

Dynamics of guessing behavior: Between-group versus within-group designs

BARBARA S. PLAKE and STEVEN L. WISE
University of Nebraska, Lincoln, Nebraska

Within-group and between-group designs have traditionally been employed in research studies that investigate test-taking behavior, particularly guessing behavior. Although within-group designs provide both control for random between-group differences and increased statistical power to detect treatment effects, they are potentially flawed due to carryover effects incurred by administering the same items to the same subjects. Although between-group designs provide a potentially cleaner assessment of test-taking behavior, they allow only inferences to group, not individual, test-taking behaviors. Since test-taking behavior has been shown to be related to individual differences in personality variables, results from between-group designs provide less than optimal information about individual-specific test-taking behavior. By utilizing modifications of the typical within-group designs, carryover effects may be estimated and controlled for, allowing for increased internal validity of within-group designs for the study of test-taking behavior.

In the assessment of the effects of test-taking behavior on multiple-choice test performance, two methodological approaches have been traditionally employed: one investigates how randomly equivalent groups of examinees respond to varying test-taking situations (between-group designs) and another considers the changes in individual examinee's test-taking behavior across varying test-taking situations (within-group designs). From a research design perspective, both of these methodologies have respective strengths and weaknesses that have important implications for their utility in the assessment and analysis of test-taking behavior.

BETWEEN-GROUP DESIGNS

By typically using large sample sizes and employing researcher control over random assignment of subjects to conditions, the between-group design has the potential for powerfully addressing some important research questions regarding test-taking behavior. However, the questions being addressed via between-group designs are group-specific, not individual-specific. Several assumptions are needed to make individual-level inferences from between-group research designs. For example, in comparing transformed total scores achieved when one group takes an examination under correction-for-guessing instructions to the scores obtained from another randomly equivalent group that takes the same examination under no-correction-for-guessing instructions, it is often assumed that the items omitted by the group under correction-for-guessing instructions are the same items that showed a guessing pattern of responses under no-correction-for-guessing instructions. Even more important, such designs

do not provide information regarding the change in an examinee's test-taking behavior across instruction change. Therefore, between-group design requirements of relatively larger sample size, coupled with the aggregate, group-level information provided by these data render between-group designs a less than optimal research paradigm in the study of individual examinee test-taking behaviors.

WITHIN-GROUP DESIGNS

Within-group designs provide controls for random between-group differences and also allow for increased statistical power for detecting within-subjects effects. More importantly, many previous studies have pointed out the individualistic character of test-taking behavior. Personal dimensions such as submissiveness (Votaw, 1936) and risk-taking behavior (Slakter, 1968) are among the characteristics which have been shown to be related to test-taking behavior. The preferred design paradigm to study test-taking behavior would be one that provides information relevant to individual behavior. Therefore, within-group designs have the potential for providing clearer and more directly relevant information in the study of test-taking behavior than their between-group counterparts. However, the within-group design has its own disadvantage. Carryover effects across conditions have the potential for threatening the internal validity of a within-group study of test-taking behavior.

STUDIES ON THE IMPACT OF SCORING METHODS ON TEST-TAKING BEHAVIOR: NEEDED DIRECTIONS

In this section, we suggest some improvements to within-subjects designs that hold promise for reducing

The authors' mailing address is: 29C Teachers College, University of Nebraska, Lincoln, NE 68588-0440.

and/or eliminating some of the primary sources of confounding in within-group studies. In addition, we address additional research questions and approaches that complement the information obtained from within-group designs.

Recommended Improvements for Within-Group Designs

For the within-group design paradigm, individual examinees are usually administered a multiple-choice examination initially using correction-for-guessing instructions. Subsequently, the examinee is given the opportunity to reconsider items previously omitted under correction-for-guessing instructions and asked to make a selection among the alternatives for each previously omitted item. Some possible improvements to this methodology could be:

1. To include a separate option for each item in a formula-scored examination that specifies, "I OMITTED THIS ITEM." This would enhance the quality of information in current within-group methodologies, because it is difficult to distinguish between items that were consciously omitted from items which were not considered and therefore not marked, possibly due to time limitations.

2. To utilize no-correction-for-guessing directions on the first administration and to subsequently (preferably immediately afterwards) have the examinees identify the items they would have omitted under correction-for-guessing instructions. Using this approach, the examinees' choices to the individual items would already be made, eliminating the possibility that increased knowledge or insight about the reconsideration item aided in the particular alternatives selected.

3. To readminister the entire examination rather than readministering only the omitted items under no-correction-for-guessing directions. The amount of score change due to increased knowledge and/or increased insight could be estimated from the change from previously attempted items. Scores on previously omitted items could thereby be adjusted by the subjects' increased performance across administrations.

Related Complementary Research Approaches

In addition to combining information regarding test-taking behavior with a consideration of specific personality traits, some related areas of research should be addressed since they could provide further illumination of factors related to test-taking behavior. For example, research is needed which inspects some of the fundamental assumptions regarding examinees' strategies for deciding which options to select on multiple-choice tests (i.e., item-consideration strategies). Furthermore, an analysis of the items omitted under correction-for-guessing directions may provide additional data regarding test-taking behaviors.

Item-consideration strategies. This research will provide essential empirical verification for some of the as-

sumptions underlying formula scoring. In particular, the algorithm for formula scoring assumes that the examinee will attempt an item if any partial knowledge exists. Partial knowledge, with correction-for-guessing directions, is often operationally defined as "being able to eliminate at least one option." This implicitly assumes an option-elimination approach to item consideration. More information is needed about the actual item-consideration strategies employed by examinees on tests with correction-for-guessing directions. Some recent work by Frary, Cross, and Sewell (1985) provides some insight into these strategies.

Item characteristics. One element of guessing behavior that has not received much attention in within-subject designs is the characteristics of the particular items that examinees tend to omit under correction-for-guessing instructions. Of interest are the items that examinees tend to omit on initial administration via correction-for-guessing directions which, upon readministration with no-correction-for-guessing instructions, examinees tend to have a better-than-chance probability of getting correct. Previous research has assigned partial knowledge on behalf of the examinee as the cause for better-than-chance probability of correct answering on readministration of these items. Perhaps, in addition, these items are excessively wordy, cumbersome, tricky, or time-consuming; under these conditions, the examinee's choice to omit certain items may be more of an exercise in concentration of intellectual energies on more reasonable items, especially on speeded tests.

CONCLUSION

Within-subject designs have the potential for providing important and useful information regarding the guessing behavior of examinees, information that is not directly available from between-subject designs. Current within-subject design research methodologies, however, can be vulnerable to confounding due to the utilization of the same subjects and the same items across administration modes. Several suggestions for improvement in methodology are made in this paper. One such improvement, having the examinees respond to *all* items initially and then identify items that they would have omitted under correction-for-guessing directions, removes the potential influence of increased knowledge or insight in the examinees' selection of item alternatives. Another variation provides for a statistical procedure that estimates the degree of increased knowledge or insight for all items in the test for an examinee at readministration. Therefore, even though within-subject research studies are traditionally flawed due to readministration using identical items, these improved research approaches allow for a reduction (or even an elimination) of these sources of confounding. Hopefully, these design improvements, complemented with additional information about examinee

test-taking strategies and about items which tend to be frequently omitted by examinees will lead to increased insight into examinee guessing behavior.

REFERENCES

- FRARY, R. B., CROSS, L. H., & SEWELL, E. H. (1985, April). *Partial information and the "correction" for guessing*. Paper presented at the NCME annual meeting, Chicago, Illinois.
- SLAKTER, M. J. (1968). The effect of guessing strategy on objective test scores. *Journal of Educational Measurement*, *29*, 115-128.
- VOTAW, D. F. (1936). The effect of do-not-guess directions upon the validity of true-false or multiple-choice tests. *Journal of Educational Psychology*, *27*, 698-703.

(Manuscript received for publication March 24, 1986.)