DOCUMENT RESUME

ED 393 932 TM 024 961

AUTHOR Shorey, Leonard
TITLE OCOD-CTTP Test Evaluation Report.
INSTITUTION Organization for Cooperation in Overseas Development, (St. Lucia).
PUB DATE Jan 91
NOTE 46p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Difficulty Level; Elementary Secondary Education; Evaluation Methods; Foreign Countries; Psychometrics; *Science Education; *Social Studies; *Standardized Tests; Testing Programs; Test Items; *Test Reliability; Test Validity
IDENTIFIERS *Discrimination Index; Dominica; Grenada; Saint Lucia; Saint Vincent; *Split Half Test Reliability

ABSTRACT

Tests in social studies and integrated science given in Saint Vincent, Saint Lucia, Grenada, and Dominica were analyzed by the Organization for Co-operation in Overseas Development (OCOD) Comprehensive Teacher Training Program (CTTP) for discrimination, difficulty, and reliability, as well as other characteristics. There were 767 examinees for the Social Studies test. Scores were placed in descending order, and high (top 27%) and low (bottom 27%) were separated to extract 414 selectees. Difficulty indices and indices of discrimination were calculated, and reliability was evaluated. Analysis indicated that although the test had some good qualities, it also had a number of weaknesses that should be improved. Four modules of the Integrated Science Test were studied by splitting the test group into halves and comparing the performance of the halves. Samples of 48, 18, 8, and 18 students were used (samples rounded to an even number). Findings indicated that three of the modules were not satisfactory in their present forms, but that, although Module 4, with a reliability coefficient of 0.60, could be improved, it performed well above the other tests of the series. Five appendixes are detailed tables of results. (Contains 17 text tables and 3 references.) (SLD)

Report

OCOD

Test Evaluation

Prepared by
Dr. Leonard Shorey
Evaluation Consultant
January 17, 1991

DOCUMENTATION INFORMATION

Author / Editor / Compiler:
Dr. Leonard Shorey, Educational Consultant
Full Title:
OCOD - CTTP TEST EVALUATION REPORT
Sub Title:

Series Title:

Corporate Publisher:
Organization for Co-operation in Overseas Development
Mr. Herb C. Edwards, Executive Director

Date: January 1991
Pagination: 41 p.

Word processing / Formatting by:

Printed by: Organization for Co-operation in Overseas Development
Regional Office
Villa Apartments, Castries, St. Lucia, West Indies
Phone: 809 - 452-7955
FAX: 809 453-0127
Mrs. Gene Walters, Office Manager

Additional copies available from:
Price:

Additional related titles, or titles in the series:

1. Academic Upgrading of Unqualified Teachers in the Windward Islands by Distance Teaching, by Errol Miller 1987
2. Facilitating Adult Learning at a Distance: OCOD - CTTP Training Manual , by Lyngrid S. Rawlings 1987
3. Comprehensive Teacher Training Project: Plan of Operation for the Project - Phase 1, by Salasan Assocaites, Inc.
4. The Comprehensive Teacher Training Program (brochure)
5. OCOD Semi-Annual Report to CIDA : Canadian International Development Agency Re: Project # 0100920 - S 22568 15 April 1990
6. OCOD Semi - Annual Report to CIDA: Canadian International Development Agency RE: Project # 010920 - S 22568 15 November 1990

ORGANIZATION FOR CO-OPERATION IN OVERSEAS DEVELOPMENT
REGIONAL OFFICE
CASTRIES, ST. LUCIA , WEST INDIES

COMPREHENSIVE TEACHER TRAINING PROGRAMME

Report

OCOD

Test Evaluation

Prepared by
Dr. Leonard Shorey
Education Consultant
January, 1991

CONTENTS

ANALYSIS OF TEST DATA

The analysis and comment presented in this Report relate to tests in Social Studies and Integrated Science administered to pupils in St Vincent, St Lucia, Grenada and Dominica.

Components of the Analysis

The data analysis focuses on the following aspects of the tests:

Index of Discrimination
Difficulty Index
Reliability

In addition the Report indicates other relevant analyses which shed light on the strengths or shortcomings of the tests, viz.:

Expected Chance Score
Ideal Mean Score
Actual Mean Score
Standard Deviation

SOCIAL STUDIES

It is generally agreed that analyses of tests can satisfactorily be made by comparing the performances of the students in the top 27 percent of the testees with the performances of students in the bottom 27 percent of the testees. (Ebel, 1972)

This approach proved satisfactory for analysis of the Social Studies test where there were some 767 testees. However, in the case of the Integrated Science tests the numbers were

considerably smaller. In these tests, therefore, an acceptable alternative such as that noted by Crocker (1974) was used, i.e. the testee group was split into halves and the performance of the two halves compared.

Index of Discrimination

An important aspect of any item in a multiple choice test is the extent to which it discriminates (or differentiates) between the "good" students and the "poor" ones. Hence items which do not discriminate at all, i.e. which are answered correctly by all pupils or are answered incorrectly by all pupils, do not help to distinguish between better and poorer performers.

Where the difference in proportions of correct responses is 0.40 or higher, the item is considered to discriminate satisfactorily. Following Ebel (p 399) Indices of Discrimination can be evaluated as follows:

Index of Discrimination	Item Evaluation
0.40 and up	Very good items
0.30 to 0.39	Reasonably good but possibly subject to improvement
0.20 to 0.29	Marginal items, usually needing and being subject to improvement
Below 0.19	Poor items, to be rejected or improved by revision

As indicated previously there were 767 pupils for whom Social Studies answer sheets were available. The scores of these

pupils were placed in descending order and two groups selected, the first, (HIGH), comprising the top 27 percent of the pupils and the second, (LOW), the bottom 27 percent. Each group therefore contained 207 respondents, giving a total of 414 selectees.

Difficulty Indices of the items and Indices of Discrimination were then calculated. The resulting data are presented in *Appendix A* which shows, as well, the number of pupils from each group, HIGH and LOW, who selected each option in each item. The Appendix also indicates, by an asterisk (*), the correct response to each item.

Subsequent to this analysis the data were further broken down into the categories suggested by Ebel (see above), and Table 1 (following) presents the breakdown showing which items are to be found in the various discrimination ranges. In this Table the items are arranged in descending order of Discrimination Indices within each Discrimination range.

TABLE 1

ANALYSIS OF SOCIAL STUDIES ITEMS
(Arranged in descending order by Discrimination Index)

(Discrimination Range 0.40 and up:
"Very good items")

ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL N=414	DIFF INDEX	INDEX OF DISCRIM
14B	139	47	186	92		0.34	**0.44**
18C	145	59	204	86		0.35	**0.42**
20B	158	73	231	85		0.38	**0.41**
11D	149	65	214	84		0.36	**0.41**
6D	173	91	264	82		0.42	**0.40**

(Discrimination Range 0.30 to 0.39: "Reasonably good but possibly subject to improvement")

ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL N=414	DIFF INDEX	INDEX OF DISCRIM
8B	109	30	139	79		0.26	**0.38**
16C	183	104	287	79		0.44	**0.38**
4A	186	109	295	77		0.45	**0.37**
10D	148	75	223	73		0.36	**0.35**
5C	191	122	313	69		0.46	**0.33**
17A	170	102	272	68		0.41	**0.33**
12B	138	72	210	66		0.33	**0.32**
19A	185	119	304	66		0.45	**0.32**

(Discrimination Range 0.20 to 0.29: "Marginal items, usually needing, and being subject to, improvement.")

ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL N=414	DIFF INDEX	INDEX OF DISCRIM
7B	108	47	155	61		0.26	**0.29**
3D	191	144	335	47		0.46	**0.23**
13A	90	44	134	46		0.22	**0.22**

(Discrimination Range below 0.20: "Poor items, to be rejected, or improved by revision")

ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL N=414	DIFF INDEX	INDEX OF DISCRIM
2B	141	107	248	34		0.34	**0.16**
15A	153	120	273	33		0.37	**0.16**
9C	57	30	87	27		0.14	**0.13**
1C	193	168	361	25		0.47	**0.12**

As will be noted, there are five items (14B, 18C, 20B, 11D, 6D) which have a Discrimination Index equal to or greater than 0.40 and which can therefore be considered "Very good" so far as discrimination is concerned.

There are another eight items (8B, 16C, 4A, 10D, 5C, 17A, 12B, 19A) whose Discrimination Indices lie between 0.30 and 0.39 and which can be deemed "Reasonably good, but possibly subject to improvement".

Three items (7B, 3D, 13A) fall into the third category, i.e. items with Discrimination Indices between 0.20 and 0.29. These must be considered "Marginal items" which will need improvement if they are to be used in future tests.

In the fourth and final category there are four items (2B, 15A, 9C, 1C) all of which have Discrimination Indices below 0.20 and which therefore fall into the category of "Poor items, to be rejected, or improved by revision."

In the event that revision of certain items is undertaken it will be useful to bear in mind that "to obtain multiple-choice test items with good discriminating power, it is more efficient to revise faulty ones on the basis of item analysis data than to design new ones." (Ahmann and Glock, p 202)

Difficulty Level

Index of Difficulty

The "Index of Difficulty" used in this analysis is based on the method recommended by Crocker (p 77) viz., **the proportion of the group getting the item right.** As Crocker also points out "The more people that get a question right, the higher will be the question's *Difficulty Index*. An odd convention, but one that is generally used."

It is also worthwhile noting Ebel's comment (p 395) on Difficulty Index:

> The index of difficulty of a test item is not solely the property of that item. It reflects also the ability of the group responding to the item. Hence, instead of saying, "The index of difficulty for this item is 56 percent," it would be better to say, "When this item was administered to that particular group, its index of difficulty was 56 percent".

Table 2 shows the test items arranged in increasing order of Difficulty Indices.

TABLE 2

ANALYSIS OF SOCIAL STUDIES ITEMS
(Arranged in ascending order of Difficulty Indices)

ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL N=414	DIFF INDEX	INDEX OF DISCRIM
9C	57	30	87	27		0.21	0.13
13A	90	44	134	46	*	0.32	0.22
8B	109	30	139	79	*	0.34	0.38
7B	108	47	155	61	*	0.37	0.29
14B	139	47	186	92	**	0.45	0.44
18C	145	59	204	86	**	0.49	0.42
12B	138	72	210	66	**	0.51	0.32
11D	149	65	214	84	**	0.52	0.41
10D	148	75	223	73	**	0.54	0.35
20B	158	73	231	85	**	0.56	0.41
2B	141	107	248	34	**	0.60	0.16
6D	173	91	264	82	*	0.64	0.40
17A	170	102	272	68	*	0.66	0.33
15A	153	120	273	33	*	0.66	0.16
16C	183	104	287	79	*	0.69	0.38
4A	186	109	295	77		0.71	0.37
19A	185	119	304	66		0.73	0.32
5C	191	122	313	69		0.76	0.33
3D	191	144	335	47		0.81	0.23
1C	193	168	361	25		0.87	0.12

Generally speaking an item of 50 percent difficulty is the ideal since it contributes the maximum amount of information about the relative levels of achievement. However, it is also common to consider as acceptable, items falling within the range 40 percent and 60 percent (Crocker) or items between 30 percent and 70 percent (Ebel).

The preceding Table shows that if the 40 percent to 60 percent range is used then there are seven items (marked **) that fall within this range. On the other hand if the 30 percent to 70 percent range is accepted than all the fourteen asterisked items would be considered to have an acceptable Difficulty Index.

Items outside the acceptable range must be considered in need of revision to improve them in this respect.

Reliability and related matters

Another way in which a multiple choice test can be evaluated is by examining the spread of the scores, and this is best done by calculating the Standard Deviation of these scores.

The reason for considering the spread of the scores is that "If items in a test tend to discriminate clearly between good and poor students, the test scores will tend to vary widely." Ebel (p 377). The spread of the scores is therefore a pointer to the effectiveness of the test, as a whole, in its power to discriminate between good and poor students.

Ebel goes on to make the point that "A standard deviation of one-sixth the range between highest possible score and the expected chance score is quite satisfactory." (p 378)

Another important aspect of test items is the relationship between the "Ideal Mean" and the "Actual Mean". The Ideal Mean is best defined as "a point midway between the maximum possible score and the expected chance score", where "The expected chance score equals the number of items in the test divided by the number of choices per item." (Ebel pp 375,376)

Moreover "If the average score is very much higher or very much lower than the midpoint of the range between highest possible and expected chance scores, the test may be inefficient. That is, it may waste the student's time trying to answer questions that almost no one can answer correctly, or reading and answering questions that almost every one answers correctly." (Ebel, p 376)

In relation to the preceding comments, Table 3 presents additional data which allow further insights into the qualities of the Social Studies test.

TABLE 3

MEANS AND STANDARD DEVIATIONS

Expected Chance Score	=	5
Ideal Mean (20+5)/2	=	12.5
Actual Mean	=	11.46
Satisfactory Standard Deviation (20+5)/6	=	4.167
Actual Standard Deviation	=	2.567

As the Table shows quite clearly the Actual Mean of the

scores approaches the Ideal Mean. This suggests that the test was reasonably efficient in that it did not require students to waste their time trying to answer questions that "almost no one (could) answer correctly, or reading and answering questions that almost every one answer(ed) correctly." Thus there were no test items which remained unanswered by all students, and there were no items which were answered correctly by all students.

Examination of the Standard Deviation of the students' scores (2.567) indicates that it is quite far from a "satisfactory" Standard Deviation. As Table 3 shows, a desirable S.D. for scores in this test would be of the order of 4.167. An even greater S.D. would have been better since "For some good tests, the standard deviation is more than one-fourth the available range." (Ebel p 378)

Consideration of the S.D. is important because "If a test is too hard, too easy or composed of too many poorly discriminating items, it will yield scores having a small standard deviation." (p 378)

Distracters

An essential element in a good multiple choice test is that the distracters, i.e. the incorrect options, should be such as would indeed "distract" students who are not confident of the correct answer. Therefore the chief characteristic of a

good "distracter" is that it should offer to the student answering the test a very "plausible" but incorrect option.

But good distracters must also satisfy other criteria. The best distracters are therefore those likely to be chosen by "poor" students rather than by "good" ones. Indeed where a distracter is more frequently chosen by good rather than by poor students it can produce what is called "inverse discrimination", i.e. it discriminates against the good students and in favour of the poor ones - an obviously undesirable result

As Ahmann and Glock put it "A good distracter should have a differential attractiveness, that is, it should be more attractive to the lower group than the upper group. In this way it contributes to the discriminating power of the test item." (p 192)

Some indication of the relative effectiveness of the distracters in this test can be gleaned by considering the data found in the column headed "HIGH - LOW". A high negative value indicates that a much higher proportion of "poor" students were attracted to this item than were "good" students, whereas small negative values indicate that the distracter was relatively less effective in this respect.

It should also be noted that where only a few students chose

a particular distracter this would suggest that that distracter was also relatively inefficient, and did not perform very well. Finally, where the "High - Low" value is **positive** this indicates negative discrimination, i.e. the distracter item "caught" good students more frequently than poor ones. This, also, is undesirable in a test item.

Some examples from Appendix A will serve to illustrate the points just made.

In Item 4 option 4D is a very good distracter as indicated by its high negative value, -73. Options 5D, 8D, 9B, 10B, 11A, 12D, 14D, 16D, 17D, 18A, 18B and 20C also have high negative values and therefore functioned well as distracters.

On the other hand distracters like options 1D, 2C, 4C, 8A and so on were very ineffective in this respect.

Item 19A is an example of an item whose distracters were quite evenly matched in their effectiveness, with values of -20, -21, and -21 respectively. In item 4 on the other hand the responses to the distracters are much more uneven as indicated by values of 0, -1, and -73; considerable variation in the effectiveness of the distracters used is also evident in, say, item 8 with distracter values of -3, -13, -59.

Since the effectiveness of a distracter depends on the number of students it "distracts", it is clear that some distracters were of little or no value, for example, 1D, 4B, 4C, 8A, 13C

etc., with values of -1, 0, -1, -3, 0 respectively. Two distracters, 9A, 9D, operated in an inverse way in that they proved more attractive to the "good" students than to the "poor" ones. As the Appendix shows these two distracters have **positive** values, viz., +2, +10 respectively.

Reliability

A test's *reliability* is an important characteristic of the test since "the term 'reliability' means the consistency with a set of test scores measure whatever they do measure." (Ebel, p 409)

Reliability is also closely connected with another desirable characteristic of any test, viz., its *validity*, where "'validity' means the accuracy with which a set of test scores measure what they ought to measure." (Ebel, p 409)

Calculation of the Reliability of the tests reported on in these analyses uses the Kuder-Richardson formula. (Ebel, p 414)

$$r = \left[\frac{k}{k-1}\right]\left[1 - \frac{\text{Sum}(pq)}{\text{Variance}}\right]$$

where r = reliability coefficient
k = number of items in the test
p = the proportion of correct responses to any one item, and
q = the proportion of incorrect responses to the same item

Analysis of the test data shows that for the Social Studies test the reliability coefficient is 0.36 (i.e. **r = 0.36**). This value must be considered undesirably low for, as Ebel notes (p 421), "most test constructors are reasonably well satisfied if their tests yield reliability coefficients in the vicinity of 0.90." However, Ebel does acknowledge that "The reliability coefficients ordinarily obtained for teacher-made tests tend to fall considerably short of this goal."

Conclusion

Analysis of the Social Studies test indicates that although it has some good qualities it also has a number of weaknesses which make it less satisfactory than it could be, and these have already been identified. In view of the weaknesses evident it would seem desirable that if the test itself or selected items in the test are to be re-used then efforts should be made to improve the quality of the weak items.

The usefulness of the information provided by this analysis is underscored by Ahmann and Glock in their book *"Evaluating Pupil Growth"*. Dealing with Diagnosing Inadequacies in Achievement, they state:

> ...the raw score of a pupil or the arithmetic mean of the raw scores of a class may be interpreted to mean that one or more pupils are having trouble, but such scores will not tell what the trouble is or where it is

> located. To locate the nature of the trouble, an item-by-item inspection of the test is necessary. (p 195)

When once this item by item analysis has been done "the strengths and weaknesses in the achievement of the pupils can be quickly found by examining the areas of achievement involved in the test items identified..." (p 195)

INTEGRATED SCIENCE

MODULE 10

There were 49 scripts available for analysis with respect to Module 10 of Integrated Science. These were divided into two groups each containing 24 students constituting the HIGH and LOW categories respectively. The test itself contained 10 items. The overall data are presented in *Appendix B*.
The correct options are indicated by an asterisk (*).

Index of Discrimination

Table 4 below shows the items rank ordered by Discrimination Indices and in Table 5, following, the items are placed in the discrimination ranges already referred to.

TABLE 4

ANALYSIS OF INTEGRATED SCIENCE - MODULE 10 ITEMS
(Arranged in descending order by Discrimination Indices)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 48	DIFF LEVEL	INDEX OF DISCRIM
4C	22	8	30	14		0.54	0.58
6B	11	1	12	10		0.77	0.42
5D	21	12	33	9		0.56	0.38
1B	15	7	22	8		0.69	0.33
7B	10	2	12	8		0.79	0.33
3B	16	8	24	8		0.67	0.33
2C	21	14	35	7		0.56	0.29
8D	24	21	45	3		0.50	0.13
9C	12	10	22	2		0.75	0.08
10A	15	16	31	-1		0.69	-0.04

TABLE 5

ITEMS ARRANGED IN DISCRIMINATION CATEGORIES

Discrimination Index > 0.40

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 48	DIFF LEVEL	INDEX OF DISCRIM
4C	2	8	30	14		0.54	**0.58**
6B	11	1	12	10		0.77	**0.42**

Discrimination Indices 0.30 to 0.39

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 48	DIFF LEVEL	INDEX OF DISCRIM
5D	21	12	33	9		0.56	**0.38**
7B	10	2	12	8		0.79	**0.33**
3B	16	8	24	8		0.67	**0.33**
1B	15	7	22	8		0.69	**0.33**

Discrimination Index 0.20 to 0.29

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 48	DIFF LEVEL	INDEX OF DISCRIM
2C	21	14	35	7		0.56	**0.29**

Discrimination Indices < 0.20

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 48	DIFF LEVEL	INDEX OF DISCRIM
8D	24	21	45	3		0.50	**0.13**
9C	12	10	22	2		0.75	**0.08**
10A	15	16	31	-1		0.69	**-0.04**

The preceding Table shows, as did Table 4, the correct responses to each item in the test together with related information such as the item's Difficulty Index and Discrimination Index.

Table 5 also shows the items placed in the categories previously identified as those to be used in this analysis. As this Table indicates, there were two items (4C and 6B) which had a Discrimination Index greater than 0.40 and which

could therefore be considered "Very good".

Four of the ten items (5D, 7B, 3B, 1B) were in the "Reasonably good" category, with Discrimination Indices ranging from 0.30 to 0.39. Of the remaining four items, one, 2C, would be considered a "Marginal item", while the last three items (8D, 9C, 10A) would be considered "Poor items to be rejected or improved by revision."

Table 6 below shows the items in the Integrated Science Module 10 test arranged by increasing Difficulty Level. If we consider as acceptable items with Difficulty Indices between 0.30 and 0.70 then six of the ten items in this test would meet these criteria. These items are marked with an asterisk.

TABLE 6

ANALYSIS OF INTEGRATED SCIENCE - MODULE 10 ITEMS
(Arranged in ascending order of Difficulty Indices)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 18	DIFF INDEX	INDEX OF DISCRIM
7C	1	1	2	0		0.11	0.00
4D	2	2	4	0		0.22	0.00
3A	3	1	4	2		0.22	0.22
9B	1	5	6	-4	*	0.33	-0.44
5C	5	1	6	4	*	0.33	0.44
2D	7	0	7	7	*	0.39	0.78
8A	6	2	8	4	*	0.44	0.44
10D	8	4	12	4	*	0.67	0.44
6B	6	6	12	0	*	0.67	0.00
1C	9	4	13	5		0.72	0.56

Distracters

On the whole the distracters in this test do not appear to have functioned very effectively. As Appendix B shows several

of them did not attract responses from students and in this sense would be considered poor distracters since they did not help to discriminate between "good" and "poor" students. For example, in item 8 distracters A, B, and C had no appeal to any student in the "High" group, and each of them attracted only one student from the "Low" group; so did some other options. Likewise in item 2 no student chose distracter D, and the same thing happened with other distracters as well.

In Module 10 the most successful distracters were 6D and 9D which attracted 30 and 22 responses respectively. It should, however, be noted that distracter 9D appealed equally to students in the "High" and "Low" groups, and was therefore not satisfactorily efficient. Careful consideration of these two distracters might well shed some light on possible reasons why they were so much more successful than the other distracters, and might consequently suggest how the phrasing or focus of less successful distracters could be improved.

Reliability and Related Matters

Further analysis provides additional insights about the test as a whole, and the relevant data are provided in Table 7 following.

TABLE 7

MEANS AND STANDARD DEVIATIONS

Expected Chance Score	=	2.5
Ideal Mean (10+2.5)/2	=	6.25
Actual Mean	=	5.53
Satisfactory Standard Deviation (10+2.5)/6	=	2.08
Actual Standard Deviation	=	1.94
r	=	0.51

Focusing attention on the Actual Mean (5.53) it is to be noted that this is lower than the Ideal Mean score. Indeed examination of Appendix B shows that nine of ten items were correctly answered by fewer than 75 percent of the pupils, and there were two items (6B) and (7B) which only 12 of the 48 students in the group (25 %) answered correctly. This bears out the inference to be drawn from the Actual Mean that, in this sense, the test is likely to have been somewhat inefficient in that "it may (have) waste(d) the student's time trying to answer questions that almost no one (could) answer correctly..." (Ebel, p 376)

It is worth noting, however, that the Mean of 5.53 means that the test still has some merit for, as Ebel indicates (p 375)

> In most classroom situations a test in which the average score is somewhat more than half the maximum possible score will be appropriate in difficulty.

The Standard Deviation of 1.94 is slightly below the desirable S.D. of 2.08, but it nonetheless indicates quite a reasonable spread of scores, and supports the view that the test has some desirable qualities.

The reliability of Module 10 turned out to be 0.51 (r = 0.51). This is lower than one would wish but as is well known the reliability of teacher-made classroom tests and similar tests tends to be relatively low.

Conclusion

The preceding analysis and comments indicate that while the Module 10 test has some good points it also has some weak ones to which attention would need to be devoted in the event of future use of the currently weak items.

INTEGRATED SCIENCE

MODULE 9

The full data for analysis of test Module 9 are presented in *Appendix C*. Table 8 shows the items arranged in descending order of Discrimination Indices, while Table 9 shows them arranged in ascending order of Difficulty Indices.

TABLE 8

ANALYSIS OF INTEGRATED SCIENCE - MODULE 9 ITEMS
(Arranged in descending order of Discrimination Indices)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 18	DIFF INDEX	DISCRIM INDEX
2D	7	0	7	7		* 0.39	**0.78**
1C	9	4	13	5		0.72	**0.56**
8A	6	2	8	4		* 0.44	**0.44**
10D	8	4	12	4		* 0.67	**0.44**
5C	5	1	6	4		* 0.33	**0.44**
3A	3	1	4	2		0.22	**0.22**
6B	6	6	12	0		* 0.67	**0.00**
7C	1	1	2	0		0.11	**0.00**
4D	2	2	4	0		0.22	**0.00**
9B	1	5	6	-4		* 0.33	**-0.44**

Data were available for 19 students. The final analysis is based on the top 9 students and the bottom 9 students, that is, a total of 18 of the 19 respondents.

For easier analysis Table 9 presents the items in the discrimination categories referred to previously.

TABLE 9

ANALYSIS OF INTEGRATED SCIENCE - MODULE 9 ITEMS
(Arranged in descending order of Discrimination Indices)

Discrimination Index > 0.40

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL	DIFF INDEX	INDEX OF DISCRIM
2D	7	0	7	7	18	0.39	**0.78**
1C	9	4	13	5	18	0.72	**0.56**
8A	6	2	8	4	18	0.44	**0.44**
10D	8	4	12	4	18	0.67	**0.44**
5C	5	1	6	4	18	0.33	**0.44**

Discrimination Index 0.20 - 0.29

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL	DIFF INDEX	INDEX OF DISCRIM
3A	3	1	4	2	18	0.22	**0.22**

Discrimination Index < 0.20

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL	DIFF INDEX	INDEX OF DISCRIM
6B	6	6	12	0	18	0.67	**0.00**
7C	1	1	2	0	18	0.11	**0.00**
4D	2	2	4	0	18	0.22	**0.00**
9B	1	5	6	-4	18	0.33	**-0.44**

As the preceding Table indicates, five of the ten items (50%) can be described as "very good", having Discrimination Indices greater than 0.40. One item (3A) would be considered a "marginal" item, having as it does a Discrimination Index between 0.20 and 0.29.

The remaining four items (6B, 7C, 4D, 9B) fare badly on this criterion, three of them showing no discrimination at all and one of them (9B) actually discriminating negatively, i.e.

being answered correctly more frequently by students in the "Low" group than by students in the "High" group. It is evident that these items will need to be discarded or will have to be improved significantly in this respect if they are to be used in future tests.

Difficulty Level

Table 10 presents the items in Module 9 arranged in ascending order of Difficulty Index. As the Table indicates six of the ten items (50%) have acceptable "Difficulty Levels", i.e. Difficulty Indices between 0.30 and 0.70. Improvement of the test in this respect would therefore require revision or replacement of the other items.

TABLE 10

ANALYSIS OF INTEGRATED SCIENCE - MODULE 9 ITEMS
(Arranged in ascending order of Difficulty Indices)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 18	DIFF INDEX	INDEX OF DISCRIM
7C	1	1	2	0		**0.11**	0.00
4D	2	2	4	0		**0.22**	0.00
3A	3	1	4	2		**0.22**	0.22
9B	1	5	6	-4		* **0.33**	-0.44
5C	5	1	6	4		* **0.33**	0.44
2D	7	0	7	7		* **0.39**	0.78
8A	6	2	8	4		* **0.44**	0.44
10D	8	4	12	4		* **0.67**	0.44
6B	6	6	12	0		* **0.67**	0.00
1C	9	4	13	5		**0.72**	0.56

Distracters

The distracters used were also several times ineffective. For

example, there were some six distracters which were not selected by any student, viz., 1B, 2A, 6C, 7A, 9C, 10C.

As in the analysis of Module 10 there are some aspects of this test which are commendable but there are also others which seriously impair the efficiency of the test. These defects will need to be corrected if the test is to be significantly improved.

Reliability and Related Matters

Table 11 below provides additional information about the items in this module.

TABLE 11

MEANS AND STANDARD DEVIATIONS

Expected chance score	= 2.5
Ideal Mean	= 6.25
Actual Mean	= 4.11
Satisfactory Standard Deviation (10 + 2.5)/6	= 2.08
Actual Standard Deviation	= 1.372
r	= -0.07

Looking first at the Actual Standard Deviation this is noticeably smaller than desirable and indicates that the test was not as effective as it should be in spreading the scores. In this area, therefore, the text exhibits significant weakness.

Comparison of the Actual Mean (4.11) with an Ideal Mean (6.25) also indicates that the test was less efficient than could have been desired. In this respect, also, test Module 9 is functioning less effectively than would be wished and its items would require improvement if the test itself is to be made more effective.

The reliability coefficient of -0.07 ($r = -0.07$) indicates quite clearly that no dependence can be placed on the results of this test and it speaks strongly for a complete revision of this test module.

Conclusion

Test Module 9 has a number of readily identifiable weaknesses, some of them major. It would seem therefore that this test would require revision and modification if it is to be used again.

INTEGRATED SCIENCE

MODULE 7

Test Module 7 was answered by nine students. For purposes of analysis these were divided into two groups of four each -the HIGH and LOW respectively. Summary data are presented in full in *Appendix D*, but relevant portions are given below for easy reference.

Index of Discrimination

As in preceding analyses items are categorized by four levels of Discrimination Index (See Table 12 below).

TABLE 12

ANALYSIS OF INTEGRATED SCIENCE - MODULE 7 ITEMS
(Arranged in descending order of Discrimination Indices)

Discrimination Indices > 0.40

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 8	DIFF INDEX	DISCRIM INDEX
6D	4	3	7	1		0.50	**0.50**
9A	4	1	5	3		0.50	**0.50**

Discrimination Indices 0.30 - 0.39

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 8	DIFF INDEX	DISCRIM INDEX
3A	3	3	6	0		0.38	**0.38**
1B	3	3	6	0		0.38	**0.38**
5C	3	1	4	2		0.38	**0.38**
7C	3		3	3		0.38	**0.38**

Discrimination Index between 0.20 and 0.29

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 8	DIFF INDEX	DISCRIM INDEX
2B	2	1	3	1		0.25	**0.25**

Discrimination Indices < 0.20

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 8	DIFF INDEX	DISCRIM INDEX
4A	1	1	2	0		0.13	**0.13**
10B	1		1	1		0.13	**0.13**
8D			0	0		0.00	**0.00**

In Module 7, two of the items (6D and 9A) lie above the 0.40 point and are therefore well described as "very good items." Four items (3A, 1B, 5C, 7C) have Discrimination Indices within the range 0.30 and 0.39 and are therefore "Reasonably good", and there is one item (2B) which, with a Discrimination Index of 0.25 must be considered "marginal" and would require improvement for effective use.

Two of the remaining three items have Indices of 0.13 (below 0.20) and must therefore be considered "Poor items". These should either be "rejected or improved by revision".

Finally there is one item, 8D, which was not answered by any of the students and which therefore played no part in helping with their assessment. It would be useful to examine this item together with the "poor items" since they all performed so badly.

Difficulty Level

Table 13 below shows the analysis of item responses arranged in ascending order of Difficulty Indices. Six of the ten items (asterisked) have Difficulty Indices within the range 0.30 to

0.70 and can therefore be considered quite acceptable in this respect.

TABLE 13

ANALYSIS OF INTEGRATED SCIENCE - MODULE 7 ITEMS
(Arranged in ascending order of Difficulty Indices)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 8	DIFF	DISCRIM INDEX
8D			0	0		0.00	0.00
4A	1	1	2	0		0.13	0.13
10B	1		1	1		0.13	0.13
2B	2	1	3	1		0.25	0.25
7C	3		3	3	*	0.38	0.38
3A	3	3	6	0	*	0.38	0.38
1B	3	3	6	0	*	0.38	0.38
5C	3	1	4	2	*	0.38	0.38
6D	4	3	7	1	*	0.50	0.50
9A	4	1	5	3	*	0.50	0.50

In connection with the above comments it must, of course, be remembered that the data analysed were drawn from only 9 students and that when such very small numbers of data are used for analysis one is likely to encounter quite extreme results. In this sense, then, the results from these nine candidates almost certainly do not provide a satisfactory estimate of the usefulness of the items in the test.

A comment from Ebel (p 391) is relevant at this point. He states that "even though one cannot determine the discrimination indices of individual items reliably without using large samples of student responses, item analysis based on small samples is still worthwhile as a means of overall test improvement."

Reliability and Related Matters

Despite the preceding comments, however, the supplementary data provided for other Modules are still presented so as to complete the picture for Module 7.

TABLE 14

MEANS AND STANDARD DEVIATIONS

Expected Chance Score	=	2.5
Ideal Mean (10+2.5)/2	=	6.25
Actual Mean	=	4.67
Satisfactory S.D. (10+2.5)/6	=	2.08
Actual S.D.	=	1.41
r	=	0.11

The data above indicate that the test must be considered "inefficient" since if falls down badly on two major counts. The first is that the test Mean is well below the Ideal Mean, indicating the possibility that the test could contain items "that almost no one can answer correctly". There was actually one such case, item 8.

The test also falls down because of its low Standard Deviation. As was pointed out by reference to Ebel, "if items in a test tend to discriminate clearly between good and poor, students, the test scores will tend to vary widely." A low S.D. therefore indicates that the spread of scores is undesirably narrow for test effectiveness.

In light of the preceding it is not surprising that this test Module should have a very low reliability coefficient, ($r = 0.11$). This merely helps to emphasize the fact that very little trust can be placed on students' results on this test.

All in all the Module 7 has not shown up well, but at the same time, as was previously acknowledged, the very limited data on which this analysis is based almost certainly do not do the test itself justice. Many more student responses would likely be needed for a reasonably accurate assessment of the test's strengths and shortcomings.

INTEGRATED SCIENCE

MODULE 4

With respect to Module 4 data were available for 18 students, and analysis was based on comparison between the 9 HIGH and the 9 LOW respondents, determined as previously indicated. The full data about responses to items are presented in *Appendix E*. However, unlike the other modules in the Science series Module 4 had only five items, instead of ten.

Index of Discrimination

The data presented in Table 15 below indicate that three of the items in test Module 4 had Discrimination Indices greater than 0.40 and are therefore in the "Very good" category while the remaining two items, 3A and 4D, are "Reasonably good".

TABLE 15

ANALYSIS OF INTEGRATED SCIENCE - MODULE 4 ITEMS (Arranged in descending order of Discrimination Indices)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 18	DIFF INDEX	DISCRIM INDEX
Discrimination Indices > 0.40							
5C	9	5	14	4		0.50	**0.50**
2C	8	10	18	-2		0.44	**0.44**
1B	8	4	12	4		0.44	**0.44**
Discrimination Indices 0.30 - 0.39							
3A	7	5	12	2		0.39	**0.39**
4D	6	2	8	4		0.33	0.33

Difficulty Index

Table 16 below shows the correct responses arranged in ascending order by Difficulty Index and, as is evident, all the items have acceptable Difficulty indices since they all lie between 0.30 and 0.70.

TABLE 16

ANALYSIS OF INTEGRATED SCIENCE - MODULE 4 ITEMS
(Arranged in ascending order of Difficulty Indices)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL N = 18	DIFF INDEX	DISCRIM INDEX
4D	6	2	8	4		0.33	0.33
3A	7	5	12	2		0.39	0.39
1B	8	4	12	4		0.44	0.44
2C	8	10	18	-2		0.44	0.44
5C	9	5	14	4		0.50	0.50

Reliability and Related Matters

With respect to the other criteria used in the analysis of the data from this series of tests, the relevant information is given below in Table 17.

TABLE 17

MEANS AND STANDARD DEVIATIONS

Expected Chance Score (5/4)	= 1.25
Ideal Mean (5+1.25)/2	= 3.125
Actual Mean	= 3.06
Satisfactory S.D. (5+1.25)/6	= 1.04
Actual S.D.	= 1.47
r	= 0.60

Conclusion

In test Module 4 the Actual Mean (3.06) is quite close to the Ideal Mean (3.125) and, in fact, is the closest relationship of this kind to be found in any of the test results analysed. This closeness suggests a high level of efficiency in the test as it easily meets the condition given by Ebel that the Mean score should be "somewhat more than half the maximum possible score".

The general quality of this test is further indicated by the finding that the Actual S.D. (1.47) is indeed higher than a merely Satisfactory S.D., viz., 1.06. This result indicates that the spread of the scores is quite good, despite the small number of respondents.

Finally, the test as a whole has a reliability coefficient of 0.60. This is of course well below the coefficient which a test should desirably have, but it is still well above the coefficients of the other tests in this series. It is also to be remembered, as was pointed out earlier, that teacher-made tests quite frequently have low coefficients.

On these grounds, then, it would seem reasonable to conclude that although this test can be improved, it deserves commendation.

APPENDIX A

ANALYSIS OF SOCIAL STUDIES DATA
(Asterisks indicate the correct item responses)

	ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL	DIFF LEVEL	INDEX OF DISCRIM
	1A	5	23	28	-18	414		-0.09
	1B	9	14	23	-5	414		-0.02
*	1C	193	168	361	25	414	0.87	0.12
	1D	0	1	1	-1	414		-0.00
	2A	61	75	136	-14	414		-0.07
*	2B	141	107	248	34	414	0.60	0.16
	2C	3	7	10	-4	414		-0.02
	2D	1	12	13	-11	414		-0.05
	3A	8	34	42	-26	414		-0.13
	3B	3	7	10	-4	414		-0.02
	3C	5	19	24	-14	414		-0.07
*	3D	191	144	335	47	414	0.81	0.23
*	4A	186	109	295	77	414	0.71	0.37
	4B	2	2	4	0	414		0.00
	4C	2	3	5	-1	414		-0.00
	4D	17	90	107	-73	414		-0.35
	5A	7	15	22	-8	414		-0.04
	5B	1	18	19	-17	414		-0.08
*	5C	191	122	313	69	414	0.76	0.33
	5D	8	47	55	-39	414		-0.19
	6A	4	25	29	-21	414		-0.10
	6B	10	52	62	-42	414		-0.20
	6C	18	31	49	-13	414		-0.06
*	6D	173	91	264	82	414	0.64	0.40
	7A	19	36	55	-17	414		-0.08
*	7B	108	47	155	61	414	0.37	0.29
	7C	51	77	128	-26	414		-0.13
	7D	27	42	69	-15	414		-0.07
	8A	7	10	17	-3	414		-0.01
*	8B	109	30	139	79	414	0.34	0.38
	8C	69	82	151	-13	414		-0.06
	8D	22	81	103	-59	414		-0.29
	9A	38	36	74	2	414		0.01
	9B	47	81	128	-34	414		-0.16
*	9C	57	30	87	27	414	0.21	0.13
	9D	58	48	106	10	414		0.05

	ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL	DIFF LEVEL	INDEX OF DISCRIM
	10A	1	5	6	-4	414		-0.02
	10B	50	97	147	-47	414		-0.23
	10C	8	29	37	-21	414		-0.10
*	10D	148	75	223	73	414	0.54	0.35
	11A	15	46	61	-31	414		-0.15
	11B	18	44	62	-26	414		-0.13
	11C	23	39	62	-16	414		-0.08
*	11D	149	65	214	84	414	0.52	0.41
	12A	28	43	71	-15	414		-0.07
*	12B	138	72	210	66	414	0.51	0.32
	12C	4	16	20	-12	414		-0.06
	12D	36	72	108	-36	414		-0.17
*	13A	90	44	134	46	414	0.32	0.22
	13B	100	124	224	-24	414		-0.12
	13C	5	5	10	0	414		0.00
	13D	11	31	42	-20	414		-0.10
	14A	10	19	29	-9	414		-0.04
*	14B	139	47	186	92	414	0.45	0.44
	14C	15	37	52	-22	414		-0.11
	14D	42	97	139	-55	414		-0.27
*	15A	153	120	273	33	414	0.66	0.16
	15B	4	12	16	-8	414		-0.04
	15C	22	38	60	-16	414		-0.08
	15D	27	32	59	-5	414		-0.02
	16A	14	38	52	-24	414		-0.12
	16B	4	25	29	-21	414		-0.10
*	16C	183	104	287	79	414	0.69	0.38
	16D	6	34	40	-28	414		-0.14
*	17A	170	102	272	68	414	0.66	0.33
	17B	14	29	43	-15	414		-0.07
	17C	6	11	17	-5	414		-0.02
	17D	17	59	76	-42	414		-0.20
	18A	38	91	129	-53	414		-0.26
	18B	13	41	54	-28	414		-0.14
*	18C	145	59	204	86	414	0.49	0.42
	18D	8	12	20	-4	414		-0.02

	ITEM	HIGH	LOW	HIGH +LOW	HIGH -LOW	TOTAL	DIFF LEVEL	INDEX OF DISCRIM
*	19A	185	119	304	66	414	0.73	0.32
	19B	3	23	26	-20	414		-0.10
	19C	10	31	41	-21	414		-0.10
	19D	9	30	39	-21	414		-0.10
	20A	8	17	25	-9	414		-0.04
*	20B	158	73	231	85	414	0.56	0.41
	20C	35	91	126	-56	414		-0.27
	20D	4	23	27	-19	414		-0.09

APPENDIX B

ANALYSIS OF INTEGRATED SCIENCE - MODULE 10
(Asterisks indicate the correct item responses)

	ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL	DIFF INDEX	INDEX OF DISCRIM
	1A	8	13	21	-5	48	0.83	-0.21
*	1B	15	7	22	8	48	0.69	0.33
	1C	0	0	0	0	48	1.00	0.00
	1D	1	4	5	-3	48	0.98	-0.13
	2A	2	8	10	-6	48	0.96	-0.25
	2B	1	2	3	-1	48	0.98	-0.04
*	2C	21	14	35	7	48	0.56	0.29
	2D	0	0	0	0	48	1.00	0.00
	3A	2	6	8	-4	48	0.96	-0.17
*	3B	16	8	24	8	48	0.67	0.33
	3C	1	3	4	-2	48	0.98	-0.08
	3D	4	6	10	-2	48	0.92	-0.08
	4A	0	3	3	-3	48	1.00	-0.13
	4B	2	10	12	-8	48	0.96	-0.33
*	4C	22	8	30	14	48	0.54	0.58
	4D	0	1	1	-1	48	1.00	-0.04
	5A	1	6	7	-5	48	0.98	-0.21
	5B	0	1	1	-1	48	1.00	-0.04
	5C	2	4	6	-2	48	0.96	-0.08
*	5D	21	12	33	9	48	0.56	0.38
	6A	2	2	4	0	48	0.96	0.00
*	6B	11	1	12	10	48	0.77	0.42
	6C	0	2	2	-2	48	1.00	-0.08
	6D	11	19	30	-8	48	0.77	-0.33
	7A	6	7	13	-1	48	0.88	-0.04
*	7B	10	2	12	8	48	0.79	0.33
	7C	3	2	5	1	48	0.94	0.04
	7D	4	11	15	-7	48	0.92	-0.29
	8A	0	1	1	-1	48	1.00	-0.04
	8B	0	1	1	-1	48	1.00	-0.04
	8C	0	1	1	-1	48	1.00	-0.04
*	8D	24	21	45	3	48	0.50	0.13
	9A	1	1	2	0	48	0.98	0.00
	9B	0	2	2	-2	48	1.00	-0.08
*	9C	12	10	22	2	48	0.75	0.08
	9D	11	11	22	0	48	0.77	0.00
*	10A	15	16	31	-1	48	0.69	-0.04
	10B	2	1	3	1	48	0.96	0.04
	10C	6	4	10	2	48	0.88	0.08
	10D	1	3	4	-2	48	0.98	-0.08

APPENDIX C

ANALYSIS OF INTEGRATED SCIENCE - MODULE 9
(Asterisks indicate the correct item responses)

ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL	DIFF LEVEL	INDEX OF DISCRIM
1A	0	4	4	-4	18	0.22	-0.44
1B	0	0	0	0	18	0.00	0.00
* 1C	9	4	13	5	18	0.72	0.56
1D	0	1	1	-1	18	0.06	-0.11
2A	0	0	0	0	18	0.00	0.00
2B	2	2	4	0	18	0.22	0.00
2C	0	6	6	-6	18	0.33	-0.67
* 2D	7	0	7	7	18	0.39	0.78
* 3A	3	1	4	2	18	0.22	0.22
3B	3	3	6	0	18	0.33	0.00
3C	1	1	2	0	18	0.11	0.00
3D	2	4	6	-2	18	0.33	-0.22
4A	3	2	5	1	18	0.28	0.11
4B	1	4	5	-3	18	0.28	-0.33
4C	3	1	4	2	18	0.22	0.22
* 4D	2	2	4	0	18	0.22	0.00
5A	1	1	2	0	18	0.11	0.00
5B	3	3	6	0	18	0.33	0.00
* 5C	5	1	6	4	18	0.33	0.44
5D	0	4	4	-4	18	0.22	-0.44
6A	0	1	1	-1	18	0.06	-0.11
* 6B	6	6	12	0	18	0.67	0.00
6C	0	0	0	0	18	0.00	0.00
6D	3	2	5	1	18	0.28	0.11
7A	0	0	0	0	18	0.00	0.00
7B	4	6	10	-2	18	0.56	-0.22
* 7C	1	1	2	0	18	0.11	0.00
7D	4	2	6	2	18	0.33	0.22
* 8A	6	2	8	4	18	0.44	0.44
8B	2	0	2	2	18	0.11	0.22
8C	1	4	5	-3	18	0.28	-0.33
8D	0	3	3	-3	18	0.17	-0.33
9A	1	0	1	1	18	0.06	0.11
* 9B	1	5	6	-4	18	0.33	-0.44
9C	0	0	0	0	18	0.00	0.00
9D	7	4	11	3	18	0.61	0.33
10A	1	4	5	-3	18	0.28	-0.33
10B	0	1	1	-1	18	0.06	-0.11
10C	0	0	0	0	18	0.00	0.00
* 10D	8	4	12	4	18	0.67	0.44

APPENDIX D

ANALYSIS OF INTEGRATED SCIENCE - MODULE 7
(Asterisks indicate the correct item responses)

	ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL	DIFF INDEX	DISCRIM INDEX
	1A			0	0	8	0.00	0.00
*	1B	3	3	6	0	8	0.38	0.00
	1C	1	1	2	0	8	0.13	0.00
	1D			0	0	8	0.00	0.00
	2A	2	2	4	0	8	0.25	0.00
*	2B	2	1	3	1	8	0.25	0.25
	2C			0	0	8	0.00	0.00
	2D		1	1	-1	8	0.00	-0.25
*	3A	3	3	6	0	8	0.38	0.00
	3B			0	0	8	0.00	0.00
	3C	1	1	2	0	8	0.13	0.00
	3D			0	0	8	0.00	0.00
*	4A	1	1	2	0	8	0.13	0.00
	4B			0	0	8	0.00	0.00
	4C	1	1	2	0	8	0.13	0.00
	4D	2	2	4	0	8	0.25	0.00
	5A		1	1	-1	8	0.00	-0.25
	5B			0	0	8	0.00	0.00
*	5C	3	1	4	2	8	0.38	0.50
	5D	1	1	2	0	8	0.13	0.00
	6A			0	0	8	0.00	0.00
	6B			0	0	8	0.00	0.00
	6C		1	1	-1	8	0.00	-0.25
*	6D	4	3	7	1	8	0.50	0.25
	7A		3	3	-3	8	0.00	-0.75
	7B			0	0	8	0.00	0.00
*	7C	3		3	3	8	0.38	0.75
	7D	1	1	2	0	8	0.13	0.00
	8A	3	3	6	0	8	0.38	0.00
	8B		1	1	-1	8	0.00	-0.25
	8C	1		1	1	8	0.13	0.25
*	8D			0	0	8	0.00	0.00
*	9A	4	1	5	3	8	0.50	0.75
	9B		3	3	-3	8	0.00	-0.75
	9C			0	0	8	0.00	0.00
	9D			0	0	8	0.00	0.00
	10A	1	1	2	0	8	0.13	0.00
*	10B	1		1	1	8	0.13	0.25
	10C	2	2	4	0	8	0.25	0.00
	10D		1	1	-1	8	0.00	-0.25

APPENDIX E

ANALYSIS OF INTEGRATED SCIENCE - MODULE 4
(Asterisks indicate the correct item responses)

	ITEM	HIGH	LOW	HIGH + LOW	HIGH - LOW	TOTAL	DIFF INDEX	DISCRIM INDEX
	1A		3	3	-3	18	0.00	-0.33
*	1B	8	4	12	4	18	0.44	0.44
	1C	1		1	1	18	0.06	0.11
	1D		2	2	-2	18	0.00	-0.22
	2A	1		1	1	18	0.06	0.11
	2B		3	3	-3	18	0.00	-0.33
*	2C	8	10	18	-2	18	0.44	-0.22
	2D		4	4	-4	18	0.00	-0.44
*	3A	7	5	12	2	18	0.39	0.22
	3B	1	2	3	-1	18	0.06	-0.11
	3C	1	2	3	-1	18	0.06	-0.11
	3D			0	0	18	0.00	0.00
	4A		2	2	-2	18	0.00	-0.22
	4B	1	2	3	-1	18	0.06	-0.11
	4C	2	3	5	-1	18	0.11	-0.11
*	4D	6	2	8	4	18	0.33	0.44
	5A		3	3	-3	18	0.00	-0.33
	5B		1	1	-1	18	0.00	-0.11
*	5C	9	5	14	4	18	0.50	0.44
	5D			0	0	18	0.00	0.00

REFERENCES

Ahmann, J. Stanley and Glock, Marvin D. **Evaluating Pupil Growth.** Boston: Allyn and Bacon, Inc., 1971.

Crocker, A.C. **Statistics for the Teacher.** England: NFER Publishing Company Ltd., 1974.

Ebel, Robert. **Essentials of Educational Measurement.** Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.