DOCUMENT RESUME

ED 161 941 TM 007 976

AUTHOR Rogers, Bruce G.
TITLE Effects of Type of Instructions on the Characteristics of an Affective Instrument.
PUB DATE Mar 78
NOTE 10p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Affective Tests; Higher Education; Overt Response; *Performance Factors; Rating Scales; *Reaction Time; *Response Style (Tests); Scores; *Testing; Test Reliability

ABSTRACT

The manner in which instructions to respondents influenced the psychometric results of an affective instrument was examined. Undergraduate students enrolled in an educational measurement course were randomly assigned to either immediate response or careful response treatments. A fifteen-item attitude scale was administered to both groups. The immediate reaction group was instructed to rely on their first impressions of the items and were advised not to go back once they had responded to a statement. The careful response group was instructed to be as sincere and accurate as possible in their answers and to feel free to change responses if necessary. For each group, coefficient alpha was computed as an estimate of internal consistency. The t-test and discriminant analysis were used to contrast the two data sets. No item types could be identified to differentiate between the two groups of respondents. In addition, instrument reliability and total scores were not significantly different under either condition of instructions. (Author/JAC)

Effects of Type of Instructions on the Characteristics of an Affective Instrument

Bruce G. Rogers
Department of Educational Psychology and Foundations
University of Northern Iowa

Abstract

This study sought to determine how instructions to respondents influenced the psychometric results of an affective instrument. Some respondents were asked to "immediately respond" to each item, while others were asked to "carefully respond". The results indicated that neither reliabilities nor total scores were adversly affected to any major degree by either set of instructions, but it was suggested that perhaps the "carefully respond" type of instruction was more ethically defensible.

A paper presented at the Annual Meeting of the National Council on Measurement in Education

NCME Session 9B
Toronto, Canada
March 27-31, 1978

Printed in the U.S.A.

Effects of Type of Instructions on the Characteristics of an Affective Instrument

Objectives of the Inquiry

A cursory review of commonly used affective measures (e.g. Robinson and Shaver, 1973; Shaw and Wright, 1967) reveals a diversity in types of instructions, particularly with respect to the amount of reflection requested of the respondent prior to recording his overt response. Some instructions request immediate responses, others encourage careful reflection, while yet others omit any reference to the matter.

Researchers in the area of cognitive test construction have investigated a somewhat related phenomenon. To answer the question, "Should examinees feel free to change their original responses to objective test items?," most researchers have focused on the effects on total scores, with reasonably consistent findings that response changes improve scores more often than not. After reviewing the research literature of the past fifty years, Mueller and Wasser (1977) concluded that "from these studies it seems clear that students taking objective tests should be encouraged to deliberate on items for which they are not sure of the correct answer" (p. 12).

Recently, when Crocker and Benson (1977) investigated the effects of examinee response changes upon other psychometric properties of cognitive tests, they found that "test reliability does not appear to be adversly affected when examinees change their answers" (p. 4).

If respondents to cognitive measures should be encouraged to deliberate on items, how should respondents to affective measures be instructed? The purpose of this investigation was to compare the effects of two extremes of instructions on the psychometric properties of an affective instrument. In particular, the following questions were addressed:

1. Is test reliability (internal consistency) affected?
2. Is the total score affected?
3. Are there particular types of items that are affected?

Sample and Instrument

Each of the eleven sections of an undergraduate educational measurement course was randomly assigned to one of two treatments, designated as either "Immediate Response" or "Careful Response." The resulting two groups contained 199 and 226 subjects, respectively.

Selected as the instrument was a well-researched attitude scale (Kerlinger, 1967; p. 62, 247), entitled Education Scale VII, Subscale B (Traditionalism). It was chosen because of its appropriate subject matter (education), brevity (15 items), and reasonable reliability (e.g. reported alpha values between

.69 and .78). The instrument was modified from its original 7 point scale to a 5 point Likert-type scale (strongly agree to strongly disagree). An illustrative item follows: "The curriculum consists of subject matter to be learned and skills to be acquired."

Methods

Each group was asked to read the instructions and follow them exactly.

"Immediately Respond" group: "Indicate your immediate reaction without thinking too much about it. Remember that first impressions are usually best in such matters. Go rapidly; do not go back once you have marked a statement."

"Carefully Respond" group: "Consider each question carefully. Be as sincere and accurate as possible within the limited time. If upon reflection, you desire to alter your response, please feel free to change it by erasing or marking it out and circling another response."

To analyze the resulting data, the five responses Strongly Disagree through Strongly Agree, were assigned the integers one through five, respectively. For each experimental group, coefficient alpha (Cronbach, 1953) was computed as an estimate of internal consistency. The two data sets were then contrasted with the t-test on total scores and with a discriminant analysis using Wilks Method (Nie, et al, 1975, p. 447).

Results

The two obtained values of coefficient alpha (Table 1) were compared using the test for independent correlation coefficients (Hays, 1973, p. 663). The resulting z value of .97 did not allow rejection of the null hypothesis, even with a significance level as large as 10% in each tail. This larger than usual significance level was chosen to reduce the Type II error. When Owen (1962, p. 518), in his well-known book of mathematical tables, desired to show that a pair of samples were essentially equivalent, he compared them using a 10% significance level in each tail. Accordingly, in this case, the use of that larger significance level increases our confidence that no important difference exists between the reliabilities of the two forms of the attitude scale.

When the groups were compared on total score, the resulting difference of 1.3 points produced a t-value of 2.03, which was significant at the .05 level.

To see if particular items were differentially affected by the instructions, a discriminant analysis was performed. Wilks method is a stepwise solution, selecting at each step that variable (item in this case) which maximizes the F ratio between group centroids. However, this procedure did not yield any function significant at the .05 level; thus, no particular item types could be identified to differentiate between the two groups of respondents.

Educational Importance of the Study

On the basis of the obtained results, it appears that the instrument reliability was not significantly different under either condition of instructions. The differences between the alpha coefficients proved to be non-significant even with a large rejection region. Further, it can be seen by inspection that these coefficients are in the vicinity of those reported by Kerlinger (1967), as was mentioned earlier in this paper.

Although the t-test computed on the difference between the mean scores of the groups proved to be statistically significant, it is not certain that this should be considered as having practical importance. The difference between means, in terms of standard deviation units, is designated as the "effect size" by Cohen (1969, p. 8). For this comparison, the effect size proved to be only .2, a value labeled as "small" by Cohen (1969, p. 23). Furthermore, the computation of omega squared indicated that the independent variable (type of instructions given) accounted for only about one percent of the variance in the scores (Hays, 1973, p. 413). There would probably be universal agreement that this, indeed, demonstrates a statistically significant yet trivial relationship between the type of instructions and the composite scores.

It might be hypothesized that certain items were affected by the instructions, but that they were not discernible by the t-test because they were only a part of the total score. But the non-significant results of the discriminant analysis failed to support this hypothesis.

The results of this study suggest that the variations in directions did not create substantial variation in the psychometric properties of the instrument or in the level of scores generated. The researcher and follower of research might be encouraged by these results, inasmuch as they can be interpreted as indicating that questionnaire results are not likely to be seriously affected by the type of instructions given to the respondents.

However, the generalizability of these results may be limited. In this study only undergraduate college students were used and only one test was administered. Perhaps the results would be different with either more mature or less mature subjects, or with different instruments. Future studies might also examine item type, instrument length, subject matter, etc.

Although the findings of this study did not show significant differences, perhaps there is an overriding ethical consideration of what is conveyed to the students responding to the instrument. Should pupils be encouraged to act upon their first impressions? One of the hallmarks of a civilized society is that the members carefully consider their thoughts before translating them into action. Perhaps students can train themselves so that their first reactions are consistent with careful reason, but it appears doubtful that immediate judgments will be better than carefully reasoned ones. Encouraging considered responses, as opposed to advocating first

reactions, is a desirable procedure even though it makes little difference psychometrically.

A perusal of commonly used affective instruments indicates that many authors have both written instructions requesting careful judgments and allocated sufficient time for this process (as is usually done with cognitive instruments). Perhaps this investigation will serve to further encourage that practice.

Table 1

Test Characteristics Classified by Type of Instruction

	"Immediately Respond"	"Carefully Respond"
Mean	42.69	43.98
SD	6.54	6.53
Alpha	.69	.74
N	199	226

References

Cohen, J. Statistical power analysis for the behavioral sciences. N.Y.: Academic Press, 1969.

Crocker, L. and Benson, J. Effects of examinee response changes on item and test characteristics. Paper presented at the annual meeting of the American Educational Research Association held in New York, N.Y., April, 1977.

Cronbach, L.J. Coefficient alpha and the internal structure of tests. Psychometrica, 1951, 16, 297-334.

Hays, W.L. Statistics for the social sciences (2nd Ed.) N.Y., Holt, Rinehart and Winston, 1973.

Kerlinger, F.N. and Pedhazur, E.J. Attitudes and Perceptions of desirable traits and behaviors of teachers. New York: New York University, Sept. 30, 1967. ERIC # ED019742.

Mueller, D.J. and Wasser, V. Implications of changing answers on objective test items. Journal of Educational Measurement, 1977, 14, 9-14.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., and Bent, D.H. SPSS: Statistical package for the social sciences. N.Y.: McGraw Hill, 1975.

Owen, D. B. Handbook of statistical tables. Reading, Mass.: Addison Wesley, 1962.

Robinson, J. P. and Shaver, P.R. Measures of social psychological attitudes. Ann Arbor, Mich.: Institute for Social Research, 1973.

Shaw, M. and Wright, J.M. Scales for the measurement of attitudes. N.Y.: McGraw Hill, 1967.

Subkoviak, M.J. and Levin, J.R. Fallibility of measurement and the power of a statistical test. Journal of Educational Measurement, 1977, 14, 47-52.