DOCUMENT RESUME

ED 334 261 TM 016 876

AUTHOR Thiede, Keith W.; And Others
TITLE Recognition Versus Recall Test Formats: A Correlational Analysis.
PUB DATE Apr 91
NOTE 15p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 3-6, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Processes; *College Students; Comparative Testing; *Correlation; Higher Education; Multiple Choice Tests; Objective Tests; *Recall (Psychology); *Recognition (Psychology); *Test Format
IDENTIFIERS Free Response Test Items

ABSTRACT

A correlational analysis was performed to examine the relationship between recognition and recall test formats. A total of 236 college students completed one of four 80-item general knowledge tests; the forms contained 20 items of each of four formats: (1) true; (2) false; (3) multiple-choice; and (4) free response. Ninety-three of the subjects attended the University of Minnesota, and 143 students attended the University of Wisconsin at River Falls. The analysis justified consideration of the true and false items of the true-false test as separate formats. The results fail to support hypotheses which suggest that recognition and recall tests require differential thought processes. Each recognition test format correlated most highly with the free-response (recall) test format. Furthermore, the multiple-choice test correlated more highly with the free-response test than did either the true or false test formats; this difference was significant beyond the 0.05 level. These findings provide evidence that a relationship exists between recognition and recall thought processes. The ability to recognize facts is likely to be a subset of the ability to recall facts, rather than a distinct thought process. Three tables and one figure present study data. (Author/SLD)

# Recognition Versus Recall Test Formats: A Correlational Analysis

by
Keith W. Thiede, Alan J. Klockars, & Gregory R. Hancock
University of Washington

Running head: RECOGNITION VERSUS RECALL

Paper presented at the annual meeting of the National Council on Measurement in Education in Chicago, April 1991.

Abstract

A correlational analysis was performed to examine the relationship between recognition and recall test formats. College students (n = 236) completed one of four eighty-item general knowledge tests; the forms contained twenty items of each of four formats. The analysis justified the consideration of the true and false items of the true-false test as separate formats. The results failed to support the hypotheses developed on the basis of the theory that recognition and recall tests require differential thought processes (Kintsch, 1970; Anderson & Bower, 1972). It was discovered that each recognition test format correlated most highly with the free-response (recall) test format. Furthermore, the multiple-choice test was found to correlate more highly with the free-response test than either the true or false test formats, and this difference was significant beyond the .05 level. These findings provide evidence that a relationship exists between recognition and recall thought processes. The results suggest that the ability to recognize facts is likely a subset of the ability to recall facts, rather than a distinct thought process.

# Recognition Versus Recall Test Formats: A Correlational Analysis

Since the development of the first objective tests, researchers have sought to determine whether such formats as multiple-choice and true-false, which generally require a recognition solution strategy, measure the same attributes as free-response, which requires a recall solution strategy. Calculating the correlation between such recognition and recall tests is one method that has been used to evaluate the similarity or dissimilarity of the attributes measured by these test formats. Investigations utilizing this direct correlation technique have reached varying conclusions.

Toops (1921) appears to have been the first to compare recognition and recall tests making use of direct correlations. Toops constructed a true-false, multiple-choice (five-options), and a free-response test using 50 general knowledge questions, with his data suggesting that these tests measure the same characteristic. Corey (1930) concurred with Toops, concluding that recognition and recall tests measure "nearly" the same thing. Ruch and Stoddard (1925) also conducted a study similar to Toops, administering free-response, multiple-choice, and true-false tests to high school students. Unlike Toops' investigation, however, the correlations reported by Ruch and Stoddard left some doubt as to whether these tests measured the same attribute. Similarly, Hurd (1932) correlated recall and recognition tests designed to cover the same content and also concluded that these formats do not measure exactly the same functions, while Hurlburt (1954) reported weak correlations between recall and recognition vocabulary tests. And more recently, Harks, Herron, and Lefter (1972) correlated free-response and multiple-choice physics tests and concluded that the multiple-choice test was an adequate substitute for the free-response test. likewise, Colgan (1977) reported a strong correlation between multiple-choice and free-response mathematics tests.

Using the direct correlations between formats as a springboard, factor analysis has been employed most recently to determine if test formats appear to be measuring common characteristics. Initial research by Traub and Fisher (1977), using confirmatory factor analysis, provided little evidence of a format effect for mathematical reasoning items, and only weak evidence that the free-response and multiple-choice items were measuring different constructs for verbal comprehension. Ward, Frederiksen, and Carlson (1980), in a comparison of machine-scored and constructed-response forms of a test to measure ability to formulate scientific hypotheses, found slight factor analytic support for the hypothesis that the two formats measure different constructs. In another study, Ward (1982) concluded that for verbal aptitude items free-response and multiple-choice formats produce much the same information and rely on essentially the same constructs. Unlike

these studies, Ackerman and Smith (1988) found that in the area of writing assessment the construct being measured does indeed appear to be a function of the format of the test. Specifically, the skill of generating topic knowledge is more accurately assessed by essay, while objective formats better assess the procedural components of writing. The authors suggest that both formats should be used to provide a complete and valid assessment of writing skills. In general, however, little evidence has been amassed to support overall different format constructs.

Cognitive psychology provides the theoretical framework for understanding the thought processes required by recall and recognition test formats. The conclusion that free-response and multiple-choice tests do not measure the same attributes is supported by cognitive psychologists, who have suggested that differential thought processes are required for recall and recognition tasks (Kintsch, 1970; Anderson & Bower, 1972). Two-phase theories state that recall tasks involve two stages: a memory search stage and a decision stage. In the memory search stage relevant information is retrieved from long-term memory and used to create viable solutions. In the decision stage the best alternative is selected from those that have been retrieved from memory. The fact that people are often able to recognize information that they were unable to recall is given as evidence to support the two-phase theories.

Two-phase theories of memory imply that a deeper knowledge is needed to recall information than is needed to recognize that same information. The memory search and decision stages are both required to find solutions to recall tasks. In contrast, recognition tasks required only the decision stage, since the alternatives are provided. Two-phase theories identify differences in thought processes required for recall and recognition tasks, however, these thought processes are not independent of each other. Many of the cognitive skills utilized to solve recall tasks are also employed in solving recognition tasks.

On a practical level, real-life problems are rarely presented in a simple multiple-choice or true (of true-false) form which require the use of pure recognition; information must be generated and applied in order to solve real-life problems. Extended to an educational setting, it can be argued that the development of knowledge that can be recalled and generated is a valid instructional goal. This increases the need for assessment instrumentation that measures recall knowledge, as opposed to recognition knowledge. The present investigation is a correlational study designed to determine which recognition test format most closely measures the knowledge measured by a recall test.

This study makes use of the free-response, two-option multiple-choice, and true-false test formats. As true items make use of recognition knowledge, while false items are believed to require recall knowledge, true and false items were considered as separate

formats. Figure 1 provides a description of the types, or levels, of knowledge required to successfully produce a correct response within each test format, it also helps illustrate characteristic differences between formats.

---

Insert Figure 1 Here

---

The information regarding the thought processes required by the various test formats assisted in the formulation of three research hypotheses: first, since the false items of the true-false test and the free-response test both require the utilization of recall memory and do not employ recognition memory, it was hypothesized that these formats would be highly correlated. Second, since the true items of the true-false test and the multiple-choice test both make use of recognition memory, it was speculated that these formats would be highly correlated. Finally, since the true items of the true-false tests provided the purest measure of recognition memory, it was hypothesized that this format would be more weakly correlated with the free-response test than other recognition formats.

## Procedures

Subjects: The subjects for this research were students from the University of Minnesota and the University of Wisconsin - River Falls. The students from the University of Minnesota (n = 93) were students enrolled in undergraduate sociology courses. The students from the University of Wisconsin - River Falls (n = 143) were a combination of undergraduates and graduates enrolled in educational measurement courses. Each student completed one of four randomly assigned tests.

Instrumentation: The present study used of four eighty-item tests, twenty items from each of the discussed formats. The items consisted of twenty general knowledge questions from each of the following areas: American history and politics; natural and physical science; geography; and art, music, and literature. The stem of each question was written as a free-response item, a multiple choice (two-option) item, a true item, and a false item. In each case the distractor from the multiple-choice item was added to the stem to create the false statement. An illustration of a question written in the various formats is provided below.

**(FORM A)** The name of the second largest continent is:
**(FORM B)** The name of the second largest continent is:
A) Africa.

B) South America.

**(FORM C)** The name of the second largest continent is Africa.

**(FORM D)** The name of the second largest continent is South America.

Following the construction of identically written items in each of the four formats, items were grouped into blocks of five items within each content area and assigned to one of four test forms (A - D). Items were randomly ordered within content areas and assigned to the various forms using a Latin Square design.

Analysis: The analysis of data involved the calculations and comparisons of the correlation coefficients between the test formats, for each of the content areas and the total test, across the individual subjects.

## Results & Discussion

The objective in having constructed tests to evaluate examinees' knowledge of so many subject areas was to attain a measure of each examinee's general knowledge. The test forms utilized in this study were designed to provide four general knowledge subscores: one for each of the different test formats.

The goal of this study was to investigate the relationship between four different test formats. It is reasonable to correlate scores between the test formats, because the individual test formats give an index of general knowledge. The Tables below present measurement information of the various test formats.

The means and standard deviations of the test formats are provided in Table 1. The means and standard deviations were obtained by summing across the four test forms. The maximum score on each subtest was 20.

---

Insert Table 1 Here

---

The reliabilities of the test formats within each form are reported in Table 2. The total test reliabilities are also given.

---

Insert Table 2 Here

---

The low reliabilities of the subtests were not entirely unexpected. Each subtest consisted of relatively few items. Additionally, incorporating four content areas into the tests contributed to the low reliabilities of the subtests. With regard to the total tests, although

the reliabilities are not extremely high for an eighty item test, they are at a level that suggests the subtests which makeup the total test are not greatly different.

The correlations between the four test formats are presented in Table 3.

---

Insert Table 3 Here

---

It should be noted that these correlations were not corrected for attenuation. While this correction does provide an indication of the strength of the correlation adjusted for the unreliability of the tests these correlations represent the relationship between the tests in an ideal setting as opposed to a real world setting.

Prior to discussing these results as they pertain to the research hypotheses it is important to note the correlation between the true and false items of the TF test. The weak correlation found between the true items and false items of the TF test justifies the consideration of these item types as separate formats.

The correlational analysis of the test formats yielded results that are not easily explained in terms of a two-phase theory of memory retrieval. The theory that recognition and recall tasks require different thought processes led to the development of three hypotheses: first, that the false items of the TF test and the FR test would be highly correlated. Second, the true items of the TF test and the MC test would be highly correlated, and third, that the FR test would have a weak correlation with both the MC and the true items of the TF test. The data do not support these relationships between the test formats.

The correlational analysis reveals that each of the recognition test formats correlates more highly with the recall test format than with any other recognition test format. The recognition format found to correlate most strongly with the recall format was the MC test format. The correlation between the MC and FR test formats was significantly higher than the correlations between either of the TF formats and the FR format. This difference was statistically significant at the .01 level.

The findings of this research lead one to suspect that the thought processes required to recognize and recall information are more similar than unique. With regard to general knowledge examinations that measure the examinees' ability to retrieve information from long-term memory, it may well be that irrespective of the format of the test, the key to obtaining the correct solution is whether the required information can be recalled.

One possible factor contributing to these counterintuitive results is the low reliabilities of the subtests. It may well be that with highly reliable subtests the resulting correlations between test formats could be more in line with the pattern predicted by the

two-phase theory of memory retrieval. This suggests a need for further research in this area.

## References

Ackerman, T.A. & Smith, P.L. (1988). A comparison of the information provided by essay, multiple-choice, and free-response writing tests. Applied Psychological Measurement, 12, 117-128.

Anderson, J.R. & Bower, G.H. (1972). Recognition and retrieval process in free recall. Psychological Review, 79, 97-132.

Colgan, L.N. (1977). Reliability of mathematics multi-choice tests. International Journal of Mathematics Education and Science Technology. 8, 237-244.

Corey, S.M. (1930). The correlation between new type and essay examination scores, and the relationship between them and intelligence as measured by army alpha. School and Society, 32, 849-850.

Harks, D., Herron, & Lefter (1972). Comparison of a randomized multiple-choice format with a written one-hour physics problem test. Science Education, 56, 563-565.

Hurd, A.W. (1932). Comparison of short answer and multiple-choice tests covering identical subject content. Journal of Educational Research, 26, 28-30.

Hurlburt, D. (1954). The relative value of recall and recognition techniques for measuring precise knowledge of word meaning - nouns, verbs, adjectives. Journal of Educational Research, 47, 561-576.

Kintsch, W. (1970). Models for free recall and recognition. In D.A. Norman (Ed.), Models of human memory (pp. 331-373). New York: Academic Press.

Ruch, G.M. & Stoddard, G.D. (1925). Comparative reliabilities of five types of objective examinations. Journal of Educational Psychology, 16, 89-103.

Toops, H.A. (1921). Trade tests in education (Teachers college contributions to education no. 115). New York: Teachers College, Columbia University.

Traub, R.E. & Fisher, C.W. (1977). On the equivalence of constructed-response and multiple-choice tests. Applied Psychological Measurement, 1, 355-369.

Ward, W.C. (1982). A comparison of free-response and multiple-choice forms of verbal aptitude tests. Applied Psychological Measurement, 6, 1-11.

Ward, W.C., Frederiksen, N., & Carlso S.B. (1980). Construct validity of free-response and machine-scorable forms of a test. Journal of Educational Measurement, 17, 11-29.

**Figure 1**
Types of knowledge required by each format to arrive at a correct answer.

| | Test formats that make use of the given knowledge level: | | | |
|---|---|---|---|---|
| | FR | MC | True | False |
| Knowledge levels: | | | | |
| The ability to generate the correct answer. | X | X | X | X |
| The ability to recognize the correct answer. | | X | X | |
| The ability to establish that a distractor is false. | | X | | X |

**Table 1**
Summary statistics total score on each format

| | T | F | MC | FR |
|---|---|---|---|---|
| Mean | 14.36 | 11.56 | 14.45 | 6.57 |
| Standard Deviation | 2.76 | 2.97 | 2.51 | 3.61 |

**Table 2**
Reliabilities of test formats within each form

| | Form A | Form B | Form C | Form D |
|---|---|---|---|---|
| True | .437 | .529 | .372 | .589 |
| False | .573 | .500 | .342 | .515 |
| Multiple-choice | .264 | .383 | .306 | .472 |
| Free-response | .692 | .517 | .344 | .756 |
| TOTAL TEST | .762 | .841 | .767 | .821 |

**Table 3**
Correlations between different test formats

| | T | F | MC | FR |
|---|---|---|---|---|
| True | 1.000 | .254 | .169 | .351 |
| False | | 1.000 | .308 | .417 |
| MC | | | 1.000 | .586 |
| FR | | | | 1.000 |

Acknowledgements

We would like to extend special thanks to Dr. William Romoser, of the University of Wisconsin - River Falls, and Mary Jo Reef, of the University of Minnesota, for their assistance in the collection of data used in this investigation.