NATIONAL LIBRARY
OTTAWA

CANADA

BIBLIOTHÈQUE NATIONALE
OTTAWA

6234

NAME OF AUTHOR....JAMES CHARLES POWELL

TITLE OF THESIS.....A STUDY OF ACHIEVEMENT INFORMATION FROM WRONG ANSWERS GIVEN TO MULTIPLE CHOICE TESTS

UNIVERSITY..THE UNIVERSITY of ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED....Ph. D.

YEAR THIS DEGREE GRANTED.....1970

Permission is hereby granted to THE NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed)....B. Powell

PERMANENT ADDRESS:

9509 - 99 B STREET

EDMONTON 81, ALBERTA

DATED..April 17.....1970

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

A STUDY OF ACHIEVEMENT INFORMATION FROM THE WRONG ANSWERS GIVEN TO MULTIPLE CHOICE TESTS

BY

© JAMES CHARLES POWELL

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1970

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "A Study of Achievement Information From the Wrong Answers Given to Multiple Choice Tests," submitted by James Charles Powell in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Supervisor

External Examiner

Date: April 3, 1970

ABSTRACT

This study began with the psychological postulate that all human performance in a choice situation tends to be made on a systematic basis.

In the setting of multiple choice achievement tests, this postulate resolved itself into three operational hypotheses which form a necessary and sufficient set to establish the possibility that it applied to both the right and the wrong answers given by examinees. Testing these three hypotheses involved the following procedures:

1) To develop and logically validate a systematic method for the construction of the foils (distractors) on a multiple choice achievement test designed to measure higher mental processes.
2) To show that the construct validity of this systematic method held up reasonably well in the results of the administration of this test.
3) To show that the validly produced foils in this context improved the predictive validity of this test with respect to other achievement tests over the more usual procedures.

The results of the study tended in general to support these three hypotheses fairly strongly if we take into account the finding that many of the foils could be classified into more than one category as evident in the low interrater reliability, the need to reclassify foils when wrong answer patterns were being interpreted, and the manner in which these interpreted foil categories cross-validated. This study would seem to have produced three fairly definite findings:

1. Human performance, when abstracted from responses to multiple choice achievement tests involving higher mental processes, would seem to be systematic, and to display evidence of multiple interpretation of the communication.
2. There would seem to be a hierarchy of foils which parallels the hierarchy of right answers and which influences the way in which each total item performs. The levels of the foils themselves seem to depend upon the ways in which this totality of each item is approached.
3. Wrong answers contain potentially useful information with respect to achievement when higher mental processes are involved.

Taxonomic tests would seem to have a number of properties not assumed to be present when the test is sufficiently homogeneous to be assumed to form a scale. The existence of these properties made it fairly evident that the more commonly used analytic procedures were probably inappropriate for the analysis and interpretation of the results from, and the criteria for, evaluating the effectiveness of this type of test. The alternative analytic procedure which the findings of the study implied were organized into a suggestion for the extension of test theory designed to deal with the problems which seem to arise where taxonomic tests are concerned.

Some implications of the findings to educational practice were drawn, and a number of suggestions for future research into this area were presented.

ACKNOWLEDGMENTS

The writer wishes to acknowledge gratefully the assistance of the members of his supervisory and candidacy committees in the preparation of this dissertation. In particular, Dr. S. M. Hunka, chairman of these committees, and much appreciated advisor during several years of graduate study providing guidance without which the writer could not have succeeded.

The constructive suggestions of Dr. A. Kaufmann, Dr. R. K. Gupta, and Dr. E. S. Shihadeh are much appreciated. The writer wishes, also, to express his gratitude for the support and encouragement of Dr. B. R. Corman, chairman of the Department of Educational Psychology.

Appreciation is expressed to Mr. Dale Burnett for his invaluable assistance during the course of this study. Also, the writer wishes to acknowledge his indebtedness to his wife, Fran, who has been a partner in the preparation of the many drafts made of this study.

Finally, the author is grateful to the University of Alberta for providing the financial assistance which enabled him to pursue graduate studies and complete this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

TABLES PAGE

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Psychological literature has been replete with studies of choice behavior. Chown (1959), Duncan (1959) and more recently Hunt (1961) and Berlyne (1965) have reviewed this literature adequately. From these studies it is fairly evident that the distribution of choices made by humans in problem-solving situations tends to exhibit some systematic trends.

These trends, however, have often been confined to discussion in terms of the patterns of "success" when compared with the nature and complexity of the task. Many studies, for example Strutz (1966) concentrate mainly on "right" answers and the relevant patterns involved. A notable second direction in this area has been the attempts to describe the nature of the procedure used by the person in his attempt to solve problems; for instance, Piaget (1953)[1] proposed a logical system for these procedures. Abelson and Rosenberg (1958) proposed a logical system which they call "psychologic" which proposes a set of "logical" procedures which lead to certain types of "wrong" answers.

Specifically, within the area of achievement testing the multiple choice test provides a good opportunity to observe choice behavior in a problem-solving setting. With the advent of Bloom's Taxonomy (1956) a considerable improvement in the classification of test items involving these "higher mental processes" became possible.

This present study will be to 1) demonstrate the presence or

---

[1] Furth (1969) discusses the Piaget model in some detail.

absence of this additional information, 2) identify the general properties, if any, of this information, and 3) speculate as to the implications of such findings. Assuming systematic choice behavior, consistencies would be expected within and between individuals for all choices made. These consistencies may not be confined to the "successes" or "right" answers. This study proposes to explore certain aspects of the possibility that the choices made among wrong answers may also be systematic and therefore contain useful information for the examiner.

Statement of the Problem

Some aspects of the possibility that wrong answers are selected systematically have been examined /Cf Fouldes and Forbes, 1965; Powell, 1968; Jacobs and Vandeventer, 1968; Powell and Isbister, 1969/, and these studies are discussed in more detail in Chapter II.

The specific concern of this study is to show to the designers of tests the significance of wrong-answer information. Any significant improvement in a test must be reflected in a corresponding improvement in the validity of the test. This study proposes to examine the construct and predictive validities of a particular method of test construction. The purpose of this study, then, is to explore the possibility that, if tests are constructed in a particular manner, "wrong" answers may add to the examiner's information about the examinee. This present study will be content to demonstrate the presence or absence of this additional information to determine the major properties of this information, and to speculate as to the implications of such findings.

CHAPTER II

BACKGROUND FOR THE STUDY

As already mentioned in Chapter I, this study was essentially exploratory. For this reason, little attempt has been made to establish a theoretical rationale upon which the hypothetical structure of the study might be built. Instead, the study was organized on the basis of procedural considerations. The absence of a theoretical rationale for the interpretation of results had the advantage that the data could be examined for consistent characteristics and the properties of these characteristics.[2]

Of course the nature of the procedures employed provide definite limitations upon the interpretations. The findings themselves would tend to provide other limitations upon the interpretation, and also the generalizability of the findings.

It is the purpose of the present chapter to review the most significant research which is relevant to the present study in order to present the research background which forms the basis for the procedural considerations which were employed.

The Problem of Measuring Academic Performance

The functions of, and therefore the outcomes of, education are a subject of debate which is beyond the scope of this dissertation. However, these functions and their corresponding outcomes have an important bearing on the nature of and the interpretations given to

---

[2] Glaser and Strauss (1967) developed a complete rationale to justify this procedure for exploratory studies.

the results of the various kinds of measuring instruments used.

Tests are formalized communications between examiner and examinee. The examiner is attempting to obtain a controlled sample of behavior to assist him in the rendering of certain judgments. These judgments may be either value judgments (continue, withdraw, certificate); or they may be procedural (concerning the nature of the appropriate treatments of programmes). The greatest possible control of the behavior sample is found in the multiple choice test.

As communications, tests involve several important considerations. First, they involve the examiner's perception of the examinees in terms of the capabilities they do have, and the capabilities they should have (educational goals). These perceptions lead to the examiner's decisions as to which information to give in the examination, in what format, and which information to withhold.

Second, on the basis of these considerations a communication is formulated. For the purpose of this study these communications will be confined to the multiple choice achievement test.

Third, the communication is presented to the examinee who is expected to interpret it and to respond to it. He will do so on the basis of the capabilities he possesses; the information he possesses, particularly that part of the information pool which was withheld by the examiner; and his sense of the importance to him, as a person, of the answers he gives including the information about himself which he wishes to withhold.

Fourth, the examiner then has the task of interpreting the responses of the examinees and making such value or procedural judgments as may be appropriate to these interpretations and to the

purpose of the test. To form these judgments the examinee's performance can be compared, as appropriate, with 1) his own past performance in similar contexts, 2) the performance of others in the same context (norm referencing) or 3) some external behavioral definition of mastery (criterion referencing).

However, where the subject matter content is itself open to disagreement, examiners themselves may not agree as to the appropriateness of the communication or its interpretation. Also, the examiner's assumptions about the capabilities and information background of the examinees may not be congruent to their actual characteristics. Furthermore, there may be little similarity between the examiner's purposes and the examinee's interpretation of these purposes. In addition, if an examinee has systematically misclassified a particular concept and this concept recurs with a high degree of frequency in a test, the examinee is likely to obtain a low total-correct score. Given the opportunity to correct this misclassification could lead to a much higher total score. How serious, then, must a misclassification be considered? Finally, suppose that the examiner misclassifies? Such an event is bound to have an adverse effect on the total correct score of the profoundly informed student as Hoffman (1962) points out. The combined effects of such considerations upon the composition of total-correct scores complicates their interpretation.

Technical Consideration in the Measurement of Choice Behavior

This study will confine itself to the choice behavior of examinees as exemplified in their responses to multiple choice achievement tests. The particular point of view to be expressed is relative to the way in which current practice tends to use wrong-answer information.

Current practice for scoring achievement tests. Present practice for scoring multiple choice achievement tests is to count the number of "right" answers selected by the examinee on a test or a subtest. The "right" answers are usually predetermined although experience with particular items may lead to subsequent revisions. In such tests the examinee is faced with several alternatives only one of which is "right." This means that he can make a wrong choice among several alternatives. In general, however, distinctions which might be made among students on the basis of differences among the wrong answers selected are not considered when the students' scores are evaluated. If wrong answers are used for any purpose it is usually to correct the scores for guessing.

Current Practice for Evaluating Achievement Tests--Reliability

There are three general areas in which the specific characteristics of a test can be improved. These are:

1. Reliability
2. Validity
3. Useability

The third characteristic of these, useability, can be dispensed with quickly because the simplicity of administration, and the simplicity and objectivity of scoring of multiple choice tests and the ease in the establishment and use of norms has been well established. The other two characteristics require more discussion.

The reliability of a test. The concept of test reliability involves how well the test measures whatever it measures. The APA Standards (1966) lists three methods of estimating reliability.

These are:

1. Internal consistency
2. Reliability between forms
3. Reliability over time

The latter two are determined by correlation coefficients either between alternative forms, or between repeated administrations of the same form on the same group of examinees. Neither of these two approaches is directly applicable to this study.

There are several possible approaches to the study of internal consistency. These are:

1. Item analysis
2. Using a correction-for-guessing
3. Designing the test to form a scale
4. Examining internal characteristics of the test
5. Part- whole comparisons

Internal consistency from item analysis. There are two schools of thought with respect to item analysis. The classical approach assumes that all of the items on the test should measure the same overall characteristic that the whole test, or that the relevant subtest, measures. To determine this similarity, the distribution of right answers on a particular item is correlated with the distribution of total-correct scores. The biserial correlation coefficient is usually used. This correlation coefficient actually correlates a dichotomous variable with a continuous variable. For the answers of a multiple choice item to be dichotomous, the plural set of wrong answers must be treated as a single variable. Similarly, the discrete distribution of

total scores (the total scores are the sum of a binary vector of "right-wrong" decisions which sum must be a whole number) must be treated as a continuous variable. The problem that score data provide discrete distribution is avoided by assuming that total scores are "best estimates" of "true scores" and true scores are assumed to form a continuous distribution. There is, of course, a multiserial correlation /Cf Jaspen, 1967/ which could be used to take account of the plurality of "wrong" answers. This latter coefficient is rarely used to evaluate multiple choice test items. Classical test theory advocates that the biserial correlation coefficient for each item should be high (significantly different from zero).

An alternative approach suggested by Lord (1952) involves the considerations necessary for addition of scores. In order to add two numbers, they must be independent, that is, the set of lattice points each represents can share no elements in common. By this approach, individual items should be relatively uncorrelated, but should collectively form a scale.

Both of these procedures tend to treat a multiple choice item as a dichotomy thus overlooking the fact that more than one choice can be made among the set of foils.

Using a correction-for-guessing. Originally, guessing corrections for scoring formulae assumed that the number of right answers which are attributable to guessing were directly related to the number of wrong answers given and inversely related to the number of alternatives per item. Because any answer could be a "guess," no meaning could be ascribed to particular answers. Meaning was thus

assumed to be confined to some form of cumulative score. This correction has the effect of increasing the variance of the total scores because a greater amount is subtracted from the low scores than from the high ones.

With respect to corrections-for-guessing, Gupta and Penfold (1961) showed that the guessing correction over-corrects in the event that the examinee is responding on the basis of misinformation. A similar argument can be presented to suggest that this correction under-corrects the partially-informed examinee. More recently, Shuford, and Massengill (1965) elaborated upon a system of "confidence scoring" in which the examinee rates every alternative on the basis of his confidence that each particular alternative is right. Honesty is encouraged on the basis that "confidently wrong" loses marks. This procedure makes it possible to classify each examinee's answer to each question as 1) well informed, 2) partially informed, 3) uninformed, and 4) misinformed. This procedure solves the guessing correction problem by identifying which items were "guessed" thus increasing the interpretability of particular items and hence the validity of the test. The scoring method these authors developed increases the internal consistency of the test by increasing the true score variance estimates proportionally more than the total score variance.

Designing a test to form a scale. The argument may be raised that the practice of distinguishing among individuals on the basis of total scores without considering the constituents of those scores may produce information loss. This argument can lead to the proliferation of subtests, or it can lead to test designs in which the scores form a

scale. For instance, Cox and Graham (1966) propose a system for designing a test which uses Gagné's task analysis /Cf Gagne, 1965, Chapter VII/, to produce a Guttmann (1954) scale. In this case, the score may indicate the level of mastery. Again, the internal consistency of the test can be increased, this time by increasing item homogeneity.

Examining internal test characteristics. One other area which has led to improvements in the reliability of tests has been through research into the improvement of the definition of the variables being measured by a test. Research toward this objective has been more extensive in the area of personality tests than in the development of achievement tests. The design of personality tests is beyond the scope of this present study. In view of the scarcity of appropriate research from the achievement testing area, only two developments in this latter testing area will be discussed here. First, Ayers (1965) attempted to validate Bloom's Taxonomy by means of factor analysis from tetrachoric correlations using programmed instruction in order to control the teacher variable. His findings in general supported Bloom's notion of a hierarchical structure. However, the results did not consistently fit the classification system in the Taxonomy. A more ambitious study to this same end was conducted by Kropp, Stoker, and Bashaw (1966). Although their findings were similar to those of Ayers (1965) because of the illumination their study provides for the construction of taxonomic tests, it is discussed in detail on page 18ff. For our present purposes, it may be sufficient to say that the validity and the reliability of achievement tests may be improved by using Bloom's

Taxonomy as a guide for developing the items.

Second, Gupta (1968) showed that the reliability in an internal consistency sense of an achievement test can be improved if the test is subdivided into subtests based on factor analytic results or on the basis of the DuBois, Loevinger and Gleser (1952) method of cluster analysis. This procedure makes subtests from relatively homogeneous items. This present study used a similar approach.

It should be noted, once again, that these methods tend to concentrate exclusively on the "right" answers.

Reliability based on part- whole comparisons. A special case of the alternative forms method of determining the reliability of the test is the group of procedures which use the correlation of one part of the test with another. The mathematical limit of the repeated use of the split-half technique when certain assumptions are made is found in the Kuder-Richardson (K-R) formulae. It is this form of reliability which increases in the DuBois et al (1952) procedure.

The Kuder-Richardson procedure is most sensitive to differences in the variance of the test. For this reason, if error variance is kept constant, increasing the test variance (as when using a correction-for-guessing), also increases the reliability. Another method of increasing the variance is to rewrite the test in such a manner as to move the difficulty (selection ratio) of each item toward .5 (50%). If the item is a dichotomy, a difficulty of .5 (50%) tended to maximize the variance assuming positive correlation because it maximizes the probability of choosing either alternative.

It is a common suggestion in evaluation texts for example, that

items should have middle range of difficulty. This suggestion assumes that each item should be treated as a "right-wrong" dichotomy. The plurality of wrong answers is being overlooked when items are treated as a dichotomy.

Current Practices for Evaluating Achievement Tests--Validity

In addition to the reliability of a test it is also necessary to be sure that a test measures the things it is intended to measure, i.e. the validity of the test. The validity and the reliability are related in that the validity of a test can never be higher than the square root of the test's reliability when the latter is defined in repeated-measures terms, hence the efforts to increase test reliability.

The APA Standards lists three types of validity. These are:

1. Content validity
2. Construct validity
3. Criterion-related validity

The concept of content validity refers to the validity of an item or test as dependent upon the appropriateness of the item or test of the information background needed to answer the test. In this present study most of the necessary information background needed is supplied in reading selections embedded in the test.

The construct validity aspect of a test. Construct validity has several aspects. In brief, construct validity refers to some psychological construct or constructs, more or less independent of content, which are included in the test. For instance, if intelligence as a construct is assumed to be manifested by intelligence measures, such as the WISC, the correlations between a new group I.Q. test and

the WISC scores on the same subjects could be support for the construct validity of the new test. In this case the construct would be "intelligence." Another approach to construct validity is to define the construct in such terms as to facilitate translation into performance terms. A good example to this approach is Bloom's Taxonomy (1956). As already indicated, this procedure should increase the reliability of the test as well as its validity. These constructs should also be identifiable in the performance of examinees when the performance data are subjected to statistical analysis.

Another procedure which strengthens support for constructs for which measures have not been standardized is "cross-validation." In cross-validation statistical analysis should reveal the same constructs in independent groups. Cross-validation is used in this study for testing the construct validity of the procedure being explored here for the development of foils (wrong alternatives) on multiple choice tests.

A final aspect of construct validity concerns the degree to which the examiner's objectives have been accomplished by the test he has developed. In the absence of standards this accomplishment is difficult to measure. One approach is to study the distribution of answers to an item to get clues to its effectiveness. Part of the discussion in Chapter III on the development of the experimental test used in this study will elaborate this procedure.

In some cases the construct may be sufficiently well defined that the different performance outcomes are indisputable. In such cases the construct validity of a test may be easily determined. Piaget's discussion of the acquisition of various aspects of conservation concepts are a case in point. Items measuring the

acquisition of these concepts must conform in their discrimination to the known characteristics of this acquisition process. Where wrong answers are concerned, as will be discussed on page 27 ff no such clear definition exists. The present study, therefore, can be no more than exploratory in nature.

Criterion-oriented aspects of a test. One of the fundamental functions of any measurement of achievement is its predictive value for future achievement. Within the context of the present study one of the concerns is the ability of the experimental test which may, by its construct characteristics, be considered as a test of strategies which may improve the prediction of other achievement test results. Popham and Husek (1969) point out that most of the statistical procedures used in current practice may be inappropriate for criterion-referenced tests.

Studies Related to Wrong Answers

At this point the concept of answering patterns becomes critical. An answering pattern will, for present purposes, be defined as some characteristic among the answers selected by a group of students which is consistent and stable under statistical analysis, and hence leads to an improvement in the validity and reliability estimates of the test to which these answers are given. The works already quoted suggest that there may be such patterns among "success" performance. The question can now be raised as to whether or not there may be answering patterns among wrong answers as well.

A possible source of findings concerning wrong answer information is diagnostic testing which has a considerable history. Schonnel (1943) discusses the cumulative results of more than twenty years of

research. His procedures showed that the nature and location of mistakes can reveal specific problems, i.e. wrong answers can be meaningful for diagnostic purposes. However, error types are usually established in advance, and are restricted to items which reflect only one type of error thus allowing the items to be scored as right or wrong. Large numbers of questions are needed for this procedure since as the complexity of the problem to be solved increases, the number of tasks required to diagnose all possible errors increases expotentially; probably explaining the absence of diagnostic tests in "subjective" subjects. The Cox and Graham (1966) procedure refines this diagnostic technique. In order to develop diagnostic tests of this sort, an interlocking pattern of items is usually designed in such a way that specific weaknesses in a particular student's performance can be inferred. This procedure identifies weaknesses on the basis of relationships between items rather than relationships between alternatives within a particular item.

It becomes evident from the fact that diagnosis can lead to the identification of specific error types that the four categories of students' responses made by Shuford et al (1965) may be an oversimplification. Furthermore, it would seem reasonable that more than one error type could be accommodated in one item if a multiple choice format were used. In this latter case, there should be evidence of answering patterns among wrong answers.

Answering patterns in foil selection. The evidence supporting the possibility that there may be answering patterns in the wrong answers as well as the right ones is sparse. Sigel (1963) reported with reference to intelligence testing that children tend to "be

consistent within themselves in the errors they make [p. 53]."

Fouldes and Forbes (1965) reported in the manual for their revision of the Advanced Set of Raven's Progressive Matrices the following finding concerning common errors:

> Four types of common errors could be identified. (A) Incomplete solutions. There were errors due to people failing to grasp all the variables determining the nature of the correct figure required to complete a test item. Instead they chose a figure which was right as far as it went but was only partly correct... (B) Arbitrary lines of reasoning. Here the figure chosen suggests that the person has used a principle of reasoning qualitatively different from that demanded by the problem... (C) Overdetermined choices. These were errors involving failure to discriminate irrelevant qualities in the figure chosen... (D) Repetitions. These are errors made by people who simply selected a figure identical with one of the three figures in the matrix immediately adjacent to the space to be filled. [p. 20]

Fouldes and Forbes (1965) did not attempt to show whether or not these common errors were more characteristic of some individuals than of others.

These types of error would seem to be more related to some form of answering procedure based on the relational characteristics of the alternatives rather than on their informational characteristics.

Powell (1968) factor analysed some wrong answers derived from an administration of Gorham's Proverbs Test (1956). This test would probably be classified as a comprehension test by Bloom's Taxonomy. A wrong answer pattern of four factors resulted. These were:

1. Reduction of information to affect a simplification of the statement
2. Addition of irrelevant information
3. Substitution of elements
4. Replacement of proverb by one largely unrelated

If this list is compared with the one by Fouldes and Forbes on page 16, we find, at least by description, Factor I remarkably like their Class C (Overdetermined choices). Possible relationships between the remainder are less certain although Factor 4 and Class B may be related. Their Class D is unlike Factor 3 but is very much like the "Word-Word Links" class present in the experimental test used in this study. A definition of this class is on page 38.

However, Sigel (1963) went on to report that there "seemed to be no relationship between type of error and total score." [p. 53] In contradiction to Sigel, Jacobs and Vandeventer (1968) showed that within the context of Raven's Coloured Progressive Matrices and by using the Guttman and Schlesinger (1967) facet design, that a relationship often does exist between right and wrong answers. Ebel (1969) has shown similar systematic characteristics among True-False items. Furthermore, Powell and Isbister (1969) showed that some wrong answers can be related to right answers so as to adversely affect the high scoring students. The type of foil involved was the "irrelevancy." A definition of this class is on page 37. An inconclusive trend in this same direction was found in Factor 4, page 17.

Thus, the available evidence, scanty though it is, suggests that neither "misinformation" nor "no information" (leading to a haphazard answer) are sufficient to account for all the wrong answers given to multiple choice achievement tests, and that in certain circumstances specific foils may influence the total-correct score.

If wrong answers contain achievement information, then the wrong answers which display systematic characteristics which acceptably support the construct characteristics of the experimental test should

improve the prediction of independent achievement scores for the same examinees.

This improvement should occur in comparison with the prediction made by either the total-correct scores on the experimental test or some reasonable subdivision of these scores into subtest scores where the subtests also fit the construct characteristics of the test.

Studies Related to Item Generation

Perhaps the most ambitious attempt to develop tests reflecting Bloom's Taxonomy (1956) was the work of Kropp, Stoker, and Bashaw (1966). These researchers encountered a number of problems in their work some of which are discussed here along with the alternative procedures used in the design of the experimental test used in this study.

The problems they encountered which are relevant to this study are:

1. Problems arising from the "Knowledge" category of the Taxonomy
2. The generation of Synthesis and Evaluation category items in multiple choice format
3. Item analysis problems
4. Problems arising from implicit assumptions in their study

Although the fourth of these problems is probably the most important for present purposes, the discussion which follows considers each problem in the order given here.

Problems arising from the "Knowledge" category of the Taxonomy. Kropp et al (1966) spent some time discussing whether the "Knowledge"

category in Bloom's Taxonomy is a legitimate category and, if so, what psychological processes other than recall this category might represent. They further compound the problem by basing their questions on reading selections supplied in the test. Thus, the legitimate question is raised as to the meaning of less than a perfect score on "Knowledge" items when all the information necessary for the answering of these items is contained in the reading selection used.

These researchers do not comment on the possibility that "Knowledge" items presented in an "open book" format may not be "Knowledge" questions in the sense of Bloom's Taxonomy at all. Instead, these questions, in order not to be obvious, produce a test of search skills more commonly known in the literature on reading skills as "reading for details" /Cf Gray, 1960, p. 11/. It is not surprising, therefore, that an important contributor to the "Knowledge" category in two of the grade levels is an unidentified factor consisting in grade nine of "Word Arrangements, Letter Sets, and Symbol Production /p. 131/, and in grade twelve of "Thing Categories, Locations, and Gestalt Transformations" /p. 134/[3]. All three of these tests were positively loaded on the unidentified factor for grade nine, and the "Locations" test is positively loaded on the unidentified factor for grade twelve. These unidentified factors add credence to the suggestion that the "Knowledge" category for the Kropp et al (1966) tests may well be more related to search-skills than to recall. There is probably no logical method of testing "Knowledge" as defined by Bloom's Taxonomy when the

---

[3]These names refer to names of specific tests from the Kit of Reference Tests (French, Ekstrom, and Price, 1963) which purport to define particular cognitive aptitudes.

information background is supplied by the test. In the case of the experimental test in this study, no "Knowledge" category items were generated.

Generating multiple choice synthesis and evaluation items. Another point made by Kropp et al (1966) was the difficulty of generating multiple choice items of the Synthesis and Evaluation Categories. One of the problems encountered in this respect is the restriction of a specific category in Bloom's Taxonomy for inductive reasoning to one subcategory of the Synthesis Category. Another subcategory adds "unique communication" requirements which are impossible to meet in a multiple choice format. The third subcategory involves producing a plan or proposed set of operations. Again, the open-endedness of this requirement restricts its employment in the multiple choice format.

Second, if the "Synthesis" category is restricted to induction, the problem remains that the internal structure of a single reading selection is usually highly organized. For this reason, the generation of a large number of items which require an inductive combination from some components of this selection or an inductive generalization from these components is very difficult because both of these possibilities are either explicit or closely implicit in the passage. However, if more than one reading selection is included in a test of this type, it would seem to be a relatively simple matter to generate items which require inductive combinations between selections or inductive generalizations between selections. This latter procedure is used in the experimental test in this study.

It is possible that the nature of the strategies employed by examinees when solving problems has an effect on the effective classification of the item by the Taxonomy. Two outcomes would be expected in this case. First, the more familiar an examinee is with the content of the problem the lower the effective classification of that problem. Second, the nature of the strategy shifts employed for generating foils may influence the strategies which the examinee has to employ to answer the problem, which in turn may also affect the classification of the problem. For instance, an item classified as synthetic on the basis of the stem alone or the stem and right-answer may become a comprehension item if the foils stress reading comprehension. Perhaps the rather surprising apparent dislocations of the Evaluation items in the Kropp et al (1966) study reflect this problem. It should be noted that the Evaluation category occurs in the second, third, and sixth positions in the ordered Simplex (Guttman, 1954) analysis and additionally in the fifth position by mean score /Cf Kropp et al, pp. 83, 87, 88/. Greater elaboration of this latter problem occurs when the implicit assumptions of the Kropp et al (1966) study are being discussed. Only three Evaluation items are used in the experimental test because of its length (30 items).

Problems of item analysis on taxonomic tests. Another problem which Kropp et al (1966) discuss at some length is the problem of item analysis for tests designed to measure levels in a Taxonomy. The Taxonomy was developed on the basis of the assumption that each higher level subsumes all lower levels and adds some unique characteristics of its own. Thus, as the level of the Taxonomy increases so does the complexity of the problems which are appropriate to this level. It

would therefore be expected that the difficulty of items designed for each category would increase as the level of the Taxonomy for which these items were designed increased. Thus, the selection of items on the basis of approximate middle difficulty at each level of the Taxonomy in order to maximize discrimination would seem to be inappropriate. This subsumption property also implies that if any item were missed at any level of the Taxonomy, all items designed for higher levels of the Taxonomy which involve the context of the item missed should also be missed. As a result, the number of items correct at any level of the Taxonomy should determine the upper limit of the possible score for the next higher level.

Kropp et al did not test this latter hypothesis in their study by examining individual performance to see whether or not individuals who answered a particular Knowledge question incorrectly tend, in general, to miss all higher level questions related to the same information background. They did, however, mention that low scorers on the Knowledge subtest tended to be low scorers on all subtests. An alternative hypothesis which might be posed is whether or not those people who misinterpreted a particular knowledge question are more likely to miss a high level item from the same background if one of the foils contained the same misinterpretation than if it did not. Although this latter alternative presents an hypothesis which is beyond the scope of this present study, it is more in keeping with the possibility of the influence of systematic choice behavior on response selection as developed here, than is the former hypothesis.

It is true that for dichotomous variables, the discrimination is maximized for items of middle difficulty. In general, if all

alternatives are to be considered, discrimination is maximized if the selection frequency for all response alternatives on any item is equal. Thus, for a four-alternative item, discrimination is maximized when the difficulty is .25 when all four categories are used. In the case of forcing a dichotomy on a polychotomous variable, the fact that 50 per cent of the examinees get the item right means that the distribution of answers on this item is not the product of chance, at least for the right answers. The same conclusions may be true for wrong answers, as Powell (1968) has shown, when higher mental processes are involved.

For these reasons, it may be reasonable to ignore item difficulty except for very easy or very difficult items as a criterion of item discrimination. At least the former argument with respect to ascending complexity, and the related ethical problem of predetermination of hypothetical results were the basis for Kropp et al (1966) ignoring item difficulty in the preparation of their tests. The latter argument with respect to the discriminative power of polychotomous items, except in extreme cases, was the basis for minimizing the importance attributed to item difficulty in the present study.

As Kropp et al (1966) point out [p. 77], an additional problem with respect to item analysis arises in the interpretation of correlations on data derived from taxonomic tests. Since the subtests are assumed to be hierarchically interdependent, the bivariate distributions of scores between subtests appear triangular, making the distribution of each higher level skewed further to the right. On this basis the total-correct score may not be normally distributed for tests of practical length used on groups of usual size, hence the use of biserial correlation for the validation of an item against the total test score

may be inappropriate. This fact, as they point out, also raises problems in the interpretation of any correlation coefficient in their study. When determining the discrimination coefficient Kropp et al (1966) used the traditional procedure.

Problems which arise from implicit assumptions in the Kropp, Stoker, and Bashaw study. The central assumption of the Taxonomy is that each higher level subsumes all lower levels and adds characteristics of its own. For this reason, Kropp et al (1966) approached their analysis with the implicit assumption that the complexity dimension was characteristic of the Taxonomy as a whole rather than being a characteristic of each level of subcategory within the Taxonomy. The results of their findings with respect to this assumption were inconclusive. Analysis of the subtest scores showed that the order of the levels of the Taxonomy as a hierarchy did not fall consistently into the order hypothesized. On the other hand, Powell and Isbister (1969) tested the assumption that hierarchical categories should be obliquely related. Their finding, however, was that the use of a promax rotation did not improve the resolution of the factors when right and wrong answers were combined, thus the expected obliqueness did not occur.

It has already be indicated on page 21 that the Evaluation category can occupy most positions above the Knowledge level. The Kropp et al (1966) study also found that on the basis of cognitive attributes no single category is consistently defined for all grade levels tested although the tasks themselves were identical for all grade levels. These two findings of Kropp et al are inconsistent with the Taxonomy as defined. Perhaps the Taxonomy is actually a description of some of the strategies employed by humans in problem-solving

situations. There may be a hierarchial order to these strategies but they may not be taxonomic in Bloom's sense of the term.

A problem which is a Synthesis level problem for a five-year old, may be a comprehension level problem for a twelve-year old. In this context two deviations from this taxonomy would be expected with Bloom's Taxonomy. First, each category of the Taxonomy, with the possible exception of "Knowledge," should be characterized by a range of complexity levels within the category in addition to an order of complexity levels between categories. In such circumstances, as Kropp et al (1966) demonstrate, a wide range of possible orders may occur among specific samples from category levels. In addition to the most common, and expected, order of the categories found in their study, the categories occur at least once in any one of three other orders under Simplex analysis.

Second, the strategies involved at different developmental stages will vary in accordance with the information and strategy backgrounds of the individuals at these stages. For this reason, striking dissimilarities in the cognitive attribute content on the basis of the Kit of Reference Tests (French, Ikstrom, and Price, 1963) for any category would be expected at different developmental levels. This is precisely what Kropp et al found. An important characteristic of this change should be its movement toward simplicity. For instance, if we reclassify the Kropp et al (1966) "Knowledge" category as a "search" category on the basis of the Undefined factor, we find that for grade nine the positively correlated cognitive aptitudes are Word Arrangement, Letter Sets, Symbol Production which suggests that the grade nines may be generating their search strategies as they proceed with the test.

For the grade twelves the positive attribute is Locations which suggests a more simple and direct approach.

Another factor which Kropp et al (1966) discussed is that the difficulty of a problem may be affected by the complexity of the problem. It also may be affected by the familiarity or obscurity of the information background and/or strategies required by the problem solver. It may also be affected by the nature and the fineness of the discriminations which the solution to the problem requires. This latter aspect may be related to the nature of the foils. Kropp et al (1966) deal only briefly with the difficulty problem. /Cf pp. 90 and 159/.

Contributions of the Kropp, Stoker, and Bashaw Study to the present study. It may be possible to assume that Bloom's Taxonomy is not a subsumptive taxonomy. In this case, the Kropp et al (1966) study more strongly supports the possibility of the transcendence of process over content than their interpretation of their findings suggests. This transcendence of process over content has also been supported by Furth (1966) in his work with the congenitally deaf.

In combination with the other research already discussed (see p. 16) there would seem to be at least three variables which contribute to the choice behavior of examinees on multiple choice achievement tests. These are: 1) content, 2) process, and 3) effective complexity. A fourth possible variable is item difficulty (see: p. 26). Since misinterpretation of content and inappropriate selection of strategies might both be expected to lead to the selection of an inappropriate response, it is reasonable to assume that at least some students will display systematic wrong-answer selection. Hence, in a forced-choice situation the nature of the alternative choice provided would

be expected to influence the nature of the selections made. If foils are deliberately designed to reflect probable misinterpretations of content, or probable inappropriate selections of strategy, more than the "right" answers might be used to determine the present achievement status of the examinee.

How can tests which meet these criteria be developed? It is fairly clear from the Kropp et al (1966) study that the use of Bloom's Taxonomy is useful as a set of guidelines for the construction of the relationship between the stem and the right answer for each item. A discussion of common recommendations for the development of the foils for each item is presented in the following section.

Recommendations for Construction of Foils

The following discussion reviews what some textbook authors have had to say to teachers about the construction of foils for multiple choice items. Among these authors, Ross and Stanley (1954) list fourteen rules for the construction of multiple choice items. Of these only two deal specifically with foil (distractor) construction.

6. Make all responses plausible
9. To measure higher levels of understanding, increase the homogeneity of the options provided [p. 185].

These authors do not define plausibility, and the example they use for increasing the homogeneity of options actually illustrates increasing the content specificity of the item. Their second suggestion involves increasing the fineness of discrimination between alternatives which may be more related to the difficulty of the item than to "higher mental processes."

As another example, Thorndike and Hagen (1961) in their second edition list ten "maxims for multiple choice items." Four of these have direct bearing on foil construction. Quoting the original we find (Italics in original):

> 4. Be sure that There is One and Only One Correct or Clearly Best Answer.
> 5. Beware of Clang Associations.
> 8. Beware of the Use of One Pair of Opposites as Options If One of the Pair is the Correct or Best Answer.
> 9. Beware of the Use of "None of These," "None of the Above," and "All of the Above" as Options./pp. 74, 75, 76, and p. 77/

Whether or not there should be more than one "correct" answer will depend upon whether or not the examiner wishes to discriminate between levels of insight into a particular problem as in the "best answer" type of test. However, to make such discriminations may require the use of information from more than one alternative of any item.

One of Hoffmann's (1962) most damning criticisms of the multiple choice types of tests arises from the arbitrary assignment of only one alternative of the response set to the "right" category in tests of this type in such a way as to discriminate against the thoughtful, well informed student. This admonition is only appropriate if we are to assume that the only answer to be taken into account for any particular item is the one designated as "right" whereas the "rightness" may be arranged on a continuum in the "best answer" type of test.

Thorndike and Hagen (1961) quite rightly point out that Clang Associations (see number 5, p. 28) between stem and right answer tend to give the answer away. However, using superficial associations between the stem and the wrong answers may in some circumstances be an effective discriminating device (see: "Word-Word Link," p. 38).

Thorndike and Hagen's (1961) alternatives, numbered eight and nine (see p. 28) are interesting in that they suggest that certain aspects of the logical relationships between answers and foils should be considered in foil construction. If a student selects an answer belonging to a set described in number five (see p. 28) that is logically opposite to the "right answer," then this selection in itself may contain useful information. Such a selection reveals at the very least which students completely misunderstand the relationship in question. Why this sort of alternative should be discarded without qualifications is therefore not clear. The criticism these authors make of the "all of these," "none of these" type of alternatives have a similar basis. They neglect to say that if "none of these" is correct it may be regarded as being logically equivalent to an omission. The "none of these" provides a noncommittal response which has the effect of making closed-choice alternatives into open-ended alternatives. For some purposes it may be useful to know if the student made one of the less common errors, if there are more possible errors than the foils account for. In addition, omissions at the end of the paper can also mean "not finished." Since there is more than one possible reason for omitting an item, interpretations of an omitted response becomes ambiguous. For these reasons, the basis upon which a student makes a noncommittal response may be a valid question for study.

More recently, Ebel (1965) lists 48 "suggestions for preparing good multiple choice test items." Of these 48 only five directly relate to foils or "distractors." He also rates these as "desirable" or "undesirable." Quoting the original:

32. Item using true statements as distractors. (Desirable)

33. Item using stereotypes in distractors. (Desirable)
34. Item using obscure distractors. (Undesirable)
35. Item using a highly implausible distractor. (Undesirable)
36. Item involving verbal trick. [pp. 183-185] (Undesirable)

The first two of these are examples of the use of errors in logic which Sanders [1966, p. 104] suggests we teach the students to recognize, but does not elaborate on, with respect to measurement.

Ebel's (1965) suggestion numbered 34 immediately above, proposes that the use of obscure or complex vocabulary is undesirable. On the contrary, if the intention of the examiner is to study responses to obscure, ambiguous or complex situations, this type of item may be desirable. Although other methods may have certain advantages when measuring complex human behavior, the multiple choice method retains two particular advantages. First, a high level of control can be maintained in the alternatives supplied so that the "controlled sample of performance" characteristic of all tests can be very explicit. Second, once the performance components of the complex behavior which is to be observed has been established, accurate counts of the frequency of the choices which fit the categories of alternative (whether right or wrong) designed to measure these components is a simple matter. Other measuring instruments have other advantages at the expense of these two. The study of such items would probably necessitate examining all responses to each item. Thus, Ebel's (1965) proposal that this type of item is undesirable can be considered valid only if the "one right answer" assumption is considered valid. Closer scrutiny of this entire problem seems reasonable.

In suggestion numbered 35, (p. 30), relating to implausibility the problem of a definition for plausibility arises once again.

Plausibility may be a function of the rationale used in determining the construct and content validity of the test. The examiner must be able to anticipate what alternatives may be plausible to the examinees. Without a definition of plausibility, implausibility is impossible to determine. In fact, plausibility is often defined on a *post hoc* basis from the item analysis with foils having a low selection ratio being classified as "implausible." However, if the purpose of discrimination is to identify individuals for differential treatment a foil which identifies ten or twelve out of 1,000 students may be more valuable than one which identifies 250 students.

Finally, many foils which seem to involve a "verbal trick" may have a valid function. These verbal tricks are probably of three kinds. The first kind could be the introduction of a peculiarity of wording designed to produce interpretive or misreading errors on the part of some students. The second kind of "verbal trick" is found in such things as Zeno's Paradoxes (*c* 340-264 B. C.) in which the "verbal tricks" involve a faulty assumption in the reasoning. The third kind of "verbal trick" introduces the possibility of detecting in the examinee an inappropriate "set" for the correct solution of the problem. Both the *Einstellung* effect and "functional fixity" may possibly be used to develop examples of foils for this type. In each of these cases it is conceivable that the information generated from response to these types of item could have discriminative value. The issue here, once again, is both content and construct validity. Does the "verbal trick" give the intended information, or interfere with the obtaining of this information.

We find this same ambiguity of advice prevailing throughout the

range of standard texts in this area. From the ETS booklet Multiple Choice Questions: A close look (1963) through to such writers as Ahmann and Glock (1963), Gronlund (1965) and Noll (1965) we find the great bulk of the suggestions about item writing discussing the functional, linguistic, and structural characteristics of the stem, and stem-right answer relationships, with only minimal and often contradictory treatment of the foils and how to construct them.

Need for a Basis for Interpreting Foil Selection

On the basis of the above discussion we can identify several general bases for foil construction as presented to constructors of multiple choice tests. These are:

1. Logical Relationships
2. Logical Errors
3. Partial Information
4. Misinformation
5. Obscure Relationships
6. Misunderstanding
7. Verbal Tricks

Not all of these are regarded favourably by the authors mentioned nor are these bases consistent with themselves or between one author and another. It would seem that the recommendations have been developed on a trial-and-error basis derived from the experience of professional test constructors during their attempts to meet the statistical criteria of an "effective item."

The Possible Value of the Experimental Test

The testing technique being explored in this study hypothesizes

that Bloom's Taxonomy adequately describes the strategies leading to right answers, and that a set of logically based guidelines for foil development effectively describes some of the possible systematic deviations from the ideal outcomes of these strategies. These two facets combine to form the construct characteristics of the testing techniques under study. Of course, any findings from a purely exploratory study must be tentative. However, wrong answers from a "strategy" test may increase the predictive power of that test for total achievement scores (found in the usual way) from independent achievement tests. In this case, more than the information background of a test may be involved in "success" on an achievement test. Such findings would strengthen the support for the hypothesis that process may transcend content. Furthermore, this study may suggest some of the typical types of errors students may make as they mature intellectually which might eventually lead to the establishment of a behavioral description of development which is independent of test content, and of educational strategies which may be appropriate to the stages and phases of this developmental sequence.

On the application side, the main advantages of guidelines for foil construction would be expected to involve the 1) simplification of item writing, 2) clarification of why a foil is wrong, and 3) possibility of producing diagnostic tests in subjective content areas.

CHAPTER III

DESIGN OF THE EXPERIMENTAL TEST

From the discussion developed in Chapter II, the usefulness of Bloom's Taxonomy for the development of process-oriented items was suggested. The possibility that Bloom's Taxonomy may not display the assumed subsumptive characteristic between categories does not minimize its role relative to the establishment of the construct validity of a test. The evidence presented suggested that there is no similar set of internally-consistent construct guidelines for the development of foils. Seven general categories of foil based upon recommendations from the literature could be established. Using these seven as a starting point the first task in this chapter is to develop a systematic set of Guidelines which may prove helpful for foil construction. The seven categories can be further reduced. Perhaps the most important category involving strategies are those foils which can be based on probable errors in logic made by the examinee. Partial information can lead to an error in logic if the wrong strategy is used to generate the missing information. It can lead, also, along with other casues, to an oversimplification of the problem. Since only the product of the choice-behavior is observable on a multiple choice test, it would be reasonable to include Logical Errors, Partial Information, some Verbal Tricks, and perhaps some Logical Relationships (like answers which are opposite to the right ones) in a list of categories of foil generation where higher mental processes are to be tested.

Misinformation and misunderstanding may be identical or they may be different in that the misunderstanding may be related to the reading

of a specific item or group of items rather than to a weakness in the information background of the examinee. If the examinee succumbs to certain kinds of verbal tricks (for example, the use of meaningless jargon in a foil) his problem may be more immediately test-related than background-related provided that he is not misled elsewhere on the test when jargon is not used. In the present context there is the possibility that in addition to the process variables there may be a class of foil related to the linguistic characteristics of the item. This class of foil may be designated as a "Misreading" class.

Another possibility is that the examinee has systematically misclassified a particular piece or set of information. In items based on the possible logical relationships among the total information background this piece of misinformation will lead to the systematic selection of specific wrong answers each time this misclassification appears in a foil. For instance, the person who confuses the work of Hebb with the work of Hull. Foils of this type, and of several others, are beyond the scope of the present study and are, therefore, classified in the "Others" class. Subsequent research may be expected to elaborate this latter problem.

The Guidelines for Foil Construction

Our earlier discussion showed that similarities can be found between common errors on nonverbal tests and verbal tests (see: p. 17). The present experimental test in its original version was based on classifications involving logical fallacies and logical relations. The test has been revised in an attempt to improve item discrimination for the present study. The same reading selections and general questions and overall format remained unchanged. The Guidelines which follow

were used to revise the foils.

Since the definitions of foil classes, as they were originally used, tended to lack precision, they were redeveloped for this study. The Guidelines are described below (see p. 37) in terms of the procedure used for constructing each type of foil. Four classes were produced:

1. Strategy class; the largest group of Guidelines to be developed for this study is based on the logical characteristics of the foil relative to the right answer and information background. Because these types of foil are suggestive of incorrect analytic procedures they are collectively referred to as the "strategy class."
2. Misreading class; this group of foils is based on semantic characteristics of the foils relative to the right answer and information background. The nature of this test, i.e. an open book test, would be expected to reduce the possible number of foil categories in the misreading class because an examinee who feels he has misread an item can refer directly back to the information background supplied. This class of foil probably has many more members which would describe different aspects of misreading where information recall is the source of information. An example of this situation is the Jargon (J) category (see p. 39).
3. Other; the foils in this class are unclassifiable, at least by the present Guidelines. Future studies are expected to reduce but not eliminate this class of foils.

4. Misinformation class; the nature of the experimental examination, (i.e. an open book examination) precludes the development of misinformation foils which would be expected to occur in the context of a test requiring information recall. These would be expected to be related to"Knowledge" level items, a level of item which was not used in this examination for reasons already discussed (see p. 19ff).

The first two of these major classes may be subdivided on the basis of a specific description of how a foil which fits any particular category is produced. This subdivision follows:

Guidelines

A. Strategy Class

1. Overgeneralization. (OG) In the development of this type of foil the author retains the correct relationship of the best answer in its entirety and adds some irrelevant information. (For example, see item 1A, p. 154).
2. Oversimplification. (OS) In this case the author omits one or more parts of the best answer. (For example, see item 2C, p. 156).
3. Inversion (Inv). In this case the author makes a statement in some way opposite to the best answer. (For example, see item 4C, p. 158).
4. Irrelevancy (Irr). In this case the author makes a true statement which is unrelated to the best answer, or a statement which could be a correct answer. (Perhaps by virtue of some restriction in the stem). (For example,

see item 1D, p. 154).[4]

5. Invalid Assumption (IA). In this case the author begins with an unwarranted assumption about the background or solution to the problem and thus writes a foil which would be correct as if this assumption were valid. (For example, see item 1C, p. 154).

6. Substitution (Sub). In this case the author replaces at least one of the elements or the relationships of the best answer by a corresponding element which is less acceptable. (For example, see item 2B, p. 156).

7. Transposition (Tr). In this case the author modifies the order of the elements in an ordinally dependent relationship. (For example, see item 30C, p. 181).

8. Common Misconception (CM). In this case the author utilizes his knowledge of the probable common misconceptions held by the examinees to write the foil. (For example, see item 5B, p. 159).

B. Misreading Class

1. Word-Word Link (WW). In this case the author produces a false statement which has strong verbal links with the stem or background information by either repetition or association. This type of foil may be similar to Foulde's (1965) Class D error, see p. 16. (For example, see item 7B, p. 162).

2. Redefining of Terms (RT). In this case the author uses a

---

[4]This type of foil misleads certain of the best students, perhaps the more imaginative ones (see p. 18).

word or words in the foil in different literal or connotative sense than it is used in the stem or background information. (For example, see item 11D, p. 166).

3. Jargon (J). In this case the author produces a quasi-meaningful statement which tenuously relates in some manner to the best answer. The use of coined "near words" may also be present. (Not used in experimental test; see p. 36).

C. Others

1. Others (O). In this case the foil is, at present, for some reason, unclassifiable.

These are the Guidelines which were used in the construction of the foils in the experimental test.

Structure of the Experimental Test

As already mentioned, Bloom's Taxonomy was used as a guide for the construction of the stem and right answers of the experimental test. An interrater reliability between judges for the advance classification of right answers was reasonably high (r = .83). The Guidelines just given on pages 37-39 were used to construct the foils. The interrater reliability for foil classification was somewhat lower (r = .62, N = 5).

As in the case of Bloom's Taxonomy for the right answers, the Guidelines presented the immediately evident advantage of increasing the number of possible foils which could be considered for any one item, making foils easier to generate than they were in the more usual "hit-and-miss" method. An additional advantage for the Guidelines became evident after the earlier administration of the test. The Guidelines

help clarify the basis for why any foil should be considered wrong. The absence of such a basis is a common weakness of teacher-made tests.

The examination consisted of five short reading selections drawn from material which was in some way related to educational psychology since this was the central topic of the course in which this examination was to be used. They were also chosen on the basis that it was relatively unlikely for the examinees to have encountered the works from which these selections were drawn in their previous training. To the extent that these selections were specifically oriented to the vocabulary of the studies of psychology and education, this test demanded information recall from the examinees. Aside from this restriction, it was assumed that all items could be answered correctly solely upon the basis of the information given in these selections. This assumption may not have been entirely warranted.

Since most of the necessary background information was assumed to have been supplied in the test, no Knowledge category items were generated. On this basis, the test was intended to be a "higher mental processes" test. Since the major emphasis of the test involved logical analysis, it was assumed that the test was essentially an "Analysis" level test. The findings of the preliminary version as reported in Powell and Isbister (1969) confirm this assumption.

Content and Construct Characteristics of the Experimental Test

A detailed item-by-item discussion of the test may be found in Appendix B (see: p. 153 ff). In brief, five reading selections related to the area of educational psychology were chosen on the basis of information density and the unlikelihood of the examinees having encountered the selections previously. These selections which are

both given and referenced in Appendix B are referred to subsequently as:

1. Stupidity
2. Awareness
3. Aggression
4. Discipline
5. Progress

The 30 items in the test were classified using Bloom's Taxonomy as a construct model as indicated in Table 1 (p. 42) and elaborated in the discussions in Appendix B. No Knowledge-level items were developed. Items were classified as Synthesis if they required the examinee to organize the material from more than one reading selection into some systematic relationship when deciding which alternative to select for an answer. A reasonably high interrater reliability ($r = .83$) was found for the classification of the items based upon the item format, the stem, and the stem-right relationship when the right answer was indicated. Disagreement occurred among several raters, and among other reviewers of this study on the keying of some of the items. This disagreement would be expected from the subjective nature of the content and the related differences among the value systems inherent in any group.

Much less agreement among raters was found for the foil classification ($r = .62$)[5]. This poor result was expected for the

---

[5] A word of caution is in order. This low interrater reliability suggests that readers following the item-by-item discussion in Appendix B may disagree with the foil classification given and with the reasoning behind it. It would be interesting for the reader to record his disagreements and to compare these with the results of the cluster analysis as given in Table II, page 73, and Table 12, page 77.

TABLE 1

CLASSIFICATION OF ITEMS
USING BLOOM'S TAXONOMY
(30 ITEMS)

	Bloom's Category				
	Comprehension	Application	Analysis	Synthesis	Evaluation
Item Numbers	3,6,12,15	5,7,11	1,2,4,8,9 17,19,20,21 22,23,24,25,27	13,14 26,28 29,30	10,16,18
Totals	4	3	14	6	3

reasons already given (see: pp. 4-5), and the findings of Kropp et al (1966) which suggest multiple interpretations of specific items and alternatives within heterogeneous groups. This multiplicity would be expected to increase with the complexity and subjectivity of the content so that a high level of agreement, even among professionals on the particular test used in this study, would be unlikely.

To illustrate the extent of this problem, a check was conducted. One of the raters of the items disagreed with the classification of three foils in particular. Of these three only one of his reclassifications was supported by the cluster analysis as given in the results of the study (see: p. 182 for details). This one-in-three success ratio was equivalent to that of the experimenter.

The overall appearance of the experimental test suggests that in the traditional sense it is a very poor one. The internal consistency value for the test was K-R 20 = .34. A review of the item difficulties and biserials from Table 40 of Appendix A (p. 150) is equally discouraging. However, the use of Bloom as a model for the right answers and the Guidelines for the wrong answers suggests that the test should not be considered homogeneous. For this reason, and the reasons given earlier when discussing this same problem relevant to the Kropp et al (1966) study, (see p. 21 ff) the use of traditional evaluative procedures on this test may be questionable. Support for this position is found in the Procrustes rotation of the factors to fit the clusters which gives six nearly orthogonal factors which display quite adequate internal consistency (see: p. 70).

The foil classification procedure differed from the item procedure in two important respects. First, although the Guidelines

were also used as a model for the possible information content of the foils, the relationship between this model and any possible characteristics of the examinees, in the almost total absence of research, was largely unknown. Second, whether these Guidelines formed mutually exclusive categories, or a hierarchy paralleling Bloom could only be inferred from the assumptions which went into their development.

These foil Guidelines were used to help develop the foils as well as to classify them. A summary of the classification of the foils is given in Table 2.

TABLE 2

CLASSIFICATION OF WRONG ANSWERS
USING THE FOIL GUIDELINES

Classification of Foil										
A Strategy								B Misreading		C Other
OG	OS	Inv	Irr	IA	Sub	Tr	CM	WW	RT	O
1A	2C	4C	1D	1C	2B	30C	5B	7B	11D	All
4A	3D	8D	2D	5A	3A		9D	10D	15D	foils
6B	5C	9B	3C	8A	6C		10A	13D	29D	in
8C	7D	10C	4D	13A	7A		12B	15C	30D	items
12C	11C	11A	6D	25D	9C		13C	30A		19-24
14D	26C	12D	14A		14C		16D			26B
15A	29C	17C	16A	27A	17D		27D			
16C			18C	28D	18A		29A			
17B			25C		25A					
18D			26D							
28C			27C							
			28B							
11*	7	7	12	7	9	1	8	5	4	19

* = Frequency of foil in each category.

The test used in this study is a revision of the one reported in Powell and Isbister (1969) which had a slightly different purpose. The present discussion supported by the item-by-item analysis given in Appendix B would seem to demonstrate that, for all the faults of the instrument, the content and construct requirements for this test as laid out in Chapter II have been met to a reasonable degree of acceptability.

In the Powell and Isbister (1969) study the advance classification was taken as given and profile scores were developed accordingly. The resulting score sets were treated as independent variables and subjected to principal axis factor analysis in order to determine relationships among these scores. In this study the advance classification was not taken as given but subjected to a comparison with a cluster analysis based upon the relationships found among each of the alternatives. In this present study the acceptability of the advance classification system as exemplified in the test was being studied.

On the basis of what has already been said about the problems that communications of this type produce, it would be reasonable to expect the advance classification systems used in this study would not hold up without the qualifications derived from the possibility of 1) multiple interpretations of the communicating stimulus, 2) multiple methods of integrating and relating the stimulus to each individual's own experience, and 3) leading to multiple interpretations of the responses.

CHAPTER IV

THE DESIGN OF THE STUDY

The success of this study is contingent upon three aspects. First, the study must stand upon the acceptance of the logic of the content and construct validity of the experimental test as given in Chapter III.

Second, the construct validity must find evidential support in the statistical results of the analysis of the examinee performance on an administration of the experimental test. This support can be found in several ways. First, the advance classification may be found to reappear in the statistical patterns. Second, the content pattern might be shown not to be an important contributor to the statistical patterns. Third, in the event that the advance classification cannot be supported, some reasonable method of modifying the advance classification which does not violate the construct assumptions, such as possible multiple interpretation of the items, may be found. Fourth, the patterns should cross-validate between equivalent independent groups. Fifth, if cross-validation fails, a reasonable explanation which fits the data and the construct assumptions must be found to explain this failure.

Third, however much the construct validity is supported, wrong answers in some form must also contribute significantly to the prediction of achievement scores obtained in the usual manner before they can be considered to contain achievement information.

These three aspects form, in combination, the necessary and sufficient conditions needed to demonstrate that the method of test construction used in this study can be used to develop tests which

contain useful information about student performance in the answers given to the foils. A further restriction to this problem arose. Since the study began with categorical data, it should end with categorical interpretations in so far as is possible.

To begin with, however, the answer selections on the experimental test cannot be assumed to have any of the usual continuous distributions. The selection pattern can be considered to be categorical, since one choice is made for each item, but not dichotomous.

An expedient method of defining categorical data mathematically is to treat categorical membership as "one" (1) and nonmembership as "zero" (0). A matrix of categorical data should have the following properties:

1. The centroids of normalized clusters from the matrix should tend to be either orthogonal or opposite each other.
2. The orthogonal projections of the members of a cluster upon its centroid should be near unity.
3. The orthogonal projections of the members of a cluster upon the centroids of all other clusters should either be non-existent, or nonsignificant.

Figure 1 illustrates a typical response matrix which displays these properties for the twelve variables included, and may display these properties for some reduction of the matrix to less than twelve variables. Figure 1, a sample response matrix is on page 48.

The usual procedure for test analysis is to use the right answer portion of Part B and Part C (the total number correct) and to treat the wrong answer division of Part B as redundant.

If all four alternatives of each item are considered the

Student Number	Part A												Part B												Part C
	Item Alternative												Right Answers			Wrong Answers									Total Number Correct
	1				2				3																
	a.	*b.	c.	d.	a.	b.	c.	*d.	*a.	b.	c.	d.	1.	2.	3.	1.	2.	3.	4.	5.	6.	7.	8.	9.	
1	0	1	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	3
2	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	2
3	0	0	1	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	2
4	0	1	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	0	2
5	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1

* = Correct response

FIGURE 1

A SAMPLE RESPONSE MATRIX

statistical problem of linear dependency arises. To illustrate what is meant by linear dependency refer to Figure 1 above. Notice that in Part A of this figure the sum of each row is always three. Only in the case of omitted items will the sum of the answers be less than the number of items.

Since all alternatives are being counted, this total is predetermined as being the number of items. If the columns are added vertically, the sum of the columns within an item is predetermined at the number of examinees. In the case of Figure 1 above, this sum is 5. In many statistical procedures, linear dependencies have the effect of rendering indeterminate or non-unique solutions.

The solution to this problem used in this study was to partition the matrix as indicated in Part B of Figure 1 (see: p. 48). Two matrices, one for the right answers and one for the wrong answers, were made from the original response matrix. The categorical property was retained within each of these two new matrices. This procedure had the effect of treating right and wrong answers as though they were independent.

Part C of Figure 1 (see p. 48) shows the row sum (horizontally) of the right answer partition of Part B. This sum, which is the total-correct score, is the usual approach to the interpretation of test results. It is with Part C that the results of the statistical analyses of Part B are being compared.

There are several possible methods of dealing with categorical data. Since this study is concerned with relations among categories the most reasonable approach is to begin with phi correlation coefficients between the category pairs. This procedure produced two correlation matrices, one 30 by 30 for the right answers, and one 90 by 90 for the wrong answers.

Since the results of these analyses were to be cross-validated, the original group of examinees were subdivided by randon assignment into two groups (Group A and Group B). The data for both groups were subjected to the same statistical treatment although most of the interpretive work was done with the results from Group A.

The result of this latter subdivision was that the analytical aspects of this study began with four phi coefficient matrices (one for right answers and one for wrong answers for each of Groups A and B). These four matrices were the basic data for much of this study. They

may be found in Tables 32 to 39 of Appendix A.

The phi matrices gave relationships among pairs of alternatives only. To proceed further, it became necessary to find relationships among these relationships. From the original structure of the experimental test there were two patterns of relationship which could be sought. The first was the pattern as defined by the advance classification based on Bloom's Taxonomy, the second was the pattern as defined by the content (information background) of the items and foils.

One of the methods of checking the data for these patterns which could be used is the Procrustes rotation solution to factor analysis. The procedure began with the principal axis factor solution and found the best rotation of this solution in a least squares sense for a given matrix.

A factor solution was used to remove as much measurement error as possible from the further analytic procedures used in this study. The phi coefficient is extremely sensitive to the marginal proportions, particularly when the selection ratios deviate considerably from .50, as in this study where four alternatives are being used. Slight changes can have a profound effect upon specific coefficients. This effect can be reduced by the factoring procedure which takes the relations among coefficients into account.

The principal axis solution was used to get as much variance as possible in as few factors as was reasonable.

A third approach normalized the principal axis matrix by rows and then found the distances between the ends of the resultant vectors by the usual distance formula. Clusters were then defined in terms of minimizing within cluster distances and maximizing between cluster

distances. The mathematical procedure used in this study is given in Appendix A (see: Table 41, p. 151).

The advantage of this procedure is that if a good fit is obtained the solution alleviates many of the problems of rotation which are otherwise inherent in factor analytic solutions.

All three of these solutions can lead to results which can be interpreted categorically. If a good fit is found with the target matrix for advance classification either by process or by content, then the categories of the original classification were to be used. If, on the other hand, good fits were not found, then the categorical solution of the cluster analysis procedure would be studied for possible interpretation on the basis of either process or content. In this latter case, the data would also have to show that there were no contradictions to the content or construct assumptions as given in Chapter II, (p. 12 ff) otherwise this study would not meet the necessary and sufficient conditions required as outlined at the beginning of the present chapter on page 46.

An additional advantage of these three procedures is that they all begin from a principal axis factor solution of a correlation matrix. Furthermore, if the same factor solution is used in all three cases, the goodness of fit of the cluster solution can also be found by the Procrustes method. Hence all three of these solutions can be subjected to the same criteria.

Since the object of this study was to support the construct characteristics of the experimental test, the analysis began with an attempt to find the best possible cluster solution to this construct criteria for the data of Group A. It was decided that the best possible solution would involve having clusters defined by the most frequently

recurring category as defined by the advance classification. The number of these identifying elements was to be as large as possible for each solution, i.e. the right answers and the wrong answers. The number of factors in the principal axis solution needed for this result was then taken as standard for all solutions involving the same kind of data. For instance, six factors gave the best solution for the right answers for Group A. Hence, six factors were used for all right answer analyses.

For cross-validation the identical statistical procedure used with Group A was repeated with Group B. Cross-validation was then established once again from a best-fit match (in terms of most frequently recurring members) between the categorical results for Group A and Group B. Several procedures were used until a satisfactory match was found. Once again, the cross-validation could not violate the construct considerations outlined in Chapter II, page 12 ff for this study to be successful.

Finally, the categories which were established as being potentially meaningful in the earlier parts of the study were used as a basis for rescoring the experimental test. The results of these subtest scores were combined in several ways and their ability to predict the total-correct scores of two independent achievement tests for the same subjects was compared with the predictive power of the total-correct score. In this latter case it would be necessary to show that the use of wrong answers consistently improved prediction over total-correct score and combinations of right answers.

If all these criteria were met, the value of wrong-answers part of performance information would be demonstrated. With so many criteria to meet, the probability that such conclusive evidence would

be found is exceedingly low. On the other hand, trends in the directions indicated could be treated as suggestive. The borderlines between undemonstrated and suggestive, and suggestive and conclusive, are unclear and subject to disagreement.

Statement of the Problem

Since this study is exploratory, attempting to demonstrate the presence of information in wrong answers and to discover the major properties of this information, an elaborate theoretical structure for formulating testable hypotheses was considered to be unnecessary. Instead, the procedures suggested for the establishment of grounded (data-based) theory as outlined by Glaser and Strauss (1967) was used.

Such theory as is used in this study comes from well established principles in psychology, communication theory, and test construction theory. Beginning with the S-O-R paradigm commonly used in problem-solving studies, it became evident from communication theory that each of the members of this paradigm may best be considered a composite. That is, any specific stimulus may be subject to a range of interpretations. If this stimulus requires the solution of a problem, the specific interpretations may be subject to a range of solution strategies some of which may lead to "correct" and some to "incorrect" solutions. In the multiple choice test, the examinee can be expected to try to match the alternatives given him in the item within the interpretation range and strategy range available to him. In this case, the most reasonable assumption would be that most, if not all, responses given by an examinee to a multiple choice achievement test would be selected on a systematic basis.

If some characteristic of particular alternatives in two

separate items are sufficiently similar to the apparent right solution in the view of the examinee, he can be expected to choose both of them. If a sufficiently large number of examinees select this same pair of alternatives this joint selection will appear as a high correlation in the phi coefficients relating the two events, thus becoming "systematic" in that it would produce a significant statistical event.

In the usual procedure used for scoring multiple choice achievement tests, only the right answers are treated as systematic in this sense, hence the requirements usually set for their performance. This study addressed itself to the exploration of the possibility that "most if not all of the answers given to multiple choice achievement tests are selected upon a systematic basis." This psychological hypothesis is the basic theoretical proposition proposed by this study.

Since it is possible that wrong answers may influence the way in which items behave, and, inferring from communication theory, the suggestion emerges that each alternative may have more than one interpretation among a group of examinees. Thus, there are four possibilities, 1) that the systematic characteristics depend upon content; 2) that the systematic characteristics depend upon the advance classification as defined by the two process models of Bloom's Taxonomy and the Guidelines; and 3) and 4) that the systematic characteristics depend upon multiple interpretations as based upon content or process. The study could have no commitment toward any of these four possibilities.

For an exploratory study, Q.E.D. can be written at this point without further interpretation attempts. The developmental characteristics of wrong answers, their relationships with personality variables,

with right answers, etc. exceeds the scope of this dissertation. These topics are, of course, legitimate areas for future research.

The Sample Used

The experimental test was administered to 277 summer school students in a one semester course in educational psychology at the senior level. The age group range of these students was from 19 to 55 with the median age about 30, and most of the students having had some teaching experience. The overall group was subdivided by random assignment into two groups (Group A of 139 students; and Group B of 138 students). A t-test for independent samples based on the total-correct scores of the experimental test designed to confirm the equivalence of the scores of these two groups is reported on page 92.

CHAPTER V

RESULTS AND THEIR INTERPRETATION

A somewhat different procedure to the one usually employed was adopted for this study. To begin with, the usual procedure for scoring and interpreting multiple choice achievement tests is to count the number of items each examinee has correct. This procedure is sometimes modified by the specification, by various methods, of subtests of the total test. One of two general procedures is usually employed. Either the experimenter establishes the categories into which the items fall in advance of the test administration and then interprets his results on this basis, or he groups his results on the basis of some analytical procedure and then endeavours to interpret these groupings. Powell and Isbister (1969) used the former procedure, and Powell (1968) used the latter. In general, only right-answer information is used.

The present study endeavours to link advance classification and statistical classification, and also endeavours to use wrong answers as well as right answers in the interpretation of test results. As has already been indicated, very little research of the type just described is present in the literature. For this reason, this study can best be described as exploratory in which negative results are more likely to be indicated than are positive results.

Each item on the experimental test had four alternatives, hence the study began with four variables for each item. The response matrix, therefore, contained a "one" (1) for each alternative selected by each examinee; otherwise "zero" (0). Since the examinee was allowed no more than one choice per item for 30 items, each examinee would have a

maximum of 30 "ones" in the vector of 120 variables which represented his selections. Because these selections were further restricted to one in each group of four, each variable was linearly dependent upon the other three in the same item. In order to remove these linear dependencies, the performance matrix was partitioned into a right-answer matrix and a wrong-answer matrix. These latter two matrices were subsequently treated as being independent.

In order to attempt to cross-validate the findings, the examinees were randomly assigned to two groups, Group A and Group B. All the statistical analysis done which was not related to cross-validation was performed on the data from Group A. The relationship between the mean total-correct scores of Group A and Group B is given in Table

In addition, since a relationship between advance classification and statistical ordering was being attempted, an advance classification system was used separately for the items as represented by their correct alternatives and their foils. These classification systems were discussed in detail in Chapters II and III.

Since an attempt to find a consistent interpretation of performance is being made, the examinees were randomly assigned to two groups so that the interpretations could be examined for cross-validation. Hence, the basic data for this study consists of two phi correlation matrices (see: Appendix A, Tables 32 to 39) for each of the two groups. The correlation matrices represent the intercorrelation between variables across examinees for the right answers and for the wrong answers in each group.

Finally, two achievement test scores were obtained for each

examinee. One of them was concurrent in the sense that the experimental test formed a subtest in the mid term examination given in a one-semester course. The other achievement score was part of the final examination in the same course. This data were collected so that the predictive validity of the various interpreted categories and their predictive cross-validation could be determined.

Several steps were taken in each phase of the analysis. For instance, attempts were made to interpret the right answers on the basis of both factor analysis and inter-point distance cluster analysis. This step was followed by a detailed logico-semantic analysis of the right answer clusters in an attempt to interpret these clusters.

A similar logico-semantic analysis was made of the wrong-answer clusters.

Attempts were then made to cross-validate the advance classification, the interpreted clusters and a particular grouping of the interpreted clusters.

Finally, the predictive validity of the advance classification, the interpreted clusters, and the grouped clusters was found. This validity was found in each case by using the right answers alone, and the combination of both right and wrong answers.

The discussions which follow adhere to this sequence.

Interpretation of Right Answers Using Factor Analysis

On the basis of the advance classification there were two possible interpretations based upon either of two independent classification systems with respect to the right answers given by the examinees. One of these interpretations could have been best described as a

"process" interpretation based upon classification of the items on the basis of Bloom's Taxonomy. The other possible interpretation was "content" in which the items were classified on the basis of the information background required to answer them.

An attempt was made to verify the possible existence of either or both of these two interpretations. The primary data for this attempt was a six-factor unrotated principal axis factor matrix derived from the phi correlations for the right answers. This matrix was rotated by a Procrustes solution to find the best fit (in a least squares sense) to two target matrices. The first of these targets specified a simple structure which indicated the way in which the items were classified using Bloom's Taxonomy. The second target matrix specified a structure which indicated which items referred to each of the several reading selections. The matrix structure was not always simple since some of the items referred to more than one selection.

Table 3 (see: p. 60) gives the target matrix and the pattern on the primary axes as related to the "process" classification of these items.

It is evident from the results that the pattern does not reproduce the target matrix in any satisfactory manner. This finding suggests the conclusion that the advance classification of items using Bloom"s Taxonomy did not give a satisfactory indication of the way in which each item performed. Table 4 gives the correlation between the primary axes in this solution. Table 4 is on page 61.

TABLE 3

PROCRUSTES ROTATION OF THE ADVANCE CLASSIFICATION OF RIGHT ANSWERS

Item No.	Bloom's Classification Target Matrix					Pattern on Primary				
	2.00	3.00	4.00	5.00	6.00	I	II	III	IV	V
3	1.00						1.87[a]		1.62[b]	
6*	1.00					2.35[c]	1.40		1.06	-1.33
12	1.00					2.03	3.46			1.99
15	1.00					1.21	4.44			3.55
5*		1.00				1.60	6.44			3.89
7*		1.00				1.03	3.63			3.51
11		1.00					3.55			2.46
1			1.00				1.36	1.04		2.50
2			1.00				1.15	1.57		2.60
4			1.00			1.53	-1.51	1.04		-2.02
8*			1.00					1.40		1.17
9			1.00				-4.95			-3.73
17			1.00			1.00			1.21	
19			1.00			1.42	4.97			3.16
20			1.00				2.40			2.75
21			1.00				1.02	1.11		1.79
23			1.00							
24			1.00				1.79			
25			1.00				-3.35	1.28		-1.38
27			1.00				-6.23	1.33		-5.55
13				1.00		2.46	1.09		1.23	-1.79
14				1.00		1.91	6.71		1.39	3.19
26				1.00				-1.15		1.82
28				1.00			1.62	1.12		3.02
29				1.00						
30				1.00			-1.97			-3.42
10*					1.00		4.91			5.22
16					1.00		4.44			3.55
18					1.00	1.07	2.60			2.40

a. The numbers in Italics had the highest loadings.
b. Only those loadings with an absolute value of 1.00 or greater are shown.
c. The items which are starred (*) approximate the target.

TABLE 4

PROCRUSTES ROTATION OF THE ADVANCE CLASSIFICATION OF RIGHT ANSWERS

Correlation Between Primary Axes					
	I	II	III	IV	V
1	1.00				
11	-.96	1.00			
111	-.93	.04	1.00		
1V	.80	-.91	-.83	1.00	
V	.95	-1.00	-.94	.90	1.00

The primary axes (Table 4) were highly correlated, suggesting that by this classification system, there may be only one factor present.

As indicated on page 88 an identical procedure was used to examine the data for the possible presence of "content" factors. Table 5 (see: p. 62) gives the target matrix and the pattern on primary for this Procrustes rotation solution.

The fit of this matrix to the target based on content is only slightly better than for process-oriented advance classification. Factor V loading with items 21, 24, and 25 show a nearly simple structure which coincides with the target matrix. These three items also formed a unique cluster on the basis of the cluster analysis conducted later in the study. It is possible, however, to give a process interpretation to the cluster which may mean that this pattern for content might be coincidental.

Aside from these items the pattern did not reproduce the target

TABLE 5

PROCRUSTES ROTATION OF THE INFORMATION CONTENT OF RIGHT ANSWERS

Item No.	Target Matrix					Pattern on Primary				
	Stupidity	Awareness	Aggression	Discipline	Progress	I	II	III	IV	V
1	1.00					.53[b]		.60[a]		
2	1.00					.55		.85		.56
3*[c]	1.00					.77	.61			
4*	1.00					.69				
5		1.00					1.13		1.24	
6*		1.00				.76	.96	.67		
7		1.00					.52	.51	.63	
8			1.00			1.05		.81	.65	
9			1.00				-.69		-1.01	
10*			1.00					.53		
11			1.00							
12*			1.00				.67	.69		
13	.70		.70			1.10	.91	.66	.66	
14	.70	.70				.65	1.30	.57	1.05	
15*				1.00			.78	-.50	1.43	
16				1.00				-1.04		
17				1.00		1.15	.73	.78	.69	
18*				1.00			.48	-.77	1.21	
19					1.00	.54	1.13		1.08	
20					1.00			.54		
21*					1.00					.73
22					1.00				.52	
23					1.00	.47		.47		
24*					1.00		.64			.66
25*					1.00					.63
26					1.00					
27					1.00	.53				
28	.50	.50	.50		.50	.85		.71	.69	
29				.70	.70					
30	.44	.44	.44	.44	.44					

a. The numbers in *Italics* had the highest loadings.
b. Only those loadings with an absolute value of .44 or greater are shown.
c. The items which are starred (*) approximate the target.

matrix in an acceptable manner. Content would seem to be only slightly better than process as a means of classifying items in advance of their use.

Table 6 shows the correlations between the primaries.

TABLE 6

PROCRUSTES ROTATION OF THE INFORMATION CONTENT OF RIGHT ANSWERS

Correlation between Primary Axes					
	I	II	III	IV	V
I	1.00				
II	.14	1.00			
III	-.76	-.39	1.00		
IV	-.52	-.81	.58	1.00	
V	.30	-.12	-.46	-.08	1.00

The Interpretation of Right Answers Using Cluster Analysis

The negative results just reported suggested the need to search for a multiple interpretation possibility. Hence, a cluster analysis procedure which normalized the same factor matrix by rows as was used in the two solutions just given. The normalization involves dividing each member of a row in the factor matrix by the square root of the communality. Since this value is the length of the vector given by the

row of factor loadings, this division raises this length to unity (one).

The procedure then calculates the interpoint distances from the ends of the vector pairs, surface-to-surface, across the hypersphere. The square of this distance is the sum of the squares of the differences across the rows taken in pairs. This interpoint distance is then used to form clusters in which the within-cluster distances are minimized as indicated by the formulae given in Table 41, p. 151. The clustering begins with as many clusters as variables and ends with all variables in one cluster. In addition, there is a unique cluster solution for each factor solution which might be used or with the inclusion of each additional factor. The experimenter, therefore, was left with the problem of determining which of many possible solutions to choose. Repeated attempts suggested an advantage to the process classification of Bloom's Taxonomy. It was decided to consider a cluster to have recapitulated the advance classification if it contained more members from one particular advance category than from any other category. The solution which gave the best recapitulation was then sought, by iteration, for both right and wrong answer clusters. The cluster was assumed to be identified on the basis of the recapitulated category.

For the right answers, in this sense, the best solution was derived from an unrotated principal axis factor matrix of six factors. In this solution twelve of the thirty items recapitulated in the clusters. This result is four times better than the Procrustes rotation to fit content just reported. Since this was 40 per cent of the items, in spite of multiple interpretation possibilities, the result was reasonably satisfactory.

An examination of the data suggested that the first unrotated factor might be a "difficulty" factor. Table 7 on page 66 expands upon this relationship.

In general, the value of the loading on Factor 1 seems to be about 50 per cent of the value of the difficulty. The correlation between these two variables was r = .65.

Does this finding seriously disrupt the use of the six factor solution? Table 8 on this page compares the interpoint distance clusters as determined with and without Factor 1.

TABLE 8

RIGHT ANSWER CLUSTERS
WITH AND WITHOUT
FACTOR 1

	With Factor 1				Without Factor 1						
C_1	1	2	8	28	1	2	8	*22*[a]	*17*	28	*23*
C_2	3	17	30		3	30					
C_3	4	13	6		4	6	13				
C_4	5	14	19		5	14	19				
C_5	7	22	23								
C_6	9	27			9	27					
C_7	10	11	16	29	10	29	11	16	*26*		
C_8	12	20	26		*7*	12	20				
C_9	15	18									
C_{10}	21	25	24		*15*	*18*	24	21	25		

a. The numbers in *Italics* displaced with the removal of Factor 1.

TABLE 7

RELATIONSHIP BETWEEN ITEM CONSISTENCY, ITEM DIFFICULTY AND ITEM FACTOR LOADING ON UNROTATED FACTOR 1

Item Number	Internal Consistency	Difficulty	Factor 1
1	.384	.086	-.007
2	.456	.245	.150
3	.304	.345	.176
4	.425	.173	.037
5	.295	.223	-.200
6	.473	.525	.392
7	.336	.144	-.185
8	.419	.295	.022
9	.135	.698	.245
10	.320	.173	.074
11	.334	.475	.216
12	.430	.345	.181
13	.378	.669	.118
14	.405	.460	.114
15	.015	.050	-.581
16	.447	.813	.314
17	.546	.079	.194
18	-.295	.014	-.732
19	.463	.309	.055
20	.269	.532	.055
21	.404	.324	.300
22	.001	.072	-.587
23	.165	.583	-.084
24	.377	.856	.402
25	.329	.547	.340
26	.140	.381	.101
27	.388	.849	.414
28	.378	.367	.048
29	.275	.827	.243
30	.158	.734	.182

The Italics in Table 8 on page 65 indicate that seven items move by dropping Factor 1. This fact does not seriously affect the replication of the advance classification in the new solution. Also, most of the items which move to a new cluster do not follow the general rule that the difficulty be roughly twice the factor loading. For these reasons, it was decided to retain the complete six factor solution throughout subsequent analyses.

Since a Procrustes rotation was used to determine how well the advance classifications fit the data, it was reasonable to use the same procedure with the cluster analysis data. Table 9 (see: p. 68) reports the target and pattern matrices in solution.

The pattern on primary in Table 9 is a good reproduction of the simple structure of the target matrix. Furthermore, if the values of "h" (i.e. the square root of the communality from the six factor matrix) are taken into account, then the pattern seemed to fit even better. The cluster analysis was produced from the interpoint distances derived from a normalized matrix. For this reason, to find the length of the largest vector approximating the overall length of the vector in the unrotated six factor solution add to the acceptability of the solution.

In short, the cluster solution was a far better fit to the data than either of the two methods of advance classification. The independence of the interpoint distance clusters from rotation problems also reinforces the acceptability of this approach. Table 10 (see: p. 69) reports the correlation between the axes.

The relatively low correlations between the axes in Table 10 further strengthens the support for the use of interpoint distances as an analytical technique for the problem under examination in this study.

TABLE 9

PROCRUSTES ROTATION OF THE INTERPRETABLE CLUSTERS OF THE RIGHT ANSWERS

Item No.	Target Matrix							Pattern on Primary					
	h[a]	C_1	C_5	C_6	C_7	C_8	C_{10}	C_1	C_5	C_6	C_7	C_8	C_{10}
1	.40	1.00						.40[b]					
2	.55	1.00						.43					
8	.62	1.00						.63					
28	.71	1.00						.62					
7	.49		1.00						.52				
22	.71		1.00						.77				
23	.51		1.00										-.43
9	.52			1.00						.48			
27	.71			1.00						.49			
10	.57				1.00						.44		
11	.48				1.00						.46		
16	.60				1.00						.49		
29	.28				1.00						.23[c]		
12	.70					1.00						.68	
20	.57					1.00						.46	
26	.61					1.00						.40	
21	.61						1.00						.60
24	.68						1.00		-.49				.58
25	.62						1.00						.54

a. The symbol "h" stands for the square root of the communality.
b. Only those loadings greater than an absolute value of .35 are reported.
c. This loading, although $\angle .35$, is reported because of its low value for 'h'.

TABLE 10

PROCRUSTES ROTATION OF THE
CLUSTER ANALYSIS OF THE
RIGHT ANSWERS

	Correlation between primary axes					
	C_1	C_5	C_6	C_7	C_8	C_{10}
C_1	1.00					
C_5	-.31	1.00				
C_6	-.07	.18	1.00			
C_7	-.17	.31	-.03	1.00		
C_8	.06	-.29	-.03	-.08	1.00	
C_{10}	.12	.20	.15	-.12	.09	1.00

The largest correlation between axes given on Table 10 was -.31 which represents an angle of more than 70° between this pair of axes. It is possible, therefore, to state that interpoint distance clusters produced a solution which was independent of the usual rotation problems and which was approximately orthogonal. The procedure gave a very satisfactory statistical representation of the data. On the other hand, the failure of the advance classification systems to render interpretability left the researcher with the problem of interpreting these clusters.

Finally, a matrix which is in fact categorical in the sense given on page 47 would be expected to have low correlations between the primary axes for a Procrustes rotation. When the dimensionality of this categorical matrix has been reduced by factor analytic techniques, before rotation, the best fit from a Procrustes rotation should display

the retained lengths of the vectors (i.e. the square roots of the communalities) as the loadings of the pattern on the primary axes. Also, the pattern on the primary axes should display a structure similar to the target matrix. All three of these properties must be present for the inference to be made that the principal axis matrix is categorical. It is evident from Tables 9 and 10 (pages 68 and 69) that all these conditions were met, strongly suggesting that the cluster analysis technique gives a good categorical solution to the six factor data. The low internal consistency was thus explained, suggesting profile scores from the clusters may better describe the data than total correct scores alone. Finally, the close match between the original vector lengths and the facdor loading in the Procrustes rotation and the near orthogonality of the factors suggests that this solution clearly identifies the homogeneous subtests of the right answers. These findings contradict the apparently poor showing of this test in the usual analytical setting.

The Meaningful Interpretation of Item Clusters

With the failure of the advance classification, the multiple interpretation hypothesis was the only alternative to investigate hence some reasonable common ground was needed for each cluster if answering was systematic. The first possibility which had to be either confirmed or eliminated was that the clusters were sufficiently strongly content oriented to suggest content as a possibility. The only cluster in which content was at least a strong contender to process was the Cluster C_{10} containing items 21, 24, and 25. All these items were based upon reading selection number five (Progress), but also, all three were classified as Analysis items. A closer look was taken at this cluster, along with all the others (see: Appendix C). This look supported the process

interpretations over the content interpretation. Thus, it was reasonable to reject content as a basis for the interpretation of any of the right answer clusters.

The misfit items in the identified clusters were examined by logico-semantic (structure-meaning) analysis to determine if they could be reasonably reclassified in common with the recapitulating category. Success by this procedure would lend support to the multiple interpretation hypothesis. All misfit items in clusters C_1, C_5, C_7 except possibly item 7 could be reclassified to fit the overall classification of these clusters, adding three or four items to the original twelve.

The logico-semantic analysis of Cluster C_8 revealed that the formation of an inductive structure within a particular reading selection could possibly be the basis for synthesis items. For this reason C_8 was classified as Synthesis, adding another three items to the support of the multiple interpretation possibility.

This procedure gave 19 of the 30 items, or 63 per cent, of the items a reasonable classification based upon process. Since the inter-rater reliability was only moderately high ($r = .83$) and since the cross-validation based upon interpretable reasons found in Powell (1968) was only 64 per cent, this level of recapitulation can be considered to be satisfactory.

For the remaining four clusters the interpretation was ambiguous. Cluster C_9 containing items 15 and 18 seemed to be simply poor items. Cluster C_4 seemed to involve implication or extrapolation from the relevant content suggesting comprehension but in the absence of better definitions for strategies it was safer to call this cluster ambiguous (see: p. 197). The remaining two clusters (C_2 and C_3) seemed to be

strongly influenced by the nature of the foils in the items which tended, in general, to lower them to Comprehension items; but each had some Analysis characteristics leaving their classification ambiguous, which seemed to further support the multiple interpretation hypothesis. These decisions are summarized on Table 11, page 73.

Trying to make Synthesis items by combining two or more reading selections proved unsuccessful for a number of reasons, suggesting the need for more research on this method.

Thus, the results of the logico-semantic analysis of the item clusters suggested reasonable support for 1) the multiple interpretation hypothesis, 2) the transcendence of process over content, and 3) the suggestion that foils influence item performance.

The summary just presented is supported by a detailed discussion given in Appendix C (see: p. 186 ff). The details are also presented, first, because it was felt that the effective reclassification represented evidential support for the multiple interpretation hypothesis, and second, because subsequent researchers might find value in an independent evaluation of the logic behind the conclusion of this study.

The Meaningful Interpretation of Wrong Answer Clusters

The tables for the factor analysis of the wrong answers were very large since they involved 90 variables (see: Appendix A, Tables 34 to 39).[6] The analysis of the right answers showed that interpoint distance cluster analysis gave a good representation, in a statistical

---

[6]The *phi* coefficients in these tables are in the following sequence: variables 1 to 30 represent $1D_1$ to $30D_1$; variables 31 to 60 represent $1D_2$ to $30D_2$; and variables 61 to 90 represent $1D_3$ to $30D_3$.

TABLE 11

CLASSIFICATION AND MEMBERSHIP OF

RIGHT ANSWER CLUSTERS AS

DERIVED FOR GROUP A

Cluster Label	Item Membership				Advance Classification	Interpreted Classification
C_1	1[a]	*2*	*8*	28	Analysis	Analysis
C_2	3	17	30			Ambiguous
C_3	4	13	6			Ambiguous
C_4	5	14	19			
C_5	7	*22*	*23*		Analysis	Analysis
C_6	*9*	*27*			Analysis	Analysis
C_7	*10*	11	*16*	29	Evaluation	Evaluation
C_8	12	20	26			Synthesis
C_9	15	18				Ambiguous
C_{10}	*21*	*25*		*24*	Analysis	Analysis

a. The numbers in *Italics* were all in the category named in the advance classification.

b. The "interpreted classification" of each cluster is given in Appendix C beginning on page 186.

sense, for those data. Finally, the interrater reliability was lower for wrong answers than right answers, suggesting that the advance classification of foils would be less likely to reappear in the data than the right answer. For these three reasons, the interpretation of the wrong answers began with the cluster analysis. Attempts to fit the unrotated principal axis factor matrix to the advance foil classification and to the foil content were not made. There was no reason on the basis of the characteristics of the results of the cluster analysis to assume that the results of these two preliminary steps would have been substantially different for the wrong answers than they were for the right answers.

The best replication of the advance foil classification which could be found in the cluster analysis involved a 25-factor solution to the phi correlation matrix of the wrong responses, and 15 clusters of foil in this solution. This replication placed 28 of the 90 foils (or 31 per cent) into clusters which might be considered equivalent to the categories of the advance classification on the basis of the most frequently occuring category of foil in that cluster. This proportion (31 per cent) was not quite as good for the wrong answers as the corresponding proportion (40 per cent) was for the right answers.

In getting this best replication, the same procedure was used for determining the number of clusters as was used for right answers.

All 90 foils were clustered. Once the best replication was found, however, those foils for which the selection ratio was less than .06 were dropped from further consideration. This precedent was established when the interpretation of the right answer clusters was being made. The result of the dropping of foils of low selection ratio

was to reduce the number of foils under consideration from 90 to 60. Of these 60 foils, 18 foils (or 30 per cent) continued to meet replication requirements. The proportion is essentially the same as before.

The interrater reliability for foil classification ($r = .62$) was not as high as for the right answers. It is possible that this figure might have been considerably lower, had the right answers to the items not been clearly indicated to the raters at the time of the rating. The procedure of rating involved comparing the foil, stem, and right answer relationships to the definitions of foil categories as listed in Chapter III. It became fairly evident that at least some of the foils might be placed quite reasonably into several different categories. The problem of multiple classification of foils will be dealt with in more detail in the interpretation of the wrong answer clusters. (see: Appendix C, p. 209 ff).

Briefly, the advance classification of foils were arranged into three or four possible general classes. These general classes were 1) Strategy Errors, 2) Misreading, 3) Misinformation, and 4) Other. The Other, (O), category was for foils which for some reason could not be readily classified into some established category. In the Experimental Test this category referred primarily to the foils for items 19 to 24 inclusive. In these items a different item format was used to that of the remainder of the items. It was not possible, in advance, to know whether or not this difference in item format would influence the way in which the foils behaved statistically. It was assumed in the interpretation of the clusters that "O" type foils which were found to be in reasonable association with foils of specified categories were not influenced in their behavior by the item format. If the Other (O) type

foils formed their own unique clusters, these clusters were assumed to represent categories of foil not identified by the advance classification of foils. Two such categories appeared in the data.

The specific categories used in the advance classification of foils were as follows:

	Name	Symbol
1.	Overgeneralization	OG
2.	Oversimplification	OS
3.	Substitution	Sub
4.	Inversion	Inv
5.	Invalid Assumption	IA
6.	Irrelevancy	Irr
7.	Common Misconception	CM
8.	Word-Word Link	WW
9.	Transposition	Tr
10.	Redefinition of Terms	RT
11.	Other	O

A cluster was identified on the basis of the most frequently occurring foil of a particular advance classification in that cluster. The identification given in Table 12 (see: p. 77) is based on all of the foils in each cluster before low selection ratio foils were eliminated. This identification was used as a starting point for the attempt at meaningful interpretation of each foil cluster. The final meaningful interpretation of each cluster is also given in Table 12.

Once again the members of each cluster were examined in an attempt to determine the common basis upon which these foils clustered together. Particular attention was paid to the foils which were not in

TABLE 12

CLASSIFICATION AND MEMBERSHIP OF WRONG ANSWER CLUSTERS AS DERIVED FOR GROUP A

Wrong Answer Cluster	Foils in Each Wrong Answer Cluster										Identification by Advance Foil Classification	Interpretation by Logico-semantic Analysis
W_1	$1D_1$	$2D_1$	$23D_1$	$22D_3$	$8D_1$	$17D_1$					OG	OG
W_2	*$3D_1$*[a]	*$20D_1$*	*$7D_1$*	*$24D_1$*	*$4D_2$*	*$18D_3$*	*$8D_2$*	*$23D_3$*	*$25D_3$*	*$29D_3$*	Sub	Eliminated
W_3	$4D_1$	$26D_1$	*$30D_1$*	$5D_2$	$6D_2$	$14D_3$	$13D_2$	$29D_2$	*$21D_2$*		U[b]	CM
W_4	*$5D_1$*	*$15D_1$*	*$11D_3$*	*$27D_1$*	*$1D_3$*	$30D_3$					U	RT
W_5	*$6D_1$*	*$16D_2$*	$12D_3$	$11D_2$	$12D_2$	*$2D_3$*	$19D_2$	$20D_3$			O	Inv
W_6	$9D_1$	*$17D_2$*	$25D_1$	$21D_3$	$7D_3$	$26D_3$					U	NS
W_7	$10D_1$	$19D_3$	$4D_3$	$13D_3$	$12D_1$	$15D_3$	$18D_2$	$26D_2$	$8D_3$		U	U
W_8	$11D_1$	$28D_3$	*$13D_1$*	$14D_2$							IA	IA
W_9	*$14D_1$*	*$1D_2$*	$3D_3$	$28D_1$	$10D_2$	$17D_3$	$22D_2$	$5D_3$			OS	OS
W_{10}	*$16D_1$*	*$20D_2$*	$29D_1$								U	RT
W_{11}	$18D_1$	$7D_2$	$6D_3$	$16D_3$	$2D_2$	$10D_3$					WW	WW
W_{12}	$19D_1$	*$22D_1$*	*$9D_3$*	$24D_2$							O	O_1
W_{13}	$21D_1$	$28D_2$	$23D_2$								O	O_2
W_{14}	$3D_2$	$25D_2$	$9D_2$	$27D_2$							Irr	Irr
W_{15}	*$15D_2$*	$30D_2$	$24D_3$	*$27D_3$*							U	Tr

a. Foils indicated by *Italics* were dropped from the analysis on the basis of a low selection ratio after the factor analysis was completed.

b. The U stands for an "unclassified" cluster.

the common advance category present in the cluster. In addition, the evident influence of foils upon the performance of an item led to the possibility that the interpretation of a foil by category might depend upon the point of view of the interpreter. The low interrater reliability suggested this possibility. For at least some foils several classifications may have been reasonable. In this case the foil categories would not be independent or unambiguous. If foils are only interpretable after a test has been given, then these interpretations become specific to the particular examinees upon whose performance the interpretations were based. That is, the expectation would be that where multidefined foils were concerned, the proportion of cross-validating foils on a cluster-for-cluster basis would be low.

A procedure similar to the one employed for item clusters was used with wrong answer clusters. For complete details see Appendix C (see: p. 209 ff).

The possibility of multiple interpretations of wrong answers was most clearly illustrated in the case of foil $2D_1$. Briefly, the classification of this foil was initially set as Substitution (see: p. 156) because it substituted a conjunctive for a disjunctive relationship. However, it is possible to look at these relationships as Overgeneralization or Oversimplification as the discussion in Appendix C indicates (see: p. 213).

Apparently, considerably closer analysis of the logico-semantic relationships within items and their components would probably reveal a range of possible interpretations for each item. This multiple interpretation effect is consistent with the findings using right answers. For foil clusters, the lower recapitulation left five clusters

containing 28 foils from the 60 that remained classifiable because of the advance classification of their members. Of these, ten foils did not have these classifications in advance, but eight of them could be reasonably reclassified into the category of the total cluster.

Only one cluster (W_7) which contained nine foils could not be classified by logico-semantic analysis although one or two more of these were shaky at best. One cluster (W_3) was classified Common Misconception on the basis of logico-semantic analysis because the characteristics of this cluster were reminiscent of the findings of Powell and Isbister (1969) thereby suggesting consistency between two independent groups on the same test. Three categories (NS, O_1, O_2) were added to the Guidelines because of the clustering as well as the logico-semantic analysis.

Clearly, the Guidelines were not exhaustive and were not mutually exclusive, as already anticipated by the establishment of the "Other" classification, and by the multiple interpretation hypothesis. Because of the support for this hypothesis it is likely that the interpretations formulated should be confined to the group upon which they were derived.

A Possible Hierarchy of Foil Categories

During the above analysis it became evident that foil categories were, to a degree, interchangeable. For instance, foil $2D_1$ could be classified as OG, OS, or Sub. In only a few cases did reclassification within a wrong answer cluster of specific foils seem unreasonable. No attempt was made to exhaustively reclassify foils. Instead, the attempt was to reclassify specific foils to fit the pattern which seemed to be evident within the entire cluster as derived from Group A. Table 13 on page 80 summarizes the reclassification which occurred for each foil class in each of the wrong answer clusters.

TABLE 13

RECLASSIFICATIONS WHICH WERE

MADE OF SPECIFIC FOILS

Cluster	Advance Classification	Reclassification
W_1	Sub	became OG
W_3	Sub, OS and OG	became CM
W_5	OS and OG	became Inv
W_6	Inv, Sub, OS, Irr	became NS
W_8	Inv and Sub	became IA
W_9	Irr, Inv, and Sub	became OS
W_{11}	Irr and OS, Sub, CM	became WW
W_{13}	OG	became O_2

Since the O class in Table 38 was treated essentially as though it were unclassified it was omitted from this table, as were foils which did not classify within their respective cluster. Table 13 lists the Advance Classification of the foils in each of the wrong answer clusters which were different from the final classification of that cluster and the final classification given. As such it summarizes Tables 53 to 66 inclusive.

Put the other way around, the substitution category of foils disappeared by reclassification as OG, CM, NS, and IA. Besides

retaining a cluster for itself, OG also reclassified as CM, Inv, and O_2. OS also retained its own category but also became WW, CM, Inv, and NS. Similarly, some of the Inv foils reclassified to NS, IA, and OS. Finally, some Irr foils became NS, OS, and WW. Figure 2 presents these changes diagrammatically.

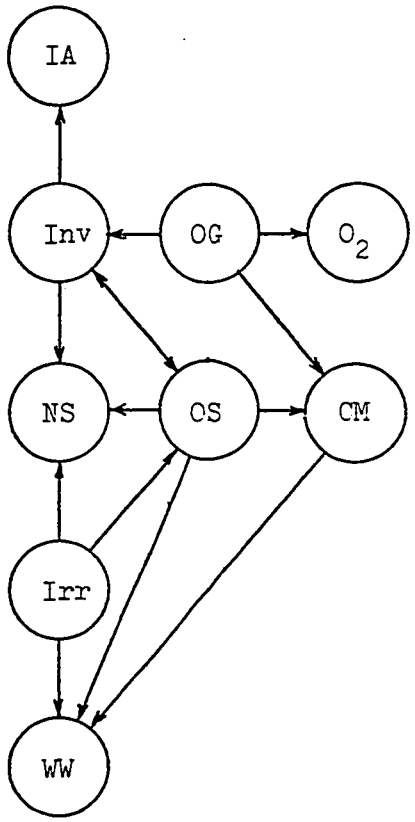

FIGURE 2

FOIL RECLASSIFICATION PATTERN

In Figure 2 there is a double headed arrow between OS (Over-simplification) and Inv (Inversion). This arrow means that at least one

OS foil was reclassified as an Inv. and at least one Inv foil was reclassified as an OS (Oversimplification). The other arrows in the figure can be interpreted in the same way.

Sub foils disappeared, and hence are dropped from this diagram. From the analysis of the right answer clusters it became evident that CM and WW type foils tended to lower the level in the taxonomy of an item while OS and OG foils tended to raise it.

From the reclassification pattern in Figure 2 OS foils seem to be pivotal in the sense that it was most frequently involved in changes (from or to). Furthermore, NS, Inv, and OG may be higher order foils, and CM, WW, and Irr of lower order foils on the basis of the changes these foils seem to cause in the reclassification of right answers.

Thus the general order of a possible foil hierarchy would seem to follow a vertical axis in Figure 2, with the lowest level at the bottom.

O_2 and IA foils are ambiguous in this pattern because they have one-way linkages only.

Other evidence to support the possibility of an hierarchy of foils should be found before this property of foils can be considered to be established from the data. Any set of random variates can be ordered on the basis of relative magnitude into an arbitrary hierarchy. Two random variables will be uncorrelated. If, however, two correlated hierarchies can be produced from two independent variables, this production would suggest that these two variables may be functionally related.

One possible source of such an arbitrary hierarchy is to consider the average total-correct score of the respondents selecting each interpreted foil as listed in the Appendix in Table 40. The average of these

averages could be found for each interpreted wrong-answer cluster. This average total-correct score for all the members of a particular cluster may reflect a systematic characteristic of the examinees which may relate the kind of "error" made to the total-correct score achieved.

If foil selection patterns reflect a functional relationship between "errors" and total-correct score, the rank of a foil (from one to three for high to low) within an item based on the average total-correct score should also reflect this functional relationship. If each item is independent of each other item, the average within item rank of the foils across a cluster should also be independent of the average total-correct scores of the foils in that same cluster. This is a reasonable assumption since only in item cluster C_3 does there not seem to be a close relationship between item clusters and foil clusters for the items in that cluster. An exception to this expectation would arise in the event that there is, in fact, a functional relationship between the kind of error and the total-correct score. In this latter case, both these procedures should produce roughly the same hierarchy.

In addition, this hierarchy would be expected to reflect the pattern which seemed to be evident on the basis of the pattern of reclassification of foils, the reclassification of items as influenced by foils, and the indications which also arise from other research into wrong-answer patterns.

Since this part of the study is exploring the possibility of a hierarchy among the foil categories, the rank order of the two variables just described was determined. That is, the rank order of the average for each interpreted foil cluster of the average total-correct score associated with each foil in that cluster was found. This procedure

established an arbitrary ordinal relationship among the interpreted foil clusters. A ranking of the average within item ranks was also established. If there is no functional relationship between foil type and total-correct score, the rank order correlation between these two ranking systems should be near zero. That is, the within item and between item characteristics would not be related to the average total-correct score on each foil in a functional manner. Table 14 shows the ranking of foil clusters by two independent methods (see: p. 85).

Each of the two ranking systems in Table 14 tend to support the more general ranking procedure suggested by the reclassification pattern. The comparison between the two ranking systems gave a rank order correlation of $\rho = .68$ which is significant ($p \simeq .01$ for two-tailed test for $N = 12$).

Other findings support this hierarchy. For instance, Powell and Isbister (1969) found that IA foils correlated negatively to Synthesis items. Foils in this category would be expected to have the fairly low rank this study suggests for IA foils. On the other hand, placing this foil type just above the content-linked misreading categories may be placing it too low in the hierarchy. Similarly, RT foils seemed to be content-linked, and would be expected to have a low rank for this reason. Support for this extrapolation was also evident. The tendency for Irr foils to distract middle and high level performers has already been noted in Powell (1968) and Powell and Isbister (1969). Thus the ranking by average total-correct score which seems to place Irr foils in sixth place would seem to be too low. Similarly, the ranking on a within item basis as sharing top place would seem to be too high.

By itself, the ranking on total-correct scores would seem to be

TABLE 14

RANKING OF FOIL CLUSTERS BY TWO INDEPENDENT METHODS

Interpreted foil cousters		Rank by total-correct averages within clusters		Rank by average within item rank by cluster[a]	
Cluster	Interpretation	Average	Rank	Average	Rank
W_1	OG	12.0	1.0	2.4	1.5
W_{13}	O_2	11.9	2.0	2.67	4.5
W_6	NS	11.8	3.0	2.6	3.0
W_5	Inv	11.7	4.0	2.7	6.0
W_9	OS	11.6	5.0	3.0	9.5
W_{14}	Irr	11.5	6.0	2.4	1.5
W_{15}	Tr	11.3	7.0	2.75	7.0
W_3	CM	11.2	8.5	2.9	8.0
W_8	IA	11.2	8.5	2.67	4.5
W_4 & W_{10}	RT	11.1	10.0	3.25	11.5
W_{11}	WW	11.0	11.0	3.0	9.5
W_{12}	O_1	10.8	12.0	3.25	11.5

a. The within item rank includes the right answer and does not drop any foils.

fairly closely related to an hypothetical foil hierarchy on a between item basis. Similarly, the within item ranking would seem to be more closely related to the influence of foil categories upon the items. However, these two variables are obviously related, as indicated by the significant rank order correlation between them. These results suggest overall systematic answering which influences the statistical outcomes

of both within and between item events.

Perhaps the most reasonable arrangement for the hierarchy would be to consider both events to be interdependent. The simplest approach in this case, is to consider the average rank of the two separate ranking systems used and to rearrange the foil categories accordingly. The resulting hierarchy would then reflect the influence of both between and and within item events. Table 15 gives the results of this procedure.

TABLE 15

RERANKING OF FOILS

BY AVERAGE RANK

Cluster	Foil Classification	New Rank
W_1	OG	1.0
W_6	NS	2.0
W_{13}	O_2	3.0
W_{14}	Irr	4.0
W_5	Inv	5.0
W_8	IA	6.0
W_9	OS	7.5
W_{15}	Tr	7.5
W_3	CM	9.0
W_{11}	WW	10.0
W_4 & W_{10}	RT	11.0
W_{12}	O_1	12.0

This rearrangement put RT nearer the bottom, OS nearer the middle (as its pivotal position suggested), and Irr nearer the top than the average total-correct rank, which seems to be reasonable relative to the available evidence. This new reordering has not been used in subsequent analysis because of the problem of the relevance of the within item ranking to this hierarchy. Further research is needed before the most probable sequence in the hierarchy has been established.

The lowering of the performance level of synthesis items in C_2 and C_3 by Tr, CM, and RT foils further supports this hierarchy. Foil $17D_3$ was reclassified as OS in the logico-semantic analysis of the wrong answer cluster. The pivotal position of the OS category would seem to support the "double-strategy" interpretation of C_2. Foil $4D_1$ was changed from OG to CM, which suggests that this foil, and $4D_3$ which was unclassifiable may have combined to lower this analysis item to a comprehension level. Right answer cluster C_2 may, therefore, be a comprehension level cluster and not a double-strategy cluster as suggested earlier. These suggestions are too tentative to alter its "undetermined" classification.

Further support for this foil hierarchy may be found by taking the new foil ranks and determining from these the average foil rank of each right answer cluster. Table 16 (see: p. 88) gives this information.

Table 16 shows three additional characteristics which may add support to the concept of a foil hierarchy. First, the order conforms to Bloom's Taxonomy with C_2, the apparent split strategy cluster falling between the analysis and undetermined categories. Second, evaluation fell out of order as occurred in the Kropp, Stoker, and Bashaw (1966)

TABLE 16

ORDERING OF RIGHT ANSWER CLUSTERS BY AVERAGE FOIL RANK

Right Answer Rank	Right Answer Cluster	Interpreted Classification	Foil Types Present in Clusters	Average Foil Rank[a]
1	C_5	Analysis	OG (2)[b], OS, O_2, NS, WW	3.8
2	C_1	Analysis	OG (3), OS, O_2, IA, WW	4.2
3	C_6	Analysis	Irr (2), NS	5.0
4	C_8	Synthesis	Inv (3), NS, CM	5.2
5	C_{10}	Analysis	NS (2), O_2, O_1, Irr, Tr	5.5
6	C_2	Undetermined (Analysis or Comprehension?)	OG, OS (2), Irr, Tr, RT	5.7
7	C_7	Evaluation	OS, IA, Inv, WW (2), CM, RT	8.3
8	C_4	Undetermined	OS, IA, Inv, CM (2), O_1	8.5
9	C_3	Undetermined (Comprehension?)	CM (3), WW	9.1
10	C_9	Undetermined	WW	11.0

a. These figures are the average of the foil ranks by total-correct averages within cluster as given in Table 14, page 85.

b. The number in parentheses indicates the frequency of recurrence of a particular foil classification within the right-answer cluster when this frequency is greater than one.

study. Third, the cluster C_9, if legitimately classified by one foil, fell at the bottom. The Procrustes rotation match-to-content suggested that this cluster might be a content-oriented cluster. In this case, its location was also reasonable.

Further support for the hierarchy can be found by determining the rank order of the right answer clusters in several ways. When the between and within item foil ranks, as given in Table 14, page 85, are used to calculate the right answer cluster rank, a highly significant correlation is found ($\rho = .90$, $p < .001$ for $N = 9$). This finding supports both the hierarchy and the apparent influence of foils upon item performance. All of the other possible rankings produced correlations which were not significant, including a comparison between ranking by average foil rank and average total-correct scores.

Results Related to the Subsumptive Property of the Taxonomy

Another interesting finding in the results just reported is the fact that the average difficulty of each cluster seems to be uncorrelated with the rank of each cluster. This correlation is $p = .05$ and does not change much if within item rank is used or if the composite rank is used. This finding would seem to contradict the subsumption characteristic assumed to be part of Bloom's Taxonomy.

From the data available in this study, another test of this subsumption hypothesis can be made. If the subsumptive property holds, successively higher members of the right answer hierarchy should be obliquely related. The Procrustes rotation gave a good fit to the target for the cluster analysis (see: p. 67). In general, the relationship among the primary axes would seem to be orthogonal.

However, the sampling distribution of these correlations is unknown so that their statistical significance of these correlations cannot be determined. In this case, the actual values of these correlations, arranged in order on the basis of the hierarchy of right answer clusters may reveal a systematic pattern. Table 17 gives this data.

TABLE 17

POSSIBLE SYSTEMATIC OBLIQUITY

BETWEEN ORDERED CLUSTERS[a]

	C_7	C_{10}	C_6	C_5	C_1
C_{10}	-.12				
C_6	-.03	.15			
C_5	.31	.20	.18	Relations among Analysis clusters	
C_1	-.17	.12	-.07	-.31	
C_8	-.08	.09	-.03	-.29	.06

a. This table is based on Table 10, page 69

If the point is stretched to the ultimate and correlations greater than or equal to an absolute value of .15 are considered

oblique ($r \geq |.15|$), there may be seven (7) of these fifteen (15) relationships that could possibly be considered as oblique. In this case three (3) of these relationships are oblique along the diagonal of Table 42, but all are within the analysis grouping. Four of these seven are found among the six of the analysis grouping. The other three are among the nine relationships outside the analysis grouping. The highest of these correlations still represented an angle larger than 70° ($r = \pm .31$) and the greatest proportion of these slightly nonorthorgonal angles are found among analysis clusters. Such slight pattern as might have existed did not seem to be too important outside of the analysis group of clusters. The relationships among the clusters did not seem to support this assumed subsumptive relationship between levels of the Taxonomy. Kropp et al (1966) found that the order from the simplex analysis did not consistently support this subsumptive property either (see: p. 21). In short, the lack of correlation between average item difficulty and cluster order would seem to be further evidence towards the probable refutation of the assumed subsumptive relationship between the levels of Bloom's Taxonomy.

The fact that the correlation between the order of the right answer clusters and the average total-correct score was negative suggests a general tendency for examinees to do better on low level items, which suggests a ceiling effect present in this difficult test. The test was also shown to be difficult by the average total-correct score, which was 12.19 out of a possible 30 items for Group A. This possibility found further support in the cross-validation part of the

Cross-validation of the Analysis

The multiple interpretation hypothesis supported thus far, and the apparent foil hierarchy, suggested two possibilities with respect to cross-validation. First, the attempts made to cross-validate these findings might not be successful. Other evidence for systematic responses may, therefore, have to be sought, such as the hierarchy, which might be supported more strongly in the cross-validation than the individual clusters.

The total-correct means and variances for each of these two groups on the experimental test were used in a t-test for independent samples in order to confirm the equivalence of test scores of these two groups. Table 18 contains this data.

TABLE 18

COMPARISON BETWEEN GROUP "A" AND GROUP "B"

ON TOTAL-CORRECT SCORES

Group A	Group B	df	t	p[a]
$X_1 = 12.19$	$X_2 = 11.91$	275	.83	.40
$S_1^2 = 8.62$	$S_2^2 = 7.60$			
$N_1 = 139$	$N_2 = 138$			

a. Probability of t equal to or larger than .83 is given for a two-tailed test.

On the basis of the results given in Table 18, page 92, the two groups would seem to belong to the same population so far as the experimental test means and variances are concerned.

Cross-validation of right answers. Three different comparisons were used in the cross-validation of item clusters. 1) Those right answers from the advance classification which were used to help identify the item clusters were checked against the clusters of Group B. 2) The clusters which occurred from the answers of Group A were compared with the clusters from Group B. 3) The right answer clusters were divided into three groups of about equal size based on their average foil rank, and this division was cross-validated. Table 19 gives the first two of these comparisons. The code C' is used to refer to Group B's clusters. (see: p. 94).

The rule used in Table 19 (see: p. 94), once again, was the most frequent repetition of items within clusters for Group A and Group B. Only four of the 12 items (33 per cent) which were grouped by the advance classification and retained this grouping in the clusters were found to cross-validate in Group B. Thus, the advance classification holds up about as well (or as badly) in cross-validation as it did in the clustering.

Comparing cluster by cluster, there were ten items (33 per cent) in the clusters from Group A which occurred in equivalent clusters for Group B. Cluster C'_5 contained two members from C_5 but also had all of C_9 in it (items 15, and 18) which leaves the definition of C'_5 ambiguous. In any case, it was evident that the clusters themselves did not cross-validate any better than the advance classification.

TABLE 19

CROSS-VALIDATION OF THE RIGHT ANSWERS OF GROUP A BY GROUP B FROM THE ADVANCE CLASSIFICATION AND THE ITEM CLUSTER

Group A					Group B					
Cluster	Item				Cluster	Item				
C_1	1*[a]	2*	8	28	C'_1	1*	2*	12		
C_2	3*[b]	17*	30		C'_2	3*	14	17*	29	
C_3	4*	13	6*		C'_3	4*	26	6*		
C_4	5	14	19		C'_4	5	10	13		
C_5	7*	22*	23		C'_5	7*	15	18	22*	
C_6	9*	27*			C'_6	8	27*	9*	30	25
C_7	10	11	16	29	C'_7	11	19			
C_8	12	20	26		C'_8	16	28			
C_9	15	18			C'_9	20	21			
C_{10}	21	25	24		C'_{10}	23	24			

a. The numbers in Italics cross-validate the advance classification.

b. The starred (*) numbers cross-validate the item clusters.

Table 20 (see: p. 95) has the right answer clusters arranged by their rank based on the average total-correct score as given in Table 16 (see: p. 88), and arranged into groups of three clusters with C_9 (items 15 and 18) dropped. The cross-validation was then repeated.

Instead of 33 per cent there is now 57 per cent cross-validation although with only three groups to match instead of nine. An increase

TABLE 20

CROSS-VALIDATION OF ITEMS, GROUPED BY AVERAGE FOIL RANK

	Group A Cluster					Proportion Cross-validating		Group B Cluster					
High Group	C_5	7[a]	*22*	23				C'_5	7	18	15	*22*	
	C_1	*1*	*2*	*8*	28			C'_1	*1*	*2*	12		
	C_6	*9*	*27*			High:	.78	C'_6	*8*	*27*	*9*	30	25
Middle Group	C_8	12	*20*	26									
	C_{10}	*21*	25	24				C'_9	*20*	*21*			
	C_2	3	17	30		Middle:	.22						
Low Group	C_7	*10*	*11*	16	29			C'_7	*11*	*19*			
	C_4	*5*	14	*19*				C'_4	*5*	*10*	*13*		
	C_3	*4*	*13*	*6*		Low:	.70	C'_3	*4*	26	*6*		
						TOTAL:	.57						

These clusters did not cross-validate:

C'_2	3	14	17	29
C'_8	16	28		
C'_{10}	23	24		

a. Items in *Italics* cross-validated by group.

of this sort would be expected. More interesting is the distribution of the shifts between Group A and Group B clusters. If the pattern for Group A is taken as a reference and the order is considered to be fixed, a shift is considered to be + 1 if the mobile item in the Group A cluster is found in a Group B cluster associated with items one cluster higher for Group A. Thus, a + 1 shift occurred for item 13 in C_3 clustered with item 5 in C_4' where the latter is in C_4. Similarly, item 10 in C_4 would represent a - 1 shift. Table 22 summarizes these shifts.

TABLE 22

MOBILITY OF ITEMS BETWEEN GROUP A AND GROUP B IN TERMS OF SHIFTS

0 Shift Items	± 1 Shift Items	± 2 Shifts Items	± 3 Shifts Items	Larger Shifts Items	Larger Shifts Shift
1 7	3 19	12	16	23	-4
2 9	8 20	14	28	24	+4
4 11	10 21	25	30	26	-5
5 22	13 29				
6 27	17				
10	9	3	3	3	

Table 22 (see: p. 96) shows that the mean absolute shifting of items between the two groups was slightly more than one place in the hierarchy ($\bar{s}$ = 1.11). This shifting was about one third of the shift expected if the items had randomly rearranged from Group A to Group B. If all possible shifts were equally probable, the mean shift would be 3.24. In fact 22 of the 28 items shifted two steps or less, which accounts for 79 per cent of the items.

A possible explanation for these shifts can be found in the hypothesized multiple interpretations. This hypothesis would suggest that in spite of the homogeneity of these two groups based upon total-correct scores, these two groups were obviously not homogeneous when it came to clustering of the items into item-homogeneous subtests. The clustering was based upon correlations which were sensitive to marginal totals. The stability of these marginal totals could be expected to be affected by the range of interpretations of the items within the groups concerned. This shifting could, therefore, be a product of the heterogeneity of the examinees and the effectiveness of the item in communicating a limited range of possible interpretations.

Cross-validation of wrong answers. An identical procedure to the one used for right answers was used with the wrong answers. Table 22 (see: p. 98) gives the cross-validation between the advance foil classification and the individual foil clusters between Group A and Group B.

In Table 22 we find that of the advance classification only three foils out of 16 (or 19 per cent) cross-validate as compared with the 16 cut of 60 foils (or 26 per cent) which help to identify the clusters. Once again, the advance classification and the clusters

TABLE 22

CROSS-VALIDATION OF ADVANCE FOIL CLASSIFICATION AND FOIL CLUSTERS

Group A		Group B	
Cluster	Foil	Cluster	Foil
W_1	$1D_1$ $2D_1$* $22D_3$ $8D_1$ $23D_1$* $\underline{17}D_1$*	W'_2	$2D_1$* $\underline{17}D_1$* $9D_1$ $25D_1$ $23D_1$* $3D_2$ $7D_2$
W_{13}	$\underline{21}D_1$[a]* $28D_2$ $\underline{23}D_2$[b]*	W'_6	$\underline{23}D_2$* $\underline{21}D_1$* $6D_2$
W_6	$9D_1$ $25D_1$ $21D_3$* $7D_3$* $26D_3$	W'_3	$28D_2$ $7D_3$* $14D_3$ $2D_2$ $21D_3$* $6D_3$
W_5	$12D_3$ $11D_2$ $12D_2$ $19D_2$* $20D_3$*	W'_5	$19D_2$* $20D_3$* $29D_2$ $26D_3$
W_9[c]	$3D_3$ $28D_1$* $10D_2$* $17D_3$* $22D_2$* $5D_3$	W'_9	$28D_1$* $5D_2$ $22D_2$*
		W'_{14}	$10D_2$* $17D_3$* $11D_2$ $8D_3$ $28D_3$ $30D_3$
W_{14}	$3D_2$ $25D_2$ $9D_2$ $27D_2$		
W_{15}	$30D_2$* $24D_3$*	W'_4	$4D_1$ $30D_2$* $19D_1$ $24D_3$*
W_3	$4D_1$ $26D_1$ $5D_2$ $6D_2$ $14D_3$ $13D_2$ $29D_2$		
W_8	$11D_1$* $28D_3$ $14D_2$	W'_8	$11D_1$*
W_4	$30D_3$	The following did not cross-validate:	
W_{10}	$29D_1$	W'_1	$1D_1$ $22D_3$ $10D_3$ $18D_1$ $26D_2$ $19D_3$
		W'_5	$16D_3$ $26D_1$
W_{11}	$18D_1$ $7D_2$ $6D_3$ $16D_3$ $2D_2$ $10D_3$	W'_7	$8D_1$ $27D_2$ $5D_3$
W_{12}	$19D_1$ $24D_2$	W'_{10}	Dropped because of low selection ratios.
W_7	$10D_1$ $19D_3$ $4D_3$ $13D_3$ $12D_1$ $15D_3$ $18D_2$ $8D_3$	W'_{11}	$9D_2$ $24D_2$ $12D_3$
W_2	Dropped because of low selection ratios. $26D_2$	W'_{12}	$29D_1$ $25D_2$ $3D_2$ $13D_2$ $14D_2$
		W'_{13}	Dropped because of low selection ratios.

a. The numbers in Italics cross-validate the advance foil classification.
b. The starred (*) numbers cross-validate the clusters.
c. Cluster W_9 splits into two clusters (W'_9 and W'_{14}) in the cross-validation.

cross-validate in about the same proportion. Wrong answers seem to cross-validate about as well as right answers.

The third comparison, once again, was between foil clusters grouped into three groups based upon the average foil rank. Table 23 (see: p. 100) gives this comparison.

In Table 22 the high group again cross-validated best using the hierarchy. With W_7 being dropped as uninterpretable, 27 of the remaining 51 foils (or 53 per cent) cross-validated compared with 16 out of 28 (or 57 per cent) for right answers, and for a combined total of 43 out of 79 alternatives (or 54 per cent).

Although the wrong answers showed a wider range of shifts, 30 foils had a shift range of $\pm$ 3 or less (or 59 per cent) compared with a probable $\bar{s}$ = 4.83 for random shifts. Foils, though less stable, showed the same trend toward stability as right answers. Once again the multiple interpretation hypothesis could account for the lack of stability.

The joint cross-validation combining right and wrong answers by group were: high group 21 out of 28 (or 75 per cent); middle group 14 out of 28 (or 50 per cent); and low group 8 out of 23 (or 35 per cent). That is, the stability increases by about the same proportion (50 per cent) from low to high. This increase in stability was also found among the wrong answers for the Proverbs Test (Cf Powell, 1968).

The Prediction Value of the Experimental Test

In addition to the scores and individual responses on the experimental test, two achievement scores were obtained for most examinees. There was some loss of data, making Group A have 125 members, and Group B have 120 members.

The first achievement test (Achievement Test I) was given with

TABLE 23

CROSS-VALIDATION BY GROUPING OF WRONG ANSWER CLUSTERS

	Group A Cluster		Proportion Cross-validating	Group B Cluster	
High Foils	W_1	$1D_1$ $\underline{2D_1}$ $\underline{23D_1}$ $22D_3$ $8D_1$ $\underline{17D_1}$		W'_2	$\underline{2D_1}$ $\underline{17D_1}$ $\underline{9D_1}$ $\underline{25D_1}$ $\underline{23D_1}$ $3D_2$ $\underline{12D_2}$ $7D_2$
	W_{13}	$\underline{21D_1}$ $\underline{28D_2}$ $\underline{23D_2}$		W'_6	$\underline{23D_2}$ $\underline{21D_1}$ $6D_2$
	W_6	$\underline{9D_1}$ $\underline{25D_1}$ $\underline{21D_3}$ $7D_3$ $\underline{26D_3}$		W'_3	$\underline{28D_2}$ $7D_3$ $14D_3$ $2D_2$ $\underline{21D_3}$ $6D_3$
	W_5	$12D_3$ $11D_2$ $\underline{12D_2}$ $\underline{19D_2}$ $\underline{20D_3}$	High : .74	W'_{15}	$\underline{19D_2}$ $\underline{20D_3}$ $29D_2$ $\underline{26D_3}$
Middle Foils	W_9	$\underline{3D_3}$ $\underline{28D_1}$ $\underline{10D_2}$ $\underline{17D_3}$ $\underline{22D_2}$ $\underline{5D_3}$		W'_9	$\underline{28D_1}$ $\underline{5D_2}$ $\underline{22D_2}$
	W_{14}	$3D_2$ $\underline{25D_2}$ $9D_2$ $\underline{27D_2}$		W'_{14}	$\underline{10D_2}$ $\underline{17D_3}$ $\underline{11D_2}$ $28D_3$ $30D_3$
	W_{15}	$\underline{30D_2}$ $\underline{24D_3}$		W'_4	$\underline{4D_1}$ $\underline{30D_2}$ $19D_1$ $\underline{24D_3}$
	W_3	$\underline{4D_1}$ $26D_1$ $\underline{5D_2}$ $6D_2$ $14D_3$ $\underline{13D_2}$ $29D_2$	Middle: .68	W'_7	$8D_1$ $\underline{27D_2}$ $\underline{5D_3}$
				W'_{12}	$29D_1$ $\underline{25D_2}$ $\underline{3D_3}$ $\underline{13D_2}$ $14D_2$
Low Foils	W_8	$\underline{11D_1}$ $28D_3$ $14D_2$		W'_8	$\underline{11D_1}$
	W_4	$30D_3$			
	W_{10}	$29D_1$			
	W_{11}	$18D_1$ $7D_2$ $6D_3$ $16D_3$ $2D_2$ $10D_3$			
	W_{12}	$19D_1$ $24D_2$	Low : .08		
			Cross-validation for all foils: .53		

	The following clusters did not cross-validate
W'_1	$1D_1$ $22D_3$ $10D_3$ $18D_1$
W'_5	$16D_3$ $26D_1$
W'_{10}	Dropped
W'_{11}	$9D_2$ $24D_2$ $12D_3$
W'_{13}	Dropped

Unclassified Foils:

W_7: $10D_1$, $19D_3$, $4D_3$, $13D_3$, $12D_1$, $15D_3$, $18D_2$, $26D_2$, $8D_3$

the experimental test as a subtest for a midterm examination. The scores used for predictive purposes do not contain the experimental test scores. The second achievement test (Achievement Test II) was the final examination in the same course. The relationship among these tests is presented in Table 24.

TABLE 24

CORRELATIONS BETWEEN THE TESTS IN THIS STUDY

	Experimental Test	Achievement Test I	Achievement Test II
Experimental Test	1.000		
Achievement Test I	.224	1.000	
Achievement Test II	.125	.414	1.000

As shown in Table 24 the two achievement tests were moderately correlated ($r = .414$). The relationship between the experimental and achievement tests was considerably less, particularly with respect to Achievement Test II.

In order to establish the predictive validity of the subdivisions of the experimental test, several comparisons were run using a step-wise multiple regression technique in all cases. Each achievement test was predicted separately for each of Group A and

Group B. The following predictions were made.

1. Total-correct score on experimental test predicting the total-correct scores of the achievement tests.
2. Right-answer subtest scores on the experimental test with the subtests defined by the advance classification of items predicting the total-correct scores on the achievement tests.
3. Combined right answer subtest scores and wrong answer subtest scores with the subtests defined by the advance classification, predicting the total-correct scores in the achievement tests.
4. Scores on right-answer subtests defined by interpreted clusters used to predict the total-correct scores on the achievement tests.
5. Scores on combined right-answer and wrong-answer subtests defined by all interpreted clusters used to predict the total-correct scores on the achievement tests.
6. Scores on right-answer subtests defined by grouping right-answer clusters by the foil hierarchy used to predict the total-correct scores on the achievement tests.
7. Combined right- and wrong-answer subtest scores defined by grouping of clusters on the basis of the foil hierarchy used to predict the total-correct scores on the achievement tests.

In each case the total-correct score of each achievement test was being predicted.

Table 25 (see: p. 103) gives the results of these predictions for Group A.

TABLE 25

STEPWISE REGRESSION OF GROUP A DATA USING SEVERAL COMBINATIONS OF VARIABLES

		Combinations of Variables[a]						
		1	2	3	4	5	6	7
Achievement Test I	R^2	.059	.142	.228	.068	.278	__	.055
	R	.243	.377	.478	.261	.525	__	.234
No. of Predictors available		1	5	14	6	18	3	6
No. of Predictors used[b]		1	5	14	3	18	None	2
Achievement Test II	R^2	.006	.029	.130	__	.156	.049	.085
	R	.073	.169	.360	__	.395	.222	.291
No. of Predictors available		1	5	14	6	18	3	6
No. of Predictors used		1	1	9	None[b]	11	2	5
Total number of right answers used		30	30	30	19	19	28	28
Total number of foils used		0	0	70	0	46	0	46

a. The combinations of variables are defined by number as given on page 102.
b. Only those predictors which made a significant contribution ($p < .06$) to the prediction were included on this table.

Table 25 gives the correlation between the total-correct score on the experimental test, the multiple correlation achieved for significant variables ($p < .06$). The squared multiple correlation (R^2) are also given to indicate the amount of variance accounted for in the predictions.

A similar set of data for Group B follows in Table 26 (see: p. 104).

TABLE 26

STEPWISE REGRESSION OF GROUP B DATA USING SEVERAL COMBINATIONS OF VARIABLES

		Combinations of Variables[a]						
		1	2	3	4	5	6	7
Achievement Test I	R^2	.046	.063	.190	.035	.189	__	.091
	R	.214	.251	.436	.188	.435	__	.301
No. of Predictors available		1	5	14	6	18	3	6
No. of Predictors used[b]		1	3	14	1	14	None	5
Achievement Test II	R^2	.046	.082	.135	.054	.237	.057	.193
	R	.214	.287	.368	.232	.487	.239	.439
No. of Predictors available		1	5	14	6	18	3	6
No. of Predictors used		1	4	9	2	18	2	6
Total number of right answers used		30	30	30	19	19	28	28
Total number of foils used		0	0	70	0	46	0	46

a. The variables used with Group B were identical in definition to those used in Group A.
b. Only those predictors which made a significant contribution ($p < .06$) to the prediction were included in this table.

There are two considerations relevant to Tables 25 and 26. First, the value of the procedure is partly determined by the amount of variance accounted for (as given by R^2). Using this criterion, there is a consistent improvement in prediction when the scores on wrong-answer subtests are included in the analysis. When this was done, the wrong-answer variables, in general, accounted for more of the variance than the right-answer variables within the same solution. The interpreted

clusters give a better solution than the advance classification. Both of these were better than the grouping by the foil hierarchy. The poorest predictor was the total-correct score on the experimental test.

The second consideration is the statistical significance of the improvement of these values of R^2 when compared with each other. The formula for this comparison is:

$$F = \frac{R_1^2 - R_2^2}{1.00 - R_1^2} \times \frac{N - K - 1}{K - L}$$

where N is the number of persons,

K is the number of independent predictions in R_1,

and L is the number of independent predictions in R_2.

This procedure gives the usual "F" test with degrees of freedom, N - K - 1 and K - L respectively. The results of these calculations derived from the data in Tables 25 and 26 (see: pp. 103 and 104) are given in Tables 27 to 30 which follow (see: pp. 106 to 109).

Table 51 gives the significance of the difference between the predictions of Test I for Group A. It shows that the sequence 1 / 2 = 3 / 4 / 5 stands clearly in the diagonal. One factor which may be involved in the equality 2 = 3 may be the fact that the number of predictors increases from 5 to 14. If all other values remain the same, an increase in the size of the sample of less than 50 per cent would make the difference between combination 2 and 3 significant. Also, no variables made a significant contribution in combination 6, and 2 variables were significant in combination 7, which makes 7 a significantly better predictor than 6. Thus, for Group A when predicting Test I there is a consistent tendency for the combined right and wrong answers to be better predictors than the right answers alone.

TABLE 27

SIGNIFICANCE OF DIFFERENCES BETWEEN R^2'S

FOR GROUP A WHEN PREDICTING TEST I

R_1^2 for variable combinations	R_2^2 for Variable Combinations[a]											
	1		2		3		4		6		7	
	F	p[b]	F	p	F	p	F	p	F	p	F	p
1									empty cell	--	-ve	--
2	2.90	.05					5.41	.01	empty cell	--	4.02	.01
3	1.86	.05	1.36	--					empty cell	--	2.05	.05
4	.58	--			2.07	.05			empty cell	--	1.69	--
5									empty cell	--	2.05	.05
7									empty cell	--		

a. Definitions for these variable combinations are given on p. 102.

b. Only the probability (p level) for significant differences are shown.

Table 28 gives the significance of the differences between the predictions of Test II for Group A (see: p. 107).

There is no similar pattern when predictions are made to the future Test II as compared with the concurrent Test I. Some of the F values in the equivalent diagonal are large enough that a larger sample might make them significant. At least all predictor combinations are better than the total-correct score. Although the grouping of alternatives (combinations 6 and 7) on the basis of the hierarchy does

TABLE 28

SIGNIFICANCE OF DIFFERENCES BETWEEN R^2'S

FOR GROUP A WHEN PREDICTING TEST II

R_1^2 for variable combinations	R_2^2 for Variable Combinations[a]											
	1		2		3		4		6		7	
	F	p[b]	F	p	F	p	F	p	F	p	F	p
1												
2	*[c]						empty cell	--				
3	2.05	.05	1.72	--			empty cell	--	1.53	--	1.49	--
5	2.01	.05	1.74	--	1.67	--	empty cell	--	1.59	--	1.57	--
6	5.52	.05	2.57	--			empty cell	--				
7	2.57	.05	1.82	--			empty cell	--	1.56	--		

a. Definitions for these variable combinations are given on page 102.

b. Only the probability (p level) for significant differences are shown.

c. * means: Denominator 0, F value indeterminate.

not yield significant differences, these two variables tend to have the largest F values. Once again, a larger sample size might have made the difference. The cross-validation of the multiple regression coefficients which follows this section casts further light on this aspect of the problem.

Table 29 (see: p. 108) gives the significance of the differences between the predictions of Test I for Group B.

TABLE 29

SIGNIFICANCE OF DIFFERENCES BETWEEN R^2'S

FOR GROUP B WHEN PREDICTING TEST I

R_1^2 for variable combinations	R_2^2 for Variable Combinations[a]											
	1		2		3		4		6		7	
	F	p[b]	F	p	F	p	F	p	F	p	F	p
1							*[c]	--	empty cell	--		
2	1.05	--					1.70	--	empty cell	--		
3	1.42	--	1.43	--					empty cell	--	1.43	--
4					1.53	--			empty cell	--		
5	1.42	--	1.43	--	*	--	1.53	--	empty cell	--	1.41	--
6												
7	1.13	--	1.76	--			1.77	--	empty cell	--		

a. Definitions for these variable combinations are given on page 102.

b. Only the probability (p level) for significant differences are shown.

c. * means: Denominator 0, F value indeterminate.

Table 29 is very similar to Table 28 except that none of the predictor combinations are significantly better than any other including the total-correct scores. These findings are not surprising considering the low level of cross-validation already found for all combinations except the grouping based upon the hierarchy. Although none of the

values for these groupings (Combinations 6 and 7) are significant, these two combinations have the highest F values. Once again a larger sample size might have made the difference.

Table 30 gives the significance of differences between predictions of Test II for Group B.

TABLE 30

SIGNIFICANCE OF DIFFERENCES BETWEEN R^2'S

FOR GROUP B WHEN PREDICTING TEST II

R_1^2 for variable combinations	R_2^2 for Variable Combinations[a]											
	1		2		3		4		6		7	
	F	p	F	p	F	p	F	p	F	p	F	p
1												
2	1.49	--					1.77	--	1.57	--		
3	1.44	--	1.35	--			1.50	--	1.42	--		
4	1.00	--										
5	1.50	--	1.52	--	1.41	--	1.56	--	1.29	--	.50	--
6	1.12	--					*[c]	--				
7	4.12	.01	7.77	.01	-ve	--	4.87	.01	4.76	.01		

a. Definitions for these variable combinations are given on page 102.

b. Only the probability (p level) for significant differences are shown.

c. * means: Denominator 0, F value indeterminate.

Table 30 shows Combination 7 to be a highly significantly better predictor in four of the six cases. When comparing Combination 3 with 7, the R^2 is larger for Combination 7; but the number of significant predictors is larger for Combination 3 and, for this reason, gives a negative F value in the formula. Logically, Combination 7 would, therefore, be the better predictor in this case as well. Combination 5 has a somewhat larger R^2 but requires many more predictor variables to achieve this, making the difference clearly insignificant.

Three trends seem to emerge from this data. First, combined right and wrong answers generally seem to yield the best predictions.

Second, the best prediction of a concurrent test seems to be found by using the interpreted clusters for the same group.

Third, when predicting remote events or the results of another group combining the predictor variables on the basis of the hierarchy would seem to give the best predictions.

Cross-validation of the multiple regression coefficients. It is possible to cross-validate multiple correlations by finding the vector product of the validity coefficients for one group and the standardized regression weights for the same variables from the other group as follows: $R^2 = V'W$

where V' is a row vector of the validity coefficients for one group,

W is a column vector of the standardized regression for the corresponding variables from the other group,

and R^2 is the resulting vector product.

Since the combining of right and wrong answers seemed to give the best results, only these combinations of variables were cross-validated in this manner. Table 31 (see: p. 112) gives the results of these calculations.

To begin with, Table 31 shows that Combination 3, the advance classification, does not survive cross-validation. Clearly, the best predictor of the concurrent test for Group A was the interpreted clusters (Combination 5), a finding consistent with the findings for the significance of differences. This combination (5) of variables did not cross-validate in any other situations.

For the remote test, Combination 7 proved to cross-validate very well, much better than the significance of the differences would suggest. For Group B, Combination 7 cross-validated about as well as the proportion of item-for-item cross-validation would suggest that it should. These findings also tend to support the suggestion that grouping on the basis of the hierarchy may be the best method for predicting future performance or performance in another group.

The reader is cautioned that the correlations between the total-correct scores of the experimental test and the achievement tests suggest that these tests may be dissimilar in the characteristics they are measuring. This situation would be expected to produce lower multiple correlations than tests of greater similarity might achieve. Second, near zero multiple correlations are easier to cross-validate than higher ones, so that the relative stability of these correlations can only supply suggestive results by themselves.

TABLE 31

CROSS-VALIDATION OF MULTIPLE REGRESSION COEFFICIENTS

FOR COMBINED ANSWERS

		Combination Number					
		3		5		7	
		Original R^2	Cross-validation R^2	Original R^2	Cross-validation R^2	Original R^2	Cross-validation R^2
Group A	Test I	.228	.001	.278	.208	.055	.045
	Test II	.130	.002	.156	.032	.085	.098
Group B	Test I	.190	.064	.189	.038	.057	.045
	Test II	.135	.026	.237	.013	.193	.100

Summary of Chapter V

Briefly, the findings as reported in this chapter were as follows:

1. Interpoint distance gave the best statistical solution to the data being considered in this study.
2. Logico-semantic analysis provided reasonable support to the construct validity of the procedure used, provided that alternative classifications are permissible, and interpretations were confined to the group under study.
3. There may be a hierarchy of foils which parallels Bloom's Taxonomy which may influence the way in which items perform.
4. None of the cross-validations were very strong, with the hierarchy tending to be somewhat better supported than other aspects of the analysis.
5. Wrong answers, in general, tended to add significantly to the predictions of both concurrent and future achievement whenever significance was found.
6. Within one group the interpreted clusters gave the best concurrent prediction, otherwise the grouping on the basis of hierarchy seemed to produce the best prediction.

CHAPTER VI

CONCLUSIONS AND IMPLICATIONS

The conclusions which can be drawn from this study are discussed in three sections; first, the conclusions which are relevant to the experimental test used in this study, second, the conclusions which are relevant to the analytical procedure, and finally, those which are relevant to the systematic response postulate.

The implications which can be drawn from this study are discussed in four sections. First, there are the implications of the results of the use of this analytic procedure to the theory of test analysis procedures. Second, there are the implications to the design, construction, and interpretation of taxonomic tests. Third, there are a number of implications of this study to educational practice. Finally, this study has opened enough possibilities for future research that these are discussed in a separate section.

Conclusions Related to the Experimental Test

Superficially, the experimental test used in this study would seem to have been a weak instrument but, as will be seen, the criteria usually used for evaluation may not have been applicable to this test. Using the usual criteria, for instance, the selection ratio for the right answers on most of the items was low; the biserial correlations of the items to the total correct scores were also low relative to the size of the corresponding difficulty ratios; the internal consistency based upon the Kuder-Richardson procedure was low, and the interrater reliabilities left a great deal to be desired.

On the other hand, the usual criteria employed for the

evaluating of tests may not be appropriate for this one. Briefly, the desirability of middle difficulty items is a criterion based upon the assumption that this level of difficulty maximizes the discrimination of the test when all the items are dichotomous variables. When all alternatives are being considered rather than when the item is being considered either right or wrong, this criterion seems no longer to apply.

The biserial correlation of the test items taken against the total correct scores is a criterion of the discriminating power of these items assuming that the test as a whole is highly homogeneous. The low internal consistency of this test suggests that the test was not homogeneous. Evidence for the lack of homogeneity in this test can be found in the logic of the construct model (Bloom's Taxonomy), and in the fact that the test subdivides in the cluster analysis into ten clusters of right answers, at least six of which were nearly orthogonal. Also, the loadings of the items on these nearly orthogonal factors were very nearly the original lengths of the vectors in the principal axis matrix from which they were derived. This latter result suggests that the internal consistency within clusters was substantially higher than within the test as a whole. For these reasons, the usual criteria may not apply to this test.

On the positive side, the average total correct score for persons selecting each alternative was higher for the right alternative than for all others used by at least 1.5 points in 37 out of 58 cases (or 64 per cent) which gives some support to the strength of the test (see : Table 40, p. 150). Although the clusters did not cross-validate very well, when the shifting of items into other clusters was considered,

it became evident that the items and foils were much more stable than would be suggested by chance alone. Also, the clear evidence for an interacting hierarchy of items and foils would seem to provide strong support for the possibility that the instrument was measuring some systematic characteristics of the examinees.

Precisely what these characteristics were would seem to be more ambiguous than the probability that they were being measured. They were, however, clearly process-oriented characteristics. The doubt about precise definitions arose from two sources, 1) the lower than desirable interrater reliabilities, and 2) the lower than desirable cross-validation of the clusters. Both these weaknesses in the interpretability of the results, however, may be a property of this type of test, and not a criticism of it.

If the proportion of the total variance used in the factor solution is considered, there was 37.7 per cent accounted for by the six factors for the items and another 48.6 per cent for the 15 factors used for the wrong answers, suggesting a higher internal consistency than the Kuder-Richardson results suggested.

Thus, the instrument displays the following properties:

1. Significantly improved prediction of independent concurrent achievement scores for the same group of examinees using interpreted clusters from both right and wrong answer clusters combined over the other combinations of scores tested.
2. A clear hierarchical pattern of both items and foils. The right answers and the foil selections seem to interact and to be relatively stable within the hierarchy under cross-validation when the range of shifts are considered.

3. An overall discrimination apparently based upon process-oriented events rather than content-oriented events.

The construct objective was to produce a process-oriented taxonomic test which had predictive value relevant to achievement variables. Whatever criticisms might be made of the test, the results made it clear that it met this construct objective to a reasonable degree, and for this reason be taken to be a valid test. Also, the indirect evidence suggested that the test was probably more reliable than was suggested by the direct evidence. Precisely which procedures should be used to establish validity and reliability estimates for tests of this type are not yet clear.

Conclusions Related to the Analytic Procedures Used

The analysis began with *phi* coefficients. The use of these coefficients can be defended on the grounds that none of the assumptions which are made for these coefficients was violated by their use. The two variables being related for each coefficient are discrete since each represents the selections made by the members of the same group for different alternatives. They were dichotomous because an accept-reject decision applies to all alternatives as a requirement of the response procedure. Linear dependencies were removed from the data by partitioning the matrix. Finally, since all values were expressed as frequencies of occurrence, the categories were amenable to appropriate representation by two-point values.

The resulting large matrices of *phi* coefficients were simplified by principal axis factor analysis in order to remove as much measurement error as possible, and to maximize the variance accounted for by any particular number of dimensions in the space being used. Beyond this

point the procedures seemed to separate into two aspects, those which are commonly used for the study of tests and their results, and those which are not commonly used but which are specifically selected to meet problems which may arise in the interpretation of the results.

The result of this study derived from the commonly employed procedures was uniformly ambiguous, inconclusive, or negative, whereas the less common procedures uniformly produced significant results. As indicated, the test itself would seem to have been of questionable value if it were evaluated by the commonly used procedures, and yet it clearly met the construct properties it was designed to meet.

The Procrustes rotations of the factor matrix to fit either content or process in the advance classification produced negative results. The results of the usual analytic rotations on the factor matrices were not reported in Chapter V because they were as uninterpretable as the Procrustes rotations. However, when interpoint distance clusters were used to avoid the problem of rotation, a set of nearly orthogonal groupings of the variables was produced. Unquestionably, the cluster approach produced a statistically satisfactory representation of the data leaving the researcher with the problem of interpretation to be resolved by non-statistical methods.

Cross-validation of clusters was equally disappointing, and contradicted the results of the t-test for uncorrelated samples based upon total correct scores. Substantial improvements in the proportion of cross-validation were found when the apparent hierarchy of items and foils was taken into account. Additional support for the cross-validation was found in the pattern of shifts of alternatives which occurred among the clusters between the two groups. The data were much more systematic among the clusters between the two groups. The data

more systematic than the usual procedures seemed to suggest.

Using total-correct scores to establish a hierarchy of foils on both a within item and a between item basis produced results with a moderate relationship. When these two procedures were used to order the right answer clusters, however, a highly significant similarity between these latter two orderings was found, leaving little doubt that an interactive ordering between the right and wrong answers was a systematic characteristic of the data. This ordering was unrelated to the total correct averages for the right answer clusters and to item difficulty, making the use of total-correct scores for the establishment of the hierarchy questionable. However, the shift patterns of the cross-validation suggested a much more stable result than did the cross-validation alone, suggesting that these shift patterns might be used to determine the ordering of the clusters instead of using total-correct scores.

Although all predictions were low, the use of the interpreted clusters did tend to give significantly better concurrent prediction of the total-correct scores on the independent concurrent achievement measure for the group on whom the interpretation was attempted. The amount of variance accounted for increased roughly three times in this case. The question of the validity of the use of the total-correct scores as adequate representations of achievement on the achievement measures was not explored. Finally, the hierarchy also proved to be more broadly stable during cross-validation than other variables.

Evidently, the more conclusive results were found by the less common procedures. Since these procedures were designed to fit the specific problems raised in this study, and since the statistical

adequacy of these procedures proved to be beyond question, they would seem to be, collectively, a more adequate method of approaching the kind of data this study produced than the more common procedures would seem to be.

Conclusions Related to the Systematic Response Hypothesis

The adequacy of the experimental test, and the analytic procedures used for the purpose of this study seem to be established to a reasonably acceptable level. Taken alone, the data were sufficiently systematic in both right and wrong answer matrices to establish that "most if not all of the answers given to multiple choice achievement tests are selected upon a systematic basis" may be a reasonable approach to human performance. The two findings most relevant were the presence of the interactive hierarchy and the increase in predictive validity evident when wrong answer clusters were included in the regression equations.

If the evidence supporting the multiple interpretation hypothesis is included, the support for this psychological postulate is greatly increased. To begin with the negative results from the Procrustes rotations for the advance classification by both content and process established the inadequacy of this approach. Logico-semantic analysis clearly established that process variables provided the best interpretation of the clusters. But the failure of the Procrustes rotation with the advance classification, the low cross-validation levels and the shift patterns clearly indicated that the classifications are not mutually exclusive.

Kropp et al (1966) found that the same subtests were differently

defined by the Kit, for different grade levels, and Powell (1968) reported only about 60 per cent cross-validation by reported reasons for the selection of particular wrong answers. Also, the higher level alternatives tend to be more stable than the lower ones when taken in combination, which is consistent with other findings (see: Powell, 1968). Thus, independent studies also report findings suggesting that the classifications of answers may not be mutually exclusive. A reasonable synthesis of these findings would be to suggest that each item may be interpreted in a variety of different ways. That is, the poor showing on cross-validation and the lower than desirable interrater reliabilities may be a product of multiple interpretations of the items. Strong support for this hypothesis was found in this study in the systematic character of many of the shifts which occurred among the alternatives. These shifts were sufficiently small in range for most items that the possibility of their use in the establishment of order among the variables could be proposed (see: p. 97). Further support was found in the fact that the prediction of the concurrent test on the interpreted group was the only case where the interpreted clusters had a distinct advantage. Predictions based upon the hierarchy, on the other hand, seemed to be less powerful, but more stable over a broader range of time and population. These findings support the multiple interpretation hypothesis as well because they suggest the short range applicability of specific interpretations.

Perhaps the range of this applicability of interpreted findings could be increased if the heterogeneity of the examinees upon whom the interpretations are made were reduced.

In summary, the conclusions of this study were:

1. Human performance, when abstracted from responses to multiple choice achievement tests involving higher mental processes, would seem to be systematic, and to display evidence of multiple interpretation of the communication.
2. There would seem to be a hierarchy of foils which parallels the hierarchy of right answers and which influences the way in which each total item performs. The levels of the foils themselves seem to depend upon the systematic ways in which this totality of each item is approached.
3. Wrong answers contain potentially useful information with respect to achievement when higher mental processes are involved.

Before the implications of this study are discussed, a statement should be made concerning the limitations to generalizability apparent in this study.

Limitations to Generalizability

There are several restrictions to the generalizability of the findings of this study which can be derived from the nature of the study and its conclusions. First, the findings of this study do not apply to multiple choice achievement tests where the simple recall of information is the only characteristic being measured. The experimental test was process rather than content oriented and Knowledge level items were considered inappropriate to its format, hence this limitation.

Second, the findings of this study do not apply where the cost of the additional effort required to obtain and interpret categorical information upon the examinees is greater than the cost of information loss, and possible misclassification attendant thereto, by using the

much simpler total-correct score method of evaluation.

Third, these findings may not apply when the most competent are being screened from already competent individuals for some specific purpose. A stronger statement in this respect cannot be made because later research may show that wrong answers may supply valid information for the purpose in question. For instance, Irr (Irrelevancy) foils may identify the most creative individuals among the high performers.

Finally, if a researcher has a valid reason to evaluate the effectiveness of a single treatment given to a heterogeneous group by using a single ordinal dimension for the particular group in question, the findings of this study clearly do not apply.

Implications of This Study to the Theory of Test Analysis

There are several situations where the findings of this study are very important to test theory. The fact that the more common procedures tended to give ambiguous, inconclusive or negative results raises a number of pertinent questions.

Classical test theory begins with the assumption that

$$X_i = T_i + E_i$$

where X_i is observed score of the i^{th} individual,

T_i is his true score, and

E_i is the measurement error

However, for multiple choice achievement tests, this observed score (X_i) itself is usually a composite entity obtained from the summation of single events as follows:

$$X_i = \sum_{j=1}^{n} x_{ij}$$

where x_{ij} is binary, being 1 if the i^{th} individual answered the item correctly, otherwise 0, and

n is the number of items in the test.

The issue then becomes -- is X_i sufficiently homogeneous for all individuals to justify the use of classical test theory? If it is not, as was evident in the present study, then an alternative approach to the data would seem to be needed, since the more common approaches proved unsatisfactory in this study.

Within the context of the present study, several considerations must be met by this alternative approach. The phi coefficients used are extremely sensitive to the magnitudes of marginal proportions. For this reason, if particular alternatives are selected for a different range of reasons among two samples of individuals, it would be expected that these alternatives would migrate to new clusters for reasons of systematic differences between groups rather than for reasons of measurement error. Also, if the range of reasons within a group of examinees were too broad, the interpretation of clusters would be expected to be difficult and possibly not applicable to specific individuals. That is, the first assumption made would be that different reasons for the selection of a particular alternative would be reflected by differences among overall patterns. These arguments suggest the need at the outset for a homogeneous group of examinees. The key to homogeneity in this study would seem to be the shifts in category which occurred upon cross-validation. Perhaps an homogeneous group should be defined in terms of minimizing the shifts which occur in the clustering of the alternatives for any, or all possible random assignments of the

group members to an arbitrary number of groups. In short, the procedure should probably begin by selecting groups of individuals with maximum cross-validity within these groups.

The clusters for such groups will be as stable as possible on the basis of the determindition of their composition. Hence, the possibility of interpreting the resulting clusters should be optimized, as should the applicability of these interpretations to the individuals within the groups.

With the clusters thus stabilized it should be possible to determine the clusters using all the data rather than a simplification of it, since the surface-to-surface interpoint distance ($\underline{d}$) between the ends of the vectors within the hypersphere can be determined quite simply by assuming that phi (ϕ) is the cosine of the angle between the arms of the isosceles triangle produced by the vector pairs. In this case the distance ($\underline{d}$) is

$$\underline{d} = \sqrt{2(1 - \phi)}$$

A tighter level of homogeneity is possible among individuals who have essentially the same response patterns but these individuals would tend to have only one meaningful cluster composed of all or most of their responses thus rendering interpretation impossible.

The dimensionality of the data from a homogeneous (by cross-validation) group of individuals would be determined by the minimum number of homogeneous clusters which can be extracted before orthogonality between the centriods of the clusters begins to disappear. Thus, the proposed procedure as just outlined as related to the problems raised by the data in this study provides a unique solution once the stop criteria are established. Each cluster would be categorical

(in the sense given on page 47), and optimally interpretable assuming that differences in interpretation of the communication by the examinee are characterized by differences in selection pattern.

There may, as the evidence from this study suggests, be an order among the categories which can be determined by the shifts which occur during cross-validation. Poor items would be unstable for the cross-validation.

So far as the categories themselves are concerned, these would be expected to be of two types 1) nominal, and 2) ordinal. Nominal categories would be expected to be bimodal with the modes tending to polarize at the extremities of the potential range within the category. Ordinal categories would be expected to display scalability characteristics within their potential range.

Relationships among categories other than the ordinal (hierarchical) one could probably be determinable by the relationships among the centroids of the categories. For instance, Powell and Isbister (1969) found a polarity between Invalid Assumptions for the wrong answers and Synthesis items among the right answers. In such a case it might be unnecessary to partition the matrix to remove linear dependencies. If this latter facilitation could be provided by this procedure, a homogeneous test in the classical sense would be one in which the right answers formed a single cluster of the ordinal type. Thus, the proposed procedure just given would seem to contain the characteristics of classical test theory as a special case.

It would be reasonable, then, to argue that the findings of this study suggest the need for alternative procedures to the ones in common use for the analysis of data from multiple choice tests, and the

findings suggest a particular procedure which contains the commonly used procedures as a special case.

Implications of This Study Concerning Taxonomic Tests

Bloom (1956) defines his Taxonomy as having three aspects:

1. It is a classification system.
2. It is hierarchically ordered on a "complexity" dimension.
3. Each higher level is formed by combinations of the lower levels.

The two properties of classification and ordering among classes combine to distinguish a taxonomy from other classification systems. Thus the evidence from this study supports the description of both Bloom's Taxonomy and the Guidelines as taxonomies. Noting the evident interrelatedness of these two taxonomies in this study suggests that they may both be part of a single taxonomy.

Concerning the "complexity" dimension, Bloom said, "Our attempt to arrange educational behavior from simple to complex was based upon the idea that a particular simple behavior may become integrated with other equally simple behaviors to form a more complex behavior [page 18]."

The findings of this study which produced no relationship between total-correct scores and the hierarchy where right answers were concerned, and no relationship between average item difficulty within clusters and the hierarchy did not help to identify the meaning of the term "complexity." Since the hierarchy could apparently have been produced through cross-validation procedures without recourse to total-correct scores, the meaning of "complexity" becomes even more vague. However, the finding that "higher" level members of the taxonomy tend to be more stable than lower levels suggests that these categories may

be the product of "more powerful strategies."

The third aspect of Bloom's Taxonomy as noted previously (see: p. 127) also deserves attention. This aspect of his definition would seem to arise as an hypothesis from his definition, the "complexity" dimension. About this subsumptive property of the Taxonomy Bloom himself said that the evidence he could collect to support this property was inconclusive (Bloom: p. 19).

Other evidence concerning this subsumptive property is meagre. Kropp et al (1966) did not find the clear reproduction of the pattern which they expected to find in the factor analysis of their tests [Kropp: p. 91 ff]. Also, their Simplex analysis did not produce the consistent order that this property would predict (see: pp. 24-25).

Powell and Isbister (1969) found that a promax rotation did not improve the resolution of factors between subtest scores based upon the advance classification for essentially the same test as used in this study. The subtests as defined in this study by cluster analysis rather than by advance classification showed this same tendency to orthogonality. It is premature to be dogmatic, but it is possible that this subsumptive property may be a hypothesis which will be refuted by the evidence. Alternative analytical procedures such as the one outlined above may be needed to settle this issue conclusively.

There are alternative theoretical positions which would predict the possibility that strategic categories may be discrete and hierarchically ordered by "power" rather than subsumptive. Piaget [1963, p. 13 ff], for instance, has suggested that development may involve shifts in the schema[7] in which case development may be expected to

[7] Alternatively, "the acquisition of new strategies," [see: Powell, 1967, p. 286 ff].

proceed in a series of discrete phases and stages, each of which would be expected to have its own distinctive properties. Such evidence as is available, in particular the difficulty in determining a data-based definition of "complexity" as just discussed; the apparent tendency for "higher" clusters to be more stable than "lower" clusters; and the broader cross-validation support for the hierarchy than for specific interpretations add suggestive support to the latter alternative over the former.

Thus, the advent of taxonomic achievement tests raise some issues in connection with the analytic procedures and the interpretive procedures used for these tests. Whatever else, the results of this study have clearly shown that these tests produce a genuine taxonomy which might be improved by the systematic development of foils, and the use of the responses to these foils as information when evaluating and interpreting these tests and when evaluating, interpreting, and predicting the performance of the individuals taking them.

Implications of This Study to Educational Practice

The findings of this study suggest that tests which are clearly homogeneous regarding internal consistency may form a special case of a broader class of tests which have taxonomic properties. This conclusion has broad implications with respect to their use in the educational setting.

To begin with, the use of the Guidelines would seem to have several practical advantages. First, they simplify item writing because they provide a systematic basis for writing a broader range of foils than can be made without them. Second, the Guidelines improve the basis for the reasons why a foil is wrong. Third, as research further extends

the range of Guideline categories and refines their definition, it may be possible to increase the precision with which such concepts as analysis may be defined, further improving the construct validity of process-oriented taxonomic tests.

Another advantage may arise from the extension of the Guidelines into the Misreading and Misclassification types of foil. Such an extension may link what is now known about diagnostic characteristics of tests in the areas of content-related performance and skill-related performance. This linkage may make it possible to extend the diagnostic aspect of testing beyond the knowledge and comprehension characteristics of of reading and arithmetic into the more abstract characteristics of mathematics and into the more esoteric subjects such as social science, and perhaps even literary appreciation where the subject matter is clearly open to multiple interpretations.

If diagnostic testing can be coupled through research with improved definitions of educational objectives and the factors involved in their attainment, teaching could be more nearly like the practice of medicine. In medicine the practioner classifies a set of characteristics (a syndrome) and uses his knowledge of the effective treatments available to remedy the condition. He then monitors the progress of the treatment. If all goes normally, the condition is corrected. If not, the practioner modifies treatment (prognosis) and/or orders further tests to modify the classification (diagnosis) of the condition, and if necessary, calls in specialists to extend his knowledge resources, moves the patient to the hospital to extend his physical resources, etc.

There are, of course, dissimilarities between medical and

educational practice. Medical men deal generally with short term problems of a clinical dysfunctional nature. The treatment range at their disposal tends to be drastic and when appropriate dramatic in its effectiveness. Educators, on the other hand, generally deal with long term developmental situations. The procedures available are less spectacular, slower acting, and much more complex. However, learning research is now providing increasingly powerful tools for the educator. Among these are CMI (Computer Monitored Instruction), and CAI (Computer Assisted Instruction). These two procedures alone, along with the meaningful interpretation of right and wrong answers in the terms just indicated might greatly extend the capabilities of education.

The essential problem with the bright picture just painted is that at the moment, it contains too many unanswered "ifs." The next section spells out some of the research which might be conducted to help to make this dream become a reality.

Further Research Suggested by the Findings of This Study

There are several areas for further research suggested by this study. One of these involves the host of problems which an extension of test theory could generate. The solution to the "multiple interpretation" problem presented in this study, although suggested and strongly supported by the findings of this study, is probably only one of a range of possible solutions some of which may be more practical than others. Those individuals interested in mathematical statistics could pursue many avenues from this single problem. At a more practical level, there are a host of numerical analyses problems in the implementation of the particular procedure proposed in this study.

Subsequent to the effective implication of an effective analytical procedure, there is the possibility of a host of studies into the characteristics of specific tests and classes of test, into the conditions under which nominal and ordinal categories form, into the types of relationship among categories which are normally found and the conditions under which these relationships occur. Second, order factoring of the centroid matrices seems a logical first step but perhaps the entire structure could be integrated into a single analytic procedure and a single model.

Another essential area for research involves the formulation and resolution of problems arising from the interpretation of clusters after their statistical characteristics have been determined. Attendant to this problem is the cross-validation of interpretation to independent samples of equivalent or nearly equivalent profile characteristics. A profile in this context is a set of clusters and its attendant statistical and logico-semantic characteristics.

The formation of a generic model for a range of type of test opens the possibility for the computer generation of a test of particular construct characteristics derived from the past performance characteristics of a large number of items in a pool of items. With an even larger pool of items, the computer could generate and administer a branching type programme tailored to a wide range of individual differences with the aid of researchable adjustments to the test construction model. In this latter case, the computer could update its performance statistics on the item pool as students take the course, and in so doing refine its own course.

Another area for research could be the reworking of many of the

studies related to educational methodology, evaluating the methods with the profile analysis procedure suggested here. Problems of matching teacher, method, student, and programme could be opened to detailed research, paving the way to the much broader use of diagnostic-prognostic practices in education than their present use. Studies into the relationships between achievement, personality, intellective and perhaps even genetic disposition variables whould also seem reasonably possible from these small beginnings.

Another area for research could be the precise definition of the developmental sequences through which children pass, the optimum ways of modifying these sequences toward specific goals and the degree to which these sequences can be modified. The charting of developmental patterns might lead to earlier and more precise identifications of specific talents. Also, the extension of the Guidelines to include the full range of academic performance might help to answer questions about the relative importance of content and of process in particular subject areas and for specific stages of development.

Finally, there is the psychological question as to whether intellectual development is continuous and cumulative, or discrete and taxonomic, or some combination of these two. In this latter case which aspects of intellectual development are continuous and which are discrete and how do they interact? Can critical phases and critical experiences be identified and matched so as to extend human capabilities?

This list is not exhaustive. It is left to the reader to extend it himself in keeping with his own special interests.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abelson, R. P., and Rosenberg, M. J. Symbolic psycho-logic: A model of attitudinal cognition. Behavioral Science, 1958, 3, 1-13.

Ahmann, S. J., and Glock, M. D. Evaluating pupil growth: Principles of tests and measurement. (3rd ed.) Boston: Allyn and Bacon, 1967.

American Psychological Association, Council of Editors. Standards for educational and psychological tests and manuals. (Rev. ed.) Washington, D. C.: APA, 1966.

Ayers, J. D. Justification of Bloom's Taxonomy by factor analysis. Paper presented at the annual convention of the American Educational Research Association, Feb. 1965.

Berlyne, D. E. Structure and direction in thinking. New York: Wiley, 1965.

Bloom, B. S. (Ed.) Taxonomy of educational objectives: Handbook I: Cognitive domain. New York: David McKay, 1956.

Chown, S. M. Rigidity - A flexible concept. Psychological Bulletin, 1959, 56, 195-223.

Cox, R. C., and Graham, G. T. The development of a sequentially scaled achievement test. Journal of Educational Measurement, 1966, 3, 147-150.

Dexter, L. A. The tyranny of schooling. New York: Basic Books, 1964.

Dinkmeyer, D. C. Child development. Englewood Cliffs: Prentice - Hall, 1965.

DuBois, P. H., Loevinger, J., and Gleser, G. C. The construction of homogeneous keys for biographical inventory. Air Training Command, Human Resources Research Center, Research Bulletin 52-18, Lackland Air Force Base, San Antonio, Texas, May 1952.

Duncan, C. P. Recent research on human problem solving. Psychological Bulletin, 1959, 56, 397-429.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, Prentice - Hall, 1965.

Ebel, R. L. Blind guessing on objective achievement tests. Journal of Educational Measurement, 5 (4), Winter 1968, 321-326.

Educational Testing Service. Multiple-choice questions: A close look. Princeton: ETS, 1963.

Foulds, G. A., and Forbes, A. R. Advanced progressive matrices: Sets I and II. (Manual), London: H. K. Lewis, 1965.

French, J. W., Ekstrom, R. B., and Price, L. A. Kit of reference tests for cognitive factors (Manual). Princeton: Educational Testing Service, 1963.

Furth, H. G. Thinking Without Language: Psychological implications of deafness, N. Y., Free Press, 1966.

Furth, H. G. Piaget and knowledge. Englewood Cliffs: Prentice - Hall, 1969.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart, Winston, 1965.

Glaser, B. G., and Strauss, A. L. The discovery of grounded theory: Strategies for qualitive research. Chicago: Aldine, 1967.

Glaser, R. Instructional technology and the measurement of learning outcomes: Some questions. American Psychologist, 1963, 18, 519-521.

Gorham, D. R. The proverbs test. Missoula, Montana: Psychological Test Specialists, 1956.

Gray, W. S. On their own in reading. (Rev. ed.) Chicago: Scott Foresman, 1960.

Gronlund, N. E. Measurement and evaluation in teaching. New York: Macmillan, 1965.

Gupta, R. K., and Penfold, D. M. E. Correction for guessing in true-false tests: An experimental approach. British Journal of Educational Psychology, 1961, 31, 249-256.

Gupta, R. K. Multivariate analyses of test responses as a pre-requisite to item analysis. Alberta Journal of Educational Research, 1968, 14, 95-100.

Guttman, L., A basis for scaling qualitative ideas, American Sociological Review, 1944, 9, 139-150.

Guttman, L., et al. A new approach to factor analysis: The radex. In Lazarsfeld, P. K. (Ed.) Mathematical thinking in the social sciences. Glencoe, Ill.: Free Press, 1954.

Guttman, L., and Schlesinger, I. M. Systematic construction of distractors for ability and achievement test items. Educational and Psychological Measurement, 1967, 27, 569-580.

Guttman, L., A general nonmetric technique for finding a smallest coordinate space for a configuration of points. Psychometrika, 33 (4), Dec. 1968, 469-506.

Hoffmann, B. The tyranny of testing. New York: Collier, 1962.

Hunt, J. McV. Intelligence and experience. New York: Roland Press, 1961.

Jacobs, P. I., and Vandeventer, M. Information in wrong responses. Research Bulletin, RB-68-25, Princeton: Educational Testing Service, June 1968.

Jaspen, N. Serial correlation. Psychometrika, 1946, 11, 23-30.

Kagan, J. and Moss, H. A. Birth to maturity: A study in psychological development. New York: Wiley, 1962.

Kropp, R. P., Stoker, H. W., and Bashaw, W.L. The construction and validation of tests of the cognitive processes as described in the taxonomy of educational objectives. Cooperative Research Project #2117, United States Office of Education, Tallahassee, Florida: Florida State University, 1966.

Lord, F. A theory of test scores. Monograph 7, Chicago: Psychometric Corporation, 1952.

Marris, P. The experience of higher education. London: Routledge & Kegan Paul, 1964.

Noll, V. H. Introduction to educational measurement. (2nd ed.) Boston: Houghton Mifflin, 1965.

Piaget, J. Logic and psychology. Manchester: University Press, 1953.

Piaget, J. The origins of intelligence in children. New York: Norton, 1963.

Prescott, D. A. The child in the educative process. New York: McGraw-Hill, 1957.

Popham, W. J. and Husek, T.R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.

Powell, J. C. A definition of experience based on a primitive learning model. Alberta Journal of Educational Research, 1967, 13, 275-289.

Powell, J. C. The interpretation of wrong answers from a multiple-choice test. Educational and Psychological Measurement, 1968, 28, 403-412.

Powell, J. C. and Isbister, A.G. A comparison between right and wrong answers on a multiple-choice test. Unpublished research paper, University of Alberta, 1969.

Ross, C. C. and Stanley, J. C. Measurement in today's schools. Englewood Cliffs: Prentice-Hall, 1954.

Sanders, Norris M. Classroom Questions: What kinds? New York: Harper and Row, 1966.

Schonell, F. J. Backwardness in basic studies. Toronto: Clarke-Irwin, 1943.

Shuford, E. H., Albert, A., and Massengill, H. E. Admissible probability measurement procedures. Psychometrika, 1965, 31, 125-145.

Sigel, I. E. How intelligence tests limit understanding of intelligence. Merrill-Palmer Quarterly, 1963, 9, 39-56.

Stoker, H. W. and Kropp, R. P. An empirical examination of the assumptions underlying the structure of cognitive processes using Guttman-Lingoes smallest space analysis. A paper presented at the annual meeting of the National Council on Measurement in Education. Los Angeles, Feb. 1969.

Strutz, P. G. A study of choice behavior of three age groups under three different treatments of a probability learning task. Unpublished Doctoral Dissertation, University of Alberta, 1966.

Thorndike, R. L. and Hagen, E. Measurement and evaluation in psychology and education. New York: Wiley, 1961.

APPENDIX A

TABLES OF BASIC DATA

PHI COEFFICIENTS FOR THE RIGHT ANSWERS
FOR GROUP 'A'
LOWER TRIANGULAR MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1000*																													
2	183	1000																												
3	-115	-061	1000																											
4	-005	050	-012	1000																										
5	020	-104	047	-062	1000																									
6	036	072	024	091	060	1000																								
7	020	053	-082	030	027	061	1000																							
8	082	146	-005	122	-005	-017	049	1000																						
9	-021	-026	-016	093	-137	065	047	-090	1000																					
10	131	094	-012	-058	075	-061	084	038	-072	1000																				
11	016	-038	128	-053	-060	010	021	-015	092	214	1000																			
12	-062	115	-018	189	011	175	047	-038	050	029	-085	1000																		
13	107	045	125	200	046	250	071	120	-063	-205	-066	061	1000																	
14	024	079	149	074	164	156	115	-059	-084	074	075	027	067	1000																
15	163	-055	-029	-018	114	-045	187	067	-207	-105	-087	-029	022	051	1000															
16	082	015	038	-025	-009	061	039	068	-034	073	087	116	055	110	-143	1000														
17	005	019	235	007	035	172	032	219	077	148	-012	011	150	103	-068	072	1000													
18	-037	-069	-088	105	226	-127	123	-078	-184	-055	-115	039	-043	010	525	-252	-035	1000												
19	016	017	038	147	128	013	036	147	-000	147	112	070	041	069	131	-078	034	050	1000											
20	083	030	-077	-030	-017	-054	014	037	105	-068	054	226	076	027	-048	142	-046	-008	-059	1000										
21	006	214	-082	-072	-038	073	023	-009	053	091	050	112	-036	009	-019	056	025	-084	169	063	1000									
22	014	101	-144	094	118	-181	124	125	001	094	-097	-027	-100	-090	191	-009	022	434	-066	038	-074	1000								
23	000	040	093	039	033	-103	097	099	-080	-077	045	001	087	-126	-072	006	032	-020	-097	-004	-069	010	1000							
24	-093	090	-090	-084	121	062	-066	-005	-047	-030	-021	168	-071	091	001	066	044	-123	141	068	196	-362	-056	1000						
25	023	047	-007	072	-068	118	003	082	030	072	055	-007	-118	000	-055	045	-001	-133	-079	045	228	-026	-126	203	1000					
26	-083	-033	053	-006	006	124	-026	-053	-032	033	114	115	048	-071	-113	111	-120	-095	051	-036	-068	-104	003	-016	-148	1000				
27	058	006	053	140	-064	202	001	053	160	-020	-041	095	088	-054	-178	-048	124	-118	022	-114	077	-116	-031	170	221	-165	1000			
28	191	227	075	-032	022	-053	028	228	-149	047	-066	-082	-131	075	029	135	109	-092	169	-034	016	077	160	057	033	-137	029	1000		
29	073	039	012	108	062	061	-084	-038	-052	007	053	012	-038	002	-156	123	-078	-105	017	-009	031	-094	-001	030	119	-033	020	071	1000	
30	-047	-036	095	017	010	047	-217	-039	-042	-069	084	-145	095	034	-010	003	-004	-064	051	-140	-001	-084	-147	078	-025	037	155	019	026	1000

*The numbers on this Table should be multiplied by 10^{-3}.

TABLE 32

PHI COEFFICIENTS FOR THE RIGHT ANSWERS

FOR GROUP 'B'

LOWER TRIANGULAR MATRIX

1	1000*																											
2	137	1000																										
3	006	-025	1000																									
4	-070	-003	-167	1000																								
5	-072	-023	-039	-051	1000																							
6	-006	007	098	125	025	1000																						
7	117	-088	-026	-005	-010	-072	1000																					
8	-080	-084	-111	-039	-030	-010	027	1000																				
9	-050	-081	221	-005	014	-024	-065	143	1000																			
10	016	003	093	063	132	014	-036	186	171	1000																		
11	009	-168	-014	-022	-084	057	-011	-029	021	-017	1000																	
12	118	144	145	-270	170	085	040	-013	-086	071	-058	1000																
13	030	-043	007	026	222	086	-033	-037	025	121	031	158	1000															
14	049	090	180	214	008	015	-017	041	153	107	001	073	119	1000														
15	128	103	-073	-008	-039	068	136	007	-187	009	-090	013	184	-139	1000													
16	060	-036	089	036	155	111	-160	007	084	042	-055	028	-093	095	-051	1000												
17	042	-004	281	-015	028	-049	182	-080	046	016	-077	024	-100	135	-083	060	1000											
18	061	-013	064	-081	-099	-089	330	-118	-280	-073	072	-020	149	-084	387	-314	190	1000										
19	065	-005	033	130	-058	037	129	010	-006	020	051	065	107	182	010	-075	158	-023	1000									
20	-114	-094	027	-015	-109	-045	128	-050	012	030	110	015	-128	103	-015	-098	058	161	-159	1000								
21	-158	-126	123	-141	-041	-057	-099	-002	152	-011	-041	-051	113	152	-061	-019	031	-018	042	128	1000							
22	-015	225	-121	104	-015	-151	080	-118	-231	-112	-111	-047	-068	033	232	006	072	304	-051	069	-043	1000						
23	-006	-057	005	-026	093	072	-027	-010	041	014	-089	-041	-090	-072	-074	042	037	-089	068	-045	006	-209	1000					
24	030	111	099	-140	131	069	-234	-043	130	052	056	173	-005	057	-121	079	-033	-307	087	-034	-062	-320	279	1000				
25	037	104	098	050	-008	-073	017	-010	269	014	-059	085	-090	-014	-074	042	037	-175	-151	101	038	-033	101	-015	1000			
26	-023	-071	-074	154	-046	080	-053	-160	-109	-026	086	-040	072	041	061	021	066	-040	046	127	038	-016	-101	039	-071	1000		
27	-011	-032	017	104	006	203	-195	086	160	119	-009	071	022	010	-064	096	-223	-335	-191	-016	063	-296	096	258	203	-092	1000	
28	078	052	-111	-007	102	059	-010	078	-024	-055	-084	-013	-084	008	-039	155	128	001	-058	060	107	121	059	-064	025	059	-035	1000
29	060	040	089	-232	-124	-198	049	-066	123	-054	048	102	011	164	-051	-139	060	-007	073	040	019	006	-061	179	042	-122	-073	-084	1000
30	-005	-020	029	-071	-025	-063	-123	059	147	125	152	081	042	168	-060	210	-005	-223	-063	038	150	-145	078	146	113	-041	073	-025	160	1000

* these numbers should be multiplied by 10^{-3}

TABLE 33

TABLES 34, 35 & 36

GIVE THE

PHI COEFFICIENTS FOR THE WRONG ANSWERS

FOR GROUP A

LOWER TRIANGULAR MATRIX

IN THREE PARTS

TABLE 34

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	1000*																																													
2	241	1000																																												
3	-271	-028	1000																																										
4	-173	022	-040	1000																																									
5	-163	049	-028	-077	1000																																								
6	-067	-005	-018	115	192	1000																																							
7	-122	-005	301	-065	132	-029	1000																																						
8	048	122	-018	-091	131	027	-069	1000																																				
9	-032	039	153	-116	026	108	057	158	1000																																			
10	-049	059	-063	057	039	163	-101	132	099	1000																																		
11	079	032	090	-049	048	-075	-001	-004	114	-002	1000																																	
12	031	037	090	-049	130	-075	-001	032	114	085	-118	1000																																
13	-109	096	493	-040	-026	-010	-023	-018	153	-063	090	-061	1000																															
14	-067	017	-018	-050	-034	-022	-029	-073	-052	-078	-075	049	-018	1000																														
15	066	118	-018	-050	192	319	-029	027	108	042	-075	-075	-018	-022	1000																													
16	054	-028	-015	-040	-026	-018	-023	-018	-042	083	-061	090	-015	-018	-018	1000																												
17	018	-032	-090	-130	116	075	097	331	001	045	-016	-016	-241	-049	075	-241	1000																											
18	030	-056	-042	018	-000	-052	-067	-077	-121	-070	-001	-117	-042	267	100	153	-172	1000																										
19	-139	-070	-037	152	-071	307	-059	005	-024	-037	-154	-027	-037	-046	-046	-037	091	-107	1000																									
20	-122	-163	625	-065	-044	-029	170	088	182	-101	-001	096	301	-029	-029	-023	001	-067	216	1000																								
21	-026	051	-034	092	-064	150	-054	126	-007	125	-001	-070	-034	-041	-041	-034	070	-007	014	-054	1000																							
22	066	-106	-016	-050	-034	319	-029	027	108	-078	-075	049	-018	-022	-022	-018	075	-052	493	503	-041	1000																						
23	127	032	-061	011	-034	049	-097	032	-001	085	061	-073	-061	049	172	-061	163	114	-091	-097	130	-075	1000																					
24	-109	-028	493	-040	246	-016	301	-018	-042	-063	090	241	-015	-018	-018	-015	061	-042	-037	301	-034	-018	-061	1000																				
25	112	098	-067	-074	026	-063	-106	000	024	036	141	-069	-067	033	-063	217	-015	024	009	-017	-024	033	-027	-067	1000																			
26	026	-076	-051	126	-097	-063	-081	-004	-017	-075	139	-062	-051	076	-063	-351	-139	112	-050	-031	-117	-063	039	-051	048	1000																		
27	-155	-128	-021	-018	157	-026	198	061	079	-090	021	021	-021	-026	-026	-021	086	079	-051	-033	-048	-026	-036	341	-096	-073	1000																	
28	169	190	-070	-059	-134	-046	066	029	012	-063	039	-126	-070	142	-000	069	-039	012	-119	-112	-033	-046	001	-070	103	-050	0.8	1000																
29	059	-040	-035	-010	-060	-044	-057	-016	-016	105	-014	-014	-035	-044	-044	412	-119	242	-090	-057	022	-044	052	-035	024	100	109	014	1000															
30	038	-107	-010	254	-020	-013	-016	-099	-030	162	-043	169	-010	-013	-013	-010	-169	-030	-026	-016	-024	-013	-043	-010	-047	202	-015	-049	290	1000														
31	-434	-034	-023	-065	134	-022	-037	010	057	-007	-001	192	-023	237	-029	-023	097	-057	-059	-037	-054	-029	-097	-023	-017	026	-033	-112	-057	-016	1000													
32	010	-403	169	-025	-074	-040	071	-024	-032	-169	-038	085	-039	-048	-048	169	038	207	077	336	-009	292	023	169	-005	002	093	041	180	-027	-062	1000												
33	-076	-055	-066	101	-018	036	-016	150	-025	045	-012	-019	-066	-001	-081	-066	-151	030	014	-014	178	-001	024	-066	-144	056	110	037	093	156	-014	000	1000											
34	-191	-107	705	-028	-020	-013	441	-099	-030	-045	-043	169	-010	-013	-013	-010	043	-030	-026	441	-024	-013	-043	705	-047	-036	-015	-049	-025	-007	-016	265	-047	1000										
35	161	-090	-028	022	-289	-025	-084	062	039	-048	-041	-005	-028	-085	017	096	078	007	035	075	051	118	032	-028	-006	007	-039	088	-040	-107	-084	053	-090	068	1000									
36	099	-105	080	-007	-044	-010	082	-064	092	013	-011	-011	-065	-080	-080	080	-033	-075	020	-011	-002	-000	-097	000	270	112	011	127	-029	159	-103	065	-047	159	101	1000								
37	-037	054	050	-017	059	008	-129	008	060	-011	052	130	050	008	008	181	-246	169	-039	038	-066	000	-064	050	-001	-022	071	078	034	127	121	052	078	-057	-041	015	1000							
38	-109	-028	493	-040	-026	-010	301	-141	153	-063	-061	241	-015	-018	-018	-015	061	-042	-037	301	200	-018	090	493	075	-051	-021	-070	-035	-010	-023	169	-066	705	096	226	-041	1000						
39	-062	-122	-054	103	-103	067	010	-244	-155	049	114	-127	-054	-066	-066	-054	-066	032	001	-066	026	-066	114	-054	025	137	-077	009	-131	-038	-046	-010	216	-038	-161	-053	-009	-054	1000					
40	027	070	230	-033	-057	-037	-048	-100	113	-130	107	-124	230	176	-037	-030	-184	113	-076	118	091	-037	-047	-030	-065	155	-043	212	042	-021	-048	027	-062	-021	007	016	035	-030	-027	1000				
41	-002	-056	-030	040	-030	118	125	090	024	074	-330	060	-000	010	-098	052	018	-027	077	041	101	226	021	-000	-073	-059	-020	-021	097	129	-043	058	057	-056	042	024	-135	-000	-0.2	039	1000			
42	-077	-030	-009	003	-032	098	019	-038	-014	009	-133	-371	-009	-006	093	037	-042	229	202	019	-031	098	000	-009	-022	067	053	-081	118	-063	-143	073	133	-063	001	003	060	-009	077	012	105	1000		
43	-101	-064	200	092	-064	-041	095	136	172	-011	-070	068	-034	-041	-041	200	-060	003	014	245	-078	150	-070	200	-024	-040	-040	-033	022	-024	-054	197	-020	306	051	105	235	200	-124	-069	059	020	1000	
44	-044	-028	160	047	-077	-050	064	-100	038	-176	190	-049	361	-050	-050	-040	-070	-116	-103	064	-093	-050	-168	-040	150	059	-050	-029	-093	-028	064	-025	-069	-026	-077	050	-017	-040	-0.5	225	-064	-147	092	1 00	
45	-001	097	-026	-071	113	-032	149	-032	-074	-111	-018	-107	-026	-032	-032	-026	-158	-074	-065	-041	-059	-032	-013	-026	048	-090	-037	121	069	-010	-041	-068	052	-018	-121	-029	011	-026	001	-052	-063	066	078	-071	1000

* these numbers should be multiplied by 10^{-3}

TABLE 35

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
46	-067*	-186	-018	115	-034	319	-029	027	-052	042	-075	-075	-018	-022	-022	-018	075	-052	131	-029	150	-022	049	-018	-083	-063	-026	028	-044	-013	-029	122	154	-013	-166	039	-099	-018	200	-037	116	-006	-041	-050	-032
47	-039	049	-021	-058	-040	-026	-033	061	218	119	-026	128	-021	-026	-026	-021	-143	079	-053	-033	-048	-026	-086	-021	106	048	-030	-100	-050	-015	-033	-055	008	-015	-039	011	-022	-021	039	-043	-020	-037	119	085	-037
48	093	094	-295	-067	094	-080	-031	063	010	016	001	053	-123	-221	-080	-123	203	-848	053	-141	-045	061	-050	-123	-012	-056	-052	002	-184	035	079	-220	-019	-208	-032	-027	-169	-123	-038	-163	002	-127	-124	001	087
49	-048	-027	128	006	004	085	148	003	064	-056	-095	-095	128	-059	085	-048	-166	064	-122	148	-111	-059	061	128	-124	008	057	203	-039	-034	036	016	180	214	059	038	096	128	048	092	012	101	091	-064	019
50	086	-009	-023	-065	-044	-029	-037	010	057	180	-001	096	-023	-029	-029	625	-096	057	-059	-037	-054	-029	096	-023	074	-081	-033	066	230	-016	-037	203	-106	-016	154	-011	038	-023	-056	-045	041	019	096	-065	-041
51	-056	006	031	087	-048	-013	010	032	-029	-008	019	131	031	-115	090	-093	055	-029	133	170	-215	192	-203	031	-082	047	045	031	-116	110	010	109	036	-066	126	-128	094	-093	056	000	-023	-073	015	-061	-091
52	-056	-057	-030	020	-057	-037	-048	-038	-086	-130	030	-124	-030	176	-037	-030	-184	213	034	-048	051	-037	-047	-030	007	-018	-043	141	042	-021	-048	133	012	-021	-183	-132	035	-030	-027	072	039	141	-069	020	100
53	050	013	-055	100	069	-068	014	046	-098	-008	-040	008	-055	-068	-068	-055	-055	025	-073	-088	094	-068	-229	-055	-165	-033	-079	-221	-063	-039	-048	-081	157	-039	052	-016	106	-055	-101	-031	-010	-058	094	-026	-097
54	-176	-121	-034	092	063	-041	-054	014	-007	-078	-070	-001	-034	-041	-041	200	-068	-097	212	245	030	342	-070	-034	172	-040	119	095	125	-024	-054	197	046	-024	051	118	115	-034	-124	-069	059	028	030	-001	078
55	-075	-168	093	077	-030	052	003	091	005	-125	-020	025	093	-073	-073	-059	-071	064	-066	100	074	052	-065	093	-274	-055	133	-117	-009	-042	003	155	293	-042	-131	-132	143	-059	026	-043	113	018	074	017	-104
56	-009	039	-086	064	-095	-002	025	-094	-151	032	-010	020	041	-106	-002	041	-020	-102	159	-056	036	-002	-056	-086	-184	-302	-123	-064	-041	-061	025	032	-030	-061	039	-164	114	-066	-073	-111	060	205	036	064	-002
57	077	003	-039	-025	-074	-048	071	076	-032	012	023	-100	-039	-048	-048	-039	038	-032	165	-062	-089	-048	-038	-039	-063	002	-055	-016	-094	-027	-062	067	118	-027	-093	125	052	-039	189	-079	-158	022	006	-107	-068
58	053	-060	048	029	055	112	035	021	-137	169	-072	-033	-032	005	112	048	188	013	-099	035	230	-101	121	048	-162	-150	-117	-395	-028	-058	-048	-006	-005	125	-060	-142	-087	170	-012	-168	-111	048	-010	-125	-145
59	110	070	-030	123	-057	-037	-048	087	014	021	030	107	-030	-037	-037	-030	047	014	-076	-048	-069	-037	-047	-030	-065	-018	-043	-001	-072	-021	-048	027	012	-021	070	016	035	-030	056	-061	-028	-182	170	-083	-052
60	-028	-035	-051	-141	-005	-063	026	-004	048	023	-012	059	118	214	-063	-051	-089	-017	-058	-001	030	-053	-012	-051	001	-066	-073	033	-124	-036	134	002	-088	-036	-076	064	-065	118	-050	-018	-015	-143	-040	-008	207
61	-334	-055	-010	115	-034	-022	-029	-073	-052	042	-075	-075	-018	-022	-022	-018	-049	-052	131	-029	-041	-022	-075	-018	-083	-063	271	142	-044	-013	-029	-048	036	-013	017	-080	115	-018	067	-037	-098	098	-041	-050	-032
62	-001	-266	-026	-071	-049	212	-041	-104	-074	-025	-107	-107	-026	-032	-032	-026	107	040	061	149	070	212	070	-026	-035	009	-037	-123	-062	-010	-041	-068	052	-018	024	056	011	-026	096	-052	014	288	-059	047	-045
63	-022	-056	-101	-036	076	077	-004	-024	-101	060	055	-127	022	178	-023	022	200	057	004	-161	-119	-023	055	-101	-018	057	-144	-113	-190	-071	075	-067	-456	-071	064	010	-115	-101	-096	045	-007	-033	051	158	-105
64	153	028	-193	-535	-076	-017	-137	079	052	133	-006	114	-059	093	093	075	006	062	021	-137	-136	093	034	-059	123	-050	107	067	005	-136	035	-019	-039	-136	-038	065	072	-059	063	-121	-042	092	-012	-110	-104
65	-082	010	-044	-046	-093	-054	-070	-055	-125	-000	-069	-069	-044	102	-054	-044	-100	093	-031	-070	-013	-054	-069	-044	011	-026	073	050	145	-031	051	039	070	-031	-453	-031	-046	-044	082	008	-041	-030	-013	104	145
66	-078	050	077	-069	-048	-031	-014	069	030	045	024	024	077	-081	-081	-066	062	250	014	170	046	154	-019	-056	-064	-040	110	-120	156	-047	070	118	066	-047	015	-293	115	-066	-060	-062	-025	133	046	044	052
67	068	076	-123	088	-038	-053	-197	033	-077	059	-011	-011	-123	046	046	-123	190	-077	045	-110	050	046	061	-123	208	132	-176	-062	-085	-007	-043	-131	-080	-087	047	040	-684	-003	087	056	080	-091	-173	-007	-075
68	039	056	-044	029	-083	-054	-070	-420	-125	030	-125	044	-044	102	-054	-044	-100	020	050	-070	-100	-024	044	-044	011	-026	-062	050	-106	-031	-070	-038	-144	-031	-083	-031	-046	-044	021	008	008	111	-100	-121	-077
69	077	049	-021	-058	-040	-026	-033	148	-060	-090	-026	021	-021	-026	-026	-021	086	079	254	198	-048	271	021	-021	005	-073	-030	098	109	-015	-033	093	008	-015	049	011	-115	-021	-077	142	074	-037	-048	085	-037
70	107	-032	-135	-084	-058	-066	090	-009	-154	-585	018	-018	-013	-066	034	-013	127	079	018	018	-142	133	-018	-135	127	-026	068	120	-035	-095	018	189	-094	-095	116	-110	037	-135	-107	-275	023	-035	-030	060	119
71	-067	118	-018	115	192	319	-029	-073	108	163	-075	-075	398	-022	319	-018	-049	-052	131	-029	-041	-022	049	-018	-083	-063	-026	-006	-044	-013	-029	-048	-031	-013	017	-080	006	-018	-066	-037	-098	098	-041	115	-032
72	020	-126	-040	047	-077	115	064	-100	038	-001	011	-160	-040	-050	-050	-040	-070	038	-010	-065	-001	-050	-049	-040	038	059	065	001	-098	-028	-065	057	-013	-028	120	165	-121	-040	044	123	-012	-247	-001	126	-071
73	112	090	-067	-010	-051	-093	-017	000	-140	071	-021	019	-067	033	033	-067	027	133	069	-108	041	-083	-027	075	-113	048	106	-207	-164	-047	074	-171	-024	-047	028	067	029	-067	-021	007	-073	119	-155	-074	048
74	050	-007	024	-031	-055	-021	078	082	-049	068	-047	063	-099	-122	-021	147	010	-002	113	078	-172	181	-083	024	-010	022	122	199	194	104	-159	240	004	104	203	265	-010	024	-050	-077	014	069	169	-275	-030
75	143	-102	-146	-029	-123	-257	-051	010	-093	-030	-044	181	-146	054	-412	-146	100	-093	031	-051	013	054	-156	044	042	089	-073	053	-145	031	070	038	-017	031	129	139	-002	044	-002	-008	090	-111	-074	046	-509
76	-093	-136	123	067	-094	-061	-079	020	-010	-066	-104	101	123	221	-061	-050	053	056	093	251	-035	221	-155	-050	-084	-058	-071	-143	-044	-035	141	210	-078	-035	116	-121	169	-050	-127	-013	132	-002	045	067	-087
77	-056	003	169	057	-074	-048	-062	-224	-032	-048	085	-038	376	122	-048	376	-640	207	-099	071	006	-048	-100	-039	-063	140	-055	098	180	265	-062	-018	235	-027	-048	125	266	-039	056	239	-050	125	102	057	175
78	-067	-085	398	-040	-034	-022	237	-073	108	042	-075	172	-018	-022	-022	-018	-049	-052	-046	537	-041	-022	-075	398	033	-063	-026	028	-044	-013	-029	122	-081	573	017	277	008	398	-066	-037	010	-110	150	115	-032
79	146	126	-113	-122	-018	-139	-104	134	-047	174	140	104	-113	059	059	129	-033	-047	-288	-181	074	-139	-003	000	053	-033	011	154	153	091	-104	-053	-067	-080	007	-017	-003	-113	-068	078	043	-148	018	-026	014
80	077	-092	-096	028	-049	-017	005	-032	056	-095	-032	041	028	-017	-118	-096	-041	008	070	-154	007	-017	041	-096	041	076	-049	-054	-070	-068	064	-003	-015	-068	-092	105	-150	028	042	-070	054	092	-163	077	-097
81	060	035	077	-069	-048	-081	170	-118	-025	-038	-062	-062	-066	-081	-081	221	-151	030	-107	-014	-152	-081	024	077	217	056	110	-002	156	-047	-106	059	-055	156	-090	241	041	077	032	012	050	062	046	-013	136
82	125	180	-279	-117	-080	-207	-002	066	-040	132	-175	-028	-133	-071	064	052	274	-103	-218	-340	-032	-343	169	-279	057	-037	-043	-021	-019	-196	084	-159	-044	-196	180	-052	-094	-113	087	-147	-058	-051	-106	014	-199
83	-434	-005	301	192	-044	237	170	-069	-067	180	-097	192	-023	-029	-029	-023	097	-067	216	170	096	-029	-097	301	-103	-081	-033	-112	-057	-016	378	071	078	441	-005	-011	038	301	016	-048	041	-062	096	064	-041
84	027	-057	-030	020	-057	176	-048	025	014	-055	030	-047	-030	176	176	-030	-030	024	034	118	051	-037	-047	-030	-065	-018	-043	070	-072	-021	-048	027	035	-021	-057	-058	-032	-030	056	072	-028	141	051	-003	252
85	-334	-186	398	115	-034	319	237	027	-052	042	-075	049	-018	-022	-022	-018	075	-052	131	237	150	-022	-075	398	-083	-063	271	028	-044	-013	-029	122	036	573	017	039	008	398	067	-037	010	-006	150	-050	-032
86	-117	-044	133	-058	009	-057	-075	-016	142	058	127	-033	133	090	-057	-047	-020	073	-042	271	141	090	020	-047	087	-163	062	-125	046	-033	041	023	044	-033	-044	051	-073	133	001	-003	-067	-060	-024	013	024
87	-039	049	-021	085	-040	271	-033	061	079	119	128	021	-021	-026	-026	-021	-021	-060	100	-033	119	-026	-086	-021	106	-073	-030	098	-050	-015	-033	-055	213	-015	-039	115	071	-021	039	-043	074	-037	119	-058	175
88	-056	007	230	020	084	-037	-048	150	014	-055	261	030	230	-037	-037	-030	-030	113	034	118	-069	-037	-047	230	-065	068	142	-143	156	-021	-048	133	012	-021	007	-058	035	-030	-110	072	-095	012	051	123	-052
89	-200	-186	398	115	-034	319	237	027	-052	042	-075	049	-018	-022	-022	-018	075	-052	131	237	150	-022	-075	398	-083	-063	-026	-085	-044	-013	-029	122	154	573	017	039	008	398	067	-037	118	-006	150	-050	-032
90	-173	-077	-040	126	142	115	-065	-003	-116	-059	130	-108	-040	-050	-050	-040	049	-039	067	-065	-001	-050	-049	160	-074	-074	228	026	079	-028	-065	057	044	-028	-077	050	-017	-040	108	-083	-064	-047	-001	047	-071

* these numbers should be multiplied by 10^{-3}

TABLE 36

	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
46	1000*																																												
47	-026	1000																																												
48	-080	-052	1000																																										
49	085	-068	-076	1000																																									
50	-029	-033	-031	-077	1000																																									
51	-013	045	063	-005	-149	1000																																								
52	-037	-043	-163	-098	-048	-063	1000																																							
53	-068	-079	025	-126	-088	119	-031	1000																																						
54	-041	-048	114	-111	096	130	051	020	1000																																				
55	052	024	-058	016	-095	184	113	112	004	1000																																			
56	103	-123	076	-063	107	-081	150	036	-081	-005	1000																																		
57	-048	-055	061	-128	071	109	087	049	006	279	-125	1000																																	
58	112	-117	-029	-046	035	-047	-035	221	-130	057	069	-059	1000																																
59	-037	-043	013	082	-048	192	-061	051	-069	113	-111	027	-035	1000																															
60	076	048	001	066	026	-077	068	-086	-117	-055	038	-067	015	-018	1000																														
61	-022	-026	061	-059	-029	192	176	-068	150	052	-002	122	-101	-037	-063	1000																													
62	212	-037	-014	019	-014	-018	-052	-097	078	164	-077	-068	160	-052	009	-032	1000																												
63	-023	-056	008	-076	075	-070	045	-110	-119	-114	022	-017	124	108	098	-023	111	1000																											
64	-017	107	018	-031	121	-060	086	-096	-074	-058	-061	091	-023	017	085	-017	054	032	1000																										
65	-054	-062	-109	-012	-070	-092	395	-045	-013	108	076	039	-052	104	226	102	-077	020	075	1000																									
66	-081	110	-214	-118	-106	107	085	-069	046	034	151	000	032	012	008	036	-032	-074	037	017	1000																								
67	-152	-004	132	-155	-043	-046	-067	028	-062	-211	-183	-032	047	-005	051	-053	-075	055	-035	082	-114	1000																							
68	102	-062	019	-012	051	-092	104	-045	-100	-063	171	-038	045	-089	-026	192	034	-026	075	-059	-090	082	1000																						
69	-026	-030	-052	057	-033	045	-043	-079	119	-085	-032	093	-117	-043	-073	-026	-037	-056	012	073	110	168	-062	1000																				
70	-066	-105	045	020	-060	096	097	-050	082	075	091	-010	-198	-089	-066	034	-023	012	-013	052	009	-067	006	068	1000																			
71	-022	-026	061	229	-029	-013	-037	-068	-041	-073	103	-048	005	-037	076	319	-032	178	-017	-054	-081	-053	102	-026	-066	1000																		
72	115	085	-135	006	064	038	-068	-026	092	-043	-138	139	-022	020	-074	-050	-071	061	-057	-046	-126	041	-046	-058	-084	-050	1000																	
73	-083	005	-060	-074	074	-047	000	058	-155	-060	030	111	020	007	142	533	-035	085	160	117	137	039	117	009	-180	033	-131	1000																
74	-021	-054	002	-071	156	153	-014	052	225	-070	-091	038	-118	-014	-101	181	-102	001	121	-020	-101	041	-066	034	029	-021	-080	-010	1000															
75	-102	062	173	012	-172	054	077	117	076	-082	-070	098	130	062	-052	-054	077	117	076	-082	-070	098	130	062	-052	-102	-029	042	066	1000													
76	-061	052	-066	-044	-079	107	-013	084	045	110	140	009	074	-013	-001	080	-087	033	-109	-019	165	-009	045	052	120	080	-001	012	-002	084	1000												
77	-048	-055	-220	160	203	-044	239	049	102	092	-021	067	-059	-079	071	182	-068	-067	036	116	-058	-180	039	-055	-060	122	-025	053	089	-194	-061	1000											
78	-022	-026	-362	085	-029	-115	-037	-068	-041	-073	-002	-048	005	-037	076	-022	-032	-124	-017	102	-081	-053	-054	-026	-116	-022	115	-083	080	054	-061	-048	1000										
79	-139	097	056	-373	129	110	140	-009	018	-059	092	-053	044	-046	007	359	-199	039	136	114	035	081	-022	-075	-058	-139	022	053	112	022	-138	093	059	1000									
80	085	128	032	198	-194	-067	-007	026	-051	-018	-070	-093	029	-007	-007	-118	121	-004	-028	083	-155	160	-010	039	062	-017	028	041	-053	195	052	-104	-017	-274	1000								
81	036	110	-068	131	078	-423	012	-205	046	-096	-102	-058	-115	-062	-040	-081	-032	-039	037	017	-096	-114	124	008	009	-081	158	-024	108	-124	-078	059	154	-067	129	1000							
82	-071	075	215	058	084	-113	-570	042	-261	-136	-023	004	126	107	018	-371	-005	041	080	-214	-044	167	033	-043	007	064	014	-036	-126	152	-047	-402	-071	-068	103	050	1000						
83	237	-033	-141	036	-037	010	-048	-088	-054	-095	189	-062	035	-048	-081	237	-041	-004	-137	051	078	-120	051	-033	-138	237	064	165	-001	-051	141	-062	237	-104	-075	-014	-128	1000					
84	176	-043	-075	-008	-048	128	078	-113	-069	035	085	027	097	-061	155	-037	100	-080	-121	201	-062	-067	008	-043	097	-037	020	-065	-077	-298	-013	133	-037	078	-070	-135	-231	118	1000				
85	319	-026	-221	085	-029	-013	-037	-068	150	-073	-002	-048	112	-037	-063	319	-032	-124	-128	-054	036	-152	-054	-026	-066	-022	115	-083	080	-102	-061	-048	319	-040	-118	036	-207	503	176	1000			
86	-057	062	-147	034	-075	012	-095	-006	058	-027	-276	023	060	-095	077	-057	024	-060	050	062	145	077	-072	062	-085	-057	013	-064	-098	-062	086	023	090	-061	-074	-007	-068	-075	089	-057	1000		
87	271	-030	-052	-068	-033	045	-043	035	-048	-085	059	-055	068	142	168	-226	-037	-144	012	073	003	-090	-062	-030	-019	-026	085	-096	-054	-208	-071	093	-026	097	-137	-094	-161	198	327	271	-066	1000	
88	-037	-043	-163	-098	-048	128	072	051	-069	191	019	027	-168	072	-104	176	-052	-018	-121	008	085	-005	008	-043	035	176	020	080	049	-008	251	027	-037	016	-070	-062	-147	118	-061	-037	-003	-043	1000	
89	319	-026	-221	085	-029	-115	-037	063	-041	052	103	-048	112	-037	-063	-322	-032	-124	-128	-054	-081	-152	-054	-026	-066	-022	115	033	-021	-102	080	-048	319	-139	-017	036	-207	503	176	659	-057	271	-037	1000	
90	279	-058	001	-064	-065	087	123	-153	092	138	064	139	-074	020	-141	779	047	110	-057	029	-069	-199	029	-058	060	115	047	-018	066	-029	-069	-025	-050	022	-120	044	-182	192	020	113	013	085	329	113	1000

* these numbers should be multiplied by 10^{-3}

TABLES 37, 38 & 39

GIVE THE

PHI COEFFICIENTS FOR THE WRONG ANSWERS

FOR GROUP B

LOWER TRIANGULAR MATRIX

IN THREE PARTS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	1000*																																												
2	327	1000																																												
3	000	000	000																																										
4	-098	-097	((	1000																																									
5	-192	-113	000	054	1000																																									
6	-226	-155	000	200	275	1000																																								
7	-080	114	000	-036	-028	-026	1000																																							
8	-056	-105	000	010	086	130	114	1000																																					
9	024	-024	000	-045	-095	215	-050	100	1000																																				
10	-155	-034	000	008	088	278	066	098	130	1000																																			
11	043	000	000	053	-017	-080	-045	-132	-029	-022	1000																																		
12	-055	000	000	-044	-056	-039	071	-033	099	043	093	1000																																	
13	-080	-129	000	-036	-028	271	-015	-007	122	204	-045	212	1000																																
14	-010	-010	000	109	157	175	-021	-097	-071	-004	198	001	-021	1000																															
15	046	081	000	-025	-020	-018	-010	081	-035	-051	228	-049	-010	-015	1000																														
16	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000																													
17	-031	136	000	-169	042	-060	-083	-075	117	033	091	-106	-230	091	045	000	1000																												
18	195	115	000	-023	-014	-077	-044	-021	-020	094	-067	100	-044	-063	-031	000	-029	1000																											
19	-089	-201	000	156	084	555	-030	172	250	276	001	074	229	142	-021	000	-170	007	1000																										
20	046	-091	000	-025	-020	-018	-010	081	-035	-051	-032	-049	-010	-015	-007	000	-162	-031	-021	1000																									
21	-050	-069	000	018	-010	193	278	094	158	147	138	073	113	043	196	000	-202	-034	231	196	1000																								
22	-080	114	000	-036	-028	-026	493	-007	-050	204	-045	071	-015	-021	-010	000	-083	145	229	-010	113	1000																							
23	022	130	000	-086	043	064	-063	-003	242	180	-139	-089	086	-089	-044	000	099	147	-128	-044	-030	-063	1000																						
24	-010	-017	000	-051	-040	-037	-021	077	-071	094	-065	102	-021	-030	-015	000	-223	072	327	494	161	340	017	1000																				
25	-001	155	000	-129	-002	020	056	-005	191	-052	-045	042	056	-110	-055	000	027	-082	-090	-055	027	056	102	-110	1000																			
26	-096	046	000	063	226	079	-058	-065	-040	-033	101	026	097	138	177	000	074	-059	039	-041	043	-058	-132	-083	019	1000																		
27	-159	-091	000	-025	370	401	-010	-091	-035	144	-032	-049	-010	-015	-007	000	045	-031	-021	-007	-037	-010	166	-017	-055	-041	1000																	
28	-163	030	000	033	-013	047	049	-006	079	020	033	160	-002	-023	127	000	051	-077	-033	-057	134	-002	037	-116	037	-044	127	1000																
29	166	034	000	004	-071	-066	-037	-069	092	-066	041	003	-037	-053	-026	000	100	-112	033	-026	-064	-037	030	-053	-034	-149	-026	-041	1000															
30	-159	-091	000	-025	-020	-010	-010	081	-035	-051	228	-049	-010	-015	-007	000	-162	-031	-021	-007	196	-010	-044	-015	-055	-041	-007	-057	-026	1000														
31	-160	020	000	-057	132	-041	-024	105	-000	061	-073	249	-004	-033	-017	000	102	-070	-048	-017	-084	-024	-005	198	047	006	-017	121	-060	-017	1000												
32	-022	-184	000	061	019	034	-044	-021	-035	043	-136	-049	-044	-063	-031	000	-029	-060	-090	236	-034	-044	-020	072	-132	-059	236	001	049	-031	-070	1000											
33	-053	-066	000	039	-042	146	-064	040	182	247	070	-016	083	014	-045	000	107	029	170	-045	010	003	030	118	051	061	-045	-089	-038	-045	-008	-136	1000										
34	-159	-091	000	-025	-020	401	-010	081	208	144	-032	-049	-010	-015	-007	000	045	-031	344	-007	196	-010	-044	-015	-055	-041	-007	127	-026	-007	-017	-031	162	1000									
35	072	028	000	-005	-312	-139	090	-093	090	-007	-090	178	090	-052	063	000	049	174	-075	-115	-091	090	011	-052	107	-107	-115	140	-093	-115	-018	-015	024	-115	1000								
36	028	-084	000	-069	063	-100	100	-034	074	020	-005	-007	-057	031	-040	000	020	006	125	-040	155	100	001	255	-052	014	-040	-032	-145	-040	009	-053	117	-040	-123	1000							
37	034	003	000	-049	017	-043	-071	-051	044	033	035	-024	-071	-101	147	000	065	101	-073	-050	019	-071	067	-101	063	017	147	-033	-003	147	-024	-107	137	-050	004	-050	1000						
38	-039	041	000	-044	-034	-032	-018	-158	080	025	-056	145	-018	-026	-013	000	079	-014	-037	-013	071	-018	-077	-026	-095	055	-013	114	-046	-013	-029	-054	152	-013	111	059	027	1000					
39	-185	-248	000	061	123	256	-044	160	-149	146	-067	-102	-044	072	-031	000	-084	-111	104	-031	028	-044	035	072	-032	057	236	098	-031	-031	051	152	029	-031	-157	-112	049	-054	1000				
40	-081	028	000	229	-045	-041	-024	028	140	-115	045	159	-024	-033	-017	000	-036	031	-048	-017	021	-024	-100	-033	047	105	-017	-130	076	-017	-018	-070	-102	-017	063	009	065	-029	-070	1000			
41	092	-259	000	-020	062	-064	-080	088	-049	-070	-248	-086	183	-114	-057	000	-262	-032	038	129	-073	-000	-032	167	-006	124	-057	-207	019	-057	-044	154	-043	-047	-069	057	121	009	000	040	1000		
42	073	093	000	005	-103	065	037	063	-004	076	-140	-424	-090	-123	-063	000	024	-030	075	115	000	037	026	052	027	-009	-063	-107	-014	-063	-144	-032	160	115	013	003	124	-111	-000	-063	069	1000	
43	-005	-093	000	-003	-075	-069	-039	-044	149	147	-045	-012	-039	-056	-028	000	-131	116	026	-028	-073	168	016	092	-097	035	-028	051	-011	-028	-063	-039	071	-000	004	106	-074	122	-039	070	036	000	1000
44	021	003	000	069	-081	040	-042	050	107	058	011	-092	-042	-060	-030	000	128	-054	013	-030	-025	-042	-009	-060	005	010	-030	016	057	-030	-063	-054	042	045	016	-043	010	100	019	111	022	033	000	1000
45	-056	059	000	068	-049	-045	-026	-012	-038	-127	026	-039	-026	-037	-018	000	-060	-077	-053	-013	101	-026	-110	-037	020	079	-010	007	-066	-010	149	-077	-026	-011	-005	-000	-013	-037	034	119	013	-010	-000	000	1000

*these numbers should be multiplied by 10^{-3}

TABLE 37

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1	1000*																																												
2	327	1000																																												
3	000	000	000																																										
4	-098	-097	000	1000																																									
5	-192	-113	000	054	1000																																									
6	-226	-155	000	200	275	1000																																								
7	-080	114	000	-036	-028	-026	1000																																							
8	-056	-105	000	010	086	130	114	1000																																						
9	024	-024	000	-045	-095	215	-050	100	1000																																					
10	-155	-034	000	008	088	278	066	098	130	1000																																				
11	043	000	000	053	-017	-050	-045	-132	-029	-022	1000																																			
12	-055	000	000	-044	-056	-039	071	-033	099	043	093	1000																																		
13	-080	-129	000	-036	-028	271	-015	-007	122	204	-045	212	1000																																	
14	-010	-010	000	109	157	175	-021	-097	-071	-004	198	001	-021	1000																																
15	046	081	000	-025	-020	-018	-010	081	-035	-051	228	-049	-010	-015	1000																															
16	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000																														
17	-031	136	000	-169	042	-060	-083	-075	117	033	091	-106	-230	091	045	000	1000																													
18	195	115	000	-023	-014	-077	-044	-021	-020	094	-067	103	-044	-063	-031	000	-029	1000																												
19	-089	-201	000	156	084	555	-030	172	250	276	001	074	229	142	-021	000	-170	007	1000																											
20	046	-091	000	-025	-020	-018	-010	081	-035	-051	-032	-049	-010	-015	-007	000	-162	-031	-021	1000																										
21	-050	-065	000	018	-010	190	278	094	158	147	138	073	113	043	196	000	-202	-034	231	196	1000																									
22	-090	114	000	-036	-028	-026	493	-007	-050	204	-045	071	-015	-021	-010	000	-083	145	229	-010	113	1000																								
23	022	130	000	-086	043	064	-063	-083	242	180	-139	-089	086	-089	-044	000	099	147	-128	-044	-030	-063	1000																							
24	-010	-017	000	-051	-040	-037	-021	077	-071	094	-065	102	-021	-030	-015	000	-223	072	327	494	161	340	017	1000																						
25	-001	155	000	-129	-002	020	056	-005	191	-052	-045	042	056	-110	-055	000	027	-082	-090	-055	027	056	102	-110	1000																					
26	-096	046	000	063	226	079	-050	-065	-040	-033	101	026	097	138	177	000	074	-059	039	-041	043	-058	-112	-083	019	1000																				
27	-159	-091	000	-025	370	401	-010	-091	-035	144	-032	-049	-010	-015	-007	000	045	-031	-021	-007	-037	-010	166	-019	-055	-041	1000																			
28	-163	039	000	033	-013	047	049	-0.6	079	026	033	160	-0.2	-023	127	000	051	-037	-033	-057	134	-062	037	-116	037	-044	127	1000																		
29	166	034	000	004	-071	-066	-037	-069	092	-066	041	003	-037	-053	-026	000	100	-112	033	-026	-064	-037	030	-053	-034	-149	-026	-041	1000																	
30	-159	-091	000	-025	-020	-018	-010	081	-035	-051	228	-049	-010	-015	-007	000	-162	-031	-021	-007	196	-010	-044	-015	-055	-041	-007	-037	-026	1000																
31	-160	026	000	-057	132	-041	-024	105	-0.0	061	-073	249	-0.4	-033	-017	000	102	-070	-048	-017	-034	-024	-005	193	047	006	-017	121	-060	-017	1000														
32	-022	-184	000	061	019	034	-044	-021	-005	043	-136	-049	-044	-063	-031	000	-029	-060	-090	236	-034	-044	-020	072	-132	-059	236	001	049	-031	-070	1000												
33	-053	-066	000	039	-042	146	-064	040	182	247	070	-016	083	014	-045	000	107	029	170	-045	010	083	030	118	051	061	-045	-089	-038	-045	-008	-136	1000											
34	-159	-091	000	-025	-020	401	-010	081	208	144	-032	-049	-010	-015	-007	000	045	-031	344	-007	196	-010	-044	-015	-055	-041	-007	127	-026	-007	-017	-031	162	1000										
35	072	028	000	-005	-312	-139	090	-093	090	-007	-090	178	090	-052	063	000	049	174	-075	-115	-091	090	011	-052	107	-107	-115	140	-093	-115	-010	-015	024	-115	1000									
36	028	-084	000	-069	063	-100	100	-034	074	020	-005	-007	-057	031	-040	000	020	006	125	-040	155	100	081	255	-052	014	-040	-032	-145	-040	009	-053	117	-040	-123	1000								
37	014	003	000	-049	017	-043	-071	-051	044	033	035	-024	-071	-101	147	000	065	101	-073	-050	019	-071	067	-101	063	017	147	-033	-003	147	-024	-107	137	-050	024	-050	1000							
38	-039	041	000	-044	-034	-032	-018	-158	080	025	-056	145	-018	-026	-013	000	079	-054	-037	-013	071	-018	-077	-026	-095	055	-013	114	-046	-013	-029	-054	162	-013	111	059	027	1000						
39	-165	-248	000	061	123	256	-044	160	-149	146	-067	-102	-044	072	-031	000	-084	-131	104	-031	028	-044	035	072	-032	057	236	098	-031	-031	051	153	029	-031	-157	-112	049	-054	1000					
40	-081	028	000	229	-045	-041	-024	028	140	-115	045	159	-024	-033	-017	000	-036	051	-048	-017	021	-024	-100	-033	047	105	-017	-130	076	-017	-038	-070	-102	-017	063	009	065	-029	-070	1000				
41	092	-259	000	-020	062	-064	-080	088	-049	-070	-248	-086	183	-114	-057	000	-262	-092	038	129	-073	-040	-032	167	-006	124	-057	-207	019	-057	-044	154	-043	-037	-069	057	121	009	036	040	1000			
42	073	093	000	005	-103	065	037	063	-004	076	-140	-424	-090	-123	-063	000	024	-080	075	115	004	037	026	052	027	-009	-063	-107	-014	-063	-144	-032	160	115	013	033	124	-111	-0.0	-003	069	1000		
43	-005	-093	000	-003	-075	-069	-039	-044	149	147	-045	-012	-039	-056	-028	000	-131	116	026	-028	-073	168	016	092	-097	035	-028	051	-011	-028	-063	-039	071	-030	004	106	-074	122	-039	070	036	0.0	1000	
44	021	003	000	069	-081	040	-042	050	107	058	011	-092	-042	-060	-030	000	128	-054	013	-030	-035	-042	-009	-060	035	010	-030	016	057	-030	-063	-056	042	045	016	-043	010	104	019	1.1	022	033	306	10.0	
45	-056	059	000	068	-049	-045	-026	-012	-038	-127	026	-039	-026	-037	-018	000	-060	-077	-053	-013	101	-026	-110	-037	020	079	-013	047	-066	-018	149	-077	-026	-013	-065	-00	-013	-039	034	119	013	-010	-069	040	1000

*these numbers should be multiplied by 10^{-3}

TABLE 38

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
46	-039*	-058	000	140	-034	212	398	141	080	025	-056	030	-018	-026	-013	000	-162	101	176	-013	207	-018	-077	-026	-095	-072	-013	114	-046	-013	-029	-054	162	573	007	059	-087	-022	-054	-029	009	097	122	108	-032
47	115	-084	000	-063	-049	-045	217	-012	-088	-127	-080	-039	-026	-037	-018	000	-405	-077	099	401	295	-026	-023	357	-058	-103	-018	010	060	-018	-041	034	-026	-018	-065	084	039	-032	145	-041	168	139	-069	-074	-045
48	-045	-014	000	-164	010	-198	053	-054	011	-147	043	-027	053	-043	-196	000	106	-832	-231	037	-135	-113	-067	-043	104	-043	037	-006	134	037	084	096	-058	-196	-116	051	-065	065	-028	-127	116	075	-063	-039	-101
49	-107	-008	000	001	148	-008	-057	-159	-033	063	110	-007	-057	031	-040	000	202	-053	-117	-040	001	-057	173	-081	-010	062	182	009	-012	-040	110	123	071	-040	-093	023	150	059	065	-091	016	-074	042	017	-008
50	-034	020	000	056	-053	113	248	020	-001	163	-087	-056	248	-040	-020	000	-118	-034	054	-020	170	248	-119	-040	216	058	-020	-084	-071	-020	-045	019	038	-020	-036	-023	-135	-034	019	132	134	105	-075	025	275
51	004	006	000	-033	-090	-148	-084	-087	-023	-061	021	-003	045	-028	-059	000	067	086	-040	-059	-303	045	021	064	-070	-217	-059	-001	114	-059	-052	-252	-142	-059	054	-086	-019	-104	038	113	010	-064	193	204	-148
52	-142	-018	000	194	046	060	-037	034	-054	051	041	003	-037	-053	-026	000	-087	-112	144	-028	006	-037	-096	100	030	049	-026	070	-004	-026	078	049	-038	-026	122	-078	116	130	129	215	131	040	-100	057	060
53	-137	-136	000	-090	-042	-026	230	040	-017	007	070	066	-064	014	-045	000	022	029	095	-045	106	093	-272	-091	-027	-029	-045	-051	-038	162	086	-081	-022	162	-013	071	-025	042	-191	086	-081	-024	-050	-071	060
54	-034	-047	000	054	248	113	-028	020	-001	088	-087	021	-028	-040	-020	000	122	019	-057	-020	-101	-028	043	-040	-075	-027	370	058	046	-020	132	226	038	-020	034	-109	093	-034	123	132	-009	-103	-075	-081	-049
55	140	-105	000	-055	-101	-093	-053	-025	-011	-033	-043	073	-053	-075	196	000	-058	028	061	196	081	-053	-030	043	-278	-058	-037	049	147	-037	-084	213	-134	-037	-091	-102	065	-065	151	021	013	-034	-005	-088	004
56	091	-019	000	-057	009	-049	063	014	-110	087	-028	034	-073	-105	-052	000	002	190	-150	141	049	-073	009	090	-134	-292	141	-019	-071	-052	-030	036	038	-052	-054	055	023	022	-015	-030	-009	-107	029	051	031
57	043	-088	000	053	-087	028	-045	221	096	-022	-006	195	139	-065	-032	000	-177	-067	190	-032	198	-045	-139	-065	-045	-068	-032	-062	041	228	045	071	-091	-032	-044	-062	026	-056	002	163	-008	-140	-121	-131	028
58	117	-044	000	001	101	-087	035	077	-053	-001	039	-221	-091	-040	-064	000	105	151	-122	113	-122	161	-093	049	-083	022	-064	-507	-019	113	-065	057	041	-064	-134	-002	-024	-009	-132	096	-040	165	-037	-021	-087
59	086	024	000	-023	-084	-077	-044	-157	-085	-009	071	213	145	-063	-031	000	-029	-060	007	-031	028	-044	035	072	018	-001	-031	098	-112	-031	-070	-080	-081	-031	032	-053	-055	101	081	-070	006	-080	116	019	034
60	-076	-112	000	144	123	145	145	070	044	-009	071	056	145	072	-031	000	-084	-060	298	-031	090	-044	-076	-063	-032	-059	-031	001	049	-031	051	-060	-081	-031	-062	123	-159	-054	081	051	-043	110	038	-054	034
61	-160	-128	000	-057	132	149	-024	-050	-080	150	-073	-021	-024	-033	-017	000	008	-070	-048	-017	-084	-024	090	-033	-124	-093	441	121	-060	-017	-038	051	-008	-017	-099	009	-024	-029	172	-038	-128	018	070	057	-041
62	-113	-128	000	176	097	113	-028	-047	092	038	215	098	248	354	-020	000	042	-084	084	-020	-010	-028	043	-040	-075	058	-020	058	-071	-020	-045	-084	118	-020	034	063	-135	-034	123	132	-009	-172	039	025	-049
63	067	064	000	003	057	-121	012	-111	001	-006	-028	126	012	018	-077	000	123	-100	-037	-077	044	012	035	-069	097	-100	095	-010	031	-077	-097	037	-476	-077	000	057	009	-035	082	059	-028	-092	-042	106	022
64	066	052	000	-486	-009	-112	073	-080	-030	024	078	004	-063	007	052	000	078	-036	-270	052	040	-063	071	-188	134	-001	052	019	-220	052	030	066	-157	-141	029	-055	052	090	-036	-233	-062	-200	027	-104	-031
65	-271	-113	000	054	-053	275	-028	086	-001	088	-087	021	-028	-040	-020	000	-118	-084	225	-020	030	-028	043	157	-075	-027	-020	058	-071	-020	132	019	118	370	-312	-023	-135	-034	123	-045	062	105	039	025	133
66	-036	098	000	069	-062	-127	-072	-034	-151	-052	-022	043	-072	-004	-051	000	-127	043	-147	-051	-123	-072	-104	-103	057	220	-051	064	-183	-051	150	043	-073	-051	096	-279	033	138	-061	-027	037	007	091	058	116
67	058	032	000	-034	-082	-117	-149	003	-043	-191	035	019	-025	052	-105	000	037	-075	-115	-105	-050	-149	-016	-036	-002	014	-105	004	094	-105	079	017	-181	-105	015	-001	-716	020	-029	-001	-105	-045	009	-001	101
68	067	021	000	-035	-090	-083	-047	-410	024	-034	182	078	-047	190	-033	000	048	061	-096	-033	-051	-047	-041	-067	037	033	-033	065	033	-033	-075	061	-204	-033	018	041	-028	-058	-140	040	-116	-063	-051	-135	-083
69	-080	-129	000	188	248	-026	-015	-007	-050	-072	-045	-069	-015	-021	-010	000	-083	-044	-030	-010	-053	-015	-063	-021	056	097	-010	-082	-037	-010	-024	-044	-064	-010	-037	-057	-071	-018	-044	301	183	-090	-039	152	271
70	142	052	000	-111	010	-169	-012	-035	-042	-658	-016	-092	-134	-018	077	000	-052	-082	-150	077	-084	-134	-070	-018	031	-066	-095	041	-083	077	-059	-082	-160	-095	030	-029	-009	035	-082	-215	028	-030	-108	-059	050
71	-114	-010	000	109	157	175	-021	-010	052	094	-065	102	-021	228	-015	000	-014	072	327	-015	-075	340	017	228	-015	027	-015	070	-053	-015	198	-063	118	-015	038	031	-101	-026	072	-033	-114	-038	092	-060	-037
72	-005	106	000	271	039	053	-039	-044	-062	-022	030	-184	-039	092	-028	000	110	-117	026	-028	-073	-039	016	-056	067	-092	-028	051	-011	-028	070	038	-050	-028	-020	-087	-131	-048	116	-063	-105	-239	-104	047	093
73	-013	-055	000	091	-014	010	049	-086	-019	-079	129	305	-082	070	-057	000	-025	-048	-100	-057	-079	-082	-078	-023	018	-084	127	054	-041	-057	121	050	-051	-057	239	-113	039	007	-048	121	-139	-140	-217	-034	010
74	-028	-014	000	112	-017	-013	161	-074	032	067	084	164	035	-130	113	000	032	010	007	-064	125	035	015	-040	017	-055	-064	-117	142	-064	096	057	114	-064	024	-002	-093	094	010	015	-073	-024	-088	-263	-087
75	203	044	000	-155	-025	-154	042	-096	012	048	-152	146	042	-078	-245	000	-015	054	-113	030	-102	042	180	060	-033	-129	030	-117	025	030	-057	126	-155	-245	033	104	043	052	-164	-181	029	-033	-047	-177	-611
76	065	-119	000	045	256	064	-063	-012	-111	-023	-085	-047	-063	335	-044	000	013	-076	024	-044	115	-063	-048	-089	062	161	-044	075	-159	-044	-005	035	-056	-044	011	034	026	-077	202	-100	007	026	-166	-180	064
77	-141	-155	000	200	-049	-045	-026	-012	013	035	-080	043	-026	-037	-018	000	-405	034	-053	-018	004	-026	064	-037	099	-012	-018	010	-066	-018	-041	034	-026	-018	024	-100	039	-032	-077	149	245	-010	296	040	129
78	-080	-007	000	188	-028	-026	-015	-007	-050	-072	139	-069	-015	-021	705	000	-083	-044	-030	-010	113	-015	-063	-021	056	252	-010	049	-037	-010	-024	-044	-064	-010	090	-057	069	-018	-044	301	052	-090	-039	152	271
79	137	152	000	-051	-142	-121	134	-023	-082	-039	060	-009	-110	018	095	000	052	128	-224	-077	-155	012	035	-069	-095	-100	-077	-010	083	-077	-019	-100	-088	-077	061	-009	-024	-035	-054	059	-091	060	058	059	022
80	-050	-078	000	084	011	-176	-100	-018	037	-155	049	-027	023	033	103	000	078	-023	-142	-071	-037	-100	045	-055	025	017	-071	018	059	103	-002	-023	030	-071	-034	-006	027	-022	-023	-002	031	004	087	043	-176
81	-095	-136	000	039	-042	-026	-064	005	033	-153	-037	-097	-064	014	-045	000	022	-026	-056	-045	-230	-064	-093	-091	051	-029	162	-089	-038	-045	-008	303	143	-045	-049	-020	-105	042	-026	-008	-043	049	-050	-015	146
82	318	162	000	-156	-162	-381	-108	-110	-037	-266	-010	-060	-108	-039	038	000	094	101	-471	038	-124	-271	087	-270	029	-033	-191	007	069	-191	-121	-020	-094	-191	-007	-042	-142	-067	-202	-121	-127	-034	078	-031	-000
83	-015	-077	000	156	225	251	-030	110	-102	135	-093	-070	-030	142	-021	000	-245	007	469	344	146	229	-129	512	-090	-040	-021	-033	-077	-021	-048	007	020	-021	-075	044	-002	-037	104	-048	105	140	026	-087	-053
84	-164	-077	000	270	034	403	-030	-014	074	276	001	002	229	142	-021	000	-020	-090	337	-021	061	-030	176	-043	-158	-040	-021	-033	033	-021	-048	-090	095	344	-140	-036	140	-037	104	-048	-029	010	-080	212	-053
85	-226	-155	000	331	275	652	-026	-012	013	197	028	-039	-026	387	-018	000	026	-077	403	-018	198	-026	-023	-037	-136	079	401	163	060	-018	-041	145	146	401	-213	-008	039	-032	256	-041	-064	065	-069	040	-045
86	-039	-045	000	012	-103	-095	108	-084	037	-089	010	015	108	-077	-038	000	-047	-040	-028	-038	018	108	056	033	100	-215	-038	-049	-069	-038	121	-040	047	-038	047	-059	007	067	-101	-087	-042	034	-011	-031	-095
87	002	-045	000	356	073	375	-032	073	058	110	-010	-083	-032	129	-023	000	-074	-004	311	323	045	214	080	304	-039	-052	-023	012	023	-023	-051	088	074	-023	-049	028	-019	-039	088	-051	-047	111	116	096	-056
88	030	010	000	111	-068	-063	-036	-097	-045	-114	-029	018	-036	109	-025	000	-039	-023	-073	-025	-055	-036	045	109	107	132	-025	-198	099	-025	086	-023	-025	-025	-005	001	-049	-044	-023	036	212	-051	-003	069	068
89	-217	-010	000	-051	157	175	-021	077	174	-004	-065	001	-021	-030	-015	000	091	-063	142	-015	043	-021	-083	-030	-110	027	-015	164	-053	-015	-033	-063	118	494	-143	-081	-001	270	072	-033	073	-038	-056	217	-037
90	104	-069	000	004	046	-066	-037	-121	092	-125	041	-057	-037	253	-026	000	-024	-112	-077	277	006	-037	030	100	086	-017	-026	-037	270	-026	-060	049	-100	-026	-040	-078	-003	-046	-112	078	075	-068	-100	057	060

* these numbers should be multiplied by 10^{-3}

TABLE 39

	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
46	1000*																																												
47	212	1000																																											
48	-207	093	1000																																									
49	-070	-100	102	1000																																								
50	-034	-049	010	-109	1000																																							
51	-104	-073	007	-006	-090	1000																																							
52	-046	060	-006	055	163	009	1000																																						
53	162	-026	-050	-020	198	046	-010	1000																																					
54	-034	-049	010	148	097	-090	163	038	1000																																				
55	-065	101	-001	052	-101	035	-134	098	-010	1000																																			
56	022	112	-139	013	-140	013	-107	-121	084	005	1000																																		
57	-096	020	043	-176	114	-073	-037	070	014	238	-028	1000																																	
58	-009	-097	-042	037	-037	-010	-072	114	-017	-081	054	-093	1000																																
59	-094	014	096	123	019	018	-031	-026	019	090	-066	-067	-226	1000																															
60	226	034	-028	006	-084	-010	-031	034	-084	028	-066	071	-085	081	1000																														
61	-029	-041	084	110	-045	113	-060	-008	132	021	145	-073	015	-070	-070	1000																													
62	-034	-049	010	-109	-093	121	-071	-122	097	-010	-065	-087	-106	123	-045	1000																													
63	-135	022	194	-009	-010	180	-124	018	-010	005	-093	060	-014	128	059	-010	1000																												
64	-134	049	139	157	-009	-137	-103	062	-004	-005	145	028	014	-036	-138	-058	-233	-045	1000																										
65	192	-049	-050	-023	097	051	046	018	-093	-010	009	014	-037	019	019	133	-093	-209	-233	1000																									
66	025	035	-012	-022	013	-202	-008	-073	-062	-123	-099	079	033	094	-112	-115	-137	027	136	013	1000																								
67	-082	-117	131	-040	121	-002	-064	-002	-082	-050	-042	-095	-002	110	064	079	-082	051	008	-014	045	1000																							
68	-058	-083	-008	153	-090	098	033	005	-090	065	-114	-145	021	061	-073	040	107	126	040	-090	-034	052	1000																				
69	-038	-026	-113	-097	248	045	178	083	-028	-053	-073	-045	035	-044	-044	-024	-028	012	-063	-028	066	-025	-047	1000																			
70	035	193	124	009	-123	038	-031	-124	-190	075	-045	028	-107	054	054	-059	-123	025	111	-057	033	098	090	-012	1000																		
71	-026	-037	-161	-081	-040	064	-053	-031	-040	-075	-105	067	-040	-063	207	-211	157	-069	-090	-040	-004	-036	-067	-021	-018	1000																	
72	-048	-069	073	042	039	-013	165	-170	265	-140	029	030	-140	038	038	070	152	058	-197	-075	-022	009	-051	-039	092	092	1000																
73	007	010	079	-032	-064	-034	014	-013	-013	-017	052	081	-182	050	001	037	130	021	-017	-004	099	004	111	049	073	-023	091	1000															
74	094	-013	-001	037	101	-106	142	005	012	-001	-048	084	059	010	057	-065	-037	-014	116	032	-038	059	111	035	-016	-130	-037	079	1000														
75	-267	074	230	043	-121	-005	-057	-098	081	025	106	060	021	054	-164	-057	-025	001	157	-131	-009	-047	135	-152	012	-078	-126	134	021	1000													
76	-077	-023	-018	-058	124	-131	030	073	124	-079	-111	023	-019	-076	035	-100	124	142	-009	-038	018	-089	011	086	073	017	016	-001	065	009	1000												
77	212	-045	-101	-008	113	003	060	060	-049	004	-049	-000	-087	-077	014	149	-049	022	-112	113	197	028	023	271	-022	-037	-069	010	-013	-154	064	1000											
78	-018	-026	-278	-057	248	-084	178	083	-028	113	-073	-045	-191	-044	-044	-024	-028	-110	-063	-028	066	-025	-047	193	-012	-021	-039	049	035	-147	-063	271	1000										
79	065	032	-044	-123	-010	086	-072	-028	-076	-135	177	-206	016	-094	-145	-019	-010	014	056	-076	027	120	039	-110	-034	-069	-042	147	077	-012	-073	022	012	1000									
80	-022	041	078	224	-191	149	007	-076	011	043	033	-040	053	049	023	-002	011	-005	-066	-124	013	070	-013	-100	123	013	-014	-174	022	004	-136	-031	023	-212	1000								
81	042	-112	038	-065	-187	024	-022	118	-038	078	-037	114	-236	-026	036	-042	-123	-038	033	247	070	-040	003	052	-091	071	139	041	-042	-013	146	033	039	-042	1000								
82	-067	000	071	008	-251	021	-090	-141	-142	142	183	-069	054	040	-020	-121	015	091	212	-162	133	192	-000	-103	143	-155	-122	-035	-067	094	040	-000	-103	130	092	-000	1000						
83	-037	231	-146	-038	-057	-040	234	-131	-057	-023	-010	-093	071	007	104	-048	034	-162	-130	034	-006	-170	-004	-030	-025	127	026	-033	-122	-013	024	-053	-030	-029	-079	-331	-305	1000					
84	176	-093	-146	-036	-057	-040	144	020	-057	-023	-150	-093	-038	201	201	-048	225	025	-200	223	-077	-305	-004	-030	-150	142	132	-033	-058	-113	024	-033	-030	-037	-079	-331	-305	204	1000				
85	212	-045	-198	176	-049	-148	060	060	273	-093	-049	-080	-161	-077	145	149	113	-050	-192	275	-127	-159	023	-026	-236	173	053	-067	-013	-154	152	-045	-026	-193	-031	-026	-286	251	403	1000			
86	087	095	089	093	074	-021	000	094	-013	018	-271	010	-054	202	142	-037	074	-052	007	074	-000	-073	-000	-054	130	039	055	077	140	094	-040	023	-054	-091	106	094	043	-028	130	-095	1000		
87	-039	088	-125	028	-061	067	127	-068	073	-035	-094	-099	106	-096	088	-051	073	-062	-171	073	-023	-209	-102	-032	-056	304	118	-114	-077	-002	030	-036	-032	-100	-030	-139	-197	437	111	375	-039	1000	
88	-044	-043	-018	001	176	138	099	039	094	-033	-057	-029	-222	-023	-107	-057	054	-021	-184	176	008	076	124	188	104	-021	100	149	056	017	111	200	138	003	-025	039	-004	-073	-073	063	136	-070	1000
89	270	-037	-043	-021	-040	064	100	014	-040	-073	-105	-065	-040	-063	198	-040	018	-090	157	-004	-124	-067	310	069	-030	092	-118	049	-217	017	-017	-021	-069	-055	014	-155	-043	142	173	-077	-046	-091	1000
90	-046	060	064	-012	046	114	087	-018	046	006	-129	-116	-078	-112	-112	-060	163	031	-045	-071	-125	-064	262	193	124	100	077	070	035	023	156	186	170	-020	007	024	-000	011	013	060	069	023	479	100	1000

* these numbers should be multiplied by 10^{-3}

TABLE 40

ITEM ANALYSIS

			Selection Ratios				Mean Scores			
C_n	Item	r_b	D*	D_1	D_2	D_3	M*	M_1	M_2	M_3
C_8	12	.345	.35	.20	.35	.10	13.5	11.4	11.4	12.1
	20	.269	.53	.04	.04	.39	12.8	9.2	10.0	11.9
	26	.140	.38	.15	.34	.13	12.6	12.1	11.9	11.9
C_1	1	.384	.09	.83	.04	.02	14.2	12.1	13.4	10.0
	2	.456	.25	.61	.09	.04	13.9	12.0	9.3	12.0
	8	.419	.29	.58	.01	.12	13.6	11.5	9.0	12.4
	28	.378	.38	.25	.32	.06	13.3	11.5	12.0	9.5
C_5	7	.336	.14	.04	.31	.51	13.7	10.8	11.6	12.2
	22	.001	.07	.02	.06	.84	12.2	8.3	11.2	12.4
	23	.165	.58	.20	.17	.04	12.5	11.8	12.4	8.8
C_6	9	.135	.70	.11	.17	.03	12.4	11.4	12.3	10.0
	27	.388	.85	.03	.09	.03	12.5	10.0	11.1	8.7
C_{10}	21	.404	.32	.07	.37	.23	13.5	11.2	11.4	11.9
	25	.329	.55	.24	.19	.02	12.9	11.5	11.5	9.0
	24	.377	.86	.01	.07	.06	12.5	8.0	10.7	10.7
C_2	3	.304	.35	.01	.23	.41	13.1	10.0	11.2	12.0
	17	.546	.08	.80	.03	.09	15.2	12.0	12.2	10.9
	30	.158	.74	.01	.15	.10	12.4	13.0	11.8	11.0
C_4	5	.295	.22	.05	.61	.12	13.4	12.9	11.8	11.5
	14	.405	.46	.02	.10	.40	13.2	12.0	12.4	11.1
	19	.463	.31	.09	.14	.47	13.7	10.8	11.5	11.6
C_7	10	.320	.17	.22	.06	.55	13.6	11.2	12.4	12.1
	11	.334	.48	.20	.30	.02	13.0	11.8	11.4	9.7
	16	.447	.81	.01	.02	.14	12.6	8.0	9.7	10.7
	29	.275	.83	.08	.06	.02	12.4	11.1	11.1	8.7
C_3	4	.425	.17	.10	.01	.72	14.0	11.9	8.0	11.8
	13	.378	.67	.01	.07	.24	12.8	12.5	9.6	11.3
	6	.473	.53	.02	.22	.23	13.2	9.3	10.9	11.3

C_9, containing items 15 and 18, is omitted from this table.

TABLE 41

METHOD OF CALCULATING INTER-POINT DISTANCE CLUSTERS

Variable		Meaning
V_{nxn}	=	A square matrix of inter-member distances
$S^{k}_{n_k \times n_k}$	=	A submatrix of V_{nxn} made up by extracting from V_{nxn} the members of Group K
N_k	=	The number of members in Group K
A_{kk}	=	$\sum_{x=1}^{N_k} \sum_{y=1}^{N_k} S^{k}_{xy} / N_k$
$TSUM_{N_g}$	=	$\sum_{k=1}^{N_g} A_{kk}$
A_{ij}	=	$\sum_{x=1}^{N_i + N_j} \sum_{y=1}^{N_i + N_j} S^{ij}_{xy} / (N_i + N_j)$
D	=	$(TSUM_{N_g}) - (TSUM_{N_{g+1}}) = A_{ij} - A_{ii} - A_{jj}$
		D was minimized in this procedure.

APPENDIX B

THE ADVANCE CLASSIFICATION OF THE ALTERNATIVES IN THE EXPERIMENTAL TEST

APPENDIX B

THE ADVANCE CLASSIFICATION OF THE ALTERNATIVES IN THE EXPERIMENTAL TEST

The discussion which follows presents a detailed item-by-item account of the procedure used in the construction of this experimental test. The format of this discussion involves:

1. Giving the reading selection as it is required.
2. Giving each item in its entirety in the format it was given to the examinees; except that in this discussion the categories of the items and the foils are indicated for the convenience of the reader.
3. The reason for classifying the item by Bloom's Taxonomy, and the foils by the Guidelines as indicated, are given following each item.
4. Items 19 to 24 inclusive form a special case and will, therefore, be dealt with as a unit.
5. In the classification of foils no one item contained, by arbitrary practice, two foils from the same category.

THE EXPERIMENTAL TEST

Directions for Examinees:

Answer all questions in Part I on the basis of the reading selections given.

First Reading Selection

Source: Dexter, Lewis Anthony; The Tyranny of Schooling, N.Y., Basic Books, 1964, p. 1.

Most people in our society at one time or another suffer humiliation, shame, or at least severe apprehension because of one great fear: they are afraid that other people may think that they are stupid. This fear of being regarded as stupid frequently underlies inferiority complexes, self-contempt,

self-depreciation, and despair.

Our society teaches contempt for stupidity and fear of being regarded as stupid through one central institution and its auxiliaries. This institution is compulsory schooling. It is aided by such auxiliary practices as compulsory written examinations for admission to many jobs, intelligence testing, and the like.

1. From the above article we may conclude that if society does not reduce its contempt for stupidity:
(Bloom's Category 4.20)

   A. Emotional problems will continue to be on the increase. (OG)

   *B. The development of creativity will continue to be restricted.

   C. Mutual co-operation will continue to be difficult to obtain. (IA)

   D. Economic power will continue to be confined to a minority group. (Irr)

This item was classified as an analysis (4.20) item on the grounds that it requires the examinee to display "skills in comprehending the interrelationship among ideas." (Bloom: p. 206). The examinee is expected to realize that, contrary to popular myth, creative people do not display "inferiority complexes, self-contempt, self-depreciation, and despair" to the same extent as is found in the population at large. For this reason, the development of creativity and the development of contempt for stupidity would be expected to be inversely related. Since, in the stem, the variable "contempt for stupidity" does not change, it follows logically that the status of any related variable should show no change. If the examinee did not know this relationship, he should have been able to arrive at it from the logic of the foils.

Foil 1A (or $1D_1$, on the basis of the symbolism used in

Chapter V)[8] suggests an increase in one variable without a corresponding increase in the other. The phrase "does not reduce" in the stem, does not validly warrant the conclusion that contempt for stupidity will increase. It is adding incorrect information to the answer to suggest a change in one variable without an explicit statement concerning change in the appropriate direction in the other variable. Hence the foil is classified as an Over Generalization (OG).

Foil 1C (or $1D_2$) assumes a functional link between co-operativeness and contempt for stupidity. Unlike the case for creativity, there is no valid reason to assume such a relationship for co-operativeness. Hence this foil involves an Invalid Assumption (IA).

Foil 1D (or $1D_3$) assumes a functional relationship between contempt for stupidity and the confining of economic power to a minority group. For this reason this foil could have been an Invalid Assumption (IA) except for the arbitrary rule used for classifying foils which allows only one foil in a category per item. On the other hand, the concentration of economic power for the purpose of maintaining economic institutions is a practical necessity independent of how the society treats the individual or how the decision makers are chosen so that this statement is true but irrelevant to the problem. Hence this foil was classified as an Irrelevancy (Irr).

2. Which of the following is the most important causitive factor of contempt for stupidity? (Bloom's Category 4.10)

---

[8] In this code used in Chapter V the subscript stands for the first distractor (foil) in item 1. The code gives a standard procedure for identifying foils without concern for which alternative is the right answer.

*A. Compulsory categorizing in school.

B. Compulsory school attendance. (SUB)

C. Compulsory written examinations. (OS)

D. Compulsory intelligence testing. (Irr)

Item 2 is asking for the "causative factor" which is not stated in the selection. This item, therefore, requires the examinee to demonstrate his "ability to recognize unstated assumptions" (Bloom: 1956, p. 205), hence the classification of this item as analysis (4.10).

Compulsory school attendance is an enabling factor in this situation, but it is neither necessory nor sufficient. In fact, compulsory school attendance is disjunctively related to the development of contempt for stupidity which can develop in universities where attendance is not compulsory. Also, contempt for stupidity need not develop in a compulsory school system. Any attempts at the institutionalization of an individual can lead to the development, on the part of the individual, of contempt for the forms of behavior considered "stupid" by that institution. (It can also have the opposite effect). In any event, the replacement of this term "any" by the term "compulsory" makes this a Substitution (Sub) foil.

With respect to foil 2C ($2D_2$) the cause of contempt for stupidity is the "pass-fail syndrome" which compulsorily classifies a certain proportion of the population as "stupid." In this case, it is not the written examinations but the use to which they are put which leads to contempt for stupidity. Furthermore, contempt for stupidity can (and does) develop in the classroom context at times other than during examination writing by the tacit acceptance by a student's peers of the assumption made by the teacher that making a mistake is "sinful."

There would be no need for compulsory examinations if there were no attempt to make classifications. However, written examinations by themselves do not cause contempt for stupidity, hence this foil is an Oversimplification (OS).

In the case of compulsory intelligence testing (Foil 2D or $2D_3$) the issue is whether or not the results of these tests are used as part of the compulsory classification system rather than whether or not the tests are given. Therefore, this foil as stated is an Irrelevancy (Irr).

3. The school acts as an agent for the continuance of contempt for stupidity by: (Bloom's Category 2.10)

   A. Placing too much emphasis on success in extra-curricular activities. (Sub)

   *B. Reflecting the attitude that personal worth is at stake.

   C. Encouraging competition between students of unequal ability. (Irr)

   D. Stressing knowledge as the only means to success. (OS)

Dexter's approach to the schools is essentially upon an emotional level. In attributing a person's inferiority complex to the school his implication is that the basic strength of contempt for stupidity is in its reflection upon self-esteem. This item tests the examinee's "ability to understand nonliteral statements (exaggeration)" (Bloom: p. 204). The classification is 2.10.

In foil 3A ($3D_1$) the phrase "in extracurricular activities" would have to be replaced by the phrase "in academic pursuits to the exclusion of success in self-corrective activities," to be correct. This foil contains a Substitution (Sub).

Competition between students of unequal ability can lead to the continuance of contempt for stupidity provided that the purpose of the competition is to make the less able appear stupid. It is the

objective and not the fact that is critical. The fact itself is irrelevant; therefore, this foil is an Irrelevancy (Irr).

In foil 3D ($3D_3$) the critical aspect of this statement is "to the exclusion of success in self-corrective activities." Hence this foil is an Oversimplification (OS). Notice that foils 3A ($3D_1$) and 3D ($3D_2$) are both related to a "correct" answer which is not given in this item. It would seem perfectly legitimate when more than one right answer is possible for a particular item to use alternative right answers for the generation of foils.

4. The author, in charging that "society teaches contempt for stupidity and a fear of being regarded as stupid" by means of the school, is assuming that: (Bloom's Category 4.20)

   A. The school should not be an enforcing arm for the customs of society. (OG)

   *B. The school is a more powerful socializing force than the home.

   C. The home is a more powerful socializing force than the school. (Inv)

   D. The school is an enforcing arm of the customs of society. (Irr)

This item asks the examinees to "recognize a hidden assumption" (Bloom: p. 206) which makes this an analysis (4.20) item.

If the school is at fault it must have more influence on the child than the home has on the child. Foil 4C ($4D_2$) must be an Inversion (Inv) by virtue of being opposite to the correct answer.

Foils 4A ($4D_1$) and 4D ($4D_3$) are related since both contain the same irrelevant premise. However, 4A ($4D_1$) also contains the additional unwarranted value judgment "should not be." By virtue of the rule which excludes category repetition, it becomes reasonable, at least superficially, to classify 4D ($4D_3$) as an Irrelevancy (Irr) and 4A ($4D_1$) as

an Overgeneralization (OG). In the case of 4A ($4D_1$), however, this Overgeneralization is an unreasonable extension of a statement which, in itself, is incorrect, suggesting that second thoughts might have led to a more reasonable classification of this foil, as the results of subsequent analysis showed.

Second Reading Selection

Source: Marris, Peter; The Experience of Higher Education, London, Routledge Kegan Paul, 1964, p. 175.

In this sense, it does not matter what subject a student studies, since each is leading towards a generalized intellectual awareness. But the starting point is still important since a student has the greatest incentive to understand whatever relates most immediately to his interests. Nor are the concepts derived from any one field of study equally relevant to any others: the ramification of insights remains biased by its roots. The intellectual content has to both guide and be guided by the purposes for which a student seeks understanding. Otherwise it is meaningless.

If, then, higher education aims to teach students how to abstract, from a particular context, principles by which they can organize the perception of their universe of thought, it requires that these students have a use for such free-ranging understanding. When they enter higher education, their aims are confused, and they may not see, or wish to see, the value of a generalized intellectual skill. Their approach to learning has been conditioned by extraneous motives: they worked to win approval or avoid blame, to pass an examination, as much as or more than for the sake of understanding. They are not used to asking themselves what they want to understand, or why, but derive enough interest to master the skills required of them from a desire to satisfy the authority who sets the task. So, I think, the function of higher education is as much to develop the autonomy of their desire to understand, as to satisfy it.

5. The author suggests that a generalized intellectual awareness can be achieved by: (Bloom's Category 3.00)

   A. Focusing on progressively more difficult topics in a subject. (IA)

   B. Teaching the students how to generalize from specific content. (CM)

C. Presenting highly abstract material which is extensive in scope. (OS)

*D. Presenting any subject matter in any predetermined sequence.

This item was treated as an application (3.00) item because it requires the examinee to make "use of abstractions in particular and concrete situations." (Bloom: p. 205). For this reason this item was classified as application (3.00). The right answer is also, in fact, an Oversimplification (OS) because Marris says "the starting point is important" (see p. 47). However, the use of this phrase in 5D would have produced a "Clang association" which Thorndike and Hagen (1961) point out should not be used. (See p. 28). Alternative 5D, is, nonetheless, the most nearly correct of the four alternatives.

The first foil (5A or $5D_1$) assumes that 1) some specific subjects are needed for the development of generalized intellectual awareness and that 2) instruction must, of necessity, begin with the "easier" topics first. Both of these assumptions are explicitly stated as invalid in the selection. This foil was classified as an Invalid Assumption (IA).

The relationship between generalized intellectual awareness and inductive reasoning as suggested in 5B ($5D_2$) is a very common oversimplification. With another oversimplification foil in the same item the classification of this foil is a Common Misconception (CM) would seem to be quite reasonable.

Similarly, the identification of generalized intellectual awareness with the transferability of content is an oversimplification of the topic of Marris' (1964) discourse. Therefore, foil 5C ($5D_3$) is classified as an Oversimplification (OS).

6. The purpose of developing a generalized intellectual awareness is to: (Bloom's Category 2.30)

   *A. Promote thinking ability which is not contextually bound.

   B. Enable an individual to master any subject area. (OG)

   C. Stimulate thinking ability within the individual's chosen field. (Sub)

   D. Give the individual an ever-widening view of his world. (Irr)

In order to answer this question, the examinee is expected to make an "extension of trends or tendencies beyond the given data." (Bloom: p. 205). On this basis this item was classified as Comprehension (2.30).

In foil 6B ($6D_1$) the mastery of "any subject area" is far too broad a statement for the purpose of Marris' (1964) discussion. Hence this foil is an Overgeneralization. (OG)

In the case of 6C ($6D_2$) the phrase "within the individual's chosen field" is substituted for the phrase in the correct manner "which is not contextually bound" making this a Substitution (Sub) foil.

For foil 6D ($6D_3$) the absence of context in the right answer renders the Weltanshauung (World view) aspect of this foil irrelevant, hence 6D ($6D_3$) was classified as an Irrelevancy (Irr). This foil could also be a Word-Word Link (WW) because of similarities in phrasiology between "ever widening view" and "free ranging understanding" (see p. 152).

7. Of the following, the best example of generalized intellectual skill is: (Bloom's Category 3.00)

   A. Thinking within the confines of particular subject areas. (Sub)

B. Generalizing from the concrete to the abstract. (WW)

*C. The widely applicable technique of logic.

D. Applying abstract principle to new situations. (OS)

The phrase "best example" in the stem led to the classification of this item as an Application (3.00) item.

There are strong similarities between this item and the previous two. For instance, the "particular subject areas" phrase is similar to the "individual's chosen field," in item 6 so that foil 7A ($7D_1$) should also be classified as a Substitution (Sub). With respect to 7B ($7D_2$) the use of induction is similar to 5B ($5D_2$). In this case, however, the Word-Word Link between "generalized" in the stem and "generalizing" in the foil is somewhat stronger because of the context than in item 5. Hence 7B ($7D_2$) was classified as a Word-Word Link. (WW)

Also, the confusion in equating transfer of training with generalized intellectual awareness found in 5C ($5D_3$) reappears in 7D ($7D_3$) which makes it reasonable to regard this foil as an Over-simplification (OS) as well.

Third Reading Selection

Source: Kagan J. and Moss, H. A.; Birth to Maturity, N.Y., Wiley, 1962, p.85.

Aggression is a second behavior system that begins its growth during the first five years. Traditionally a response was labeled aggressive if the goal of the behavior was assumed to be psychological or physical injury to a person or person surrogate. We have adhered to this definition. As with dependency the display of aggressive acts is a regular concomitant of development. The slapping or pushing of an age mate, the destruction of a sibling's new fort, and the stinging verbal attack are regularly observed in the behavior of many children.

Aggression, like dependency, is subject to socialization pressures, for the child does not have complete license to unleash his anger when he chooses. In addition, as with

dependency, the occurrence of overt aggression is a function of both the threshold for motive arousal and the intensity of anxiety associated with direct expression of this behavior.

In contrast to dependency, however, the potential for conflict over aggression is greater for females than for males. The pattern of social rewards and traditional sex-role standards act in concert to discourage the direct expression of aggression in girls and women. It might be anticipated, therefore, that aspects of aggression would be more stable for males than for females. This is precisely what occurred, for overt aggression to mother and frequent tantrums during childhood predicted adult aggressivity for men but not for women.

8. If the school were to encourage tolerance for honest mistakes, we would expect aggression to: (Bloom's Category 4.10)

   A. Diminish somewhat. (IA)

   *B. Take different forms.

   C. Disappear completely. (OG)

   D. Remain unchanged. (Inv)

Although this item makes a reference to the first reading selection (Stupidity), the question can be answered within its own context. For this reason this item was not classified as synthesis but as analysis (4.10). The logic of this item revolves around the assumption of the author that aggression is an innate characteristic of human beings which can be modified but not diminished. Thus 8A ($8D_1$) can only be true if this assumption is violated. Foil 8A ($8D_1$) would seem to be an Invalid Assumption (IA) foil in the sense that the examinee must make an invalid assumption to select this answer as "correct." Foil 8C ($8D_2$) strongly overstates the same error as found in 8A ($8D_1$). For the same reason as 4A ($4D_1$) this foil was classified as Overgeneralization (OG).

Changes in the psychological climate will lead to changes in the modes of expression of aggression which makes foil 8D ($8D_3$) an Inversion

(Inv).

9. The basic position of the author in writing about aggression is that it: (Bloom's Category 4.20)

*A. Is inevitable but can be direct through socialization.

B. Can be eliminated through the process of socialization. (Inv).

C. Will result in internal conflict independent of the environment. (Sub)

D. Is crippling to the individual by wasting considerable energy. (CM)

This item requires the recognition of the assumption of the authors which was indicated in relation to item 8. For this reason the item was classified as analysis (4.20).

Foil 9B ($9D_1$) is an Inversion (Inv) for the same reason as 8D ($8D_3$).

Aggression, as an innate behavior system, will produce internal conflict only if its modes of expression are frustrated. Hence, internal conflict is "dependent upon environmental conditions" rather than being "independent of the environment." Since there is already an Inversion (Inv) foil in this item, another category had to be found for this foil. Comparing the two statements "dependent upon..." and "independent of..." the latter phrase can be treated as a replacement for the former. Therefore, this foil 9C or $9D_2$) was classified as a Substitution (Sub).

Aggression can be harmful, but as one basis for intrinsic motivation it can be constructive as well. Foil 9D ($9D_3$) by treating aggression as being exclusively harmful oversimplifies the situation. This oversimplification is so commonly held that this foil was classified as a Common Misconception (CM).

10. With which of the following statements concerning aggression would the author be most likely to agree? (Bloom's Category 6.10)

A. Aggression is like dependency in that it is harmful to personality development. (CM)

B. Aggression generally interferes with the attainment of educational goals. (Inv)

*C. Aggression is potentially useful for educational purpose.

D. Aggression is considered to be a response to threats to a person or person surrogate. (WW)

This item asks the examinees to evaluate the statements made in the alternatives against the information given by the authors about the topic, therefore this item was classified as Evaluation (6.10).

Foil 10A ($10D_1$) was classified as a Common Misconception (CM) on the same basis as foil 9D ($9D_3$).

As pointed out with reference to item 9, aggression can be one of the bases for intrinsic motivation. Hence the possibility that aggression may be potentially useful for educational purposes may be inferred from the selection. In this case the opposite statement as in 10B ($10D_2$) must be an Inversion (Inv).

Foil 10D ($10D_3$) is best classified as a Word-Word Link (WW) on the basis of the phrase "person or person surrogate." This foil is wrong because it contains the phrase "a response to" which is extraneous to the definition of aggression. (see: p. 51)

11. Overt aggression would likely be decreased by: (Bloom's Category 3.00)

A. Blocking of many modes of aggression. (Inv)

B. Lessening the threat of punishment. (OS)

*C. Increasing the threshold of motive arousal.

10. With which of the following statements concerning aggression would the author be most likely to agree? (Bloom's Category 6.10)

A. Aggression is like dependency in that it is harmful to personality development. (CM)

B. Aggression generally interferes with the attainment of educational goals. (Inv)

*C. Aggression is potentially useful for educational purpose.

D. Aggression is considered to be a response to threats to a person or person surrogate. (WW)

This item asks the examinees to evaluate the statements made in the alternatives against the information given by the authors about the topic, therefore this item was classified as Evaluation (6.10).

Foil 10A ($10D_1$) was classified as a Common Misconception (CM) on the same basis as foil 9D ($9D_3$).

As pointed out with reference to item 9, aggression can be one of the bases for intrinsic motivation. Hence the possibility that aggression may be potentially useful for educational purposes may be inferred from the selection. In this case the opposite statement as in 10B ($10D_2$) must be an Inversion (Inv).

Foil 10D ($10D_3$) is best classified as a Word-Word Link (WW) on the basis of the phrase "person or person surrogate." This foil is wrong because it contains the phrase "a response to" which is extraneous to the definition of aggression. (see: p. 51)

11. Overt aggression would likely be decreased by: (Bloom's Category 3.00)

A. Blocking of many modes of aggression. (Inv)

B. Lessening the threat of punishment. (OS)

*C. Increasing the threshold of motive arousal.

D. Motivating people to rise above their peers. (RT)

This item was classified as an application (3.00) item because it asks for a practical method of behavior change.

Blocking of modes of aggression (11A or $11D_1$) would be expected to intensify responses in the remaining available directions, hence this would not necessarily decrease overt aggression. This foil was classified, therefore, as an Inversion (Inv).

If the threat of punishment is lessened, overt aggression may or may not temporarily <u>increase</u>, depending upon the amount of frustration which has previously developed and the way in which the threat is lessened. If the release leads to an increased frustration the increase in overt aggression would continue. On the other hand, if lessening threat also lessened frustration <u>and</u> provided for alternative modes of expression, overt aggression could decrease. Foil 11B ($11D_2$) must be considered an Oversimplification (OS) in this context.

Rising above one's peers can involve the use of aggression as intrinsic motivation but it can also be accomplished by the use of overt aggression or the threat of its use(i.e. intimidation). The term "motivating" in this sense refers to "behavior modification" rather than motivation in the sense used by psychologists. Foil 11D ($11D_3$) was classified as a Redefinition of Terms (RT).

12. Aggressive behavior in female children is:
(Bloom's Category 2.30)

*A. More likely to produce guilt feelings than in males.

B. A less likely occurrence than in males of the same age.(CM)

C. More unpredictable and is expressed differently than in males of the same age group. (OG)

D. Less differentiated in expression than in males of the same age. (Inv)

Once again the examinee is expected to go beyond the given information in order to recognize the role of guilt in child rearing practices. For this reason this item was classified as comprehension (2.30).

Since aggression in males and females tends to take different forms because of sex differences in child rearing practices, guilt is more likely with females. Both sexes show aggression. The difference in mode of expression leads to the common misconception that girls are less aggressive than males, hence the classification of 12B ($12D_1$) as a Common Misconception (CM).

The first two words in 12C ($12D_2$) make this statement false. The lack of predictability is from childhood to adulthood and not across peer groups. This foil is therefore classified as an Overgeneralization (OG).

Foil 12D ($12D_3$) is exactly opposite to the true state of affairs making this foil an Inversion (Inv).

13. If we assume that the school increased its use of contempt for stupidity as a motivating device, we would expect that: (Bloom's Category 5.30)

    A. Parental pressure would intervene to prevent the school from making this change. (IA)

    *B. Overt aggressive behavior would increase and autonomous thinking would decrease.

    C. Both academic success and generalized intellectual awareness would increase. (CM)

    D. The level of student motivation would decrease rather than increase. (RT)

This item is classified as a Synthesis item (5.30) because it

involves more than one reading selection in order to achieve the answer.

Foil 13A ($13D_1$) involves the assumption that parents would oppose this move. However, contempt for stupidity could not be used at present if it did not have at least support by implication from parents at the present time. The major supporters of the school system, the middle class, want to keep the "riff-raff" out of the professions as unwanted competition for the aspirations they have for their own children. Contempt for stupidity as Dexter (1964) implies is an extremely effective method of destroying the academically unfit. It is unlikely that powerful parents would oppose this move. This foil was, therefore, classified as an Invalid Assumption (IA).

Contempt for Stupidity has the effect of maintaining the dichotomy between the academically successful and the others. Increasing this pressure would sharpen the dichotomy and would not necessarily increase the academic success of the survivors and would most certainly not increase the academic success of those who did not survive. On the other hand, the overt effect of this increase would be to produce an apparent increase in academic standards which would be expected to lead to the common misconception that making school achievement more difficult improves the quality of schooling. For these reasons foil 13C ($13D_2$) was classified as a Common Misconception (CM).

Foil 13D ($13D_3$) redefines motivation in the narrow sense of positive intrinsic motivation (i.e. interest). The use of contempt for stupidity is, in fact, increasing the level of extrinsic aversive motivation. This foil was classified, therefore, as a Redefinition of Terms (RT) foil.

14. Which of the following best describes the probable relationship between contempt for stupidity and generalized intellectual awareness? (Bloom's Category 5.30)

A. Changes in either will have no effect on the other. (Irr)

*B. As one increases the other will decrease.

C. Either will increase with an increase in the other. (Sub)

D. Contempt for stupidity should be reduced and awareness should be increased. (OG)

This item, once again, involves two selections (Stupidity and Awareness). For this reason it was classified as a synthesis (5.30) item. A person who is motivated by contempt for stupidity (his own and others') would be expected to be constantly en guarde against making mistakes. Such an orientation toward his own behavior would tend to make him intellectually cautious and hence less inclined to the expansive thinking needed to develop a generalized intellectual awareness. These two variables would be most likely to be inversely related thus explaining the correct answer.

Treating these two variables as unrelated as in Foil 14A ($14D_1$) is contradicted by the information in these two passages. This statement could be true in another (independent) context, hence the Irrelevancy (Irr) classification of this foil.

One part of the statement in foil 14C ($14D_2$) is correct, the other incorrect, hence this foil was classified as a Substitution (Sub) foil.

Foil 14D ($14D_3$) contains an unwarranted value judgment when only the relationship and not its psychological importance is asked for. This foil was classified as an Overgeneralization (OG).

The classification of the foils in this item is difficult because three of the foils are based to a large degree upon logical relations rather than errors in logic. The three possible relationships direct (14D) inverse (14B) and unrelated 14A) form the basis for three of the four foils. It might have been more reasonable to have classified 14A ($14D_1$) and 14C ($14D_2$) as Other (O) than to attempt to establish classifications on the basis of the rather tenuous arguments given here.

Fourth Reading Selection

Source: Dinkmeyer, D. C.; Child Development, Englewood Cliffs, N. J., Prentice-Hall, 1965, p. 59.

The social studies committees were working on their reports. Doris was chairman of the southern states committee which included Jack, Susan, and Bill. There seemed to be confusion in this group so I decided to investigate. "Jack won't co-operate," complained Susan. "What do you want him to do?" I asked. Jack was frowning. "They say I have to study economic conditions in the states, and I am interested in state capitals," said Jack. "Did you volunteer to take economic conditions?" I asked. "There wasn't a chance to volunteer. We were just told her plans," answered Jack. "Is anyone investigating the state capitals?" I asked. The children indicated this job had not been assigned. "In that case, does the group mind having Jack study the capitals?" No one seemed to care. "What about the rest of you--are you all satisfied with your jobs?" They were. Jack went to the reference shelf and started to read.

"B. H."

15. From this report we may infer that the:
(Bloom's Category 2.20)

A. Classroom is very well equipped with instructional materials. (OG)

*B. Classroom probably has moveable seats.

C. Class is studying the Southern United States. (WW)

D. Teacher favours voluntary participation. (RT)

This item involves a "reordering, rearrangement, or new view of

the material (Bloom: p. 205) which explains its classification (2.20) level.

Foil 15A ($15D_1$) overstates the situation in the phrase "very well equipped" proposing an inference which goes far beyond the information in the passage than is reasonable. This foil was classified, therefore, as an Overgeneralization (OG).

In foil 15C ($15D_2$) only one committee and not the whole class seems to have been studying the southern United States. Since this item already has an OG foil, the next most reasonable is a Word-Word Link (WW) on the grounds that careless reading might lead to this word association.

In foil 15D ($15D_3$) the term voluntary is used in more than one sense. Actually, the teacher has substituted her own arbitrary decision for Doris' decision. This foil could also have been a WW foil except that 15C ($15D_2$) fits this category better. For these reasons foil 15D ($15D_3$) was classified as Redefinition of Terms (RT).

16. If the teacher had written "Doesn't work well with others," as an anecdotal record for the above incident, this would have been: (Bloom's Category 6.10)

    A. Better; it says the same thing with less words. (Irr)

    *B. Worse; it fails to indicate the circumstances of the incident.

    C. Better; the details of the event are unnecessary when judging Jack's behavior. (OG)

    D. Worse; teachers are failing in their obligations in not supplying complete information. (CM)

This item clearly asks for a value judgment based upon the evidence in this reading selection which makes it an Evaluation (6.10). item.

Since it should be evident by comparison of the two alternative descriptions given for this same incident that the function of an anecdotal record is to give a clear picture of an event for future reference, a goal of "saying the same things in less words" is irrelevant to the task at hand. For this reason foil 16A ($16D_1$) was classified as an Irrelevancy (Irr).

Adding unwarranted statements concerning judgment of behavior makes this foil, 16C ($16D_2$), an Overgeneralization (OG).

Once again, inappropriate value judgments are involved in 16D ($16D_3$), this time directed at the teacher rather than the student. This attitude is so common that this foil was treated as a Common Misconception (CM).

17. From the above passage we can infer that Doris' leadership of the group was: (Bloom's Category 4.20)

*A. Coercive.

B. Autocratic. (OG)

C. Destructive. (Inv)

D. Laissez-faire. (Sub)

In item 17 the examiner is expecting the examinee to "comprehend the interrelationships among ideas in a passage (Bloom: p. 206)." Hence the analysis (4.20) classification.

On the basis of the argument that the successful autocrat would not tolerate contradiction and therefore have no overt objection to his or her decisions, Doris' leadership was considered as "coercive" rather than "autocratic" for the best answer. Notice, by the way, that the teacher is a successful autocrat in this passage. For these reasons foil 17B ($17D_1$) was regarded as an Overgeneralization (OG).

It is not evident from the passage that Doris' leadership was destructive. In fact, she apparently had the support of the two members of the committee one of whom reported the problem to the teacher. Being opposite to a possible best answer, foil 17C ($17D_2$) was classified as an Inversion (Inv).

Since Doris' attempts to coerce Jack were ineffective, she permitted the teacher to intervene. As a result, her later leadership was laissez-faire, but only under the arbitrary intervention of the teacher. The replacement of her later performance for her former performance led to the classification of this foil (17D or $17D_3$) as a Substitution (Sub).

Once again, the classification of these foils is tenuous and open to disagreement. The format of these foils also deviates from the usual format of foils in this test, as in the case of item 13 and items 19 to 24 inclusive. It is possible that an "Other" (O) classification of these foils would have been more reasonable.

18. From the description of the incident, we can conclude that the teacher's handling of the incident was:
(Bloom's Category 6.10)

    A. Good; she intervened to prevent a serious conflict from continuing. (Sub)

    *B Poor; she allowed Jack to use her authority as a lever to get his own way.

    C. Good; she resolved the problem to the mutual satisfaction of the group. (Irr)

    D. Poor; she failed to collect sufficient information before enforcing a decision. (OG)

This item created a good deal of consternation upon its first administration. At issue seemed to be philosophical differences between the examiner and the examinees. It is probable that this problem may be

a complication which is possibly inherent in any multiple choice Evaluation (6.10) item where the evaluative criteria is not supplied. The problem arose essentially because many examinees insisted that the function of the teacher was to prevent or to eliminate conflict. In this case either 18A or 18C would be correct depending upon the interpretation given to the phrase in the reading selection "no one seemed to care."

The examiner, on the other hand, took the stand that the function of the teacher is to educate. If conflict arises the conflict should be used in an educational manner. In this case, the teacher should have found out why the topic of economics was important enough to the group to have engendered the conflict. Once Jack understands its importance he may agree to do it. That is, the teacher helps to improve communication. If, on the other hand, Jack remains adamant, forcing him to do something disagreeable to him may not help. In this case, the reorganization of committees more nearly upon sociometric lines might improve the situation. There may be some personality reasons for Jack's behavior. In this case, the teacher's long-term role is to help Jack cope with his own and others' personalities. Letting Jack have his own way does not help meet this latter goal. Hence the keyed answer.

Some of the examinees argued that they had insufficient information to answer this question. This argument was discounted because the differences still seemed to be philosophical. In any case, the few people who chose the keyed answer had the highest average total-correct score which meant the retention of this item.

The prevention of conflict rather than the educational use of conflict led foil 18A ($18D_1$) to be classified as a Substitution (Sub).

With education as a goal, the mutual satisfaction of the group (as in 18C or $18D_2$) is irrelevant, hence the foil was considered an Irrelevance (Irr).

The classification of 18D ($18D_3$) as an Overgeneralization (OG) is somewhat arbitrary. The teacher could have sought more information, but the essential problem is her use of the information which she obtained. Since 18C ($18D_2$) was already classified as an Irrelevancy (Irr) some other classification is needed.

Fifth Reading Selection

Source: Prescott, D. A.; The Child in the Educative Process, N. Y., McGraw-Hill, 1957, pp. 125-126.

Progress Report

X Attendance Area Y County Schools

Name: Chester M Teacher: Miss C. Grade: 6
Days Absent: 0 Days Tardy: 0

Reading: Is reading independently on the third-grade level and instructionally on the fourth-grade level. Does not enjoy reading. Finds many excuses to leave reading to do something else. Has trouble understanding what he reads. Is better able to find facts than to interpret facts. Has trouble finding words in context when meaning is given.

English Language: Has a wide speaking vocabulary. Uses correct English. Does not enjoy story writing. Understands sentence construction.

Spelling: Learns words in spelling lessons and uses them in written work. Enjoys spelling.

Writing: Spaces words well. Is practicing again on the formation of letters. Is not neat in written work. Erases often.

Arithmetic: Has worked again this year on addition, subtraction and multiplication. Had some trouble with subtraction. Is not ready for division. Has had experience with problem solving. Enjoys arithmetic.

Social Studies: (History, geography, and civics). Has worked

with maps. Takes part in discussion. Showed interest in a study of his community. Shared materials. Is trying for a better relationship with classmates.

Science: Experimented with the force of air. Has become interested in cloud formation. Likes dogs.

Music and Art: Listens to music. Takes part in singing and rhythms. Enjoys all phases of music. Works with clay, wood, paints, and fingerpaints. Enjoys all media of art expression.

Instruction for Questions 19 to 24

Based on the above progress report answer the next six questions by marking:

A. If the hypothesis is supported by the facts.

B. If the hypothesis is implied by the facts.

C. If the hypothesis is refuted by the facts.

D. If the hypothesis cannot be tested by the facts.

(Bloom's Category 4.20)

Hypotheses to test:

19. Chester is not liked by the other children; he avoids trying to read because he doesn't want them to see him fail.

20. Chester lacks character. He does all sorts of bad things and will not discipline himself to learn to read because he has not been punished enough.

21. Chester is growing very slowly and really is quite immature for his grade. Everyone expects too much of him.

22. Chester has no real reason to want to read, since no one ever reads at home.

23. Chester's reading deficiency has not yet begun to affect seriously his performance in other areas.

24. Chester's mother has kept after him about reading until he hates it.

All six of these items were classified as Analysis items (4.20) because of the hypothesis testing characteristics of their format. This format was used as a marker for analysis subtests. However,

because of the format, the classification of the foils became problematic. The simple expedient was used of classifying all the foils for these items as Other (O).

25. The most useful suggestion to help Chester is: (Bloom's Category 4.20)

A. To give Chester personal warmth, acceptance and support wherever it is appropriate. (Sub)

*B. To give Chester concrete help in getting started on specific tasks, especially in reading.

C. To give Chester responsibilities and roles of acknowledged importance in the daily life of the classroom. (Irr)

D. To try to get Chester's mother to take the pressure off him and offer him more opportunities for self-direction. (IA)

The examinee is expected to comprehend interrelationships in the answering of this item, hence its analysis (4.20) classification. Foil 25A ($25D_1$) substitutes emotional support for corrective instruction, hence the Substitution (Sub) classification of this foil.

The treatment suggested in 25C ($25D_1$) has no bearing on his academic needs; it was therefore classified as an Irrelevancy (Irr).

There is no evidence in this selection that there is an unreasonable pressure on Chester by his mother; hence this foil 25D ($25D_3$) involves an Invalid Assumption (IA).

26. If additional information on Chester is desired, and none of the following had been attempted, which one would provide the greatest amount of immediately useful information? (Bloom's Category 5.20)

A. An interview with Chester's previous teacher. (O)

*B. An interview with the parents.

C. A diagnostic test in reading skills. (OS)

D. A request for the assistance of a guidance counselor. (Irr)

This item involves the examinee generating a structure to represent Chester's entire situation by inductive reasoning prior to answering the question. For this reason, this item was regarded as a synthetic (5.20) item.

The best first hand source of information about Chester is his parents. The next best is his previous teacher. Since the Guidelines do not make any provision for this kind of relationship, foil 26A ($26D_1$) is best classified as "Other" (O).

A diagnostic test in reading skills is only useful to a teacher who know enough about these tests and the reading problems they diagnose to be able to use them effectively. Also, administering and interpreting such tests can be time consuming. It is an Oversimplification (OS) for foil 26C ($26D_2$) to suggest this course of action to be superior to any other.

Most of the examinees were experienced teachers, hence it was reasonable to assume that they would know from experience that guidance personnel rarely can give a teacher information they do not already know. This effect occurs because test batteries are rarely more reliable than a month or two of sensitive observation by a teacher. Therefore, the course of action suggested in foil 26D ($26D_3$) is classified as Irrelevancy (Irr).

27. A reasonable conclusion which can be drawn from this report is that: (Bloom's Category 4.10)

    A. Chester's problem stems essentially from his poor relationship with his mother. (IA)

    B. Chester's problem stems essentially from his poor relationships with his peers. (Irr)

*C. Chester's problem has no single cause and no simple solution.

D. Chester's problem stems from such a wide range of sources that a classroom solution is impossible. (CM)

This item is somewhat difficult to classify because the "drawing conclusions" is not part of Bloom's Taxonomy. However, this item also involves the examinee's "skill in distinguishing facts from hypotheses" (Bloom; p. 205), hence the analysis (4.10) classification of this item.

Once again, the mythical poor relationship with his mother is introduced in foil 27A ($27D_1$). In this context the most reasonable classification of this foil would be an Invalid Assumption. (IA).

Chester's problem seems to be centered upon his reading difficulty. His relationship with his peers may influence his motivation to attempt improvement, but is irrelevant to his problem. Therefore, foil 12B ($12D_2$) was classified as an Irrelevance (Irr).

Foil 27D ($27D_3$) overgeneralizes to an extreme level which made the most reasonable classification for this foil to be Common Misconception (CM).

28. In Chester's progress report, which one of the following is the most important factor contributing to his difficulty with school achievement? (Bloom's Category 5.30)

*A. Aggression which is building up due to frustration over his reading development.

B. His inability to develop a generalized intellectual awareness. (Irr)

C. His weakness in reading which is affecting all areas of learning. (OG)

D. The teacher has been using "contempt for stupidity" as a motivating device. (IA)

This item, once again, involves more than one reading selection and was therefore treated as a synthesis (5.30) item.

Foil 28B turned out to be somewhat unreasonable because the author assumed that the examinee would be able to identify the fact that generalized intellectual awareness is an adult phenomena. The evidence is tenuous since the Awareness selection from its title is related to university education, and Chester in this (Progress) selection is in Grade six. This foil would be an Irrelevancy (Irr) but it may have been unreasonable to expect so tenuous a connection to be made if it could not be assumed that the examinees would know this fact.

The progress report indicates that there are some areas of Chester's development which are not out of step which makes foil 28C ($28D_2$) an Overgeneralization (OG).

Foil 28D ($28D_3$) suggests that this teacher is using contempt for stupidity for motivation, which may be an Invalid Assumption (IA) since the tone of the report is supportive rather than condemnatory.

29. On the basis of the foregoing which of the following seems to be the most important consideration when preparing anecdotal records or progress reports? (Bloom's Category 5.30)

    A. Make no attempts at interpretation since your judgments are probably biased. (CM)

    *B. Present as much information as possible about all the salient aspects of the situation.

    C. Be as brief as possible, giving no information which may cloud the central problem. (OS)

    D. Do not put anything into these reports which might antagonize the child's parents. (RT)

This item also involves more than one reading selection and therefore it was classified as a synthesis (5.30) item.

In the case of foil 29A ($29D_1$), it is impossible to avoid judgments in any reporting, hence this advice is a Common Misconception

(CM). Because of the problem of observer bias, as much pertinent information as is possible should be supplied so that alternative interpretations can be considered by others. This latter statement makes foil 29C ($29D_2$) an Oversimplification (OS).

In foil 29D ($29D_3$) the term report referring to "progress report" is redefined by the suggestion that such a report may become public property, i.e. it will be part of the "report card" to the parents. This approach involves a Redefinition of Terms (RT).

30. The most important principle illustrated in this set of questions is that the teacher should: (Bloom's Category 5.30)

    A. Promote generalized intellectual awareness in aggressive children by using contempt for stupidity as a motivating force. (WW)

    *B. Recognize that developmental deficiencies arise from complex circumstances, requiring multiple-strategy solutions.

    C. Seek professional assistance from the school counselor in the identification of developmental problems. (Tr)

    D. Recognize that "contempt for stupidity" is not necessarily an effective way of generating motivation in pupils. (RT)

This item synthesizes the previous twenty-nine which led to its synthesis (5.30) classification.

Foil 30A ($30D_1$) is a good example of a contrary-to-fact statement developed by the glib use of the repetion of phrases from the reading selections. It illustrates very well the way in which Word Word Link (WW) foils might be generated.

The discussion concerning the role of the counselor which occurred on page 66 suggests that it would be more reasonable tc get the teacher to identify the problem and then get the professional's help in

its treatment. However, this approach means changing the role of the counselor from diagnostician to prognostician. The relationship recommended by foil 30C ($30D_2$) may, therefore, be reversed to the ideal, and this foil was classified as a Transposition (Tr) for these reasons.

One of the important characteristics of the so-called "Puritan Ethic" is its heavy reliance on adversive extrinsic motivation (i.e. punishment) as a means of regulating behavior. Contempt for stupidity is an extremely effective method of motivation in this context if the ill effects of aversive motivation are ignored. Furthermore, there is a tendency, for political reasons, as already suggested that the strongest supporters of the school also support this method of motivation. On this basis, foil 30D ($30D_3$) is clearly incorrect. For it to be considered correct, the term "motivation" must be confined to intrinsic positive motivation (or interest) making this foil a Redefinition of Terms (RT).

In general terms, the overall classification of the foils as reported here is probably open to considerable disagreement as suggested by the low interrater reliability. Would other researchers have produced superior results? One of the raters of the items made the following alternative classifications:

1. $6D_3$ was classified as a Word Word Link (WW).
2. $10D_3$ was classified as the right answer.
3. $15D_1$ was classified as an Invlaid Assumption (IA).

In this case his classification of $6D_3$ turned out to be supported by the cluster analysis (see: p. 77); his classification of $10D_3$ was incorrect for the reason given (see: p. 165); and his classification of $15D_1$ was not supported by the cluster analysis. (see: p. 77).

This success ratio of one out of three is equivalent to that of the experimenter (see: p. 75).

It should also be noted that the nature of the examination was, at least in part, predetermined by the examiner's philosophy of education. This characteristic of the examination is evident in the first place in the nature of the reading selections upon which the examination is based. Second, it is evident in the nature of the questions asked concerning these reading selections. Third, it is evident in the reasons given for the classification of items and foils. In general, it is hoped that the major portion of this bias is confined to the nature of the reading selections used and that once these are given the astute reader should be able to infer the bias, and answer accordingly. The possible exceptions which are clearly evident are item 18 and item 28, particularly with respect to foil 28B ($28D_1$).

It is being argued that bias is unavoidable and, hopefully, can only be minimized in its adverse effects upon student performance. The clear thinking student should be able to recognize and adopt a number of points of view concerning any particular subject matter and apply logic, once the point of view is assumed, in order to arrive at reasonable conclusions. So long as the logic which follows from the assumptions cannot be faulted the system itself can remain intact. The purpose of presenting the development of the experimental test in such detail was to expose the logic of the test including the reasons why the foils are considered to be wrong (i.e. its construct validity) to such reasonable attacks as may be made. If the construct validity of the test is supported, on both logical and evidential grounds, then the experimental test may be regarded as an effective measuring

instrument of higher mental processes independent of the assumptions upon which the particular items are based. In this case the refutation of these assumptions by subsequent evidence will not diminish the value of this procedure as a method for the construction of measuring instruments for the evaluation of student performance in the cognitive domain (i.e. in the use of higher mental processes).

APPENDIX C

LOGICO-SEMANTIC ANALYSIS OF RIGHT AND WRONG ANSWER CLUSTERS

APPENDIX C

LOGICO-SEMANTIC ANALYSIS OF RIGHT AND WRONG ANSWER CLUSTERS

This appendix presents a detailed discussion of the items and foils which differed in their advance classification from the classification of the clusters in which they occurred. A cluster was classified by the most frequently recurring advance classification in the cluster.

The findings for this part of the study were summarized on pages 70 to 71 for the right answer clusters and on pages 71 to 79 for the wrong answer clusters. This detailed analysis is given here for two reasons. First, it was felt that the effective reclassification of alternatives represented evidence in support for the multiple interpretation hypothesis. Second, it was felt that subsequent researchers might find value in an independent evaluation of the logic which led to the conclusions this study has presented.

The Meaningful Interpretation of Item Clusters

In an exploratory study into a new area of research, the relative relevance of characteristics can be expected, in general, to be unknown. For this reason, the failure of the advance classification of items to provide much assistance towards a meaningful interpretation of the data was disappointing, but not surprising.

On the other hand, the cluster solution used, replicated the advance classification by Bloom's Taxonomy to the extent that 40 per cent of the items which appeared in a single cluster also held a common advance classification. Table II (see: p. 73) gives the clusters from Group A and indicates which items were in a common classification, the

classification of these items, and the final interpretation given to each cluster.

Where the interpretation remained ambiguous in Table 11 (see: p. 73) the uncertainty is indicated. It would be a fairly simple matter with clusters like C_6 and C_{10}, for instance, to assume that the advance classification of these items adequately interprets the cluster.

In other cases such as C_1, C_5 or C_7, the majority of the items were in a common class. If the class of the majority is used as an interpretation, the members which did not share this common classification must be explained.

Finally, there were several clusters, C_2, C_3, C_4, C_8, and C_9 which did not contain even two members which shared a common advance classification. The interpretation of these clusters would seem most problematic, but must be attempted.

Several considerations were used in an attempt to arrive at an unambiguous meaningful interpretation of each cluster. The first and most obvious one was the advance classification of the items by Bloom's Taxonomy.

Second, the possibility that some clusters (C_9 for instance) might be content clusters could not be entirely discounted.

Third, the possibility that items might have been misclassified in either of two possible ways. The aspect of the item which leads to the misclassification might be related to some obvious but irrelevant format characteristic. This problem has already been illustrated by the case of items 19 to 24 inclusive. (see: Appendix B pp. 175-177). Alternatively, there may be a discrepancy between the way in which the examiner intended that the item be interpreted and the way it was, in

fact, interpreted by the examinees. For instance, an item which is a comprehension item for some students may well be an analysis item for others. This latter possibility would suggest that the classification of items might be better after their characteristic clustering has been determined than before the test is given.

Fourth, since this study is postulating that the foils may have some effect upon the interpretation of the item and, therefore, the way in which it is answered, the nature and selection ratio of the foils should also be taken into account when an attempt is being made to interpret the clusters. In this respect, foils with a selection ratio of .05 or less were dropped from this and subsequent analyses, since these foils were selected by too few people for the statistics pertinent to these foils to be stable.

Finally, there were other sources of information about these items, such as the interitem phi correlation coefficient matrix and the item consistency which might have proved useful in the attempt to make an unambiguous interpretation of each of the item clusters.

In the discussion which follows each of the ten item clusters are dealt with in turn in an attempt to establish an unambiguous interpretation for each cluster. In advance of each discussion a table appears which supplies the following information:

1. The numbers of items in the cluster.
2. The subject matter content from which each item is drawn.
3. The advance classification of each item.
4. The biserial correlation (r_b) of each item with the total test score.
5. The difficulty of the items.

6. The selection ratio for each foil and the advance classification of each foil for the items in the cluster. The foils which are dropped are also indicated.

Any other information needed for the discussion is supplied in the context. Table 42 follows on page 190.

Three of the four items in Table 12 have their content clearly drawn from the Stupidity reading selection on pages 41 and 42, and the fourth one, item 8, has a reference to this selection in its stem. However, item 8 can be answered without having read this selection since a good student should be able to infer what is meant by the phrase "contempt for stupidity" from the context. Also, foil $28D_3$ was the only part of item 28 which contained a reference to this selection but, once again, it should be possible to infer the meaning from the context. Furthermore, foil $28D_3$ is classified as an IA (Invalid Assumption) foil, the invalid assumption for which can be arrived at without reference to the "Stupidity" selection. In addition, this cluster does not exhaust the items in which a reference to this selection is made. There are five other such items. For these reasons, it is not possible to interpret this cluster unambiguously on the basis of content.

The relative magnitude (.378 to .456) of the r_b (biserial correlations with total scores) varies sufficiently to suggest that they were probably not related to the statistical artifacts which caused these items to form a cluster. Also, since none of the difficulties or selection ratios (D_*) are large enough that these items must (of necessity) overlap, the D_* ratios cannot be considered to be relevant in this event either.

Since three of the items were classified in advance as analysis

TABLE 42

SUMMARY OF DATA FOR CLUSTER C_1

Item No.	Item Content	Advance Classification	r_b	Selection Ratio				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
1	Stupidity	Analysis	.384	.09	.83	.04	.02	OG	X	X
2	Stupidity	Analysis	.456	.25	.61	.09	.04	Sub	OG	X
8	Aggression	Analysis	.419	.29	.58	.01	.12	OG	X	Inv
28	Stupidity Awareness Aggression Progress	Synthesis	.378	.37	.25	.32	.06	Irr	OG	IA

D_* is for the correct answer.

X means foil is dropped.

items, whereas the fourth one (item 28) is Synthesis as this class was defined, the reasonableness of retaining the Analysis classification for this cluster is greater than for changing it to a content-oriented classification. Also, as has already been noted, item 28 has identical foils by class with item one. Furthermore, in three of these items (1, 8, 28) the most commonly selected foil has an advance foil classification of OG (Overgeneralization). In this case, if $2D_1$ could be reasonable reclassified from Substitution to OG an alternative to the reclassification procedure is to suggest that the categories of foil given above may not be independent. In this case, the common element to these items would be the common classification of the most commonly selected foil. This argument would be strongly supported if all of these foils fell into a common wrong answer cluster, which they did not do ($28D_2$ is the exception). It is reasonable to be reluctant to classify a cluster derived from the right answer correlation matrix of Group A on the basis of the performance of foils to the items when performance on specific foils is not part of the statistical basis from which this cluster is derived. In this cluster, the classification was (reluctantly) retained as Analysis, even though this meant reclassifying item 28. Table 43 follows on page 192.

To begin with, in Table 43 there was no consistency between the items in Cluster C_2 with respect to their content (information background) which might have accounted for the formation of this cluster. A similar statement can be made for the advance item classification, for the relative magnitude of the r_b and the D_* coefficients, and for the advance classification of the foils.

Superficially, then, there would seem to be no basis for the

TABLE 43

SUMMARY OF DATA FOR CLUSTER C_2

Item No.	Item Content	Advance Classification	r_b	Selection Ratios				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
3	Stupidity	Comprehension	.304	.33	.01	.23	.41	X	Irr	OS
17	Discipline	Analysis	.541	.08	.80	.03	.09	OG	X	Sub
30	Summary of all items	Synthesis	.158	.74	.01	.15	.10	X	Tr	RT

D_* is for the correct answer.

X means foil is dropped.

interpretation of this cluster. However, an examination of the foil classification for item 30 is revealing (see: pp. 181,182); and $30D_3$ was an RT (Redefinition of Terms) foil. Two and possibly all three of these foils are related to Comprehension-type operations (i.e. they are Misreading type foils). If the foils of an item could be eliminated by comprehension-type strategies, the fact that the stem-right-answer relationship involved a synthesis-type relationship may have been irrelevant. Similarly, if the stem-right answer relationship can be recognized by comprehension-type strategies without having to eliminate foils, an item may be a comprehension item with high level foils. Such a combination of arguments could account for item 3 and item 30 occurring in this group. In order to account for the presence of item 17 in this cluster, it is necessary to suggest that the OG, OS, etc., type foils are related to analysis-type strategies. For some individuals, then, item 3 could have been treated as an analysis item because of the high level of the foils. In this case, item 17 would have to have most of the examinees who selected the correct answer in the top one-third of the group as defined by total score correct, and there would have to be a high phi coefficient between items 3 and 17. The results support this contention. In the first place, about 80 per cent of the individuals who answered the item correctly were in the top 40 per cent of the group. In addition, the phi correlation coefficient between item 3 and item 17 is .281 which is significant at a probability level of $.02 > p > .01$.

These results did not make possible the unambiguous interpretation of Cluster C_2. On the contrary, the interpretation would seem to be that this cluster involves multiple strategies, some at the

comprehension level, and some at the analysis level. Hence C_2 cannot be defined in terms of a unitary category from Bloom's Taxonomy.

Once again, the data suggests right- wrong answer interactions. In addition, there appeared to be a multiple-strategy level involvement in the cluster.

Cluster C_3, Table 44 (see: p. 195) proved to be somewhat similar to C_2 in that the same phenomena occurred once again. All the initial bases for interpreting this cluster failed to provide for an unambiguous decision. In addition, the high level (Synthesis) item seemed to be lowered by virtue of the low level foils.

It may be reasonable to relate items 4 and 6 because of the fact that $4D_1$ was reclassified a CM (Common Misconception) in the interpretation of the wrong answer clusters and $4D_3$ became unclassifiable. It is possible that these are both low level foils which might lower the analysis classification of item 4 to comprehension. This argument concerning the reclassification of item 4 must remain inconclusive since a classification of $4D_3$ was not established.

The fact that these two clusters (C_2 and C_3) proved to be similar raises the question as to why they did not form a single cluster. Obviously, the items from one cluster did not correlate highly with the items from the other, but this fact does not add any information since it is this fact which is the statistical basis for the information of clusters. A look at the wrong answer clusters (indicated by W_n) into which the foils fell proved illuminating.

Table 45 (see: p. 196) shows that there is a degree of similarity within C_2 of the wrong answer clusters (W_9 occurs in two items). C_3 has a high level of similarity among the foils (W_3 occurs in

TABLE 44

SUMMARY OF DATA FOR CLUSTER C_3

Item No.	Item Content	Advance Classification	r_b	Selection Ratios				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
4	Stupidity	Analysis	.425	.17	.10	.01	.72	OG	X	Irr
13	Stupidity Aggression	Synthesis	.378	.67	.01	.07	.24	X	CM	RT
6	Awareness	Comprehension	.473	.53	.02	.22	.23	X	Sub	Irr

D_* is for the correct answer.

X means foil is dropped.

two items). C_3 has a high level of similarity among the foils (W_3 occurs in all three items and W_7 in two). There are no common wrong answer clusters between the two groups. This evidence would have been much more conclusive as to why these clusters were distinct if the foil groups in C_2 were more strongly similar. Once again, however, there is some indication that foils may influence the formation of right answer clusters.

In any event, the fact that there seems, once again, to be multiple strategies involved in this cluster means that it cannot be unambiguously classified.

TABLE 45

WRONG ANSWER CLUSTER MEMBERSHIP WITHIN AND BETWEEN RIGHT ANSWER CLUSTERS C_2 AND C_3

C_2			C_3		
Item No.	Wrong Answer Clusters		Item No.	Wrong Answer Clusters	
3	W_9	W_{14}	4	W_3	W_7
17	W_1	W_9	13	W_3	W_7
30	W_{15}	W_4	6	W_3	W_{11}

As Table 46 shows (see: p. 198), there is no clear basis for the unambiguous classification of Cluster C_4 from any of the sources of data being used. This cluster, therefore, remained unclassified.

On the other hand, item 5, in addition to involving a practical application, also involves "going beyond the given data to determine implications...which are in accordance with the conditions described in the 'original communication'" (Bloom's Taxonomy, 1956, p. 205). This could mean that the best Bloom's Taxonomy classification for this item, if the application aspect is ignored, is comprehension (2.30: Extrapolation). This item may, therefore, be capable of a dual classification. Item 19 is the only one in this series in which the correct answer involves the implication of the statement by the reading selection. Item 14 also involves extrapolation except that the extrapolation is from two selections rather than one, making this item (arbitrarily) a Synthesis item. It is too early in the development of this testing technique to be dogmatic on a post hoc basis about the interpolation on a strategy basis, of any cluster. This statement is particularly reasonable since this clustering does not cross-validate (see: p. 94). However, it does suggest that a better definition of the multiple strategies which may be involved in the answering of multiple choice items for the types represented on this test might improve the effectiveness of the advance classification of the item, and most certainly would improve the interpretation of the clusters which were found to be peculiar to a particular group of examinees.

In cluster C_5, Table 47 (see: p. 199), the common element, on the basis of the decision rule that the cluster be identified by the most frequent process category from the advance class, would be

TABLE 46

SUMMARY OF DATA FOR CLUSTER C_4

Item No.	Item Content	Advance Classification	r_b	Selection Ratios				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
5	Awareness	Application	.295	.22	.05	.61	.22	X	CM	OS
14	Stupidity Awareness	Synthesis	.405	.46	.02	.10	.40	X	Sub	OG
19	Discipline	Analysis	.463	.31	.09	.14	.47	O	O	O

D_* is for the correct answer.

X means foil is dropped.

TABLE 47

SUMMARY OF DATA FOR CLUSTER C_5

Item No.	Item Content	Advance Classification	r_b	D_*	Selection Ratios D_1	D_2	D_3	Advance Foil Classification D_1	D_2	D_3
7	Awareness	Application	.336	.14	.04	.31	.51	X	WW	OS
22	Progress	Analysis	.001	.07	.02	.06	.84	X	0	0
23	Progress	Analysis	.165	.58	.20	.17	.04	0	0	X

D_* is for the correct answer.

X means foil is dropped.

Analysis. Item 7 was originally classified as an Application item because the stem asked for "the best example." In other items the lowering of the level of performance of that item by low level foils was observed. In this case the popularity of the analysis-related OS ($7D_3$) foil may have had the opposite effect. Furthermore, foil $7D_3$ (an OS foil) was classified during the interpretation of the wrong answer clusters as NS (Non Sequitur) (see: p. 225 for definition) which was one of the three new foil categories which came out of these discussions. Since both of these categories (OS and NS) seemed to be more related to the logic of the item than to its semantics, these types of foil may help to define analysis type items. This possibility was strengthened by the fact that about 73 per cent of the examinees who chose $7D_2$ (the WW foil) were in the bottom 60 per cent of the group, whereas about 65 per cent of the examinees who chose $7D_3$ (the OS foil) were in the top 60 per cent of the group. These figures suggested a moderate but definite trend on the part of these foils to move the performance of this item upward in level. The same trend is evident in the average total-correct values for each of these foils and for the right answer. Those who chose $7D_2$ had an average total correct score of 11.6, while those who chose $7D_3$ had an average total correct score of 12.2, and those who chose the correct answer had an average total correct score of 13.7. The average score on the entire test was 12.2.

The basis for the upgrading of item 7 to the Analysis level is somewhat tenuous. The fact that it would otherwise be the only anomaly in this cluster strengthens the use of the decision rule. The use of the rule is further strengthened by the fact that none of the possible bases being used for interpretation give a more reasonable explanation

for this cluster. The Analysis classification of this cluster was retained.

Cluster C_6, as summarized on Table 48 on page 262, was classified as Analysis since both its members were so classified in advance. However, the procedures being used suggest at least two other bases for interpreting this cluster. First, the difficulties (D_* selection ratios) were high. Second, the most commonly selected foils fell into the same wrong answer cluster (W_{14}). These two events could be related to each other since the advance classification of the foils with their respective items are different.

In any event, C_6 was treated as an Analysis cluster in subsequent statistical analysis.

In Table 49 on page 203 none of the bases being used for interpretation assist in the explanation of the formation of cluster C_7 except the arbitary rule that the majority of the items shared a common classification in advance. In both these items (Item 10 and Item 16) a value judgment is specifically asked for in the stem and explicitly included in each of the alternatives. The other item (Item 18) which had these same characteristics fell into another cluster. Item 29 asks for "the most important consideration" in the stem which makes the stem contain an explicit request for a value judgment; however, this value judgment is not explicit in the alternatives. The term "likely" occurs in the stem of Item 11 suggesting, perhaps, an implicit value judgment may be involved in this stem. Should the definition of the Evaluation level of items as used in this study be extended to include implicit as well as explicit value judgments? The results of the cross-validation part of the study

TABLE 48

SUMMARY OF DATA FOR CLUSTER C_6

Item No.	Item Content	Advance Classification	r_b	Selection Ratios				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
9	Aggression	Analysis	.135	.70	.11	.17	.03	Inv	Sub	X
27	Progress	Analysis	.388	.85	.03	.09	.03	X	Irr	X

D_* is for the correct answer.

X means foil was dropped.

TABLE 49

SUMMARY OF DATA FOR CLUSTER C_7

Item No.	Item Content	Advance Classification	r_b	Selection Ratios				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
10	Aggression	Evaluation	.320	.17	.22	.06	.55	CM	Inv	WW
11	Aggression	Application	.334	.48	.20	.30	.02	Inv	OS	X
16	Discipline	Evaluation	.447	.81	.01	.02	.14	X	X	CM
29	Discipline Progress	Synthesis	.275	.83	.08	.06	.02	CM	OS	X

D_* is for the correct answer.

X means foil is dropped.

suggested that Group B responded quite differently to these items.

In Appendix B where Item 26 from Cluster C_8 in Table 50 (see: p. 205) was discussed (see: pp. 177,178) it was pointed out that an inductive structure had to be generated in the mind of the examinee in order to answer this question, led to its classification as a Synthesis item. All other Synthesis items had the additional characteristic of involving more than one reading selection. The use of the device of having more than one reading selection as a basis for Synthesis items did not generate a unique cluster. On the other hand, if the logic of Item 12 and Item 20 are examined, it becomes evident that a similar process of reasoning to Item 26 may have been involved in these two items as well.

The classification of Item 20 as Analysis was made because it is clearly structured so as to involve a hypothesis testing procedure. If each of the alternatives in Item 12 and Item 26 were also regarded as hypotheses which were to be tested against the inductive structure employed by the examinees in Group A, the answering of these items may have been relatively homogeneous.

This cluster (C_8) is classified as Synthesis on the grounds used for the advance classification of Item 26, or it could be Analysis on the grounds used for the classification of Item 20. However, in multiple choice format it is impossible to avoid hypothesis testing aspects of a Synthesis item when a specific set of alternatives is given in the item. The type of item in this cluster came about as close to a Synthesis level as it may be possible to come in multiple choice items. Bloom (1956) suggests that if ambiguity of classification occurs, it should be resolved in favour of classifying to the highest possible

TABLE 50

SUMMARY OF DATA FOR CLUSTER C_8

Item No.	Item Content	Advance Classification	r_b	Selection Ratios				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
12	Aggression	Comprehension	.430	.35	.20	.35	.10	CM	OG	Inv
20	Progress	Analysis	.269	.53	.04	.04	.39	X	X	O
26	Progress	Synthesis	.140	.38	.15	.34	.13	O	OS	Irr

D_* is for the correct answer.

X means foil is dropped.

level. For this reason, on the basis of the performance of these items (at least within Group A) the items in this cluster have been reclassified as Synthesis items.

Both Cluster C_9 items, as summarized in Table 51 (see: p. 207), are from the same reading selection. Also, the Procrustes rotation to content suggested a possible content basis for this cluster. However, both of them also proved to be very poor items on the basis of the combination of their r_b and difficulty (D_* selection ratios). In addition to this their most common foils occurred in the same wrong answer cluster (W_7), and the D_n selection ratios of these two foils were very high. It is probable that these two items form a cluster, at least partly, on the basis of the relationship between these two foils. This cluster remained unclassified in subsequent statistical analysis.

There would seem to be two possible bases for the interpretation of Cluster C_{10} as summarized in Table 52 (see: p. 208) content and advance classification. It is possible that both of these factors were operative to make this cluster distinct from other Analysis clusters. Some support for this argument may be found in the fact that C_{10} was positively correlated with all the other classified clusters in the Procrustes rotation analysis (see: Table 10, p. 69) except C_7, although all of these correlations may have been too small to have much meaning.

In summary, then, this interpretation attempt upheld the Analysis level classification of Clusters C_1, C_5, C_6, and C_{10} as they emerged in the comparison between the results of the minimal interpoint distance cluster analysis and the advance classification of the items. It similarly upheld the classification of C_7 as Evaluation. The procedure led to the reclassification of one cluster (C_8) from

TABLE 51

SUMMARY OF DATA FOR CLUSTER C_9

Item No.	Item Content	Advance Classification	r_b	Selection Ratio				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
15	Discipline	Comprehension	.015	.05	.02	.04	.88	X	X	RT
18	Discipline	Evaluation	.295	.01	.11	.86	.02	Sub	Irr	X

D_* is for the correct answer.

X means foil is dropped.

TABLE 52

SUMMARY OF DATA FOR CLUSTER C_{10}

Item No.	Item Content	Advance Classification	r_b	Selection Ratios				Advance Foil Classification		
				D_*	D_1	D_2	D_3	D_1	D_2	D_3
21	Progress	Analysis	.404	.32	.07	.37	.23	0	0	0
25	Progress	Analysis	.329	.55	.24	.19	.02	Sub	Irr	X
24	Progress	Analysis	.317	.86	.01	.07	.06	X	0	0

D_* is for the correct answer.

X means foil is dropped.

undetermined to Synthesis. In the remaining clusters (C_2, C_3, C_4, and C_9) it was impossible to provide an unambiguous basis for classifying these clusters from the available data. Hence these clusters remained unclassified. Clusters C_2 and C_3 seemed to involve some form of multiple strategies and C_4 seemed to be a Comprehension level cluster for Group A if the superficial characteristics of the items which led to their advance classification were ignored. Since this cluster did not reappear in Group B, multiple strategies between groups of examinees may be involved.

Only in the case of C_9 can content be said to be a more reasonable interpretation of these clusters than some form of single or multiple strategy. Even in this case, the content interpretation is in some doubt, suggesting that this test is essentially "process oriented," as intended.

From the original 30 items on the test 12 items (40 per cent) formed at least pairs in the clusters which emerged. This "interpretability" figure was improved to 19 items (63 per cent) on the basis of the interpretation procedure used.

The Meaningful Interpretation of Wrong Answer Clusters

A similar interpretative procedure was used for wrong answers as for the items. The considerations were taken in the following order, 1) advance classification, 2) information background (content), 3) statistics of foils, 4) logico-semantic analysis. A cluster was again assumed to be identified by the most frequently recurring advance classification. Foils of the 0 (Other) category were assumed to be part of this common classification.

Of these four, the only one related to the logico-semantic characteristics of the foils was their advance classification. If these

clusters could not be accounted for in other ways, and the logico-semantic characteristics of the foils can be shown as a possible basis of cluster membership, then this latter basis may be the best available interpretation. It has already been shown that the advance classification of both right answers and foils did not survive very well in the cluster analysis. It has also been shown that this classification can be improved by examining the clusters for some relatively unambiguous basis for interpretation. In the case of foils this involved a re-examination of their logico-semantic structure. The five bases used in an attempt to interpret the wrong answer clusters were, once again:

1. The advance classification of the foil.
2. The content of the foil.
3. The selection ratio of each foil.
4. The relationship between right answer and wrong answer clusters.
5. The reconsideration of the logico-semantic structure of the foils.

In some cases, material from other sources such as Powell and Isbister (1969) were used to assist in this interpretative process. Finally, one cluster (W_2) was completely lost by virtue of the low selection ratio among all of its members. In addition, two other clusters were reduced to single members by this procedure. In the discussions which follow some tentative attempts are made to account for the "special cases" as well as the "general trends" in each cluster.

For the convenience of the reader, whenever a particular foil is being discussed for the purpose of reclassification, the stem of the item and the particular foil are both given ahead of the pertinent discussions.

Table 53 on this page gives information for the interpretation of wrong answer cluster W_1.

TABLE 53

INFORMATION FOR THE INTERPRETATION OF WRONG ANSWER CLUSTER W_1

Foil	Content	Advance Classification	D_n [a]
$1D_1$	Stupidity	OG [b]	.83
$2D_1$	Stupidity	Sub	.61
$22D_3$	Progress	O	.84
$8D_1$	Aggression	OG	.58
$17D_1$	Discipline	OG	.80

a. The symbol D_n refers to the selection ratio.

b. Foils which had appeared in a common category in the advance foil classification.

The content from which these foils in wrong answer cluster W_1 are drawn comes from a broad spectrum of the test ruling out content as a possible basis for the interpretation of this wrong answer cluster.

The phi coefficients upon which this cluster is based are dependent upon the size of the overlap between particular pairs and upon the marginal totals. When the selection ratios for two alternatives both exceed .50 there is tendency for the range of phi to be shifted positively. In this case, all of the selection ratios in the cluster

exceed .50 and for this reason the sizes of the selection ratios could be a contributing factor to the statistical formation of this cluster. If this event were the only factor, however, it would be reasonable to have expected more of the other six foils which have selection ratios of greater than .50 in this cluster, or in a limited number of other clusters. In fact, they occurred in four of the clusters.

Of the five foils in this cluster three of them were in items which occur in a common right answer cluster (C_1). This finding was suggestive, once again, of a right-answer wrong-answer interpretation, but insufficient to lead to an unambiguous interpretation of the present cluster under discussion.

Also, three of the five foils in Table 24 were classified as OG (Overgeneralization, i.e. $1D_1$, $8D_1$, $17D_1$) and a fourth one as O (Other, i.e. $22D_3$) which means it could be treated as an OG as well. Thus, the advance classification seemed to be the most promising basis for interpretation, making it necessary to re-examine the classification of $2D_1$ for its logico-semantic relationship with the stem.

Item 2: Which of the following factors is the most important causitive factor of contempt for stupidity?

$2D_1$ Compulsory school attendance.(Sub)

Probably the best procedure in the analysis of this foil is to use a Venn Diagram, as given in Figure 3, page 213.

Figure 3 shows clearly that "contempt for stupidity" and "compulsory school attendence" (Foil $2D_1$) are disjunctively related. For a factor to be "causitive" it must be either conjunctively or implicatively related to the other factor. Either conjunction or implication is a necessary but not a sufficient condition for causation.

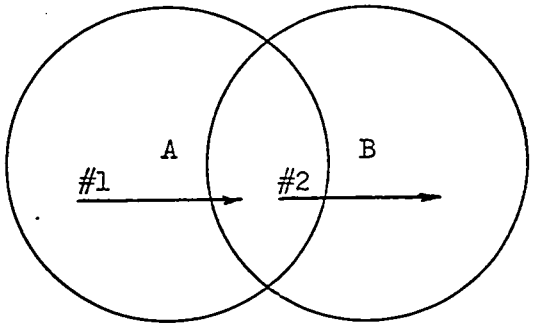

FIGURE 3

VENN DIAGRAM ILLUSTRATING FOIL $2D_1$

A. Compulsory school attendance.

B. Contempt for stupidity.

It would be reasonable to suggest that the student who replaced this disjunctive relationahip with a conjunctive relationship, (ignoring the fact that either can occur without the other) is substituting one category for another. It was upon this basis that this foil ($2D_1$) was originally classified as a Sub (Substitution).

On the other hand, if the student considers the relationship as implicative, (i.e. compulsory school attendance can occur without contempt for stupidity but not vice versa) this interpretation could be considered an OS (Oversimplification). The thinking in this case involves proceeding from an entire set to a subset of that entire set as indicated in the use of Arrow #1 in Figure 3.

If the student begins with the conjunctive subset and extends this to include all cases of contempt for stupidity, the thinking process would follow the path of Arrow #2 in Figure 3. In such a case, (proceeding from a subset to an entire set) the most reasonable

interpretation of the thinking is OG (Overgeneralization). These arguments suggest the importance of the way in which an item is interpreted with respect to the way in which a foil should be classified. As just shown, this foil ($2D_1$) can be reasonably argued to have at least three possible classifications depending upon the interpretation placed upon the foil by the examinee.

A decision rule was needed to deal with foils which might reasonably be classified in several different categories. Where a cluster seemed, in general, to reflect one category of foil, and the same interpretation was one of the possible classifications of the ambiguous foil, then this classification was assumed to be an appropriate interpretation of the ambiguous foil for the particular group on which the cluster analysis was conducted. The most important characteristic of this rule was the requirement that the interpretation of a foil should probably not be generalized beyond the group of examinees upon whom the interpretation was established.

Hence, since $2D_1$ could be an OG, this wrong answer cluster could be interpreted as an OG cluster for Group A.

Wrong answer cluster W_2 was eliminated before logico-semantic analysis on the basis of the fact that all of the foils in this cluster had a selection ratio of less than .06.

For Cluster W_3 (see: p. 215) there is no common content. Items 4, 13, and 6 comprise all the items in right answer cluster C_3 but these items represented less than half of the members of W_3. Only one of the foils has a selection ratio of more than .50, and both OG and CM have two representatives from the advance classification of foils. Hence, the interpretation of this cluster could not be established by any of these

TABLE 54

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_3

Foil	Content	Advance Classification	D_n
$4D_1$	Stupidity	OG	.10
$26D_1$	Progress	O	.15
$5D_2$	Awareness	CM	.61
$6D_2$	Awareness	Sub	.22
$14D_3$	Stupifity, Awareness	OG	.40
$13D_2$	Awareness	CM	.07
$29D_2$	Discipline, Progress	OS	.15

methods.

In the ensuing discussions a possible common element among the foils in a cluster is suggested. The clue which was used in this case was the presence of OG, OS, and CM (Common Misconception) in the same cluster of which OG and CM were the most frequent. Powell and Isbister (1969) found a polarity between OG and OS on the one hand, and CM on the other. This polarity in a factor matrix indicates a significant negative correlation between the poles. Since this polarity, at least in part, could have been an artifact of the mutually exclusive selection of responses within items, it is reasonable to suggest that OS, OG, and CM foils may be related. In addition to using the "most frequent" rule, it might be possible to reinforce the interpretability of this cluster as

CM if most of the remaining foils had CM as one of their possible alternative classifications.

To begin with, foil $26D_1$ as a member of the O category can have any interpretation which is reasonable for the remainder of the foils.

A discussion of the others follows:

Item 4: The author, in charging that "society teaches contempt for stupidity and fear of being regarded as stupid" by means of the school, is assuming that:

$4D_1$. The school should not be an enforcing arm for the customs of society. (OG)

This foil is wrong on two counts. In the first place it is a conclusion rather than an assumption made by the author. Second, it is overstated by containing a value judgment which may be unwarranted. This second reason led to its OG classification. It is, however, in addition, one of the ways of phrasing a very common argument against the establishment of parochial schools. In this latter context it could be a CM foil as well. However, this argument is shaky at best, and would not be likely to extend beyond the context of the group upon which this cluster was established.

Item 29: On the basis of the foregoing which of the following seems to be the most important consideration when preparing anecdotal records on progress reports?

$29D_1$ Make no attempts at interpretation since your judgments are probably biased. (CM)

$29D_2$ Be as brief as possible, giving no information which may cloud the central problem. (OS)

Foil $29D_1$ was dropped on the basis of its low selection ratio. Since in the advance classification two foils in the same class in the same item were not entertained, and since the "brief as possible" part of $29D_2$ is an Oversimplification, foil $29D_2$ was originally classified

as an OS foil. However, the phrase "which may cloud the central problem" in foil $29D_2$ is similar in its central idea to "your judgments are probably biased." This idea being another way of saying that a person should be as objective as possible. Students who emphasized this idea in their thinking could be responding more to the second part of the statement than to the first (OS) part, making CM (Common Misconception) a reasonable alternative classification for this foil. Of course, there is the problem of the "relevancy of details" which this foil may also raise. It is, however, better to put in details which the writer may think irrelevant and an independent observer may not than to require the independent observer to infer these details from the context because of their omissions. This aspect of the discussion leads to another common misconception, namely that the simple act of speaking or writing has produced a successful communication.

Item 6: The purpose of developing a generalized intellectual awareness is to:

$6D_2$ Stimulate thinking ability within the individual's chosen field.

The confining of thinking ability to "the individual's chosen field" is false. It is a substitution for the phrase "which is not contextually bound." Once again, however, the frequency with which this misconception is encountered suggests that the foil could be regarded as a CM foil as well as a Sub foil.

Item 14: Which of the following best describes the probable relationship between contempt for stupidity and generalized intellectual awareness?

$14D_3$ Contempt for stupidity should be reduced and awareness should be increased. (OG)

This foil is clearly an OG foil since it adds an unwarranted

value judgment to an otherwise correct statement. Should the definition of CM foils be extended to incorporate this characteristic? In any event, six of the seven foils in this cluster could be assigned fairly reasonably to the CM class, making the CM interpretation of the entire cluster plausible, if not reasonable. Table 55 on Cluster W_4 follows.

TABLE 55

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_4

Foil	Content	Advance Classification	D_n
$30D_3$	Summary of all passages	RT	.10

W_4, as a single member cluster, should probably be dropped unless a good reason for retaining it can be found. As an approach to the problem of classifying this Cluster W_4 the fate of other RT foils proved helpful. Foil $11D_3$ was originally in this same wrong answer cluster (W_4) but was dropped because of its low selection ratio. Foils $13D_3$ and $15D_3$ were both in wrong answer cluster W_7, and this cluster remained uninterpreted. The separating factor between $15D_3$, $29D_3$, and $30D_3$ may have been on content lines since each of these are from different reading selections. Foil $29D_3$ was also dropped because of its

low selection ratio. Foil $13D_3$ had similar but not identical background to $30D_3$ but was dissimilar to $15D_3$ as to content, which was part of the classification problem of W_7. The other single member cluster (W_{10}) contains foil $29D_1$ which was originally classified as CM but could also be an RT. If the foil in W_{10} is an RT, the separation is on a content basis. The low average total correct scores for these two foils (11.1 and 11.0) may be taken as equivalent except for content; hence, this cluster (W_4) was retained as an RT. Table 56 on Cluster W_5 follows.

TABLE 56

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_5

Foil	Content	Advance Classification	D_n
$12D_3$	Aggression	Inv	.10
$11D_2$	Aggression	OS	.30
$12D_2$	Aggression	OG	.35
$19D_2$	Progress	O	.14
$20D_3$	Progress	O	.37

By the rules used thus far, except for the logico-semantic interpretation, Cluster W_5 should be classified as O (Other) since this is the most frequently occurring equivalent advance foil class. The

unusual element in this cluster, however, is the fact that there are two foils from the same item in this cluster. This event violates the assumption behind the rule for classification discussed earlier on page 153, number 5.

Since the reclassification of any one of foils $12D_3$, $11D_2$, or $12D_2$ to the same category as either of the others would give that classification to four out of the five foils in this cluster, and since content, the size of the D_n, and the distribution in right answer clusters of the corresponding items do not account for this cluster, a logico-semantic analysis of the two foils in item 12 would seem reasonable.

Item 12: Aggressive behavior in female children is:

$12D_2$ More unpredictable and expressed differently than in males of the same age group. (OG)

$12D_3$ Less differentiated in expression than in males of the same age. (Inv)

Foil $12D_3$ was classified as Inv (Inversion) because it is opposite to a true statement. It is not opposite to the actual correct answer but to a statement that could have been used as an alternative correct answer if the examiner had so chosen. On the other hand, $12D_2$ was classified as OG because the first two words (more unpredictable) form an incorrect statement added to a correct statement. However, these two words are incorrect by virtue of being opposite to the truth (Inv) when the restriction "of the same age group" was applied to this statement.

This last property reinforces the Inv (Inversion) classification of this foil. It may be argued that two Inv foils were possible in Item 12 because of the complex logical structure which this item

required before it could be answered. (See: the discussion of right answer cluster C_8 on pages 118 to 120).

Item 11: Overt aggression would likely be decreased by:

$11D_1$ Blocking many modes of Aggression. (Inv)

$11D_2$ Lessening the threat of punishment. (OS)

Lessening the threat of punishment, or permissiveness, does not, by itself, either increase or decrease overt aggression. It is on these grounds that this foil was classified as an OS. Overt aggression will be likely to increase if permissiveness develops frustration. The energy of the children must be channelled into alternative directions if overt aggression is to decrease in a permissive setting. Thus, it is an oversimplification to say that aggression is likely to either increase or decrease in a permissive setting. However, lessening the threat of punishment, in the absence of alternatives, will probably increase overt aggression. This foil could have been classified as an Inv if it were not for the fact that $11D_1$ was already so classified. An alternative possible classification for $11D_1$ is given with the discussion of W_8 (see: p. 227).

Perhaps the reclassification of $11D_2$ as Inv is a bit tenuous. The other four foils in this cluster are not on such shaky ground, so that it is reasonable to reclassify wrong answer cluster W_5 as Inv. Information for the interpretation of Wrong Answer Cluster W_6 appears in Table 57 on page 222.

Short of a close logico-semantic analysis of the characteristics of the foils in Table 57, there is no clear basis for the interpretation of cluster W_6.

It should be pointed out that the fact that these foils have

TABLE 57

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_6

Foil	Content	Advance Classification	D_n
$9D_1$	Aggression	Inv	.11
$25D_1$	Progress	Sub	.24
$21D_3$	Progress	O	.23
$7D_3$	Awareness	OS	.51
$26D_3$	Progress	Irr	.13

formed a cluster has led to the assumption that there must have been a common logico-semantic element in these foils, at least so far as the members of Group A were concerned. The argument for the formation of a new class of foil (namely: Non Sequitur--NS) which follows should not be construed to deny the plausibility of the original classification of the foils in this cluster but only as to the inappropriateness of these classifications for this particular group of examinees. By this point in the study, it had become clear that the foil categories were not mutually exclusive, and that it seemed possible to classify at least some foils in several different ways (see: $2D_1$, p. 131 ff). Thus, the foil interpretations presented here most probably apply only to Group A.

Item 9: The basic position of the author in writing about aggression is that it:

$9D_1$ Can be eliminated through the process of socialization. (Inv)

On the contrary, Kagan and Moss (1962) assume that aggression is one of several innate behavior systems. Being innate makes its elimination impossible. However, the socialization process can channel aggression away from its more destructive aspects. A number of classifications of this foil are possible. Once again, these classifications are dependent upon several possible logico-semantic distinctions which can be made. Most simply, the foil does not follow from the data, i.e. it is a non sequitur relationship. There was, however, no such classification in the advance classification. This NS (non sequitur) category was eliminated, as indicated above, on the basis that it displayed experimental dependencies with CM foils in the Powell and Isbister (1969) study.

Item 25: The most useful suggestion to help Chester is:

$29D_1$ To give Chester personal warmth, acceptance, and support wherever it is appropriate. (Sub)

As the right answer indicates, Chester needs "concrete help in getting started on specific tasks." In other words his problems would seem, from the progress report, more developmental than emotional. The procedure given in $25D_1$ substitutes a treatment procedure designed to deal with emotional problems for a procedure designed for developmental problems. By itself, the use of $25D_1$ is simply inappropriate. Reclassifying $25D_1$ into a new NS category would seem to be reasonable.

Item 21: Chester is growing very slowly and really quite immature for his grade. Everyone expects too much of him.

$21D_3$ This hypothesis is refuted by the facts. (O)

On the contrary, the only information directly available about Chester's physical development is the number of days he has been absent

or tardy. This limited information is insufficient concerning the physical areas of development given in Item 21 to form any conclusions. He shows some indications of specific academic immaturity, and perhaps some social immaturity, but this is all. As to whether or not the expectations made of him are unreasonable, it cannot be decided from the information given. The only expectation statement made about his work is "Is not ready for division," which does not sound like a statement of overexpectation.

Of course, $21D_3$ could have been reclassified without this analysis because of its O (Other) advance classification. However, its relationship to the correct answer supported the use of an NS category for this foil. $21D_2$ was dropped for underselection, and $21D_1$ formed part of one of the two unnamed new categories to emerge from this study. (W_{13}). This information also implies the reasonableness of the formation of a new NS category for wrong answer cluster W_6.

Item 7: Of the following the best example of generalized intellectual skill is:

$7D_3$ Applying abstract principles to new situations. (OS)

Comparing $7D_3$ with the "correct" answer "The widely applicable technique of logic" showed this foil to be without question an OS foil, since it is a true statement of narrower generality than the correct answer. As a true statement it cannot be classified as an NS, which leaves the interpretation of this cluster based on this item in some doubt. Any alternative interpretations such as involving its large selection ratio (.51) or suggesting a tenuous link between NS (Non Sequitur) and OS (Oversimplification) would be premature at this stage. Cross validation data helps to clarify this cluster somewhat. The fact

that this foil moves to wrong answer cluster W_3 in Group B which is most prominantly composed of members from W_7 (the wrong answer cluster from Group A which proved to be unclassifiable) relieves the problem of interpreting this cluster somewhat, but does not solve it. Foils $9D_1$ and $25D_1$ moved together, as did $21D_3$ and $7D_3$, while $26D_3$ migrated by itself to a new cluster. This cluster held together better than any other in cross validation with the exception of W_{13}.

Item 26: If additional information on Chester is desired and none of the following has been attempted, which one would provide the greatest amount of immediately useful information:

$26D_3$ A request for the assistance of a guidance counselor. (Irr)

The D_n (selection ratio) on this foil is .13. It might be interesting to know who made these selections. Most of the examinees in this group were practicing teachers and they would know from experience that the usual information from the guidance counselor would merely reinforce what they already knew and not add much further useful information. In the absence of an NS category, this foil is clearly an Irr.

Of the five foils in this cluster, four could have been classified quite reasonably as NS foils. The fifth one ($7D_3$) could not have been so classified meaning that this cluster could not be unambiguously classified. However, the best interpretation for this cluster within Group A is to establish a new class of foil, namely: Non Sequitur (NS): This type of foil is a wrong answer foil by virtue of the fact that it simply does not follow from the given information. As a rule some part of the foil stands in direct contradiction to the logical structure or connotative meaning of the background information (or part thereof) required to answer the question.

Wrong answer cluster W_6 was classified as containing NS foils, at least so far as Group A was concerned. Table 58 giving information for the interpretation of wrong answer Cluster W_7 follows.

TABLE 58

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_7

Foil	Content	Advance Classification	D_n
$10D_1$	Aggression	CM	.22
$19D_3$	Progress	O	.47
$4D_3$	Stupidity	Irr	.72
$13D_3$	Stupidity, Awareness Aggression	RT	.24
$12D_1$	Aggression	CM	.35
$15D_3$	Discipline	RT	.88
$18D_2$	Discipline	Irr	.86
$26D_2$	Progress	OS	.34
$8D_3$	Aggression	Inv	.12

The information in Table 58 provides no clear basis upon which to interpret wrong answer cluster W_7. It contains two CM's, two RT's, and two Irr's. In general, the lower D_n's relate to the selection on aggression ($26D_2$ is the exception).

No detailed discussion will be given for this group. One

illustration is sufficient. Foils $10D_1$ and $12D_1$ could conceivably be reclassified as RT foils on the basis of logico-semantic analysis, as could $19D_3$ by the "O" rule. However, such a reclassification is unreasonable for foils $4D_3$, $18D_2$, $26D_2$, and $8D_3$. Similar findings occurred for other pairs in this wrong answer cluster.

The inability to interpret this cluster led to its being dropped from further relevant analysis. This was somewhat unfortunate since several of these foils have a high selection ratio meaning that a fair amount of information was lost by this decision. Nonetheless, it is reasonable to argue that if a wrong answer cluster cannot be given an adequate label, it should not be used. Table 59 giving information for the interpretation of wrong answer cluster W_8 follows.

TABLE 59

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_8

Foil	Content	Advance Classification	D_n
$11D_1$	Aggression	Inv	.20
$28D_3$	Stupidity, Progress	IA	.06
$14D_2$	Stupidity, Awareness	Sub	.10

Table 59 shows that Cluster W_8 also seems to be ambiguous.

Item 11: Overt aggression would likely be decreased by:

$11D_1$ Blocking of many modes of aggression. (Inv)

There are several ways in which this foil can be interpreted. The blocking of modes of aggression would not reduce overt aggression except in those specific areas where the blocking occurred. Overt aggression would increase in areas where the blocking was absent or less effective, or more socially acceptable to the peer group. If the blocking increased the frustration level, the absolute incidence or overt aggression would also increase. For this latter reason, (i.e. since blocking would probably have the opposite to the desired effect) this foil was classified as an Inv. However, to arrive at this conclusion as being correct the examinee must make the invalid assumption that attempts to regulate overt behavior also regulate innate drives. This foil could, therefore, be classified as an IA foil. This decision gave the cluster a majority of IA foils. If IA was also a reasonable alternative for $14D_2$, then IA would be a reasonable classification for this cluster so far as Group A was concerned.

Item 14: Which of the following best describes the probable relationship between contempt for stupidity and generalized intellectual awareness?

$14D_2$ Either will increase with an increase in the other. (Sub)

Foil $14D_2$ was one of the foils about which the raters showed considerable disagreement. It was classified as Sub because part of the relationship was wrong, and it did not seem to relate to the other Inv foils in the pilot study. This relationship could also be an Inv, because a direct relationship is logically opposite to an inverse relationship. Powell and Isbister (1969) encountered the same problem in logical relations (as compared to logical fallacies) type foils. On

the other hand, this foil could wrongly be considered correct if the examinee forms an invalid assumption relating "critical" thinking with contempt for stupidity by defining stupidity in terms of uncritical thinking.

Other possible classifications were given to this foil with similarly tangled arguments justifying each rater's conclusions.

Two of these foils can be reasonably regarded as IA foils, the third could well be stretching the point. In any case, a reasonable over-all classification for this wrong answer cluster would seem to be IA. Table 60 giving information for the interpretation of wrong answer cluster W_9 follows.

TABLE 60

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_9

Foil	Content	Advance Classification	D_n
$3D_3$	Stupidity	OS	.41
$28D_1$	Progress	Irr	.25
$10D_2$	Aggression	Inv	.06
$17D_3$	Discipline	Sub	.09
$27D_2$	Progress	O	.06
$5D_3$	Awareness	OS	.12

In Table 60 the two OS's and the O foil account for half of the foils in Cluster W_9 making an OS classification of this cluster within the rules based on the advance classification. It would be better if the other three could be alternatively classified as OS.

Item 28: In Chester's progress report, which one of the following is the most important factor contributing to his difficulty with school achievement?

$28D_1$ His ability to develop a generalized intellectual awareness. (Irr)

This foil was classified as Irr on the basis that in Grade 6 most children are still too young to have progressed very far into "Formal Operations" which form the basis of generalized intellectual awareness. This classification assumes that the examinees know this information about development which is an unreasonable assumption. In the absence of this information, Chester's problems also involve decoding skills in reading and to a lesser extent in arithmetic. Hence his problems involve more than this foil suggests, and the foil might, for this reason, be reclassified as an OS foil.

Item 10: With which of the following statements concerning aggression would the author be most likely to agree?

$10D_2$ Aggression generally interferes with the attainment of educational goals. (Inv)

This foil was classified as an Inv because the best answer was "Aggression is potentially useful for educational purposes." Superficially $10D_2$ would seem to be opposite to the right answer. On the other hand, aggression can interfere with the educational process. Aggression expressed in the form of competition may be a useful form of intrinsic motivation. The term "generally" in the foil overstates the case for the negative aspects of aggression whereas the foil itself understates

the total picture. This foil might be classified as either OS or OG depending on how the examinee looks at the item.

Item 17: From the above passage we can infer that Doris' leadership of the group was:

$17D_3$ Laissez-faire. (Sub)

The reason for classifying this foil as Sub was that Doris' attempts to coerce were ineffective, so she let the teacher take over. As a result her later leadership was laissez-faire but only under the arbitrary intervention of the teacher. As pointed out in the original discussion of this item, even this argument is stretching the point. (See: pp. 172,173). There is no similar plausible argument which might make this foil an OS.

Five of the six foils in this cluster can be included in the OS categories hence the advance classification of this wrong answer cluster is retained. Information for the interpretation of wrong answer cluster follows in Table 61.

TABLE 61

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_{10}

Foil	Content	Advance Classification	D_n
$29D_1$	Discipline, Progress	CM	.08

Cluster W_{10} as given in Table 61 is the second of two single member clusters. It contains only $29D_1$ which was originally classified as a CM for the reasons already given (see: pp. 180,181). However, the fact that it did not occur in a common cluster with the other CM foils raises some doubts over this classification within the confines of Group A. It is possible, however, to classify this foil into other categories.

> Item 29: On the basis of the foregoing which of the following seems to be the most important consideration when preparing anecdotal records or progress reports?
>
> $29D_1$ Make no attempts at interpretation since your judgments are probably biased. (CM)

The suggestion in this foil that all interpretations are sufficiently biased so as to be of little value has the effect of redefining the term "interpretation" to mean "biased interpretation."

In this case, it would be necessary to assume that the other single member group (W_4) which is also an RT (Redefinition of Terms) split from this one along content lines. The obvious vocabulary-content linkage in both of these foils makes this conclusion reasonable. Foil $29D_1$ was reclassified, therefore, as RT (Redefinition of Terms) making Cluster W_{10} an RT cluster. The apparent content binding of some misreading type foils would seem reasonable, since there seems to be a parallel group of foil levels to the right answer levels in Bloom's Taxonomy. Bloom's description of the levels in the Taxonomy suggest a steady progression away from context as the level of the categories increase. For this reason, it seemed reasonable to combine W_{10} with W_4 as an RT cluster in subsequent analysis, rather than to discard both clusters because of their small size.

Table 62 gives information for the interpretation of wrong answer cluster W_{11}.

TABLE 62

INFORMATION FOR THE INTERPRETATION OF WRONG ANSWER CLUSTER W_{11}

Foil	Content	Advance Classification	D_n
$18D_1$	Discipline	Sub	.11
$7D_2$	Awareness	WW	.31
$6D_3$	Awareness	Irr	.23
$16D_3$	Discipline	CM	.14
$2D_2$	Stupidity	OS	.09
$10D_3$	Aggression	WW	.55

In Table 62 the most frequent foil category in Cluster W_{11} was WW (Word-Word Link) which serves by the decision rules to identify this cluster.

A logico-semantic analysis of the other foils might reinforce this classification.

Item 18: From the description of the incident, we can conclude that the teacher's handling of the incident was:

$18D_1$ Good: she intervened to prevent a serious conflict from continuing. (Sub)

The reading selection states: "There seemed to be confusion in

this group so I decided to investigate." The similarity between this statement and the phrase "intervened to prevent" in the foil is self-evident. WW would seem to be a reasonable alternative classification for this foil.

Item 6: The purpose of developing a generalized intellectual awareness is to:

$6D_3$ Give the individual an ever-widening view of his world. (Irr)

The similarity between the phrase "ever-widening view" in this foil and the phrase "free-ranging understanding" in the reading selection warranted the use of the alternative class of WW for this foil.

Item 2: Which of the following is the most important causative factor of contempt for stupidity.

$2D_2$ Compulsory written examinations. (OS)

The phrase "compulsory written examinations" occurs in both this foil and the reading selection.

Item 16: If the teacher had written "Doesn't work well with others," as an anecdotal record for the above incident, this would have been:

$16D_3$ Worse: Teachers are failing in their obligations in not supplying complete information. (CM)

There is no similar connection in foil $16D_3$ between the stem or the reading selections to the ones presented above. The above discussion, nonetheless, in general supports the retention of WW as a reasonable interpretation for this wrong answer cluster.

Since, as shown in Table 63 (see: p. 235) both of these foils ($19D_1$ and $24D_2$) are classified as O and both come from the same content, there may be some doubt about the interpretation of this cluster.

One obvious course of action with this foil cluster would be to drop it from further analysis. On the other hand, the foil classes in

TABLE 63

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_{12}

Foil	Content	Advance Classification	D_n
$19D_1$	Progress	0	.09
$24D_2$	Progress	0	.07

the Guidelines made no pretence at being exhaustive. Since the experience with other foil clusters has been that logico-semantic snalysis has often revealed a possible common base for interpreting foils within a particular cluster, such an analysis for this cluster may assist in the illumination as to how the list of Guidelines might be extended. For this reason, these foils were also analysed.

To begin with, the fact that they stood separate from any of the interpreted categories suggests that these two foils may form a foil class, the basis of which has not yet been determined, but possibly related to the special format of items 19 to 24 inclusive.

Items 19 to 24 inclusive used a classification protocol designed to get the examinees to treat the statements represented by these items as hypotheses to be tested on the basis of the information in the reading selection from Prescott (1957) concerning Chester's Progress Report. The categories were:

1. supported.

2. implied.

3. refuted.

4. insufficient evidence.

In these categories there is a hierarchy of inferential support from insufficient evidence, to implied, to support.

The "correct" answer from item 19 is "implied" and the corresponding answer given in $19D_1$ was "supported." Similarly, the "correct" answer for item 24 was "insufficient evidence" and the corresponding answer in this cluster was "implied." Each answer given in these foils was one step higher in the hierarchy than the "correct" answer. This "overstatement" was not the same as "overgeneralization" as defined in this study. Whether or not such a relationship is exclusive to this type of question remains undetermined. Rather than premature naming, the O (Other) classification of this cluster was retained. Since another cluster was interpreted as O (i.e. W_{13}), a subscript was applied for purposes of distinguishing between these two clusters.

Once again, Cluster W_{13} as given in Table 64 (see: p. 156) is an O classification based on the advance foil classification. The logico-semantic analysis of this cluster proved to be very interesting. A slight change of format will be used in this discussion with all three items being presented before the discussion of them as a cluster.

Item 21: Chester is growing very slowly and is quite immature for his grade. Everyone expects too much of him.

$21D_1$ The hypothesis is supported by the facts. (O)

Item 28: In Chester's progress report, which one of the following is the most important factor contributing to his difficulty with school achievement.

$28D_2$ His weakness in reading which is affecting all areas of

learning. (OG)

Item 23: Chester's reading deficiency has not yet begun to affect seriously his performance in other areas.

$23D_2$ The hypothesis is refuted by the facts. (O)

TABLE 64

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_{13}

Foil	Content	Advance Classification	D_n
$21D_1$	Progress	O	.07
$28D_2$	Progress	OG	.32
$23D_2$	Progress	O	.17

As it happens, the hypothesis in item 23 is supported rather than refuted by the facts. Notice, however, that this same contrary-to-fact conclusion is stated in foil $28D_2$ and implied in the response to item 21. The positive statement of this false conclusion was most frequently selected, (i.e. $28D_2$ for which the D_n was .32). The negative statement ($21D_2$) was less frequently selected (D_n = .17) and the implied statement ($21D_1$) least frequent (D_n = .07). This would seem to be reasonable. Also, this cluster holds together better than any other in cross validation (two of the three form a new three-member cluster).

It would seem, then, that this cluster is content bound, not so much on the basis of a common reading selection, but rather upon a single common contrary-to-fact conclusion formed by some of the examinees.

The O (Other) designation was therefore, retained making this wrong answer cluster O_2. Table 65 giving information on Cluster W_{14} follows.

TABLE 65

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_{14}

Foil	Content	Advance Classification	D_n
$3D_2$	Stupidity	Irr	.23
$25D_2$	Progress	Irr	.19
$9D_2$	Aggression	Sub	.17
$27D_2$	Progress	Irr	.09

Wrong answer Cluster W_{14} contained Irr foils in three out of four cases. Foil $9D_2$ cannot be reclassified as an Irr foil since it is not a true statement. However, the Irr classification of this cluster was retained on the basis of the decision rules already discussed (see: p. 237). Table 66 giving information for the interpretation of wrong answer cluster W_{15} follows (see: p. 239).

TABLE 66

INFORMATION FOR THE INTERPRETATION

OF WRONG ANSWER CLUSTER W_{15}

Foil	Content	Advance Classification	D_n
$30D_2$	Summary of all Selections	Tr	.15
$24D_3$	Progress	0	.06

Wrong answer cluster W_{15}, as given in Table 66, could not be interpreted without logico-semantic analysis because it did not contain a most frequent category by the advance foil classification. A reasonable approach would be to investigate the possibility that $24D_3$ might also be classified as Tr.

Item 24: Chester's mother has kept after him about his reading until he hates it.

$24D_3$ The hypothesis is refuted by the facts.

Several facts can be derived from Chester's "Progress Report" which have a bearing on Item 24. The three most important are:

1. "Does not enjoy reading."
2. "Has a wide speaking vocabulary."
3. "Enjoys spelling."

From these three facts two conclusions can be drawn:

1) Chester's background must be fairly verbal, hence his reading deficiency is probably not the direct product of a

background disadvantage.

2) Since he likes spelling and he "is better able to find facts than to interpret facts," his reading problem would seem to be a decoding problem.

In addition, the report gives no direct evidence about Chester's home background. Hence the best answer for this question is "insufficient evidence." In order to have the proposition presented in Item 24 refuted by the facts more would have to be known about the probable sources of the decoding problems and the side speaking vocabulary. Only if these two considerations were emphasized beyond reason could the refutation be acceptable. This conclusion might be expected to have arisen, then, from a reordering of the emphasis given to parts of the reading selection vis á vis other parts of the selection. The classification of $24D_3$ as Tr would seem to be appropriate in this case, giving the entire cluster this same interpretation.