DOCUMENT RESUME

ED 058 284 TM 000 999

AUTHOR	Garvin, Alfred D.
TITLE	Non-Chance Results from a Pure-Chance Test: A Study in Response Position Selection Set.
PUB DATE	Sep 71
NOTE	8p.; Paper presented at the Annual Meeting of the American Psychological Association, Washington, D. C., September 1971
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Conditioned Response; *Objective Tests; *Patterned Responses; *Response Style (Tests); Test Bias; *Test Construction; Testing; Test Interpretation; Test Results; *Test Wiseness

ABSTRACT

The objectives of this study were to ascertain the existence of any widely held, systematic sets in response position selection (RPS) and to evaluate the potential biasing effects of such sets on multiple choice and true-false test results. It is concluded that a sudden change in the accustomed pattern of keyed response positions can shift observed scores downward by a significant amount. (DLG)

Non-chance results from a pure-chance test: A study in response position selection set

Alfred D. Garvin

University of Cincinnati

There are at least three ways a testee may arrive at the correct answer or, more operationally, response position selection (RPS) for a given item in a conventional objective achievement test: 1.) he may think he knows or recognizes the correct response and make his RPS accordingly; 2.) he may believe he has discovered a specific determiner--some salient anomoly of text or format that provides a clue to the desired response--and base his RPS on this clue; or 3.) he may disregard the content of the item entirely and make an arbitrary RPS, either randomly or according to some systematic set. Of course, any of these RPS strategies can also lead to a wrong answer but we are not concerned with such outcomes here.

All correct responses to a given item look alike on the test protocol; there is no way to know for sure how a given testee arrived as his RPS for a given item. However, it is reasonable to assume that a rational testee attempts these RPS strategies in the order given above, proceeding down the cognitive hierarchy until, on one basis or another, he is prepared to make his RPS. Unless a testee is absolutely certain of his RPS on the basis of the substantive content of the item, i.e., he knows the answer and knows that he knows it, or his uncertainty is wholly resolved by an unmistakable specific determiner, his RPS is more or less arbitrary, in proportion to his unresolved uncertainty. Since there is almost always <u>some</u> unresolved uncertainty, and very often there is a great deal of it, almost every RPS is arbitrary to some degree and many are almost wholly so. It follows that any psychological factor that influences the arbitrary component in each RPS also influences objectives test scores. Given the foregoing rationale, the following research questions commend themselves to our urgent attention:

1. Do any of the psychological factors that influence RPS's operate systematically within individuals?

2. If so, do these factors operate similarly in most individuals?

3. If so, does the operation of these factors effect a consisent bias on the results of objective achievement tests?

4. If so, would an inadvertant reversal in these factors effect a misinterpretable discontinuity in test results?

Stated very simply, the objectives of this study were to ascertain the existance of any widely held, systematic sets in RPS's and to evaluate the potential biasing effects of such sets on the results of such tests. More specifically, this study sought to:

1. Ascertain experimentally, in the absence of <u>any</u> RPS clues at all, the degree to which test-wise students exhibit any common, systematic patterns of RPS preferences on short multiple-choice (M-C) and true-false (T-F) tests, and

2. Evaluate empirically the maximum probable bias that might be introduced into the results of such tests if the patterns of keyed responses were inadvertantly made to deviate maximally from these "natural" RPS patterns.

Method

The subjects (Ss) were 73 undergraduates enrolled in a colleague's course in Developmental Psychology. None of these Ss had ever taken an objective test written and keyed by the experimentor (E).

At the beginning of her class session just before midterm, the course instructor introduced E merely as a colleague who would conduct a special "pre-midterm exam." Answer sheet forms familiar to these Ss were distributed. These forms provide spaces in which the S prints (capital) letter response symbols. First of all, E announced that this "pre-test" would comprise 25 M-C items and 10 T-F items. He added that the pattern of the answers on the scoring key embodied "good measurement practice." Then, E announced that the "questions" for the pre-test were not available then--but he had to have their "answers" anyway. A brief, remarkably well-taken explanation served to dispel the Ss' understandable dismay and, in most of the cases, engaged their good-natured cooperation. The gist of it was that E wanted to know--and to let them know--how well they could do when they tried, openly and earnestly, to "outguess" an objective test. When it was clear to all Ss that they were merely to produce a "psuedo-random" RPS pattern, the "pre-test" was begun.

When all Ss had completed the "pre-test," answer sheets were exchanged among Ss (to reduce the effect of ego involvement on scoring accuracy) and E read the "answers" from a key while Ss scored each other's tests, recording separate scores for the M-C and T-F subtests. The keyed M-C "answer" sequence was: D, A, A, C, B, E, C, E, D, B, A, C, B, B, D, E, A, D, C, E, E, C, D, B, A. It can be seen that the frequency distribution of the five response positions available is rectilinear and that no response occurs more than twice in a row. Subject to these constraints, the response pattern is at least quasi-random. The keyed T-F "answer" sequence was: F, F, T, F, T, T, F, T, T, F. This, too, conforms to "good measurement practice" in that there are five T's and five F's and no response occurs more than twice in a row. However, this "answer" pattern was deliberately designed--<u>before</u> the test was administered--to deviate maximally from an hypothesized "natural" T-F response sequence.

A provisional score distribution for each subtest was obtained by calling for shows of hands to provide some immediate feedback. A detailed feedback session was conducted one week later when the data were fully analysed.

Results

Of the 73 original Ss, 10 submitted systematic, as opposed to psuedo-random, RPS patterns and these were eliminated from the study. The most common systematic M-C pattern was A, B, C, D, E, A, B, C, D, E, etc. The most common systematic T-F pattern was all T's. Every S who gave a systematic response pattern on the M-C subtest also gave one on the T-F subtest.

The remaining 63 Ss comprised 18 men and 45 women. All data were analysed separately by sex and no significant differences were found. Thus, all results reported here are pooled across sexes.

The frequency distribution of scores on the M-C subtest was:

X:	11	10	9	8	7	6	5	4	3	2	1
f:	1	1	3	3	3	11	9	12	10	8	2

The mean M-C score was 4.76; this did not differ significantly from the expected value of 5.0 ($t < 1.0$). However, against an expected mean of 5.0 for each response position, the observed means were:

A	B	C	D	E
5.67	5.59	6.00	4.12	3.59

A Friedman two-way analysis of variance on rank order of RPS preference gave a chi-square (ranks) of 75.8 with 4 d.f. ($p < .001$).

The frequency distribution of scores on the T-F subtest was:

X:	7	6	5	4	3	2	1	0
f:	3	5	13	12	19	5	3	3

The mean T-F score was 3.71; against an expected value of 5.0, the t-test yielded a statistic of 6.23 ($p < .001$). The mean number of "T" responses was 5.71. A Friedman test of equal preference for T and F yielded a chi-square (ranks) of 18.65 with 1 d.f. ($p < .001$). It is interesting to note the frequency distribution of consecutive errors, beginning with item #1:

Errors:	1	2	3	4	5	6	7	8	9	10
f:	55	33	31	22	18	14	11	5	3	3

The K-R (20) reliability of the M-C subtest was +.26; for the T-F subtest, it was +.24; and for the total test, it was +.23. The product-moment correlation between sub-test scores was -.001.

Discussion

When testees attempt to produce a psuedo-random distribution of RPS's over five response positions, they tend to favor positions C, A, and B, in that order, and to avoid position E. Since these Ss were presumably test-wise, we may conclude that this is largely due to conditioning; these are the response positions where they have previously found correct answers.

Students acquire most of their M-C "test-wiseness" on teacher-made tests and these typically comprise four-choice items; if a fifth choice (E) is offered at all, it is often "All (or None) of the above."--and is inevitably wrong.

It is just possible that a second factor may contribute to the skewness of arbitrary RPS's on M-C tests. It will be remembered that Ss were asked to print capital letter response position symbols in a column of blanks on their answer sheet--a common classroom testing procedure. The familiar use of the capital letters A-F as grading symbols has attached an affective order of preference to them. This could account, in part, for the relative attractiveness of "A" where a simple central-tendency theory would have it be equal to "E".

Finally, it should be reported that not one S produced a rectilinear (5-5-5-5-5) psuedo-random distribution and that not one selected "D" for the first item.

In this experiment, the distribution of keyed response positions was rectilinear and the mean pure-chance score was not significantly different from that theoretically expected. However, if the "conditioning" theory advanced herein is sound, the keyed response position patterns previously encountered by these Ss must have substantially coincided, in most cases, with that displayed in this experiment. The effect of this congruence of keyed response position patterns and students' RPS patterns is to inflate all M-C scores, particularly those nearest the chance end of the score distribution. A shift to an oppositely skewed keyed response position pattern, whether inadverant or capricious, would result in a sudden drop in M-C test scores that would very probably be misinterpreted by both testers and testees.

In the case of T-F tests, particularly very short ones, the effects of conditioning on RPS behavior are even more pronounced. The term "Acquiescence Set" has been applied to the common tendency of respondents to agree with any reasonably plausible, grammatically correct proposition. In the context of T-F achievement tests, this is manifested in a predisposition to favor T over F. In this experiment, 55 of the 63 Ss gave a T response to the first item and 33 of these gave a T response to the second--where there was nothing at all to agree or disagree with! This is more than twice the number of T, T responses to be expected by chance alone. Here, too, we must conclude that students have typically found more True statements than False ones on the T-F tests they have taken.

The sequential dependence inherent in T-F RPS's makes "getting off on the wrong foot" especially disasterous (and 55 out of 63 Ss did). In this experiment, 31 of the 33 Ss who "missed" the first two items also missed the third; 11 of the 14 who "missed" the first six items also missed the seventh. The number of Ss who missed all 10 items--three--was nearly 50 times the number to be expected by chance alone.

It is clear that students have "learned," consciously or unconsciously, to reproduce the T-F response patterns that their teacher-test constructors have, consciously or unconsciously, tended to produce. This tendency for inexpert test constructors to offer more True than False propositions and, so, to begin with a True one is more likely to persist than the characteristic bias found in M-C keying pattern; M-C alternatives can be rearranged almost at will while plausible False propositions are difficult to compose. Nevertheless, a test constructor could, either inadvertantly or capriciously, reverse the expected proportions of True and False propositions and cause a misinterpretable drop in apparent T-F test performance.

Conclusion

Consistent inflation of objective achievement test scores does little real harm. However, it is shown here that a sudden change in the accustomed pattern of keyed response positions can shift observed scores downward by a significant amount and, if this is done unwittingly, the resulting score discontinuities are likely to be misinterpreted by both testers and testees. It is important that teacher-test constructors be alerted to this reliable and powerful psychological phenomenon and be made to appreciate its implications for the interpretation of test results.