DOCUMENT RESUME

ED 110 492 TM 004 765

AUTHOR Williams, David V.; And Others
TITLE Memory vs, Inference: A Preliminary Study of Process-Referenced Test Items.
SPONS AGENCY Spencer Foundation, Chicago, Ill.
PUB DATE [Apr 75]
NOTE 6p.; Paper presented at the Annual Meeting of American Educational Research Association (Washington, D.C., March 30-April 3, 1975); For a related document, see TM 004 771

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Cognitive Processes; College Students; Comparative Analysis; *Memory; *Multiple Choice Tests; Recall (Psychological); *Response Style (Tests); Test Construction; Testing

ABSTRACT

Seven undergraduate volunteers studied a written passage on Atomic Structure and then, while answering a set of 24 multiple-choice items, talked aloud about the strategies they were using for option selection. The tape recordings of their verbal responses were analyzed for latency, memory references, and inference references. The items testing knowledge required a shorter time to answer, and the verbal reports contained more words and phrases associated with memory processes, fewer associated with inference, than did those for the items testing higher-order skills. The results suggest the usefulness of a more complex definition of item difficulty. (Author)

MEMORY VS. INFERENCE: A PRELIMINARY STUDY OF PROCESS-REFERENCED TEST ITEMS[1]

David V. Williams, James J. Diamond,
Elizabeth Kreher and Jude Braucher
Ithaca College University of Pennsylvania

Recent interest in performance-based evaluation and criterion-referenced tests has recalled attention to the need to explore process[2] as well as content variables in the establishment of criteria.

Recall of specific facts is relatively easy to measure by means of present testing techniques, and it might be suggested that items of this sort are the dominant type in multiple-choice based evaluation. While there are several extensive systems describing complex levels of cognitive functioning together with instructional suggestions (e.g., Bloom et al, 1956), none has been demonstrated to be useful in the construction of multiple-choice items measuring the attainment of these levels. Rather, it is generally assumed that the item writer's intent alone is sufficient to produce an item which will successfully engage the student in a given cognitive function. It would be desirable to approach the development of instruments which are empirically valid measures of given processes.

Kropp (1956) investigated the relationship between solution process and item performance. However, since his items were from a standardized instrument, he was not in a position to know the solution process intended by the item writer. Connolly and Wantman (1964) also worked with standardized instruments. Little and D'Asaro (1965) studied solution processes of a single student on a biology examination. However, they were more concerned with test scoring issues than with cognitive processes per se.

A previous study (Diamond & Williams, 1972) compared the item writer's process-intent directly with the student's report of process. No consistent relation was found between global ratings of written descriptions of item-solution processes, and the four Bloom processes of Knowledge (K), Comprehension (C), Application (Ap), and Analysis (An); intended by the item writer.

Thus the present study sought to test the hypothesis that "process" distinctions would be found between two types of items: those requiring responses in terms of the materials read (K) and those reflecting "higher order skills" (HOS) (i.e., C. Ap. An), involving some transformation and/or recombination of the materials presented. Specifically it was hypothesized that K items would have shorter response-latencies than HOS items and that there would be recognizable semantic and syntactic differences in the verbal reports associated with the two item-types.

---

[1] Paper presented at meetings of the American Educational Research Association, Washington, April 1975. This research was supported by a grant from the Spencer Foundation.

[2] The differences among behavioral, psychometric and neurological definitions of "process" are being laid aside for the present purpose.

METHOD

Two modules each consisting of a passage ("Atomic Structure"; "Glaciers") and 24 multiple-choice items were adapted from materials developed by Kropp & Stoker (1956). From items judged by them to involve the Bloom skills of Knowledge, Comprehension, Application and Analysis, six of each type were selected and randomly arranged in booklet form, one to a page (with the constraint that three of each type occur in each half of the booklet and no more than two items of one type be adjacent).

Subjects were undergraduate volunteers who worked individually with both passages on separate occasions, in a counterbalanced sequence. Each student was asked to read the passage as (s)he would if studying for a test (i.e., to make notes, underline, etc.) for as long as (s)he wished. The passage was then removed and (s)he was asked to answer the multiple-choice questions while talking about the strategies (s)he was using ("...what you are thinking about... the basis for choosing an answer..."). Each was continuously tape-recorded while answering the items. An observer was present to whom the student could direct these statements who functioned only in the role of an attentive listener. Both were blind to the specific hypothesis of the study.

Tapes were blindly and independently surveyed and key words and phrases associated with "memory" and "inference" were identified, and collated. "Memory references" consisted of (1) explicit references to the passage ("...it said...mentions...was on page...") and (2) words whose ordinary meanings relate to the reporting of events ("...recognize...remember..."). "Inference references" consisted of (1) words and phrases whose ordinary meanings relate to the concept of inference ("...suppose...infer...figure out...") and (2) subjunctives and logical conjunctions suggesting combination, transformation, or predication (...would have...therefore...but...if..."). An experimenter, blind to the categorization of items, then recorded the number of references of each type and choice-latency for each item from the tapes.

RESULTS

Each of the three measures, response latency, memory and inferences references, was examined in each set of items by treatment x subjects analyses of variance. Analyses were made in terms of the K vs.HOS classification of processes as well as the Bloom categories. These results are provided as Table 1.

DISCUSSION

The hypothesis that K vs. HOS items would be distinguished was confirmed. Both sets of K items had significantly shorter response latancies than HOS items consistant with the presumably greater complexity of the higher-order-skills. (This measure of item complexity would seem thus to be a useful compliment to traditional difficulty indices.) One set of K items showed significantly more memory references than the HOS items consistent with the hypothesized relation between conceptualizations of the "Knowledge" process-intent of the item writer and the "memory reference" of the test-takers. In both sets of items, there were significantly more inference references to HOS items than to K items showing a relation between items intended to elicit "higher order skills" and the subjects' reports of inferential processes.

While the memory/inference measures were developed to distinguish between K and HOS items, they were also examined within each of the four original Bloom item-groups. Memory references yielded no pattern across the two sets of items, consistent with the 1972 study's findings. Latencies and inference references, however, showed three significantly different levels corresponding with K, C/Ap and An. (This suggests a means of empirically developing a taxonomy of "process".)

These findings confirm the inappropriatness of applying the Bloom taxonomy directly to the construction of multiple-choice items. The item-writer's intent is demonstrably an insufficient guide to determining the "process" elicited by an item. In addition, the results point to the need for reconsidering such process descriptions in light of analyses of verbal solution strategies. Further, information about these "extra-cognitive" aspects of multiple-choice item performance (e.g., choice-latency, process description and demand characteristics) seems useful as an adjunct to the use of traditional structural characteristics (e.g., difficulty & discrimination indices) in the construction of multiple-choice instruments capable of assessing process as well as content.

TABLE 1

	Mean Latency/Item by Item Type		Memory References/Item by Item Type		Inference References/Item by Item Type	
	Glacier	Mendeleef	Glacier	Mendeleef	Glacier	Mendeleef
K	20.26	24.96	.52	.84	1.18	1.32
HOS	49.56	50.91	.61	.41	3.40	2.90
[$F_{(1, 17)}$	78.4026	71.7556	0.9165	28.5501	58.6285	32.8744]
K	20.26	24.96	.52	.84	1.18	1.32
C	42.08	48.97	.83	.68	2.68	2.79
Ap	40.35	43.26	.58	.30	2.49	2.81
An	67.31	56.98	.41	.32	4.18	3.07
[$F_{(3,51)}$	44.6508	31.4629	4.8889	7.8378	33.7125	14.9309]
[H.S.D.	10.77	9.07	0.30	0.36	0.78	0.77]

General References

Bloom, B.S. (ed.) Taxonomy of educational objectives, Handbook I: cognitive domain. New York: David McKay & Co., 1956

Bloom, B.S. & Broder, L.J. Problem-solving processes of college students: an exploratory investigation. Supplementary Educational Monographs, 1950, No. 73.

Campbell, A.C. Some determinants of the difficulty of non-verbal classification items. Educational and Psychological Measurement, 1961, 21 899-913.

Connolly, J.A. & Wantman, M.J. An exploration of oral reasoning processes in responding to objective test items. Journal of Educational Measurement, 1964, 1, 59-64.

Cronbach, L.J. Essentials of psychological testing. (2nd ed.) New York Harper, 1960.

Diamond, J.J. & Williams, D.V. An identification of the process response to Taxonomy-type test items. Paper presented at meetings of the American Educational Research Association 1972.

Diamond, J.J. & Williams, D.V. Demand characteristics of multiple-choice test items. Paper presented at meetings of the National Council for Measurement in Education, 1975.

Johnson, E.S. An information-processing model of one kind of problem solving. Psychological Monographs, 1967, 78.

Kaiser, H.F. Scaling a simplex. Psychometrika, 1962, 27 155-162.

Kropp, R.P. The relationship between process and content item-response. Journal of Educational Research, 1956, 49, 385-388

Kropp, R.P., Stoker, H.W., & Bashaw, W.L. The construction and validation of tests of the cognitive processes as described in the Taxonomy of Educational Objectives. USOE, CRP No 2117, 1966. (a)

Kropp, R.P., Stoker, H.W., & Bashaw, W.L. The validation of the Taxonmy of Educational Objectives. Journal of Experimental Education, 1966, 34, 69-76. (b)

Little, E.B. & D'Agaro, G.E. Inside a student's mind during a final examination: Thoughts on wrong responses only. Peabody Journal of Education, 1956, 43, 113-119.

Miller, G.A., Galanter, E. & Pribram, K.H. Plans and the structure of behavior. New York: Holt, 1960.

Scates, D. Complexity of test items. Journal of Educational Research, 1936, 30, 77-92.

Smedslund, J. Educational Psychology. In P.R., Farnsworth, Olga McNemar, and W. McNemar. (Eds.) Annual Review of Psychology, 1964, 15, 251-276.

Williams, D.V., Diamond, J.J.; Kreher, L. and Braucher, J. Memory vs. Inference: A preliminary study of process-referenced test items. Paper presented at meetings of the American Educational Research Association, 1975.