THE EFFECTIVENESS OF A MEDIATING STRUCTURE FOR WRITING ANALYSIS LEVEL TEST ITEMS FROM TEXT BASED INSTRUCTION

DISSERTATION

Presented to the Graduate Council of the University of North Texas in Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Michael D. Brasel, A.A., B.A., M.A.
Denton, Texas
August, 1989
Brasel, Michael D., The Effectiveness of a Mediating Structure for Writing Analysis Level Test Items from Text Based Instruction. Doctor of Philosophy (Educational Research), August, 1989, 144 pp., 7 tables, bibliography, 64 titles.

This study is concerned with the effect of placing text into a mediated structure form upon the generation of test items for analysis level domain referenced test construction. The item writing methodology used is the linguistic (operationally defined) item writing technology developed by Bormuth, Finn, Roid, Haladyna and others. This item writing methodology is compared to 1) the intuitive method based on Bloom's definition of analysis level test questions and 2) the intuitive with keywords identified method of item writing.

A mediated structure was developed by coordinating or subordinating sentences in an essay by following five simple grammatical rules. Three test writers each composed a ten-item test using each of the three methodologies based on a common essay. Tests were administered to 102 Composition 1 community college students. Students were asked to read the essay and complete one test form. Test forms by writer and method were randomly delivered.
Analysis of variance showed no significant differences among either methods or writers. Item analysis showed no method of item writing resulting in items of consistent difficulty among test item writers. While the results of this study show no significant difference from the intuitive, traditional methods of item writing, analysis level test item generation using a mediating structure may yet prove useful to the classroom teacher with access to a computer. All three test writers agree that test items were easier to write using the generative rules and mediated structure. Also, some relief was felt by the writers in that the method theoretically assured that an analysis level item was written.
ACKNOWLEDGMENTS

The author would like to thank Dr. Kathleen Brasel, New Mexico Department of Corrections; Dr. Jane Close Cololey, University of Nebraska, Lincoln; Dr. Thomas Haladyna, Arizona State University; Mr. Jim Hardy, doctoral candidate, University of North Texas; Dr. Harold F. O'Neil, Jr., University of Southern California; and Dr. Gale Roid, Assessment Research, for their valuable assistance.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Overview</td>
<td></td>
</tr>
<tr>
<td>Background</td>
<td></td>
</tr>
<tr>
<td>Need</td>
<td></td>
</tr>
<tr>
<td>Statement of Problem</td>
<td></td>
</tr>
<tr>
<td>Research Question and Hypotheses</td>
<td></td>
</tr>
<tr>
<td>Significance of Study</td>
<td></td>
</tr>
<tr>
<td>II. REVIEW OF RELATED LITERATURE</td>
<td>15</td>
</tr>
<tr>
<td>III. METHODOLOGY</td>
<td>36</td>
</tr>
<tr>
<td>Subjects</td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
</tr>
<tr>
<td>Instrumentation</td>
<td></td>
</tr>
<tr>
<td>Procedures</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
</tr>
<tr>
<td>IV. RESULTS</td>
<td>53</td>
</tr>
<tr>
<td>Overview</td>
<td></td>
</tr>
<tr>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>Summary of Findings</td>
<td></td>
</tr>
<tr>
<td>V. CONCLUSIONS AND RECOMMENDATIONS</td>
<td>62</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td></td>
</tr>
<tr>
<td>Implications for Practice</td>
<td></td>
</tr>
<tr>
<td>Implications for Further Research</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>APPENDIX 1</td>
<td>69</td>
</tr>
<tr>
<td>APPENDIX 2</td>
<td>91</td>
</tr>
<tr>
<td>APPENDIX 3</td>
<td>98</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE 1:</td>
<td>Items by Writer and Method</td>
<td>39</td>
</tr>
<tr>
<td>TABLE 2:</td>
<td>Noun Classification by Standard Frequency Index</td>
<td>40</td>
</tr>
<tr>
<td>TABLE 3:</td>
<td>ANOVA TABLE (Adjusted Means)</td>
<td>48</td>
</tr>
<tr>
<td>TABLE 4:</td>
<td>Test Score Means, Standard Deviations, and Number of Participants by Writer</td>
<td>56</td>
</tr>
<tr>
<td>TABLE 5:</td>
<td>ANOVA Table--Multiple Regression with Effect Coding</td>
<td>57</td>
</tr>
<tr>
<td>TABLE 6:</td>
<td>ANOVA Table (Adjusted Means)</td>
<td>58</td>
</tr>
<tr>
<td>TABLE 7:</td>
<td>Proportion of Variance Accounted for</td>
<td>59</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Overview

Persuasive arguments have been offered for the development of item writing "technologies" for domain referenced test construction (Roid & Haladyna, 1980). Most arguments for an item writing technology center on the problem of reproducibility of results for a given measure by other researchers. Roid and Haladyna (1982) hold that the use of the intuitive skills of a test writer is, today, still the most widely used method of writing test questions for either criterion referenced or norm referenced tests, though "some quantitative indices for summarizing judges' decisions on individual items" (Crocker & Algina, 1986) have been developed, notably item-objective congruence (Ravinelli & Hambleton, 1977; Hambleton, 1980) and weighting objectives and items for correlation (Klein & Kosecoff, 1975)

Studies show, however, that individual test writers do not develop, intuitively, test items of the same difficulty level when writing test items for the same material to be tested (Roid & Haladyna, 1978). Anderson (1972) found that 50 percent of the studies he surveyed
did not provide minimal item definition, a deficiency inhibiting the reproducibility of research. Thus, Roid and Haladyna (1980) state, "If the researcher and test-item writer use mental processes that cannot be described and communicated to another educator, the process of item writing remains a private event which is not defined and, hence, not easily replicable by other researchers" (pp. 294, 295). This echoes Bormuth (1970):

> Unless the original experimenter can verify that his test items are indeed of the type he claims, and unless other experimenters can construct items which they can certify are of the same type, other experimenters cannot independently claim that they have refuted or verified the original results; and so the original study is worthless. (pp. 5, 6)

Thus, while a scientifically based technology for use in the classroom is a long-term goal of those concerned with item writing technologies, of immediate concern is the "design [of] criterion-referenced instruments for sophisticated interpretive or diagnostic purposes" (Roid, 1984, p. 49).

The purpose for test item writing technologies is to allow for the uniform generation of test items by test item writers. To date, item writing technologies have been successfully used for test item generation, mainly in mathematics and science, through the application level of
Bloom's (1956) taxonomy. While some exploration of an analysis level item writing technology has been offered by Conoley and O'Neil (1979), the exploration is tentative. Difficulties arise especially when the instruction to be tested is presented as written text.

Background

Methods of criterion referenced test item writing fall broadly into one of two categories: (1) methods derived from instructional intent and (2) methods derived from "criterion referenced item creations involving domain specifications" (Roid & Haladyna, 1980, p. 296). Contrasted to item writing methods derived from instructional intent are item writing methods derived from criterion referencing involving domain specification. This dichotomy is initially confusing, as criterion referenced (CR) testing and domain referenced testing (DR) are often used synonymously (Nitko, 1984; Berk, 1984; Roid & Haladyna, 1982). The critical difference developed by these authors is that "the objective-based CR test is viewed as allowing a weaker form of interpretation of test results in contrast to the domain-based CR test" (Roid & Haladyna, 1980, p. 296).

The search to find a method for item writing has led to the development of several item writing technologies. Item writing technologies most clearly set themselves
apart from the traditional objective-based methods not
only in the insistence upon domains, but upon test
specification constraints. Popham (1984) explains:

Once truly constraining test specifications have been
devised, then a necessary but not sufficient
condition for accurate test interpretation is
present. An additional necessary condition is that
writers follow those specifications to the letter.
Following to-the-letter writing of items is
inconsistent with creative writing of items. (p. 32)

The Linguistic Method

Of the various item writing technologies, the one
of interest in this study is the linguistic method first
developed by Bormuth (1970). Bormuth's method was capable
of assessing "learning of each of several ideas in one
sentence, either as completion or multiple choice" (Roid &
Haladyna, 1980, p. 301). The methodology uses keyword
(high frequency) nouns and singleton (high information)
nouns. Both noun types are defined as nouns with a
Standard Frequency Index of less than 60 as reported in
Carroll, Davies, and Richman (1971), but keywords appear
often in the text, whereas rare singletons would appear
only once in a passage.

Finn (1978b) confirmed the hypothesis that questions
could be written from sentences in which high information
nouns (singletons) occur. Roid (1979) states, "The Bormuth and Finn methods are domain-based rather than objective-based, in that sentences (or types of words such as rare singletons) constitute the domain of elements to be learned. This domain specification takes the place of written objectives" (emphasis added, p. 90). Roid, Haladyna, Shaughnessy, and Finn (1979) comment:
The use of sentences as a basic element of instruction is based on the logic that these are countable units of instruction which define a domain of content used in a course of instruction. Also, as Bormuth (1970) has described, by transforming sentences from instructional materials, it is possible for the difficulty of tests to be matched exactly to the readability level of the instructional materials. (p. 3)

Need

Presently, despite Bormuth's (1970) exhortation that "in principle there is no reason why any item type whatever cannot be operationally defined" (p. 56), there seems to be agreement among the experts that an item writing technology cannot be developed for testing at Bloom's evaluation and synthesis levels (Roid & Haladyna, 1980; 1982). Experts also generally agree that lower taxonomic levels through the application level may be
written (Roid & Haladyna, 1982), though Roid and Haladyna believe that technology at the comprehension level is still "primitive" (Roid & Haladyna, p. 97). "The very edge" (Roid & Haladyna, p. 110), however, of development is at the analysis level: "Higher cognitive levels of testing, such as the interrelationship between sentences, themes in a prose passage, or the integration or summation of ideas in prose...have not yet been developed, and have not been widely discussed in the literature but for a few exceptions" (Roid & Haladyna, 1980, p. 310).

Conoley & O'Neil (1979) rephrase Bloom (1956) to say: Analysis results in the student's ability to break down material into its constituent parts, to detect relationships among the parts and the way the parts are organized. It necessitates that the student recognize unstated assumptions, distinguish facts from hypotheses, and determine which are the pivotal points in an argument. Analysis is distinct from comprehension and application in its emphasis on understanding and generalizing from the underlying structure of instruction. (p. 101)

This is comparable to Gagne's (1971) higher order rules in which the student "generates" a problem solving strategy by combining two or more lower order rules or principles or, similarly, to Williams-Miller's higher levels (Miller, Williams, & Haladyna, 1978).
Rhetoricians Cleanth Brooks and Robert Penn Warren (1979) explain that while analysis is "the method of dividing into component parts", to stop here would merely be classification. Analysis occurs only if the student "grasps the principle of the relation among the parts" (pp. 86, 87).

Conoley and O'Neil (1979) present a methodology for writing test items at the analysis level. They call for the necessary identification of a "mediating structure," for which they propose a simple chapter outline. A mediating structure is a means for the test item writer to place a prose passage in symbolic form to show the relationship of the sentences to each other.

It is proposed here that a mediating structure for the development of analysis level test items will be tested using the mediating structure suggested by Conoley and O'Neil (1979) against traditional item writing methods.

Statement of Problem

The problem addressed in this study is the effect of mediating structures upon the writing of analysis level test items. Without this information, it is unknown whether the use of a mediating structure is more advantageous than the traditional approach for writing analysis level test items. Since linguistic algorithms for item development at the knowledge, comprehension, and
application levels have been written and tested (Roid, et al., 1978, 1978b, 1979, 1980, 1983), this study will explore the effects of using a mediating structure for an analysis level test item writing technology for instruction presented as written text compared to the traditional method and to the traditional method with keywords identified.

Research Question and Hypotheses

Research Question: Does the use of a mediating structure give any advantage over traditional approaches to writing analysis level test items?

Hypothesis 1

Null hypothesis: All interaction effects for treatments equal zero.

$$\mu_{ab} - \mu_{a'b} - \mu_{ab'} + \mu_{ab'} = 0$$

(Kirk, pp. 358)

Alternative hypothesis: All interaction effects for treatments do not equal zero.

$$\mu_{ab} - \mu_{a'b} - \mu_{ab'} + \mu_{ab'} \neq 0$$
Hypothesis 2

Null Hypothesis: Variance among item writing methodologies does not differ significantly.
\[ \mu_i = \mu_j \]
(where \( i \neq j \)).

Alternative Hypothesis: Variance among item-writing methodologies does differ significantly.
\[ \mu_i \neq \mu_j \]
(where \( i \neq j \)).

Hypothesis 3

Null Hypothesis: Variance among test writers does not differ significantly.
\[ \mu_i = \mu_j \]
(where \( i \neq j \)).

Alternative Hypothesis: Variance among test writers does differ significantly.
\[ \mu_i \neq \mu_j \]
(where \( i \neq j \)).

Significance of Study

Roid and Haladyna (1982, p. 110) call the development of an analysis level algorithm the "very edge" of current
research in the field. The authors also state that development might take 10 years or more. It would be presumptuous to attempt to develop the algorithm here; however, it is appropriate to explore the question: Is a mediating structure helpful?

This study will have theoretical significance for educational research in the area of testing and measurement by answering the question of whether the presence of a mediating structure in the generation of analysis level test items for prose passages has any advantage over traditional methods. It should establish whether a need for a mediating structure exists for use in some future item writing algorithm.

**Delimitations**

1. Subjects were 102 college freshmen enrolled in Composition I level writing classes at a southwestern community college.

2. Subjects were drawn from contained classrooms.

3. Research was conducted in the classroom at the instructors' permission. While all instructors were happy to allow their class to participate, some instructors felt that they could not allow class time for the administration of the research instrument. Thus, 6 classes read the essay and completed the test during class. Three classes completed the research instrument at
home and returned it at the next class meeting. Analysis of variance was conducted on the two groups. As no difference in results was found, the two groups were combined for further analysis.
CHAPTER REFERENCES


CHAPTER 2

REVIEW OF RELATED LITERATURE

Traditional Item Writing Methodology

The traditional, informal method of item writing takes item writing as an introspective art. In criterion referenced testing, traditional item development is usually based upon either (1) instructional objectives (Roid & Haladyna, 1980), (2) amplified objectives (Popham, 1978, 1984), or (3) some formal structure to aid the writer, such as the Instructional Quality Inventory (Ellis, Wulfeck, Richards, Wood, & Merrill, 1978; Roid & Haladyna, 1982, p. 187).

Roid and Haladyna (1982) hold that the intuitive skills of a test writer is still, today, the most widely used method of writing test questions for either criterion referenced or norm referenced tests. "Some quantitative indices for summarizing judges' decisions on individual items" have been developed (Crocker & Algina, 1986), notably item-objective congruence (Ravinelli & Hambleton, 1977; Hambleton, 1980) and weighting objectives and items for correlation (Klein & Kosecoff, 1975). Studies show, however, that test writers do not develop, intuitively, test items of the same difficulty (Anderson, 1972; Roid & Haladyna, 1978).
Criterion Referenced Tests

Glaser (1963) is credited with the first use of the term "criterion-referenced test" (Roid & Haladyna, 1982). "Criterion referenced tests" is a generic term for tests constructed to "tell us what it is that examinees can or can't do" (Popham, 1984) within a specified range of specific performance standards (Nitko, 1984), as opposed to norm-referenced tests, which place an examinee along a continuum of how the examinee performs on certain objectives compared with a "norming sample."

Domain Referenced Tests

Hively (1974) is credited with development of the concept of a "domain referenced test" (Roid & Haladyna, 1982). Criterion referenced (CR) testing and domain referenced (DR) testing are often used synonymously (Nitko, 1984; Berk, 1984; Roid & Haladyna, 1982), and, thus, the terms cause some confusion. The difference is that "the objective-based CR test is viewed as allowing a weaker form of interpretation of test results in contrast to the domain-based CR test (Roid & Haladyna, 1980, p. 296). Anthony Nitko (1984) elaborates: "A criterion referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specific performance standards" (p. 12). Hively (1974) states, "The basic notion underlying domain
referenced achievement testing is that certain important classes of behavior in the repertories of experts (or amateurs) can be exhaustively defined in terms of structured sets or domains of test items" (p. 5). These domains may be defined through an algorithm, i.e., an item writing "technology." Thus, all item writing technologies are domain referenced tests, as opposed to the criterion referenced test, as distinguished above.

**Item Writing Technologies**

Criterion referenced test writers attempt to define a domain by specifying what is included in the domain; domain referenced test writers operationally define the domain, exhausting the membership within the domain. Item writing "technologies" are the outgrowth of the concept of domain referenced testing.

Most arguments for an item writing technology center on the problem of reproducibility of results for a given measure by other researchers. The purpose of test item writing technologies is to allow for the uniform generation of test items by test item writers. Roid and Haladyna (1980) state, "If the researcher and test-item writer use mental processes that cannot be described and communicated to another educator, the process of item writing remains a private event which is not defined and, hence, not easily replicable by other researchers" (pp.
294, 295). This echoed Bormuth (1970), "Unless the original experimenter can verify that his test items are indeed of the type he claims, and unless other experimenters can construct items which they can certify are of the same type, other experimenters cannot independently claim that they have refuted or verified the original results; and so the original study is worthless" (pp. 5, 6).

Six item writing technologies have been developed: item forms (Hively, et al., 1968; Hively, 1974), facet-design (Foia, 1968; Engel & Martuza, 1976; Berk, 1978), concept based (Tiemann & Markle, 1978); algorithmic (Scandura, 1970; Durnin & Scandura, 1973), factor based (Guilford, 1967; Meeker & Meeker, 1975), and linguistic (Bormuth, 1970; Finn, 1975; Roid & Haladyna, 1982).

**Item Forms**

Item forms, also referred to as domain referenced (Hively, 1974; Roid, 1984), was the first item writing technology developed (Hively, Patterson, & Page, 1968) in the attempt to "generate" test items, rather that have an item writer rely on intuition. Item forms would generate items of a fixed syntactical structure or "shell," containing one or more variable elements and replacement sets for the variable elements (Osburn, 1968; Roid &
Haladyna, 1980), defining a domain of all the particular questions for a specific task (Roid & Haladyna, 1982).

**Facet Design**

Roid & Haladyna (1982) note that facet design is rooted in Foa's (1965) personality theory, but application of facet theory for criterion-referenced testing is credited to Guttman (1969). A facet design begins with a "mapping sentence" similar to an item form shell. The variable elements, however, are called "facets." The attempt here is to have a set of mapping sentences, the facets of which cover the domain to be tested. Once achieved, the writer has a "facet design" of the entire domain (Roid & Haladyna, 1982).

**Concept Based**

Tieman and Markle (1978) developed a means for item generation through analysis of a concept. Given a concept (eg., antonyms), the item writer identifies the "critical attributes" and the "variable attributes" of the concept. These attributes allow for the identification of both examples and nonexamples of the concept. Students' understanding of the concept is tested through the use of the examples and nonexamples. The testing domain is the listing of the examples and nonexamples of the concept,
and test items can be generated by substituting these examples and nonexamples (Roid & Haladyna, 1982).

Algorithmic

Scandura (1970) and Durnin and Scandura (1973) criticize item forms for not mapping student problem solving rules as part of the domain. Their methodology, then, creates item forms from a decision making algorithm that looks very much like a computer programmer's flow chart (Durnin & Scandura, 1973). By sampling from any step of the flowchart, the authors believe they provide "a theoretical basis for categorizing classes of problems" and assure "this categorization partitions the domain of problems into equivalence classes. It also provides a theoretical basis for the hierarchical relationship between tasks and takes into account the different ways in which a domain of tasks may be solved" (p. 266).

The question of the "cognitive strategy" employed by a test taker upon the results of a domain referenced test has been explored by several authors (Resnick, 1976; Brown & Burton, 1978; Birenbaum & Tatsuoka, 1982, 1983; Webb, Herman, & Cabello, 1984, 1986). Cognitive strategies assume the algorithmic approach, and these authors address the problem that the algorithm taught to students in class is not always the algorithm learned, but a "bug" (Birenbaum & Tatsuoka, 1982), and that an incorrect
algorithm for problem solving often yields a correct answer. Webb, Herman, and Cabello have done for the language arts what Birenbaum and Tatsuoka have done for mathematics in the area of diagnosis of student error.

Computers

If an algorithm for item writing can be developed, then the algorithm should make computer generation of test items feasible. To date, computer generated test item writing programs have not been feasible for the classroom teacher. However, large mainframe programs have been written; for instance, Hively's item forms have been adapted for the computer by Johnson (1973), Millman & Outlaw (1977), and Millman (1980). Roid (1986) gives a complete overview of computer technology in testing.

Linguistic

The item writing technology of interest in this study is the linguistic method, also referred to as the operationally defined method (Roid, 1984). The literature of an item writing technology based on the linguistic method grows out of the seminal work for the linguistic method, John R. Bormuth's On the Theory of Achievement Test Items (1970). There is one important antecedent in the literature of a related methodology, item forms
(described above), in which Hively, Patterson and Page (1968) demonstrate the internal consistency of transformations.

Bormuth's work is a loose adaptation of Noam Chomsky's (1957) transformational generative grammar (TG), and for those familiar with TG, many aspects are similar, such as the use of WH-transformations, base sentences, derived sentences, and so forth. Through the use of anaphoric transformations (the relationship between pronoun antecedents and postcedents [Bormuth, 1970, pp. 50-54]), Bormuth claims to be able to isolate intersentence syntax, i.e., test items may be derived not only from individual sentences but from the relationships between sentences as well. To quote Roid and Haladyna (1980), this is useful because items may be written "to assess learning of each of several ideas in one sentence, either as completion or multiple choice" (p. 301).

Cronbach (1970) agrees that "One might conclude, wrongly, that Bormuth is only interested in low level responses...Bormuth does show how intersentence syntax is used to form questions on paragraphs" (p. 510).

Bormuth (1970) argued that "in principle there is no reason why any item type whatever cannot be operationally defined" (p. 56) and, thus, a methodology designed to sample that domain. Roid and Haladyna (1982) report that "[Anderson (1972)] added a new dimension to the
measurement of prose learning by emphasizing the importance of testing at the comprehension level rather than at the level of recall" through the use of paraphrasing in which all original substantive words are replaced with synonyms of equivalent meaning (p. 91).

Finn (1975) codified the Bormuth-Anderson method in an 82-step algorithm. Finn's algorithm was modified by Finn (1978b), Roid and Finn (1978), and Roid, Haladyna, and Finn (1978a) "to make the method feasible for teachers and other test developers who may not have linguistic training" (Roid & Haladyna, 1982, pp. 94, 95).

Finn (1978b) then suggested an approach to objectively analyze a text through a computer program that utilized Carroll's, et al., (1971) Word Frequency Book. A count of the frequency of words used in the text is made, thus allowing test writers to identify high frequency words (words that appeared infrequently in the language but often in the text) and high information words (words that appeared both infrequently in the language and only once in the text). Finn hypothesized that high information singleton words would be likely candidates for linguistic transformation.

Empirical studies were conducted to test the effectiveness of the modifications on the Bormuth method. Roid, Haladyna, and Finn (1978b) showed that adjectives and nouns--as opposed to verbs--proved to be more suited
for item transformation. Roid and Haladyna (1978) found that two test writers, when given only guidelines (as opposed to an algorithm) for item construction, wrote questions of differing difficulty. When rules (an algorithm) for the way sentences were to be written were used (Roid, Haladyna, & Finn, 1978b), four different writers wrote questions of the same difficulty. In this same study, it was determined that high frequency keyword nouns were not appropriate for transformation; rather, high information rare singletons served better. It was also found that the use of paraphrasing allowed for greater instructional sensitivity (Crocker & Algina, 1986; Haladyna & Roid, 1981), i.e., it "matched the content of tests to the content of the courses in very direct ways" (Roid & Haladyna, 1982, p. 93).

A subsequent study, similar to the one above, was conducted by Roid, Haladyna, Shaughnessy, and Finn (1979) utilizing more subjects. Results confirmed the original study but showed that with the use of paraphrasing on the part of the item writer, item bias by writer seemed to rise with the "information density" (amount of testable material) of the passage, and instructional sensitivity (difference in item difficulty from pretest to posttest) decreased, perhaps because of changes in reading level. Informal, objective based, and linguistic methods were compared by Roid, Haladyna, and Shaughnessy (1980) in yet
another study; there were no surprising conclusions. Roid and Wendler (1983) demonstrated that the use of a linguistic algorithm controlled item bias.

"The very edge of the state of the art" (Roid & Haladyna, 1982) is now the development of an item writing technology at the analysis level, and the only work done has been by the team of Conoley and O'Neil (1979). Conoley and O'Neil suggest that questions at the analysis level are formed by identifying a mediating structure:

Mediating (or underlying) structures of instruction are general statements that put the instruction in symbolic language. The purpose of discovering and making explicit the underlying structure of instruction is to give the test-item writer a clean look at the relationships contained within the instruction. Once these relationships are made explicit, questions concerning constituent parts of the instruction can be generated. (p. 124)

As a mediating structure, the authors suggest the use of an outline, which may "result in a clear picture of the hierarchical relationships within instruction and may facilitate item generation" (p. 124).

Throughout the literature of the linguistic method of item writing, the "analysis level" is consistently referred to in terms of Bloom's (1956) taxonomy. The exception is a discussion in Roid and Haladyna (1982) of
"higher levels of cognition" developed by Williams (1977) and Miller, Williams, and Haladyna (1978). Conoley and O'Neil (1979) did initial analysis level work based on Bloom's (1956) taxonomy. This study will also use Bloom's (1956, 1971) definition and descriptions of analysis level performance.

Bloom (1956) defines analysis level ability as:

The breakdown of a communication into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between ideas expressed are made explicit. Such analyses are intended to clarify the communication, to indicate how the communication is organized, and the way in which it manages to convey its effects, as well as its basis and arrangement. (p. 205)

More simply stated, perhaps, the student must be presented with "new problems, materials, or situations," which the student must comprehend and "separate himself from the message to view it in terms of how it does what it does" by:

- classify[ing] words, phrases, and statements...using given analytic criteria;
- infer[ring] particular qualities or characteristics not directly stated...;
infer[ring]...underlying qualities, assumptions, or conditions...;
using...relevance, causation, and sequence...to discern pattern, order, or arrangement...;
recogniz[ing]...organizational principles or patterns...
infer[ring] the particular framework, purpose, and point of view....

(Bloom, Hastings, & Madaus, 1971, pp. 177 - 181)

Summary

Roid and Haladyna (1980) complain that "If the researcher and test-item writer use mental processes that cannot be described and communicated to another educator, the process of item writing remains a private event which is not defined and, hence, not easily replicable by other researchers" (pp. 294, 295). This study should provide a practical, significant step toward the understanding of a means for replication of analysis level test items.
CHAPTER REFERENCES


CHAPTER 3

METHODOLOGY

Subjects

Subjects were 102 community college freshmen students. Subject sample reflected the college-wide student body composition of 49 percent white non-Hispanic, 49 percent white Hispanic, and 2 percent Native American. Student population is roughly 65 percent female, 35 percent male. Traditional students (ages 18 to 24 years) represent the minority of the student body at 29 percent. Age group 25 to 35 years represents 38 percent of the student body. Age 36 years and older represents 32 percent of the student body. Ninety-one (91) percent of the student body attends college part-time.

Subjects were taken from the entry level Composition 1 classrooms. Composition 1 is a 16 week credit course required of all entering freshmen. Emphasis in the course is on the teaching and practice of rhetorical techniques. Models of such techniques are presented in a reader textbook of collected published essays. Analysis of these published essays is standard classroom practice. In as much as this research is concerned with the writing of analysis level test questions when instruction is based on
the reading of a text such as an essay, Composition 1 classes were selected as the setting for conducting this research.

Design

This study tested two treatments: item writing methodologies (treatment A) and test item writers (treatment B), with three levels each (a two-way, 3 X 3 [Kirk, 1982, Ch. 8]; Table 1).

Treatment A: Item Writing Methodologies

Level 1

Method 1 was the intuitive method of item writing. Item writers were given no directions for writing items, other than general rules for construction of good question stems and foils (Haladyna & Downing, 1988), and guidelines for analysis level questions from Bloom (1956, 1971; Appendix 1).

Level 2

Method 2 again used the intuitive method but with "keywords" identified in the text, and the writers were given a new definition of analysis level questions.
TABLE 1

Items by Writer and Method

<table>
<thead>
<tr>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>__</td>
<td>__</td>
<td>__</td>
<td>__</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Key:
Method 1 = traditional, based on Bloom's guidelines.
Method 2 = traditional, plus keywords identified and Conoley & O'Neil's definition.
Method 3 = generative, based on Roid & Haladyna's (1982) rules, plus keywords identified and rare singletons highlighted, presented in mediated structure.

(Conoley & O'Neil, 1979; Appendix 1). The use of sentences as "countable units of instruction which define a domain of content used in a course of instruction...the purest form of a criterion-referenced test" has already been established in the literature (Roid, Haladyna, Shaughnessy, & Finn, 1979). Likewise, the use of
"keyword" nouns has been established (Roid, Haladyna, & Finn, 1978b).

Identification of keywords, based on the Standard Frequency Index (SFI) found in Carroll, Davies, and Richman's (1971) Word Frequency Book is consistent throughout the literature, and this study did not deviate from that practice. Historically, identification of keywords has relied on a criterion value of less than 60 as given by the Standard Frequency Index (Roid, et al., 1978, 1978b, 1979, 1980, 1983). This study conformed to this procedure (Appendix 2). Thus, a "keyword search" was accomplished by: (a) entering the essay text into a database for a listing of frequency of noun usage in the essay and (b) classifying nouns as presented in Table 2.

TABLE 2

Noun Classification by Standard Frequency Index

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>greater than 60</th>
<th>less than 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>in text</td>
<td>HIGH</td>
<td>SINGLETON</td>
</tr>
<tr>
<td>more than</td>
<td>FREQUENCY</td>
<td>INFORMATION</td>
</tr>
<tr>
<td>once</td>
<td>RARE SINGLETON</td>
<td></td>
</tr>
<tr>
<td>only once</td>
<td>SINGLETON</td>
<td>RARE SINGLETON</td>
</tr>
</tbody>
</table>
Level 3

The third item writing methodology used a mediating structure (Conoley & O'Neil, 1979) and item generation rules (Roid & Haladyna, 1982). Keywords were again identified, but rare singleton nouns with the ten lowest SFI's were highlighted, and writers were instructed to concentrate on these rare singletons when writing a test question (Appendix 1). Finally, the essay was presented in an outline mediated structure.

Development of the Mediating Structure

Researchers argue that advantages of the linguistic approach to item writing allow the item writer to: (a) identify important sentences either through operationally defining the sentences to be chosen or as a substitute for the keyword search (Finn, 1978a; Roid & Haladyna, 1982) and (b) clarify intersentence and anaphoric (pronoun ante- and postcedent) relationships (Bormuth, 1970; Roid & Haladyna, 1982). A mediating structure should not negate these advantages. In this study, a keyword search was used to operationally define the sentences to be used for test items. Paraphrasing (see Instruction Set 3, Appendix 1) was used to clarify intersentence pronoun reference. Thus, criteria (a) and (b) above were satisfied.
A mediating structure is left undefined in the literature save for Conoley and O'Neil's "general statements that put the instruction in symbolic language" (1979, p. 124). The authors then suggest an outline as a possible mediating structure. In their description of a mediating structure, Conoley and O'Neil call for a superimposed structure beyond the simple paragraphing of the essay to allow the test writer to look anew upon the "hierarchical relationships" within the text (1979, p. 124).

A sentence outline of the essay was developed as a mediating structure (Appendix 1). For purposes of this study, the outline was completed by indenting and outdenting superordinate and subordinate sentences. Five simple rules were developed for the placement of sentences:

1. Paragraph topic sentences are always superordinate.

2. Coordinate sentences—sentences that are, or could be, connected to the previous sentence by a coordinating conjunction (and, or, but, so, for, nor, yet, and the semicolon punctuation mark)—were given the same level of ordination as the sentence that preceded it.

3. Parallel sentences are coordinate with each other.
4. Sentences containing a pronoun whose antecedent is in a preceding sentence are subordinate to the sentence containing the antecedent.

5. Illustrative examples are always subordinate.

Treatment B: Test Writers

Three test item writers were used, each with different levels of skill in writing test items. The effect of writer skill on subsequent test items can therefore be assessed. As all writers wrote items within all three methodologies, variance is controlled across item writing methodologies (Treatment A).

Level 1

Writer 1 was the only writer versed in the linguistic-generative methodology of test writing.

Level 2

Writer 2 has a doctorate in adult and higher education administration and experience with multiple-choice test item writing.

Level 3

Writer 3 is a doctoral candidate in educational research and is considered a naive writer.
Instrumentation

A sample essay of approximately 1500 words in length was selected at random from a current college level Composition 1 textbook. The textbook was also selected at random from a representative number available. The textbook selected was Patterns for College Writing: Rhetorical reader and guide (Kirszner & Mandell, 1986). The essay selected was Neil Postman's "Euphemism," an excerpt from Postman's (1976) Crazy Talk, Stupid Talk.

Analysis level test items were then developed from the essay. Postman argues in his essay that, while the abuse of euphemisms lies in their unscrupulous usage to hide an ugly reality, the use of euphemism is still to be preferred over the impolite use of "earthy" language. The essay requires the student to grasp the concept of "euphemism" and to identify both the legitimate and illegitimate use of euphemisms. As this study is concerned with the writing of analysis level questions when instruction is presented as written text, this essay was considered appropriate and representative, since the object lesson would be the essay (written text) as presented in a popular, widely used freshman Composition 1 reader.

No in-service training was given to the test writers. It was hoped that method 3 with its mediated structure and rules for item generation would afford classroom teachers
a methodology for writing analysis level questions. Given
that the average classroom instructor would not seek out a
professional test developer to write objective test items,
all test writers received the same set of instructions but
otherwise were told only to write what they believed to be
analysis level questions. If method 3 held any advantage
over methods 1 and 2, then the difference would be
apparent in the statistical testing of mean differences.
It follows, then, that no field testing of questions was
undertaken. The 10 questions for each method were
presented to the students as originally written by each
writer in test form comparable to the average essay quiz.

Procedures

Step 1

1. Each writer received a copy of the essay
reproduced as presented in the original text.

2. Each writer received instructions for (a)
traditional item writing (Appendix 1; Haladyna & Downing,
1988) and (b) directions for the use of Bloom's
analysis-level question guidelines (Bloom, et al., 1956,
1971).

3. Each writer composed ten (10) multiple-choice test
items.

4. Test items from each writer become test forms A,
B, and C, respectively (Appendix 3).
Step 2

1. Each writer received (a) the essay with keywords identified (Appendix 1) and (b) a new definition (Conoley & O'Neil, 1979) of analysis-level ability.

2. Each writer composed ten (10) multiple-choice test items.

3. Test items from each writer became test forms D, E, and F, respectively (Appendix 3).

Step 3

1. Each writer received (a) the essay with keywords identified and all rare singleton nouns highlighted in an outline "mediated structure" format (Appendix 1) and (b) Roid & Haladyna's (1982), rules for generation of multiple-choice test items (Appendix 1).

2. Each writer composed ten (10) multiple-choice test items.

3. Test items from each writer became test forms G, H, and I, respectively (Appendix 3).

The result of Steps 1 through 3 is nine different 10 item tests. Each writer wrote one test for each of the three treatment conditions (Table 1). A table of random numbers was used to effect random placement of items and item foils. Fifteen (15) reproductions of each of the tests by
writer were made for distribution, a total of 135 tests (3 forms X 3 writers X 15 copies = 135 tests).

Consent forms were read to the students and their permission to participate in the study was obtained. The consent forms guaranteed anonymity and assured the students that their participation or nonparticipation in the project would in no way affect either their rights at the college or their grade in the class.

Subjects were assigned to treatment groups by the random distribution of the tests. Tests were distributed one per student, beginning on one side of the room and working row-by-row to the other side of the room. The essay was attached to the front of the test packet. Tests had been previously arranged sequentially by writer and method (test method 1, writer 1; test method 2, writer 1;...test method 3, writer 3).

Analysis

Analysis of variance was completed through the multiple regression approach using effect coding (Pedhazur, 1982, pp. 316-323). Analysis of variance table is presented in Table 3.
Table 3

ANOVA Table (Adjusted Means)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Squares</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method (A)</td>
<td>Sum $y^2(R_{y.a,b}^2 - R_{y.b}^2)$</td>
<td>p-1</td>
<td>1/4</td>
<td></td>
</tr>
<tr>
<td>Writer (B)</td>
<td>Sum $y^2(R_{y.a,b}^2 - R_{y.a}^2)$</td>
<td>q-1</td>
<td>2/4</td>
<td></td>
</tr>
<tr>
<td>A * B</td>
<td>Sum $y^2(R_{y.a,b}^2 - R_{y.a,b}^2)$</td>
<td>(p-1)(q-1)</td>
<td>3/4</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>Sum $y^2(1 - R_{y.a,b,b}^2)$</td>
<td>pq(n-1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>Sum $y$ - Sum:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item analysis was performed to determine if any significant differences existed between student responses to individual items. Because of small sample size, an adjustment to the point biserial was used (Crocker & Algina, 1986, p. 317):

$$p_i = \frac{x - i}{\sqrt{d_i^2 + t_i^2}}$$

$$p_i(x - i) = \frac{\sigma_i - \sigma_i}{\sqrt{\sigma_i^2 + \sigma_i^2 - 2p_i \sigma_i \sigma_i}}$$
CHAPTER REFERENCES


CHAPTER 4

RESULTS

Overview

This study was an attempt to explore the effects of using a "mediating structure" (Conoley & O'Neil, 1979) for writing analysis level test items when instruction is presented as written text. Three test writers wrote three separate tests, each following three separate instruction sheets: 1) items written traditionally (i.e., intuitively) based on Bloom's (1956, 1971) guidelines for analysis level questions, 2) items written traditionally with "keywords" identified, and 3) items written following (a) Roid and Haladyna's (1982) rules for item generation, with (b) low Standard Frequency Index keywords (rare singletons) identified (Carrol, et al., 1971), and (c) the instructional text presented in an outline form as a mediated structure (Conoley & O'Neil, 1979).

Findings

Writers

Writers 1 and 3 agreed that method 2 (traditional, with keywords identified) was the easiest to use.
"Keywords" are words from the text identified to have a Standard Frequency Index (SFI) of less than 60 (Carroll, et al., 1971). These writers felt that the identification of keywords assisted or assured that "important material" was tested and, perhaps, this attitude was a reflection of their inexperience as test writers.

All writers agreed that method 3 (mediated structure with the 10 lowest SFI keywords identified) was initially the most difficult to use, though after "the first couple of questions," the writing became easier. Method 3 involved the use of paraphrase (Appendix 1), and the necessary identification of appropriate synonyms was often a challenge. Also, the low SFI for the rare singletons aside, all writers often questioned the usefulness of the words so identified. However, all writers agreed that once the technique became familiar to them, the "questions wrote themselves," and there was some general relief felt in that, as the method was in theory a guarantee of an analysis level question, the responsibility for assuring as much was removed from them.

ANOVA

Analysis of variance was completed through the multiple regression approach using effect coding. This method was preferred because unequal numbers of subjects (Pedhazur, 1982, pp. 316 - 323) completed the various test
forms designated by writer and method. No significant
differences were found at any stage of the analysis,
though the difference between the experienced writer
(writer 2) and naive writer (writer 3) neared significance
for method 3 (generative mediated) at \( p = .0509 \).

Cell means for writers by method, standard
deviations, and number of participants are given in Table
4. The mean correct test score was the dependent variable
tested. Cell means represent the average score earned on
a ten-item test by all students \((n)\) for a given test
writer using a given test writing methodology. Weighted
means for the rows represent the overall average by method
across writers. The weighted means for the columns
represent the overall average by writer across item
writing methodologies.

Unweighted means are also given and represent the
"average of the averages" for a given row or column. All
comparisons of column and row means are based on the
unweighted mean (Penhauzer, 1982, p. 323). In the case of
a particular cell, for instance, the unweighted mean is
computed when the cell is compared to the average,
disregarding cell size \((n)\). When making an analysis based
on unequal \(n\)'s, the advantage of effect coding seen in the
orthogonality (independence) of effects is lost. Thus,
the grand mean of the analysis is an unweighted grand
mean.
Table 4

Test Score Means, Standard Deviations, and Number of Participants by Writer and Method

<table>
<thead>
<tr>
<th>Writer</th>
<th>Method mean 1</th>
<th>s.d. 1</th>
<th>n 1</th>
<th>Method mean</th>
<th>s.d.</th>
<th>n 2</th>
<th>Method mean</th>
<th>s.d.</th>
<th>n 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.909</td>
<td>1.0445</td>
<td>11</td>
<td>6.0909</td>
<td>1.4459</td>
<td>11</td>
<td>5.6</td>
<td>1.5055</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>s.d. 1.4459</td>
<td>2.644</td>
<td>12</td>
<td>2.065</td>
<td>2.1044</td>
<td>12</td>
<td>6.4545</td>
<td>4.8334</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>s.d. 1.5055</td>
<td>2.464</td>
<td>12</td>
<td>1.6967</td>
<td>1.999</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

means:

Weighted: 5.875  6.0556  5.353  \text{GM: 5.765}

Unweighted: 5.8666  6.0640  5.4056  \text{5.7787}

s.d.: 1.3137  2.3169  2.1017  1.981

n: 32  36  34  N: 102

Given the low $R^2$ (proportion of variance accounted for - Table 5), it was deemed inappropriate to conduct

Table 5 shows the regression analysis of variance. Of interest here is the nonsignificant F-ratio and the overall $R^2$. This analysis indicates that only about 7 percent ($R^2 = 0.0709 \times 100 = 7\%$) of the variance is accounted for by the data.

Analysis of variance based on adjusted means is shown in Table 6. For nonorthogonal designs (designs with unequal sample sizes), analysis of variance based on adjusted means is preferred because "each main effect [treatment] is tested after it has been adjusted for its correlation with the other main effect" (Pedhauzer, 1982, p. 377). Overall F-ratios were found to be nonsignificant.

Table 5

ANOVA Table--Multiple Regression with Effect Coding

<table>
<thead>
<tr>
<th>Effect</th>
<th>SS</th>
<th>df</th>
<th>ms</th>
<th>F</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>28.15</td>
<td>8</td>
<td>3.514</td>
<td>.888</td>
<td>.0709</td>
</tr>
<tr>
<td>Residual</td>
<td>368.238</td>
<td>93</td>
<td>3.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>396.353</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proportions of variance accounted for by the interaction of effects and, then, the main effects were
tested (Pedhauzer, 1982, p. 377). Main effects interaction (A*B) tested nonsignificant. Tests for main effects followed. Neither the effect of item writing methodology (A) nor the effect for writer (B) tested significant. \( R^2 \) and analysis is given in Table 7.

Table 6

ANOVA Table (Adjusted Means)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Sum of Squares</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method(A)</td>
<td>5.51575</td>
<td>2</td>
<td>2.75788</td>
<td>.69651</td>
</tr>
<tr>
<td>Writer(B)</td>
<td>8.34178</td>
<td>2</td>
<td>4.17089</td>
<td>1.05338</td>
</tr>
<tr>
<td>A * B</td>
<td>13.4001</td>
<td>4</td>
<td>3.35002</td>
<td>.84606</td>
</tr>
<tr>
<td>Within</td>
<td>368.238</td>
<td>93</td>
<td>3.95955</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>101</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sum \( y \)-sqr: 396.353  Sum: 395.495

As no main effect proved significant, comparisons of individual means were not completed.

Item Analysis

Item means (item difficulty \( p \)), variance, and point biserial for each test by writer and method are given in Appendix 4. Point biserial correlation was adjusted for small sample size (Crocker & Algina, 1986, p. 371). None
Table 7

Proportion of Variance Accounted for

| R-square |
|------------------|---|---|---|---|
| y.a,b,ab         | 0.07093 | 0.03713 | 0.01608 | 0.02321 |
| y.a,b            |       |       |       |       |
| y.a              |       |       |       |       |
| y.b              |       |       |       |       |

Interaction:

\[
\frac{(R^2_\text{y.a,b,ab} - R^2_\text{y.ab})}{(df_1 - df_2)}
\]

\[F = \frac{(1 - R^2_\text{y.a,b,ab})}{(N - df_1 - 1)} = 0.84753 \text{ (n.s.)}\]

Main Effect A (Methods):

\[
\frac{(R^2_\text{y.a,b} - R^2_\text{y.b})}{(df_1 - df_2)}
\]

\[F = \frac{(1 - R^2_\text{y.a,b})}{(N - df_1 - 1)} = 0.70095 \text{ (n.s.)}\]

Main Effect B (Writers):

\[
\frac{(R^2_\text{y.a,b} - R^2_\text{y.a})}{(df_1 - df_2)}
\]

\[F = \frac{(1 - R^2_\text{y.a,b})}{(N - df_1 - 1)} = 1.06011 \text{ (n.s.)}\]

of the point biserial correlation coefficients met the critical value (Ferguson, 1981, Table D, p. 523).
Summary of Findings

Writers agreed that both method 2 and method 3 were preferable over method 1, the traditional approach to item writing, with a marked preference for method 2, the traditional approach with keywords identified. Method 3, mediated structure with low SFI words identified, required considerably more effort on the part of the item writer, notably in the search for appropriate synonyms for paraphrasing, and was not well received.

Data analyzed accounted for only 7 percent of variance. No test proved significant, though writer 2 and writer 3 approached significance on method 3 (p = .0509). All null hypotheses advanced were retained as not significant at the .05 level.
CHAPTER REFERENCES


61
CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Introduction

Conoley and O'Neil (1979) have suggested that if an algorithm for analysis level test item writing were to be developed, that algorithm would require a mediating structure. The purpose of this research was to contribute to the understanding of (1) a mediating structure in test item writing and (2) the necessity for such a mediating structure in the future development of an item writing algorithm for analysis level questions when instruction is based on a text, such as an essay.

Conclusions

The major conclusion to be drawn from this study is that a mediating structure as advanced here affords no significant advantage over the traditional item writing methodology for analysis level items. This may be a result of an apparent dilemma present in the mediated structure developed for this study: While the outline and identification of rare singletons may assist the test writer to perceive "hierarchical relationships," how the relationships are perceived by the writer still remains.

62
apparently, filtered by the writer's intuitive skills with no guarantee that any two writers would write equivalent questions that were any better than those written traditionally. Neither the analysis of variance based on the mean correct student score nor the item analysis based on the mean correct item score showed any difference between methods.

Second, an outline, at least as defined in this study, is not a sufficient mediating structure. While the outline used here may, indeed, assist a writer to write a test item, there is no evidence given to suggest that another researcher would reproduce similar questions. This conclusion is based on the one difference that approached statistical significance in this study (p = .0509), the difference between writers 2 and 3 when both were using method 3, the mediated method.

Third, the use of a mediated structure, as defined here, with low SPI keywords identified may assist the individual classroom teacher who is experienced in item writing, or who receives in-service training, to write more consistent parallel or equivalent test forms. This conclusion is based on the coefficient alpha (KR20) for writer 2 (experienced) for methods 2 and 3.

Fourth, the simple identification of keywords with a Standard Frequency Index of less than 60 may be a sufficient mediating structure for the experienced test
writer. Again, coefficient alpha (KR20) for writer 2 (experienced) for methods 2 and 3 would suggest as much.

Fifth, method 2, the simple identification of keywords, may be sufficient to meet the needs of the inexperienced classroom teacher. This conclusion is based on the lack of any significant difference between method 2 and method 3 for any writer.

Finally, the methodology, given current computer technology available to the average classroom teacher, probably would not be used. To use either method 2 or method 3, the instructor would need access to the mainframe computer version of the Word Frequency Book. To attempt to do this manually each time is simply unrealistic.

Implications for Practice

Professional test writers presently have access to software and computers that will give them the Standard Frequency Index of words in text, as well as make appropriate paraphrasing of a question through synonym substitution. This, however, only assures an application level question (Roid & Haladyna, 1982). A mediating structure, because of its reliance on the ability (intuition) of a writer to see "hierarchical relationships" within the text, would be of little use given current computer capabilities.
Software able to give at least the identification of keywords through some means like the Standard Frequency Index could very well be of value to the classroom teacher familiar with the requisites of writing an analysis level test item. Software is available for outlining. However, for the classroom instructor who would have to do the SFI by hand, the pragmatics of such an undertaking on any regular basis would be prohibitive.

Implications for Further Research

Future studies should concentrate closely upon the definition and development of a mediating structure. It would be incorrect to conclude from this study that a mediating structure cannot be developed, defined, or that one is inappropriate at the analysis level. This study only attempted to give one definition, and based on that definition, found no reason to believe it to be effective. One means of approaching a definition of a mediating structure, if one wished to pursue the outline as a mediating structure, would be to use software that will produce an outline of text to give the researcher a replicable structure. It would not be appropriate, however, to assume that the first sentence of a paragraph is always a topic sentence, or that even the identification of all topic sentences would produce a mediating structure.
One is still left with the problem of finding a way of forcing writers to perceive "hierarchical relationships" in a similar way. In this study, the attempt was made by forcing the writers to concentrate (in method 3) on the ten keywords with the lowest SFI. There is no reason to believe this was successful; indeed, the only difference in the study that approached significance was between two writers using method 3. Future studies might, for instance, restrict the writer to incorporating references to the sentence immediately preceding or following the sentence with the identified keyword, rather than the entire paragraph.

Future studies, too, may want to find more equally trained writers and afford these writers with in-service training in method 3. This study did not have an in-service training session because the methodology under study was presumed to provide analysis level questions, and it was hoped that the comparison between this method and more traditional methods would show this advantage. It did not.

Field testing more items before final test development would be helpful. Item analysis could then be used to remove inappropriate or ineffective items before a final test is constructed. In this study, for instance, one writer used analogy to test the concept of a euphemism. Students, through analogy, were to identify
the euphemism of a "computer virus" as a type of computer program. Not one student answered this question correctly. While the question may or may not have been a "good" analysis level item, it was obviously outside the experience of the experimental sample and thus inappropriate as a test question.

Summary

The following conclusions were derived from the results of this study:

1. An outline, as defined in this study, is not a sufficient mediating structure for forcing item writers to perceive similar relationships in the text.

2. The use of a mediated structure, as defined here, with low SFI keywords identified may assist the individual classroom teacher to write more consistent parallel or equivalent test forms, though not necessarily at the analysis level.

3. Method 2, the simple identification of keywords, may be sufficient to meet the needs of the classroom teacher.

4. Neither methodology 2 or 3, given current computer technology, would be of any practical use to the individual classroom teacher.
CHAPTER REFERENCES


APPENDIX 1

TEST WRITERS' INSTRUCTIONS
INSTRUCTIONS SET #1

You are attempting to write analysis-level multiple-choice test items (Bloom, 1956). Please type your test items. Thank you.

Directions

1. Read the attached essay "Euphemisms" by Neil Postman.

2. Write ten multiple choice questions:

   A. Each question set should consist of

      1) an item stem (question prompt)
      2) four (4) responses (the correct answer and three foils)
      3) an asterick (*) marking the correct response
      4) random placement of the correct response.

   B. Check each question set against the guidelines for correctness (attached).

   C. Questions should be written to elicit from the student evidence that the student can do one of the following:

      1) uncover unique characteristics,
      2) distinguish between facts and inferences,
      3) evaluate relevancy of data,
      4) recognize logical fallacies of reasoning,
      5) recognize unstated assumptions,
      6) analyse organizational structure,
      7) compare/contrast ideas, themes, data, etc.,
      8) recognize subordination of examples to main theme/topic of essay or paragraph,
      9) identify problems,
     10) identify attributes,
     11) analyse morphology (i.e., differences among parts of speech: eg.: "discern" and "discernment").
Item Writing Guidelines/Rules/Suggestions/Advice
as Derived From Textbooks

General Item-Writing (Procedural)

1. Use either the best answer or the correct answer format.

2. Avoid complex multiple-choice (Type K) items.

3. Format the item vertically not horizontally.

4. Allow time for editing and other types of item revisions.

5. Use good grammar, punctuation, and spelling consistently.

6. Minimize examinee reading time in phrasing each item.

7. Avoid trick items, those which mislead or deceive examinees into answering incorrectly.

General Item-Writing (Content Concerns)

8. Base each item on an educational or instructional objective.


10. Keep the vocabulary consistent with the examinee's level of understanding.

11. Avoid cuing one item with another; keep items independent of one another.

12. Use the author's examples as a basis for developing your items.

13. Avoid overspecific knowledge when developing the item.

14. Avoid textbook, verbatim phrasing when developing the item.

15. Avoid items based on opinions.

16. Use multiple-choice to measure higher level thinking.

17. Test for important or significant material; avoid trivial material.
Stem Construction

18. State the stem in question form.

19. When using the completion format, don't leave a blank for completion in the beginning or middle of the stem.

20. Ensure that the directions in the stem are clear, and that wording lets the examinee know exactly what is being tested.

21. Avoid window dressing (excessive verbiage) in the stem.

22. Word the stem positively; avoid negative phrasing.

23. Include the central idea and most of the phrasing in the stem.

General Option Development

24. Use as many functional distractors as are feasible.

25. Place options in logical or numerical order.

26. Keep options independent; options should not be overlapping.

27. Keep all options in an item homogeneous in content.

28. Keep the length of options fairly consistent.

29. Avoid, or use sparingly, the phrase "all of the above."

30. Avoid, or use sparingly, the phrase "none of the above."

31. Avoid the use of the phrase "I don't know."

32. Phrase options positively, not negatively.

33. Avoid distractors that can clue testwise examinees; for example, avoid slang associations, absurd options, formal prompts, or semantic (overly specific or overly general) clues.

34. Avoid giving clues through the use of faulty grammatical construction.

35. Avoid specific determiners, such as "never" and "always."
Correct Option Development.

36. Position the correct options so that it appears about the same number of times in each set of items.

37. Make sure there is one and only one correct option.

Distractor Development

38. Use plausible distractors; avoid illogical distractors.

39. Incorporate common errors of students in distractors.

40. Avoid technically phrased distractors.

41. Use familiar yet incorrect phrases as distractors.

42. Use true statements that do not correctly answer the item.

43. Avoid use of humor when developing options.

1 Haladyna and Downing, 1988.
INSTRUCTIONS SET #2

You are again attempting to write "analysis-level" multiple-choice questions, but based on a different set of criteria.

1. Please look at the attached essay.

2. You will notice it is the same essay you worked with in Instruction Set #1. "Keywords" have been identified for you. These keywords have a low standard frequency index in textbook usage. Because they only appear once (in noun form) in the essay, a high amount of instruction information is assumed to be associated with each.

3. Consider the following definition and brief explanation of analysis given by Conoley & O'Neil, Jr. (1979):

Analysis results in the student's ability to break down material into its constituent parts, to detect relationships among the parts and the way the parts are organized. It necessitates that the student recognize unstated assumptions, distinguish facts from hypotheses, and determine which are the pivotal points in an argument. Analysis is distinct from comprehension and application in its emphasis on understanding and generalizing from the underlying structure of [textual] instruction. (p. 101)

4. Considering the definition given above, write ten (10) multiple-choice analysis-level questions using ten (10) keywords of your choice, one keyword per question. Do not refer to or feel constrained by Instructions Set #1. Select those words based on your intuition that best meet the definition given above.

5. Please type your question set, mark correct responses, check for item-writing violation against guidelines for correctness.
"Euphemism"
by
Neil Postman

A euphemism is commonly defined as an auspicious or exalted term (like "sanitation engineer") that is used in place of a more down-to-earth term (like "garbage man"). People who are partial to euphemisms stand accused of being "phony" or of trying to hide what it is they are really talking about. And there is no doubt that in some situations the accusation is entirely proper. For example, one of the more detestable euphemisms I have come across is recent years is the term "Operation Sunshine," which is the name the U.S. Government gave to some experiments it conducted with the hydrogen bomb in the South Pacific. It is obvious that the government, in choosing the name, was trying to expunge the hideous imagery that the bomb evokes and in so doing committed, as I see it, an immoral act. This sort of process--giving pretty names to essentially ugly realities--is what has given euphemizing such a bad name. And people like George Orwell have done valuable work for all of us in calling attention to how the process works. But there is another side to euphemizing that is worth mentioning, and a few words here in its defense will not be amiss.
To begin with, we must keep in mind that things do not have "real" names, although many people believe that they do. A garbage man is not "really" a "garbage man," any more than he is really a "sanitation engineer." And a pig is not called a "pig" because it is so dirty, nor a shrimp a "shrimp" because it is so small. There are things, and then there are the names of things, and it is considered a fundamental error in all branches of semantics to assume that a name and a thing are one and the same. It is true, of course, that a name is usually so firmly associated with the thing it denotes that it is extremely difficult to separate one from the other. That is why, for example, advertising is so effective. Perfumes are not given names like "Bronx Odor," and an automobile will never be called "The Lumbering Elephant." Shakespeare was only half right in saying that a rose by any other name would smell as sweet. What we call things affects how we will perceive them. It is not only harder to sell someone a "horse mackerel" sandwich than a "tuna fish" sandwich, but even though they are the "same" thing, we are likely to enjoy the taste of tuna more than that of the horse mackerel. It would appear that human beings almost naturally come to identify names with things, which is one of our more fascinating illusions. But there is some substance to this illusion. For if you change the names of things, you change how people will regard them,
and that is as good as changing the nature of the thing itself.

Now, all sorts of scoundrels know this perfectly well and can make us love almost anything by getting us to transfer the charm of a name to whatever worthless thing they are promoting. But at the same time and in the same vein, euphemizing is a perfectly intelligent method of generating new and useful ways of perceiving things. The man who wants us to call him a "sanitation engineer" instead of a "garbage man" is hoping we will treat him with more respect than we presently do. He wants us to see that he is of some importance to our society. His euphemism is laughable only if we think that he is not deserving of such notice or respect. The teacher who prefers us to use the term "culturally different children" instead of "slum children" is euphemizing, all right, but is doing it to encourage us to see aspects of a situation that might otherwise not be attended to.

The point I am making is that there is nothing in the process of euphemizing itself that is contemptible. Euphemizing is contemptible when a name makes us see something that is not true or diverts our attention from something that is. The hydrogen bomb kills. There is nothing else that it does. And when you experiment with it, you are trying to find out how widely and well it kills. Therefore, to call such an experiment "Operation
Sunshine" is to suggest a purpose for the bomb that simply does not exist. But to call "slum children" "culturally different" is something else. It calls attention, for example, to legitimate reasons why such children might feel alienated from what goes on in school.

I grant that sometimes such euphemizing does not have the intended effect. It is possible for a teacher to use the term "culturally different" but still be controlled by the term "slum children" (which the teacher may believe is their "real" name). "Old people" may be called "senior citizens," and nothing might change. And "lunatic asylums" may still be filthy, primitive prisons though they are called "mental institutions." Nonetheless, euphemizing may be regarded as one of our more important intellectual resources for creating new perspectives on a subject. The attempt to rename "old people" "senior citizens" was obviously motivated by a desire to give them a political identity, which they not only warrant but which may yet have important consequences. In fact, the fate of euphemisms is very hard to predict. A new and seemingly silly name may replace an old one (let us say, "chairperson" for "chairman") and for years no one will think or act any differently because of it. And then, gradually, as people begin to assume that "chairperson" is the "real" and proper name (or "senior citizen" or "tuna fish" or "sanitation engineer"), their attitudes begin to
shift, and they will approach things is a slightly different frame of mind. There is a danger, of course, in supposing that a new name can change attitudes quickly or always. There must be some authentic tendency or drift in the culture to lend support to the change, or the name will remain incongruous and may even appear ridiculous. To call a teacher a "facilitator" would be such an example. To eliminate the distinction between "boys" and "girls" by calling them "childpersons" would be another.

But to suppose that such changes never "amount to anything" is to underestimate the power of names. I have been astounded not only by how rapidly the name "blacks" has replaced "Negroes" (a kind of euphemizing in reverse) but also by how significantly perceptions and attitudes have shifted as an accompaniment to the change.

The key idea here is that euphemisms are a means through which a culture may alter its imagery and by so doing subtly change its style, its priorities, and its values. I reject categorically the idea that people who use "earthy" language are speaking more directly or with more authenticity than people who employ euphemisms. Saying that someone is "dead" is not to speak more plainly or honestly than saying he has "passed away." It is, rather, to suggest a different conception of what the event means. To ask where the "shithouse" is, is no more to the point than to ask where the "restroom" is. But in
the difference between the two words, there is expressed a vast difference in one's attitude toward privacy and propriety. What I am saying is that the process of euphemizing has no moral content. The moral dimensions are supplied by what the words in question express, what they want us to value and to see. A nation that calls experiments with bombs "Operation Sunshine" is very frightening. On the other hand, a people who call "garbage men" "sanitation engineers" can't be all bad.
INSTRUCTIONS SET #3

1. Please look at the attached essay.

2. Keywords again have been identified, but ten (10) of these have been identified with BOLD CAPITALIZATION.

3. The essay has also been placed in outline form for you to show the relationship among the passages. Consider the outline structure of the essay. The outline structure is given to you as an aid to graphically demonstrate the "hierarchical relationships within [the] instruction." It is hoped the outline structure will assist you "to devise items that present the students with novel problems while maintaining logical relevance to the textual material."

4. Using the ten especially identified keywords (BOLD CAP'S), write one question for each, following the procedure below:

I. CONSTRUCT A BASE SENTENCE

A. Copy the sentence in which the keyword (BOLD CAP'S) appears,

B. Clarify pronoun references to other sentences by replacing the pronoun with the word to which it refers,

C. Simplify the sentence by taking out unnecessary clauses IF doing so
   1) does not limit meaning
   2) does not make the sentence ungrammatical
   3) does not make the sentence an incomplete thought.

D. You have now made the base sentence.

II. CONSTRUCT A PARAPHRASE SENTENCE

Paraphrase (restate) the base sentence

A. replace every general term (eg.: "tool") in the base sentence with a specific term (eg: "hammer") OR replace every particular term with a general term (eg., replace "hammer" with "tool");

NOTE: proper nouns cannot be paraphrased;
B. substitute synonyms for every remaining substantive word (noun, verb, adjective or adverb),

C. check your paraphrase

1) the paraphrased and base sentences should have no substantive word in common--i.e., nouns, verbs, and modifiers,

2) the paraphrased and base sentences should be equivalent in meaning.

III. CONSTRUCT THE ITEM TEST QUESTION STEM

either 1) delete a segment from the paraphrase (i.e., fill-in-the-blank with multiple choice answers),

or 2) change the paraphrase into a WH-question by deleting a substantive word and making the appropriate selection below:

A. use "who" or "what" to replace human and nonhuman nouns, respectively,

B. use "which" to replace noun modifiers,

C. use "how", "when", "where", and "why" to replace verb modifiers,

D. use a form of "what do" when the verb is deleted.

EXAMPLE

(from "A Slice of Life" by Russell Baker)

Sample original sentence: If you do this correctly, which is improbable, the knife will almost immediately encounter a barrier of bone and gristle.

Sample base sentence: If the student inserts the knife between the thigh and torso correctly, the knife will almost immediately encounter a barrier of bone and gristle.

Sample paraphrased sentence: The cutting edge of the blade will strike the joint when properly inserted between the thigh and torso.
Sample test item (Method 1)

Q: The cutting edge of the blade will strike the [ ] when properly inserted between the thigh and torso.
   a. thumb
   b. thighbone
* c. joint
   d. barrier

Sample test item (Method 2)

Q: What should the cutting edge of the blade strike if properly inserted between the thigh and torso?
   a. thumb
   b. thighbone
* c. joint
   d. barrier

IV. USE OF OUTLINE STRUCTURE

The above example does NOT take into consideration an outline structure, and is used only to demonstrate the mechanics involved. Thus, as far as it goes, the example is one of the cognitive, or, at best, the application, level. Given the following outline of the paragraph:

You are now ready to begin carving.

Sharpen the knife on the stone and insert it where the thigh joins the torso.

If you do this correctly, which is improbable, the knife will almost immediately encounter a barrier of bone and GRISTLE.

This may very well be the joint.

It could, however, be your thumb.

If not, execute a vigorous sawing motion until satisfied that the knife has been defeated.

Withdraw the knife and ask someone nearby, in as testy a manner as possible, why the knives at your house are not kept in better carving condition.
Given the ironic tone of the passage, you might wish to write a very different type of "base" sentence that accounts for this, such as:

Sample base sentence: If the student inserts the knife between the thigh and torso correctly, the knife will almost immediately encounter a barrier, which may be the joint or the thumb.

Sample paraphrased sentence: Ironically, the cutting edge of the blade will strike either the turkey's joint or the carver's thumb when properly inserted between the thigh and torso.

Thus, an analysis-level question would take into account this structure:

Q: If properly inserted, the cutting blade will strike the joint or, _______ the carver's thumb.
   a. sadly
   b. luckily
   * c. ironically
   d. tragically

OR

Q: The cutting blade striking either the joint or the carver's thumb is like a chainsaw hitting either a ______ or the lumberjack's leg.
   a. limb
   * b. knot
   c. rock
   d. bird

OR

Q: How does the author set a satiric tone for carving the turkey?
   a. By suggesting all carving knives are dull.
   b. By suggesting turkey carving is easy.
   c. By suggesting we enjoy carving turkeys.
   * d. By suggesting we might cut ourselves along with the turkey.

The point is: USE both the paraphrased sentence AND the outline structure to help you come up with unique
situations for the student to analyse.

5. Check foils for correctness, etc.
"Euphemism"
by
Neil Postman

A euphemism is commonly defined as an auspicious or exalted term (like "sanitation engineer") that is used in place of a more down-to-earth term (like "garbage man").

People who are partial to euphemisms stand accused of being "phony" or of trying to hide what it is they are really talking about.

And there is no doubt that in some situations the ACCUSATION is entirely proper.

For example, one of the more detestable euphemisms I have come across is recent years is the term "Operation Sunshine," which is the name the U.S. Government gave to some experiments it conducted with the hydrogen bomb in the South Pacific.

It is obvious that the government, in choosing the name, was trying to expunge the hideous imagery that the bomb evokes and in so doing committed, as I see it, an immoral act.

This sort of process—giving pretty names to essentially ugly realities—is what has given euphemizing such a bad name.

And people like George Orwell have done valuable work for all of us in calling attention to how the process works.

But there is another side to euphemizing that is worth mentioning, and a few words here in its defense will not be amiss.

To begin with, we must keep in mind that things do not have "real" names, although many people believe that they do.

A garbage man is not "really" a "garbage man," any more than he is really a "sanitation engineer."

And a pig is not called a "pig" because it is so dirty, nor a shrimp a "shrimp" because it is so small.

There are things, and then there are the names of things, and it is considered a fundamental error in all branches
of **SEMANTICS** to assume that a name and a thing are one and the same.

It is true, of course, that a name is usually so firmly associated with the thing it denotes that it is extremely difficult to separate one from the other.

That is why, for example, advertising is so effective.

Perfumes are not given names like "Bronx Odor," and an automobile will never be called "The Lumbering Elephant."

Shakespeare was only half right in saying that a rose by any other name would smell as sweet.

What we call things affects how we will perceive them.

It is not only harder to sell someone a "horse mackerel" sandwich than a "tuna fish" sandwich, but even though they are the "same" thing, we are likely to enjoy the taste of tuna more than that of the horse mackerel.

It would appear that human beings almost naturally come to identify names with things, which is one of our more fascinating ILLUSIONS.

But there is some substance to this illusion.

For if you change the names of things, you change how people will regard them, and that is as good as changing the nature of the thing itself.

Now, all sorts of SCOUNDRELS know this perfectly well and can make us love almost anything by getting us to transfer the charm of a name to whatever worthless thing they are promoting.

But at the same time and in the same vein, euphemizing is a perfectly intelligent method of generating new and useful ways of perceiving things.

The man who wants us to call him a "sanitation engineer" instead of a "garbage man" is hoping we will treat him with more respect than we presently do.
He wants us to see that he is of some importance to our society.

His euphemism is laughable only if we think that he is not deserving of such notice or respect.

The teacher who prefers us to use the term "culturally different children" instead of "slum children" is euphemizing, all right, but is doing it to encourage us to see aspects of a situation that might otherwise not be attended to.

The point I am making is that there is nothing in the process of euphemizing itself that is contemptible.

Euphemizing is contemptible when a name makes us see something that is not true or diverts our attention from something that is.

The hydrogen bomb kills.

There is nothing else that it does.

And when you experiment with it, you are trying to find out how widely and well it kills.

Therefore, to call such an experiment "Operation Sunshine" is to suggest a purpose for the bomb that simply does not exist.

But to call "slum children" "culturally different" is something else.

It calls attention, for example, to legitimate reasons why such children might feel alienated from what goes on in school.

I grant that sometimes such euphemizing does not have the intended effect.

It is possible for a teacher to use the term "culturally different" but still be controlled by the term "slum children" (which the teacher may believe is their "real" name).

"Old people" may be called "senior citizens," and nothing might change.

And "lunatic asylums" may still be filthy, primitive PRISONS though they are called "mental institutions."
Nonetheless, euphemizing may be regarded as one of our more important intellectual resources for creating new PERSPECTIVES on a subject.

The attempt to rename "old people" "senior citizens" was obviously motivated by a desire to give them a political identity, which they not only warrant but which may yet have important consequences.

In fact, the fate of euphemisms is very hard to predict.

A new and seemingly silly name may replace an old one (let us say, "chairperson" for "chairman") and for years no one will think or act any differently because of it.

And then, gradually, as people begin to assume that "chairperson" is the "real" and proper name (or "senior citizen" or "tuna fish" or "sanitation engineer"), their attitudes begin to shift, and they will approach things in a slightly different frame of mind.

There is a danger, of course, in supposing that a new name can change attitudes quickly or always.

There must be some authentic tendency or drift in the culture to lend support to the change, or the name will remain incongruous and may even appear ridiculous.

To call a teacher a "facilitator" would be such an example.

To eliminate the DISTINCTION between "boys" and "girls" by calling them "childpersons" would be another.

But to suppose that such changes never "amount to anything" is to underestimate the power of names.

I have been astounded not only by how rapidly the name "blacks" has replaced "Negroes" (a kind of euphemizing in reverse) but also by how significantly PERCEPTIONS and attitudes have shifted as an accompaniment to the change.

The key idea here is that euphemisms are a means through which a culture may alter its imagery and by so doing subtly change its style, its PRIORITIES, and its values.
I reject categorically the idea that people who use "earthy" language are speaking more directly or with more AUTHENTICITY than people who employ euphemisms.

Saying that someone is "dead" is not to speak more plainly or honestly than saying he has "passed away."

It is, rather, to suggest a different conception of what the event means.

To ask where the "shithouse" is, is no more to the point than to ask where the "restroom" is.

But in the difference between the two words, there is expressed a vast difference in one's attitude toward privacy and propriety.

What I am saying is that the process of euphemizing has no moral content.

The moral dimensions are supplied by what the words in question express, what they want us to value and to see.

A nation that calls experiments with bombs "Operation Sunshine" is very frightening.

On the other hand, a people who call "garbage men" "sanitation engineers" can't be all bad.
APPENDIX 2

KEYWORD FREQUENCY
## Noun Listing

HF = High freq; SING = Singleton; HI = High info; RARE = Rare sing

<table>
<thead>
<tr>
<th>SFI:</th>
<th>&gt;60</th>
<th>&lt;60</th>
<th>=1</th>
<th>=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>FREQ:</td>
<td>&gt;1</td>
<td>=1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Noun</th>
<th>SFI</th>
<th>Occurs</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE</td>
<td>75.9</td>
<td>6</td>
<td>HF</td>
</tr>
<tr>
<td>PEOPLE</td>
<td>71.3</td>
<td>10</td>
<td>HF</td>
</tr>
<tr>
<td>WORDS</td>
<td>70.5</td>
<td>3</td>
<td>HF</td>
</tr>
<tr>
<td>MAN</td>
<td>69.9</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>THINGS</td>
<td>68.8</td>
<td>8</td>
<td>HF</td>
</tr>
<tr>
<td>YEARS</td>
<td>68.6</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>NAME</td>
<td>68.0</td>
<td>14</td>
<td>HF</td>
</tr>
<tr>
<td>SOMETHING</td>
<td>67.1</td>
<td>3</td>
<td>HF</td>
</tr>
<tr>
<td>CHILDREN</td>
<td>66.7</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>THING</td>
<td>65.4</td>
<td>5</td>
<td>HF</td>
</tr>
<tr>
<td>CHANGE</td>
<td>65.3</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>EXAMPLE</td>
<td>65.3</td>
<td>4</td>
<td>HF</td>
</tr>
<tr>
<td>POINT</td>
<td>65.0</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>FISH</td>
<td>64.5</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>NOTHING</td>
<td>64.0</td>
<td>3</td>
<td>HF</td>
</tr>
<tr>
<td>IDEA</td>
<td>63.9</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>ANYTHING</td>
<td>63.4</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>MIND</td>
<td>62.7</td>
<td>2</td>
<td>HF</td>
</tr>
<tr>
<td>NAMES</td>
<td>62.3</td>
<td>6</td>
<td>HF</td>
</tr>
<tr>
<td>TEACHER</td>
<td>61.5</td>
<td>4</td>
<td>HF</td>
</tr>
<tr>
<td>DIFFERENCE</td>
<td>61.1</td>
<td>2</td>
<td>HF</td>
</tr>
</tbody>
</table>
### Noun Listing

**HF**=HIGH FREQ; **SING**=SINGLETON; **HI**=HIGH INFO; **RARE**=RARE SING

<table>
<thead>
<tr>
<th>NOUN</th>
<th>SFI</th>
<th>OCCURS</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OTHER</td>
<td>73.2</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>TIME</td>
<td>72.1</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>SAME</td>
<td>69.7</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>WORK</td>
<td>69.2</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>PLACE</td>
<td>69.0</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>66.9</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>SIDE</td>
<td>66.6</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>KIND</td>
<td>66.4</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>HAND</td>
<td>66.3</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>BOYS</td>
<td>65.9</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>MEANS</td>
<td>65.5</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>WAYS</td>
<td>65.5</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>COURSE</td>
<td>64.1</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>NOTICE</td>
<td>63.2</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>GIRLS</td>
<td>62.9</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>POWER</td>
<td>62.7</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>FACT</td>
<td>62.4</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>QUESTION</td>
<td>62.1</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>CHANGES</td>
<td>61.8</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>LANGUAGE</td>
<td>61.8</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>GOVERNMENT</td>
<td>61.3</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>SUBJECT</td>
<td>60.5</td>
<td>1</td>
<td>SING</td>
</tr>
<tr>
<td>NOUN</td>
<td>SFI</td>
<td>OCCURS</td>
<td>TYPE</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>ATTENTION</td>
<td>59.4</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>PROCESS</td>
<td>58.9</td>
<td>4</td>
<td>HI</td>
</tr>
<tr>
<td>EXPERIMENTS</td>
<td>55.8</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>TERM</td>
<td>55.8</td>
<td>6</td>
<td>HI</td>
</tr>
<tr>
<td>CITIZENS</td>
<td>55.3</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>RESPECT</td>
<td>55.1</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>PIG</td>
<td>53.4</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>CULTURE</td>
<td>53.1</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>ATTITUDES</td>
<td>50.3</td>
<td>3</td>
<td>HI</td>
</tr>
<tr>
<td>SANDWICH</td>
<td>49.9</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>BOMB</td>
<td>49.5</td>
<td>3</td>
<td>HI</td>
</tr>
<tr>
<td>SHRIMP</td>
<td>45.8</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>MACKEREL</td>
<td>42.9</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>IMAGERY</td>
<td>37.5</td>
<td>2</td>
<td>HI</td>
</tr>
<tr>
<td>EUPHEMISMS</td>
<td>30.6</td>
<td>5</td>
<td>HI</td>
</tr>
<tr>
<td>EUPHEMISM</td>
<td>20.8</td>
<td>2</td>
<td>HI</td>
</tr>
</tbody>
</table>
### Noun Listing

**HF=*/HIGH FREQ; SING=_SINGLETON; HI=*/HIGH INFO; RARE=RARE SING

<table>
<thead>
<tr>
<th>Noun</th>
<th>SFI</th>
<th>Occurs</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>59.8</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Act</td>
<td>59.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Rose</td>
<td>58.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Nature</td>
<td>58.8</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Experiment</td>
<td>58.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Nation</td>
<td>58.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Effect</td>
<td>58.3</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Purpose</td>
<td>58.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Danger</td>
<td>58.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Reasons</td>
<td>57.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Branches</td>
<td>57.8</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Support</td>
<td>57.6</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Style</td>
<td>56.8</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Taste</td>
<td>56.6</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Automobile</td>
<td>56.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Frame</td>
<td>56.1</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Situation</td>
<td>56.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Importance</td>
<td>56.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Doubt</td>
<td>56.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Substance</td>
<td>55.4</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Resources</td>
<td>54.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Society</td>
<td>54.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Event</td>
<td>54.6</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Values</td>
<td>54.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Attempt</td>
<td>53.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Desire</td>
<td>53.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Beings</td>
<td>53.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Content</td>
<td>52.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Defense</td>
<td>52.3</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Citizen</td>
<td>52.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Negroes</td>
<td>52.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Attitude</td>
<td>52.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Error</td>
<td>51.3</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Sorts</td>
<td>51.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Situations</td>
<td>50.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Drift</td>
<td>50.6</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Fate</td>
<td>49.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Charm</td>
<td>49.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Reverse</td>
<td>49.3</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Identity</td>
<td>49.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Chairman</td>
<td>49.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Advertising</td>
<td>48.8</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Aspects</td>
<td>48.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>Dimensions</td>
<td>48.6</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>NOUN</td>
<td>SFI</td>
<td>OCCURS</td>
<td>TYPE</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>TENDENCY</td>
<td>48.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>BOMBS</td>
<td>47.4</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>INSTITUTIONS</td>
<td>46.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>CONSEQUENCES</td>
<td>46.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>ACCOMPANIMENT</td>
<td>46.4</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>TUNA</td>
<td>45.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>VEIN</td>
<td>45.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>PRIVACY</td>
<td>43.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>BLACKS</td>
<td>43.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>ILLUSION</td>
<td>42.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>REALITIES</td>
<td>41.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>CONCEPTION</td>
<td>41.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>PERFUMES</td>
<td>40.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>ILLUSIONS</td>
<td>38.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>DISTINCTION</td>
<td>37.3</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>ACCUSATION</td>
<td>35.2</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>PRISONS</td>
<td>34.8</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>SEMANTICS</td>
<td>34.5</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>PERSPECTIVES</td>
<td>34.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>PERCEPTIONS</td>
<td>26.0</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>PRIORITIES</td>
<td>23.9</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>SCOUNDRELS</td>
<td>20.7</td>
<td>1</td>
<td>RARE</td>
</tr>
<tr>
<td>AUTHENTICITY</td>
<td>20.2</td>
<td>1</td>
<td>RARE</td>
</tr>
</tbody>
</table>
### NOUN LISTING

HF = HIGH FREQ; SING = SINGLETON; HI = HIGH INFO; RARE = RARE SING

<table>
<thead>
<tr>
<th>NOUN</th>
<th>SFI</th>
<th>OCCURS</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASYLUMS</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>CHAIRPERSON</td>
<td>NA</td>
<td>2</td>
<td>NA</td>
</tr>
<tr>
<td>CHILDPERSONS</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>EUPHEMIZING</td>
<td>NA</td>
<td>7</td>
<td>NA</td>
</tr>
<tr>
<td>FACILITATOR</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>GARBAGE MAN (MEN)</td>
<td>NA</td>
<td>4</td>
<td>NA</td>
</tr>
<tr>
<td>PROPRIETY</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>RESTROOM</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>SANITATION ENGINEER(S)</td>
<td>NA</td>
<td>5</td>
<td>NA</td>
</tr>
<tr>
<td>SHITHOUSE</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>SLUM CHILDREN</td>
<td>NA</td>
<td>3</td>
<td>NA</td>
</tr>
</tbody>
</table>
APPENDIX 3

TEST FORMS
Form WMM1T(A)

DIRECTIONS: Please circle the one best answer.

1. Horse mackerel is to tuna fish as:
   A. feline is to housecat.
   * B. pigeon is to dove.
   C. Honda is to Rolls Royce.
   D. rose is to flower.

2. A euphemism is most like a(n):
   A. oxymoron.
   B. alliteration.
   C. gerund.
   * D. metaphor.

3. A computer "virus" is:
   A. a computer bug.
   B. a computer disease.
   * C. a computer program.
   D. a programming error.

4. "Shakespeare was only half right in saying that a rose by any other name would smell as sweet" because:
   * A. society could agree that a rose does not smell "sweet."
   B. a rose thorn could hurt someone.
   C. a rose isn't really a "rose" and therefore doesn't have an odor.
   D. not all roses have the same odor.
5. Who might prefer to call janitors "maintenance engineers"?
   A. The building superintendent filling a job vacancy.
   B. The major supply dealer of electrical parts.
   * C. The union leader negotiating a pay contract.
   D. The newspaper editor complaining about ice on the school's sidewalk.

6. We may call a person a "pig" if he is dirty, a "shrimp" if he is small, or a:
   * A. "mule" if he is stubborn.
   B. "bean" if he is tough.
   C. "rock" if he is flaky.
   D. "tree" if he is old.

7. Rapid eye movement (R.E.M.) happens while a person dreams. A teenage girl says to her friend, "Oh look! There's John! He's a REM!" John is a(n):
   A. "doll."
   * B. "dream."
   C. "real dog."
   D. fantasy.
8. We euphemize when we:

* A. name political assassination a "termination with prejudice."
B. name a small animal a "shrimp."
C. name a church official a "deacon."
D. name a presidential candidate a "liberal."

9. The author would probably agree that propaganda:

* A. is used both for honorable and dishonorable purposes.
B. is an error of semantics.
C. is a false perception.
D. is a euphemism for advertising.

10. An example of a legitimate use of euphemism is:

A. body count.
* B. head of household.
C. tax-and-spend liberal.
D. surgical strike.
DIRECTIONS: Please circle the one best answer.

1. To use the phrase "Operation Sunshine" as a name for hydrogen bomb experiments is an example of:

   A. the most common form of euphemism.
   * B. an attempt to change public opinion towards bomb testing.
   C. an attempt to keep the experiment secret.
   D. giving euphemizing credibility.

2. If we compare the different effects of the euphemisms "sanitary engineer" and "Operation Sunshine," it is apparent that these phrases represent:

   A. similar attitude trends in American society.
   B. two very hypocritical attempts to hide an ugly reality.
   C. a meaningless attempt to twist thinking.
   * D. qualitatively different attitude trends in American society.

3. The writer assumes that people who use euphemisms in advertising:

   A. are sincere in their intentions.
   * B. are intending to deceive us.
   C. are honestly mistaken about the true nature of the thing they are promoting.
   D. are confused about the nature of reality.
4. Euphemisms, according to this writer, are best described as:
   A. a frightening trend in social thought.
   B. an expression of hypocrisy.
   C. an expression of respect for social events.
   * D. an expression of the values and perceptions of society.

5. To say that there must be a pre-existant trend in society for a euphemism to not appear "ridiculous" is to assume that:
   A. language causes ideas to change.
   B. there is no connection between the use of language and thinking.
   C. if an euphemism is seen as ridiculous, it will eventually be accepted.
   * D. ideas change before language changes.

6. This essay is really a look into the very heart of:
   A. the secrets of advertising.
   B. the use of literary allusions.
   C. the source of government propaganda.
   * D. the relationship of language and reality.
7. To say that "the process of euphemizing has no moral content" is to suggest that:

A. euphemizing has a clearly positive value.
* B. euphemizing is a neutral power of expression.
C. euphemizing has a clearly negative value.
D. euphemizing has little effect on our behavior.

8. The writer perceives the euphemism "black" as "a kind of euphemizing in reverse" because:

A. euphemisms are an attempt to get closer to the true nature of a thing.
* B. euphemisms typically are a glorified version of things.
C. euphemisms make us more clearly aware of basics.
D. the term "black" is not close to how "white" America perceives "people of color."

9. How important is it to our thinking to be able to recognize that "things" and the names we give them are not the same?

A. It doesn't effect the way we experience reality.
B. Names are an illusion that doesn't effect reality.
* C. It is important enough to cause us to be victims of propaganda if we don't know the difference.
D. Names do not effect the true nature of things.
10. When the writer suggests that the use of the term "dead" is not more honest than the phrase "passed away," he:

* A. creates a fault in his logic because the phrase "passed away" is generally used to avoid the reality of death.

B. continues the logic that all euphemisms express a change in values.

C. does not assume that people who use the phrase "passed away" are certain of a life after death.

D. assumes that there is no afterlife.
DIRECTIONS: Please circle the one best answer.

1. According to Postman, a proper use of a euphemism would be to

   * A. elevate the way in which we think about a person, place or event.
   B. distract from shortcomings.
   C. create negative stereotyping.
   D. assist in the marketing of defective merchandise.

2. Postman views euphemisms in a negative sense when

   A. names are changed to restrict negative stereotyping.
   B. they are used to change perspectives.
   C. they are used to market products.
   * D. innocuous names are given to hideous realities.

3. Which of the following statements concerning the moral basis of euphemizing does Postman subscribe to?

   * A. The moral basis is supplied by what the euphemism itself expresses.
   B. The moral basis is not supplied by the expression of the euphemism.
   C. The process of euphemizing contains the moral content.
   D. None of the above.
4. Which name-euphemism combination would Postman find contemptible?

A. Chairman - chairperson.
B. Janitor - custodian.
* C. Texas electric chair - Ol' Sparky.
D. Tornado - Texas twister.

5. Euphemizing is contemptible when

A. it is used to sell products.
B. it is used to alter negative stereotyping.
* C. the euphemism masks reality.
D. none of the above.

6. A euphemism which is perceived as incongruent is likely to

A. gradually change attitudes.
B. elicit the intended effect.
C. succeed in making people see new perspectives.
* D. not succeed in altering attitudes.

7. According to Postman, a value of euphemizing includes

A. the way it masks reality.
* B. encouraging alternate ways of looking at someone or something.
C. the way it diverts attention.
D. promoting retail products.
8. Postman believes it is unfortunate that

* A. names and things are thought of as being the same.
B. euphemisms are not accepted.
C. more euphemisms are not used to hide reality.
D. euphemisms are used to sell products.

9. It is difficult to predict the outcome of a euphemism because

A. behavior can be changed with the use of euphemisms.
* B. the success of a euphemism depends upon cultural support and acceptance.
C. euphemisms change perceptions.
D. euphemisms seldom change anything.

10. According to Postman, which of the following name-euphemism combinations would most likely not succeed from a retailing perspective?

A. Refrigerator - Kool King.
* B. All examples should succeed.
C. Bed - Serene Sleep.
D. Professional wrestler - Dr. Damage.
Form WMM2T(D)

DIRECTIONS: Please circle the one best answer.

1. The author's main purpose in this essay is to

   A. argue for the correctness of speaking "plainly."
   * B. defend the use of euphemisms.
   C. argue for more use of euphemisms.
   D. condemn the use of euphemisms.

2. The author argues that the charm of a name

   A. is a fundamental semantic error.
   B. displays the moral content of the speaker.
   C. logically follows from the object named.
   * D. rests on our image of the object named.

3. Euphemisms are used to highlight a(n)

   A. purpose.
   * B. aspect.
   C. effect.
   D. experiment.

4. The author assumes that perceptions, styles, priorities, attitudes and values are

   A. representations of reality.
   B. void of moral content.
   C. the result of the charm of a name.
   * D. verbal illusions of a society.
5. Negroes adopting the label "black" is "a kind of euphemism in reverse" because

A. civil rights were set back a decade.
B. "black" is a morally corrupt term.
C. "black" is usually associated with death and mourning.
* D. "black" was not seen as a more auspicious term.

6. A danger of the intelligent use of euphemisms is

A. euphemisms may start a social drift.
B. distinctions between the sexes will disappear.
C. we might change the physical characteristics of the object being euphemized.
* D. to presume attitudes will change quickly.

7. The fate of euphemizing depends on

A. the support of television advertisers.
* B. the support of society for the change of image.
C. the support of senior citizens, heads of households, displaced homemakers, and other such successfully renamed social groups.
D. the support of teachers to teach the art of euphemizing.
8. An attempt to rename something may be
   
   A. inherently morally corrupt.
   
   * B. an intelligent method to gain a new perspective.
   
   C. an accusation.
   
   D. a correct portrayal of reality.

9. The author might agree that our perceptions of objects are based on
   
   A. scientific fact.
   
   B. personal experience.
   
   * C. illusions.
   
   D. the five senses.

10. The moral dimension of a euphemism is presented by
    
    * A. the speaker.
    
    B. the object being euphemized.
    
    C. society.
    
    D. advertisers.
DIRECTIONS: Please circle the one best answer.

1. To confuse a thing with the name we call it is:

* A. an error in thinking.
B. a euphemism.
C. advertising.
D. government propaganda.

2. How are the phrases "slum children" and "culturally different children" contrasted in meaning:

A. one indicates a down-to-earth attitude, the other is a ridiculous euphemism.
B. one indicates a failure to face the harsh realities of the slums, the other confronts that reality.
C. one is a pretense, the other is factual.
* D. one provides reasons for children's behavior, the other narrows our understanding of these children.

3. The writer assumes that the use of the euphemism "childpersons"

* A. ignores real distinctions between boys and girls.
B. is consistent with current social trends.
C. correctly expresses that there are no real differences between boys and girls.
D. correctly expresses that children should all be treated the same regardless of sex.
4. A central theme in this essay is that euphemisms:

A. are generally used to distort reality.
B. are generally used to express important social change.
* C. can be defended as a legitimate expression of changing social values.
D. are more useful than "down-to-earth" terms.

5. "Tuna fish tastes better" than horse mackerel because:

A. the name is an illusion that doesn't effect the thing.
B. we can't really taste the thing without a name for it.
* C. perception is strongly influenced by names.
D. if we can't identify a thing, it is not real to us.

6. According to the writer, people who use euphemisms to use charming names for worthless things are:

A. government agents.
B. advertisers.
C. phonies.
* D. scoundrels.
7. A logical fallacy of the writer's argument concerning the "true" nature of things is that:

A. we perceive things the same no matter what we call them.
* B. even if you can change the way people perceive something because of the name you call it, this does not mean you have changed its nature.
C. things are real, names are not.
D. there is an independent reality that stands apart from names, but we are not capable of perceiving it.

8. Euphemisms are explored in this essay and shown to be a valuable resource for:

* A. providing new understanding of things.
B. people becoming politically important.
C. workers getting new respect.
D. children getting better education.

9. "A rose by any other name would smell as sweet" is an expression of a false assumption that:

A. the name of a thing does effect how we react to it.
* B. the name of a thing does not effect how we react to it.
C. a thing has no independent reality apart from its name.
D. we only react to a name, not a thing
10. An interesting aspect of this essay is that the writer, by exploring the various uses of euphemisms is also:

A. making euphemisms seem more dangerous.

* B. increasing our perception of euphemisms as a "method" of understanding.

C. increasing our perception of euphemisms as a "method" of distortion.

D. decreasing our perception of euphemisms.
Form WJM2T(F)

DIRECTIONS: Please circle the one best answer.

1. A negative effect of euphemizing would be
   A. creating an understanding for a person, event or thing.
   * B. instilling prejudice or fear.
   C. creating a fresh, new perspective on a problem.
   D. promoting a greater understanding.

2. The accusation Postman makes about euphemisms is that they
   A. sound suspicious.
   B. are pretentious.
   C. are congruent.
   * D. control reality.

3. The fate of a euphemism is difficult to predict because
   A. euphemisms alter perceptions.
   B. euphemisms rarely change perceptions.
   * C. the success of a euphemism is dependent upon cultural support and acceptance.
   D. behavior can be changed by euphemisms.

4. Postman believes that the authenticity of a person
   * A. is not compromised when one employs euphemisms.
   B. is compromised when people employ euphemisms.
   C. is greater when people employ euphemisms.
   D. depends upon the result of the euphemism.
5. What does Postman mean by the "charm" of a name?

A. The sound of a name when one hears it.
B. The magical qualities in the name itself.
* C. The positive association evoked by the use of the name.
D. None of the above.

6. Recognizing in certain situations euphemisms do not have the intended effect, Postman believes improper euphemisms

* A. all of the below.
B. mask the reality of the situation.
C. detract from the ugliness of the situation.
D. elevate the importance of the situation beyond what is necessary.

7. The defense of euphemizing is not illustrated by which of the following name-euphemism combinations:

A. janitor - custodian.
* B. commuter plane - tree top express.
C. realtor - broker.
D. doctor - physician.
8. What does Postman mean by giving "substance" to the illusion of identifying names with things?

* A. The frequent use of a name serves to give credibility to that name.
B. The name will be highly recognizable.
C. The frequent use of a name seldom creates an illusion.
D. The illusion depends upon substance.

9. Which of the following name-euphemism combinations reflect an attempt to create new perspectives on a subject?

* A. All of the below.
B. Blind - visually impaired.
C. Mentally retarded - mentally handicapped.
D. Slow learner - learning disabled.

10. According to Postman, which of the following name-euphemism combinations would not succeed from an advertising perspective?

A. Bed - Serene Sleep.
* B. All should succeed.
C. Refrigerator - King Kool.
D. Professional Wrestler - Dr. Damage.
1. Though the author does not define the term, we may infer "semantics" is

* A. a field of language study.
B. the study of euphemisms.
C. the study of "correct" word usage.
D. the study of advertising.

2. Essential qualities of an object must be maintained or else

A. essential differences will be lost.
B. unscrupulous people will win the war of correct word usage.
C. changes in meaning will create a second "Tower of Babel."
* D. society will not accept the name change.

3. The author refutes the theory that

* A. coarse language is more genuine.
B. changing the name of an object changes the perceived essence of the object.
C. Shakespeare was only half correct.
D. Advertisers and the government sometimes attempt to purposefully deceive us.
4. The author claims that, in certain instances, we should

A. forgive the government for hiding ugly reality from the civilian population.
B. use mystique to misrepresent an object.
* C. charge people who use "auspicious" terms as being insincere.
D. prefer coarse language.

5. Some unscrupulous people may dupe us by

A. making us underestimate the power of words.
B. making all euphemisms laughable.
* C. manipulating word meaning to alter how we perceive an object.
D. using earthy language rather than a euphemism.

6. Purposefully modifying what we call something can

A. change the reality of the object.
B. take away essential characteristics of the object.
C. force society to change its priorities.
* D. give us a fresh outlook on the object.
7. An ancient jail that was actually a psychopathic hospital would also be called a(n)

A. monkey house.
*B. madhouse.
C. fun house.
D. penitentiary.

8. An example of public awareness and sentiment turning as a product of name-changing would be:

* A. unwed mother to single parent.
B. pig to porker.
C. tiny to shrimp.
D. autumn to fall.

9. The author would view confusing the word with the object to which the word refers as

A. good advertising.
*B. an intriguing self-deception.
C. a subject that should be controlled.
D. a ridiculous drift in society.

10. A major concept discussed in the essay is

A. name-changing is immoral.
*B. name-changing could change social goals.
C. advertisers are unethical.
D. governments use name-changing to lie to its citizens.
Form WKM3T(H)

DIRECTIONS: Please circle the one best answer.

1. Using euphemisms can be a significant source of creative understanding of a topic by:

   A. distorting our view of reality.
   * B. making us aware of a new way to look at things.
   C. making us get back to hard facts.
   D. narrowing our choices in behavior.

2. When people confuse the beauty of a word for the object being sold, the advertisers are:

   A. changing the nature of the thing.
   B. using euphemisms for our benefit.
   C. reflecting changing social trends.
   * D. changing how people perceive a thing.

3. Using euphemisms, according to the writer, instead of more pragmatic language is:

   A. more sincere.
   B. more genuine.
   * C. expressing a different understanding.
   D. being artificial in your speech.

4. A central concept of this essay is that euphemisms provide society with a method of changing that is:

   * A. gradual.
   B. rapid.
   C. shocking.
   D. insignificant.
5. What do people do that creates a self-deception?

A. Misuse euphemisms.
B. Avoid euphemisms.
*C. Think that words and objects are identical.
D. Think that words and objects are separate.

6. In what case is it appropriate to complain that using euphemisms is hypocritical?

* A. When it is an attempt to hide something ugly behind a beautiful name.
B. When it is an attempt to change our values.
C. When it is an attempt to give new respect to individuals.
D. When it is an attempt to hide something beautiful behind an ugly name.

7. It is a basic mistake in the study of meaning to suppose that a word and an object are:

A. different.
* B. identical.
C. related.
D. unrelated.

8. If you refer to a "nut house" as a "mental health haven," but your perception does not change, then:

* A. the euphemism does not have a positive effect.
B. the euphemism more closely reflects the reality.
C. there is no euphemism.
D. you have changed your perception.
9. To ignore the real differences in males and females by using the term "childpersons" to refer to children of both sexes is a problem because:

* A. social change must precede use of euphemisms.
B. there are insurmountable differences between boys and girls.
C. there are no important differences between boys and girls.
D. names change attitudes quickly.

10. What was remarkable when the term "blacks" superceded "Negroes" in acceptable usage?

A. People's attitudes didn't change.
* B. People's attitudes also changed.
C. "Blacks" is a kind of euphemizing in reverse.
D. Names have little power to change behavior.
DIRECTIONS: Please circle the one best answer.

1. The complaint that a person who uses euphemisms is a "fake" is valid

   A. when the euphemisms are used to help people see new perspectives on a topic.
   * B. when the choice of a euphemism is dictated by a desire to hide certain qualities associated with a name.
   C. when the choice of a euphemism is dictated by marketing trends.
   D. in most situations.

2. In all semantic categories, the belief that a name and a thing are one and the same is

   A. not a problem.
   B. a minor error in judgment.
   C. a major assumption.
   * D. a basic misinterpretation.

3. By shifting the appeal of a name to whatever inferior product they are marketing, certain "scoundrels"

   A. change the way in which people perceive a product.
   * B. A, C, and D are all correct.
   C. are encouraged to continue the practice due to frequent success.
   D. know the process works as well as if the product itself were changed.
4. The belief that individuals who speak without utilizing euphemisms are more authentic than those who do

A. is weakly supported by Postman.
B. is strongly supported by Postman.
C. is not discussed by Postman.
* D. is strongly denied by Postman.

5. Postman believes that euphemizing

A. is seldom used appropriately.
B. is one of the least critical methods for creating new outlooks on a subject.
C. is one of the most important intellectual methods for establishing better mental institutions.
* D. is one of the most critical intellectual methods for establishing new outlooks on a subject.

6. In certain situations, euphemisms

A. B, C, and D are all incorrect.
B. may be accepted, even though they are perceived as ridiculous by society.
C. may not be accepted, however benefits are realized.
* D. may be accepted, but expected benefits are not achieved.
7. Euphemisms are a method by which a group of people may change their imagery of an object,

A. thereby quickly and overtly altering cultural style, priorities, and values.
* B. thereby gradually altering cultural style, priorities, and values.
C. however, the change is temporary.
D. but not change cultural style, priorities, and values.

8. Using the euphemism "childpersons" in place of "boys" and "girls" would

A. slowly be accepted with continued usage.
* B. not be accepted due to a lack of group corroboration.
C. slowly be accepted because of a natural drift in cultural support.
D. not be accepted due to increased cultural support.

9. One of the most interesting illusions harbored by people

A. is that a euphemism will sometimes change perceptions and attitudes.
B. is the identification of euphemisms with clear speaking.
* C. is the identification of names with things.
D. is that a euphemism must be supported by society in order for it to succeed.
10. In many situations, successful euphemisms

* A. also result in altered cognizance and viewpoints.
B. lack cultural support.
C. are used despite being perceived as inconsequential.
D. none of the above.
APPENDIX 4

ITEM DIFFICULTY, VARIANCE, AND POINT BISERIAL

VALUES BY WRITER AND METHOD
<table>
<thead>
<tr>
<th>Item</th>
<th>Difficulty (mean)</th>
<th>Item</th>
<th>Variance</th>
<th>Point Biserial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>.2</td>
<td>Item 1</td>
<td>.16</td>
<td>.15</td>
</tr>
<tr>
<td>Item 2</td>
<td>.6</td>
<td>Item 2</td>
<td>.25</td>
<td>-.18</td>
</tr>
<tr>
<td>Item 3</td>
<td>0.00</td>
<td>Item 4</td>
<td>.46</td>
<td>-.06</td>
</tr>
<tr>
<td>Item 4</td>
<td>.46</td>
<td>Item 5</td>
<td>.27</td>
<td>.52</td>
</tr>
<tr>
<td>Item 5</td>
<td>.46</td>
<td>Item 6</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td>1.0</td>
<td>Item 7</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td>.7</td>
<td>Item 8</td>
<td>.22</td>
<td>-.1</td>
</tr>
<tr>
<td>Item 8</td>
<td>.8</td>
<td>Item 9</td>
<td>.16</td>
<td>-.4</td>
</tr>
<tr>
<td>Item 9</td>
<td>.6</td>
<td>Item 10</td>
<td>.25</td>
<td>.02</td>
</tr>
</tbody>
</table>

**Method 2**

<table>
<thead>
<tr>
<th>Item</th>
<th>Difficulty (mean)</th>
<th>Item</th>
<th>Variance</th>
<th>Point Biserial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>.4</td>
<td>Item 1</td>
<td>.25</td>
<td>-.13</td>
</tr>
<tr>
<td>Item 2</td>
<td>.6</td>
<td>Item 2</td>
<td>.27</td>
<td>.11</td>
</tr>
<tr>
<td>Item 3</td>
<td>.7</td>
<td>Item 3</td>
<td>.22</td>
<td>.18</td>
</tr>
<tr>
<td>Item 4</td>
<td>.5</td>
<td>Item 4</td>
<td>.27</td>
<td>-.28</td>
</tr>
<tr>
<td>Item 5</td>
<td>.8</td>
<td>Item 5</td>
<td>.16</td>
<td>-.08</td>
</tr>
<tr>
<td>Item 6</td>
<td>.7</td>
<td>Item 6</td>
<td>.22</td>
<td>.01</td>
</tr>
<tr>
<td>Item 7</td>
<td>1.0</td>
<td>Item 7</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Item 8</td>
<td>.8</td>
<td>Item 8</td>
<td>.16</td>
<td>.3</td>
</tr>
<tr>
<td>Item 9</td>
<td>.5</td>
<td>Item 9</td>
<td>.27</td>
<td>-.03</td>
</tr>
<tr>
<td>Item 10</td>
<td>.2</td>
<td>Item 10</td>
<td>.16</td>
<td>.23</td>
</tr>
</tbody>
</table>
Method 3

<table>
<thead>
<tr>
<th>Item</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>.5</td>
<td>.28</td>
<td>.25</td>
</tr>
<tr>
<td>Item 2</td>
<td>.4</td>
<td>.27</td>
<td>.2</td>
</tr>
<tr>
<td>Item 3</td>
<td>.3</td>
<td>.23</td>
<td>.19</td>
</tr>
<tr>
<td>Item 4</td>
<td>.3</td>
<td>.23</td>
<td>.62</td>
</tr>
<tr>
<td>Item 5</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td>.5</td>
<td>.28</td>
<td>.44</td>
</tr>
<tr>
<td>Item 7</td>
<td>.9</td>
<td>.1</td>
<td>-.49</td>
</tr>
<tr>
<td>Item 8</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 9</td>
<td>.4</td>
<td>.27</td>
<td>-.37</td>
</tr>
<tr>
<td>Item 10</td>
<td>.3</td>
<td>.23</td>
<td>-.14</td>
</tr>
<tr>
<td>Item Writer 2</td>
<td>Item</td>
<td>Difficulty (mean)</td>
<td>Item</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>------------------</td>
<td>------</td>
</tr>
<tr>
<td>Method 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td></td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td></td>
<td>.6</td>
<td></td>
</tr>
<tr>
<td>Item 3</td>
<td></td>
<td>.8</td>
<td></td>
</tr>
<tr>
<td>Item 4</td>
<td></td>
<td>.9</td>
<td></td>
</tr>
<tr>
<td>Item 5</td>
<td></td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td></td>
<td>.6</td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td></td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>Item 8</td>
<td></td>
<td>.8</td>
<td></td>
</tr>
<tr>
<td>Item 9</td>
<td></td>
<td>.8</td>
<td></td>
</tr>
<tr>
<td>Item 10</td>
<td></td>
<td>.3</td>
<td></td>
</tr>
<tr>
<td>Method 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td></td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td></td>
<td>.4</td>
<td></td>
</tr>
<tr>
<td>Item 3</td>
<td></td>
<td>.8</td>
<td></td>
</tr>
<tr>
<td>Item 4</td>
<td></td>
<td>.7</td>
<td></td>
</tr>
<tr>
<td>Item 5</td>
<td></td>
<td>.8</td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td></td>
<td>.3</td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td></td>
<td>.4</td>
<td></td>
</tr>
<tr>
<td>Item 8</td>
<td></td>
<td>.8</td>
<td></td>
</tr>
<tr>
<td>Item 9</td>
<td></td>
<td>.5</td>
<td></td>
</tr>
<tr>
<td>Item 10</td>
<td></td>
<td>.3</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>Method 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 3</td>
<td>.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 4</td>
<td>.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 5</td>
<td>.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td>.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td>.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 8</td>
<td>.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 9</td>
<td>.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 10</td>
<td>.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Item Writer 3

<table>
<thead>
<tr>
<th>Item</th>
<th>Difficulty (mean)</th>
<th>Item Variance</th>
<th>Point Biserial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>.7</td>
<td>.23</td>
<td>.76</td>
</tr>
<tr>
<td>Item 3</td>
<td>.4</td>
<td>.27</td>
<td>-.3</td>
</tr>
<tr>
<td>Item 4</td>
<td>.7</td>
<td>.23</td>
<td>.51</td>
</tr>
<tr>
<td>Item 5</td>
<td>.7</td>
<td>.23</td>
<td>.28</td>
</tr>
<tr>
<td>Item 6</td>
<td>.3</td>
<td>.23</td>
<td>.66</td>
</tr>
<tr>
<td>Item 7</td>
<td>.9</td>
<td>.1</td>
<td>.53</td>
</tr>
<tr>
<td>Item 8</td>
<td>.7</td>
<td>.23</td>
<td>.76</td>
</tr>
<tr>
<td>Item 9</td>
<td>.2</td>
<td>.18</td>
<td>.58</td>
</tr>
<tr>
<td>Item 10</td>
<td>.7</td>
<td>.23</td>
<td>.51</td>
</tr>
</tbody>
</table>

#### Method 2

| Item 1 | .7 | .01   | -.12 |
| Item 2 | .2 | .15   | .02  |
| Item 3 | .4 | .27   | .36  |
| Item 4 | .2 | .15   | .39  |
| Item 5 | .3 | .24   | .65  |
| Item 6 | .6 | .27   | .34  |
| Item 7 | .8 | .15   | .06  |
| Item 8 | .4 | .27   | .58  |
| Item 9 | .9 | .08   | .03  |
| Item 10 | .6 | .27   | .23  |
### Method 3

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>.9</td>
<td>.08</td>
<td>-.2</td>
</tr>
<tr>
<td>Item 2</td>
<td>.4</td>
<td>.27</td>
<td>.36</td>
</tr>
<tr>
<td>Item 3</td>
<td>.6</td>
<td>.27</td>
<td>-.18</td>
</tr>
<tr>
<td>Item 4</td>
<td>.4</td>
<td>.27</td>
<td>.36</td>
</tr>
<tr>
<td>Item 5</td>
<td>.5</td>
<td>.27</td>
<td>.11</td>
</tr>
<tr>
<td>Item 6</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td>.8</td>
<td>.2</td>
<td>.16</td>
</tr>
<tr>
<td>Item 8</td>
<td>.7</td>
<td>.24</td>
<td>.64</td>
</tr>
<tr>
<td>Item 9</td>
<td>.2</td>
<td>.15</td>
<td>-.18</td>
</tr>
<tr>
<td>Item 10</td>
<td>.3</td>
<td>.24</td>
<td>-.11</td>
</tr>
</tbody>
</table>
REFERENCES


Ellis, J. A., Wulfeck, W. H., Richards, R. E.,
Wood, N. D., & Merrill, M. D. (1978). The
instructional quality inventory: I. Introduction and
overview (Navy Personnel Research & Development Center
Research & Development Cntr.

A systematic approach to the construction of domain-
referenced multiple-choice test items. Paper presented
at the meeting of the American Psychological
Association, Washington, D.C.


of Reading Behavior, 7, 341-376.

Finn, P. J. (1978a, March). Generating domain-referenced,
multiple-choice test items from prose passages. Paper
presented at the meetings of the American Educational
Research Association, Toronto.

Finn, P. J. (1978b). Word frequency, information theory,
and cloze performance: A lexical marker-transfer
feature theory of processing in reading. Reading
Research Quarterly, 13, 508-537.

Psychological Bulletin, 70, 460-473.


In R. A. Berk (Ed.), A guide to criterion-referenced
test construction. Baltimore, Maryland: John Hopkins
University Press.

Osburn, H. G. (1968). Item sampling for achievement
testing. Educational and Psychological Measurement,
28, 95-104.

Pedhazur, E. J. (1982). Multiple regression in behavioral
research: Explanation and prediction (2nd. ed.). NY:
Holt, Rinehart, & Winston.

NJ: Prentice-Hall.

Popham, W. J. (1984). Specifying the domain of content or
behavior. In R. A. Berk (Ed.), A guide to criterion-
referred test construction. Baltimore, Maryland:
John Hopkins University Press.

Postman, N. (1976). Crazy talk, stupid talk. NY:
Delacorte Press.

of content specialists in the assessment of criterion-
referred test item validity. Dutch Journal of
Educational Research, 2, 49-60.

design: Some cases from mathematics. In D. Klahr (Ed.),
Cognition and instruction. Hillsdale, NJ: Lawrence
Erlbaum Associates.


