	0.8	Poor-Unacceptable

unavoidable in a test that was designed primarily for curriculum coverage rather than for discriminating between pupils. Some of the easier items could be omitted from the test without reducing its effectiveness as a test of individual differences, although such omissions would of course affect content coverage.

The discrimination index of an item is a measure of the extent to which the item discriminates between high and low scorers on the test as a whole. The discrimination index used in this study is the point-biserial correlation co-efficient. Ebel (1972) considers discrimination levels above 0.4 to be very good, those above 0.3 and below 0.4 to be reasonably good, and those below 0.3 to be marginal to unacceptable. According to these criteria, the present test performs very well, with 91.9% of the items having 'very good' or 'reasonably good' discrimination levels (76.4% are 'very good').

The good discrimination levels are reflected in a Kuder-Richardson reliability coefficient of 0.97, which is above the minimum required for the use of individual test results. The reliability of the test is in turn related to the standard error of measurement, which is an index of the error involved in using the test to measure the performance of an individual pupil. In the present test, it is 4.61

raw-score units (i e , out of a range of 0 to 123) or 3 75 percentage units (out of a score of 0% to 100% correct)

A problem with the use of the number of correct items as a total score is that it is difficult to impute meaning to an individual's particular score on a test With the present test, this problem can be partially circumvented by the use of number of objectives mastered rather than number of items correct Both indices are closely related (r= 992) The number of objectives mastered (out of 41) seems to be a more meaningful measure

GENERAL MATHEMATICAL ACHIEVEMENT AND ACHIEVEMENT IN CURRICULAR AREAS

The analyses reported in the previous section show that it is reasonable to use the test as a whole as a measure of mathematical ability, but that it may be desirable to use area scores for certain purposes In the present section, both total number of correct items and number of objectives mastered are used as measures of general mathematical achievement, while average number of objectives mastered per domain is used as the measure of achievement within a curricular area The section contains an analysis of the relationship between general mathematical ability and performance in the different curricular areas

General Mathematical Achievement

Figure 1 shows the percentage of the total sample which mastered varying numbers of objectives The graph reveals a wide range of performances on the test as a whole The median performance level is mastery on 24 out of the 41 objectives (the exact percentage of the sample mastering 24 objectives is 52 1) There is a considerable difference between the performance levels of those who do well on the test and those who do poorly At the high end of the performance continuum, those at the 90th percentile (the 10% of pupils with the highest mastery levels) master on average 35 objectives In contrast, those at the 10th percentile (the 10% of pupils with the lowest mastery levels) master on average eight objectives

In Figure 2, mathematics achievement is expressed in terms of the average level of mastery for the sample as a whole within each of the curricular domains covered by the test The figure shows that the highest mastery levels are to be found in the area of Operations with Whole Numbers, which is not surprising since this area is really more suited to the third and fourth standard curriculum Next come Operations with Fractions, Fractional Number Structure, Decimals, and Charts and Graphs with similar mastery levels (ranging from 59% to 71% on average) Whole Number Structure, Problems, Algebra, and Metric Measure are also similar to each other in mastery level (ranging from 49% to 51%). The lowest level of mastery is found in Geometry.

FIGURE 1

THE PERCENTAGES OF PUPILS WHO MASTERED VARYING NUMBERS OF OBJECTIVES

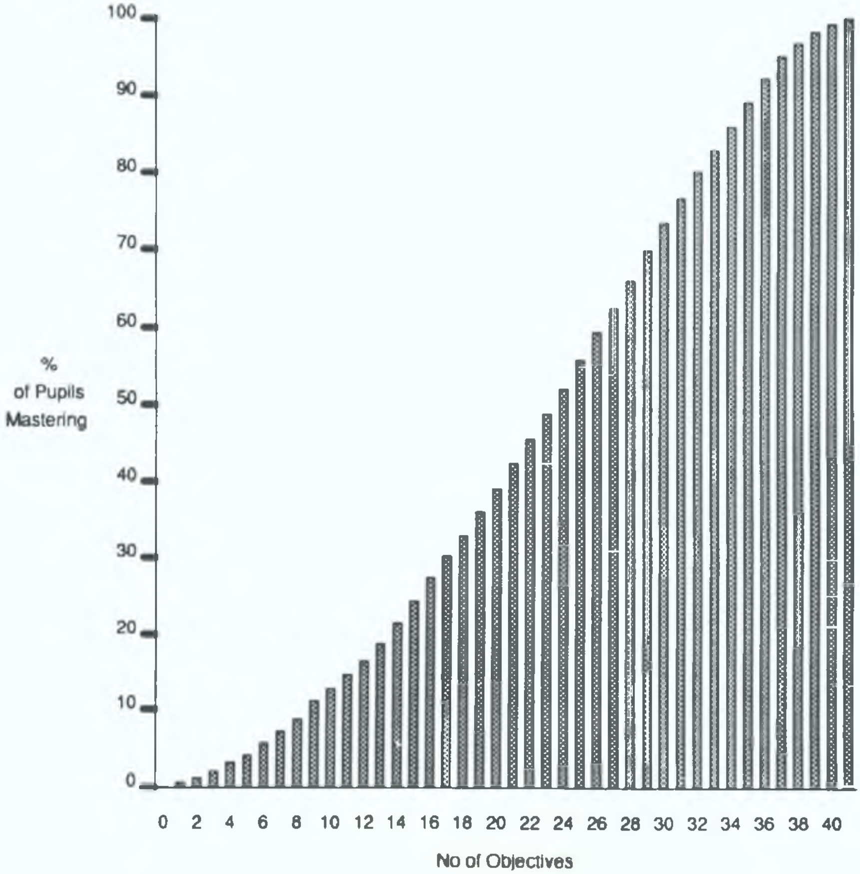


FIGURE 2

AVERAGE PERCENTAGE MASTERY BY CURRICULAR DOMAIN

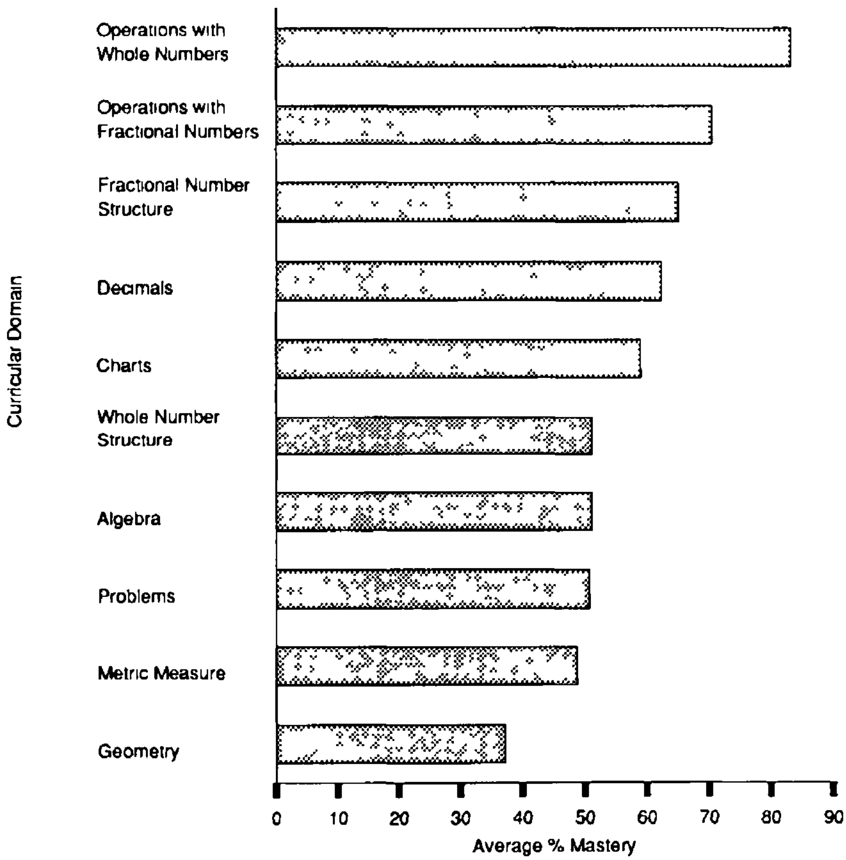
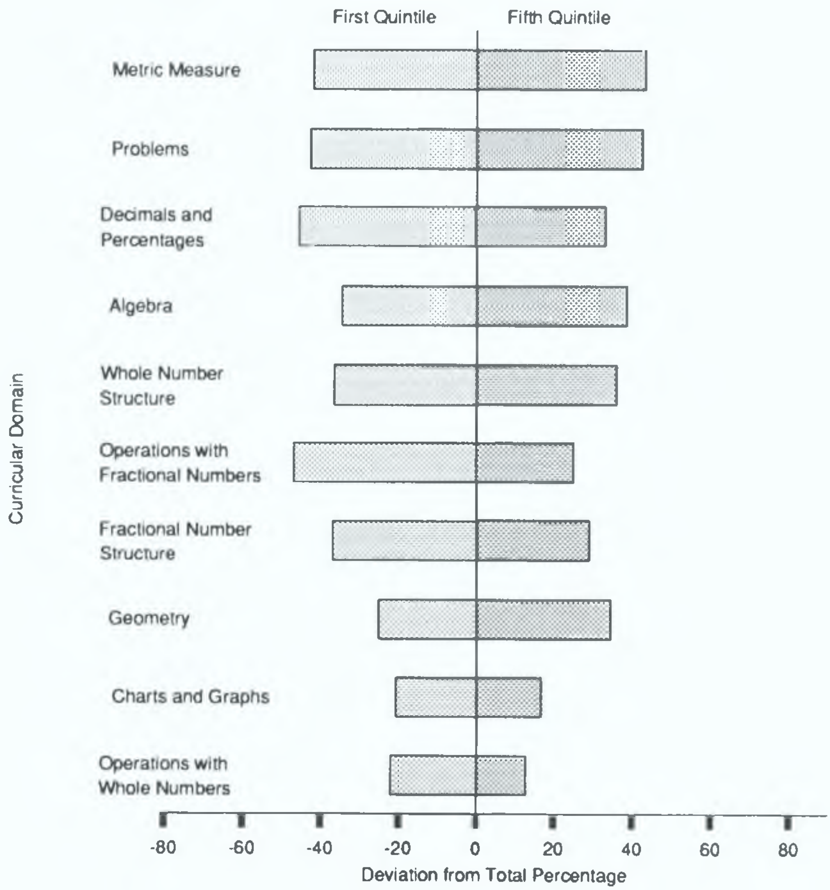


FIGURE 3

CURRICULAR DOMAIN BY LEVEL OF MATHEMATICS ACHIEVEMENT



Level of Mathematical Ability

To explore further the differences between high and low mathematics achievers, the sample was divided into three groups on the basis of performance on the test as a whole (using all 123 items) The first quintile group (n=498) was made up of pupils scoring at or below the 20th percentile on the total test (the 20% of pupils in the sample with the lowest total scores) The second to fourth quintile group (n=1,384) was made up of pupils between the 20th and the 80th percentile (the 60% of the sample with intermediate total scores) The fifth quintile group (n=495) consisted of all those scoring above the 80th percentile (the top-scoring 20%)

Figure 3 shows the difference in mastery levels between the first and fifth quintile groups for each curricular domain, expressed as a deviation from the average for the total sample The purpose of this graph is to show the domains where there is the greatest difference between the high (fifth quintile) and low (first quintile) performing groups The domains with the greatest difference are Metric Measure and Problems, while those with the least difference are Operations with Whole Numbers and Charts and Graphs

CONCLUSION

An examination of the dimensionality of the criterion-referenced mathematics test used in a survey of the achievements of sixth-standard pupils revealed that performance on the test could be explained in terms of a single latent trait of mathematical ability This finding finds support in the examination of the relationship between general mathematics and achievement in curricular areas described in this paper Since, however, the general trait identified as underlying performance on the test fell well short of explaining all the variance in performance, factors other than the major trait are clearly implicated

When the psychometric properties of items comprising the test were examined, the average difficulty level of the items was found to be about the same as that of items normally used in norm-referenced tests while the discriminations indices of over 90% of items could, according to criteria established by Ebel (1972), be regarded as 'good' or 'reasonably good' Furthermore, the internal consistency of the test as a whole was found to be above the minimum recommended for the use of test results at the level of the individual student We may conclude from these findings that it would be reasonable to interpret scores on the test as a whole as a measure of mathematical ability, though an inspection of area scores could be helpful in providing a more differentiated picture of pupils' performance

The analysis of the distribution of mathematical achievement among sixth class pupils revealed a very large difference between the most able and least able pupils. While the high performance levels of the more able is obviously a source of satisfaction, the fact that substantial numbers of pupils are completing their primary schooling without mastering the objectives of the mathematics curriculum must be a cause for serious concern.

REFERENCES

Ebel, R.E., (1972). *Essentials of educational measurement.* Englewood Cliff, New Jersey: Prentice-Hall.

Éire. An Roinn Oideachais, Aonad Curaclam, Brainse an Bhunoideachais (1985). *Tuairisc ar theagasc na matamaitice i rang VI (1984).* Áth Cliath: Údar.

Harman, H.H. (1976). *Modern factor analysis.* Chicago: University of Chicago Press.