

Reply to van der Linden's "Thoughts on the Use of Decision Theory to Set Cutoff Scores"

Dato N. M. de Gruijter
University of Leyden

Ronald K. Hambleton
University of Massachusetts

We appreciated the opportunity to provide a few comments on the thoughtful paper prepared by van der Linden (1984), a paper that was, in part, a response to our initial paper (de Gruijter & Hambleton, 1984). Van der Linden's paper addresses not only (1) the problems we raised about using decision theory to set cutoff scores but also (2) a comparison of decision theory with other common standard-setting methods, and (3) uses of decision theory with criterion-referenced tests. Our response here will be limited to several of the points raised by van der Linden in area 1 since our paper was only concerned with the use of decision theory in setting standards. At the outset we would like to emphasize that we agree with van der Linden about the merits of decision theory for providing a framework for considering many problems that arise in criterion-referenced measurement. In fact, the second author was one of the initial proponents of such an approach (Hambleton & Novick, 1973). We will consider the problems in the same order that they are considered by van der Linden, although that order is different from the order in which the problems were presented in our paper.

In our paper we argued (problem 4) that setting optimal cutoff scores using the test scores obtained from a particular test administration *might* result

in losses. To us, it seemed important to draw attention to this possibility. We did not, however, suggest that such losses *inevitably* result from the use of decision theory. Further, we argued, that within the decision-theoretic framework, the problem could be circumvented by optimizing the cutoff score under the condition that examinees know beforehand which cutoff score was going to be used. We therefore do not understand van der Linden's statement that our conclusion with respect to losses is wrong and that more loss than necessary will be incurred if decision-theoretic optimization is omitted.

With respect to the potential harm resulting from discrepancies between expected and actual cutoff scores (problem 5), van der Linden seems to agree with us. In fact, his third solution under the heading "The Cutoff Score As a Target for Examinees" is quite similar to our abovementioned proposal for a fixed cutoff score.

With respect to the choice of the relevant population (problem 3), we disagree with van der Linden. It seems incorrect simply to equate the population of interest with the sample of examinees who took the test, instead of generalizing over examinee samples. The narrow definition of a population results in a sample dependency of the probability of passing. When a person happens to take a test together with a relatively low performing group of examinees, his/her chances of passing are lowered. Such an outcome is inconsistent with the

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 8, No. 1, Winter 1984, pp. 19-20
© Copyright 1984 Applied Psychological Measurement Inc.
0146-6216/84/010019-02\$1.35

spirit of criterion-referenced testing, i.e., examinees should be assessed in relation to domains of content and *not* other candidates.

Also, we note that varying difficulties between test forms can be a real nuisance. When, for example, the performance on a test is relatively low, this might be due to test difficulty as well as to the characteristics of the examinee sample. If the test is more difficult, the cutoff score should be lowered. A superficial analysis of the decision problem would suggest a raised cutoff score if the target population had been equated with the group of examinees.

Van der Linden suggests that some of the problems related to different test models and parameter estimates (problem 2) can be solved by using carefully calibrated item pools. This is correct, but how common is it for test developers to build their tests from calibrated item pools?

With respect to the problem of inaccurate parameter estimates (problem 1), it should be noted that the Angoff and Nedelsky methods are methods for setting standards on the true scale. So, when test forms have different true score scales, these methods cannot give accurate estimates of π_0 , which is defined by a generalization over test forms. Also, in other aspects, the methods have sometimes led to disappointing results. Nevertheless, the Angoff and Nedelsky methods can provide first estimates of π_0 . We did not deny this possibility. We stressed, however, that uncertainty with respect to π_0 should

be acknowledged and built into the decision-theoretic analysis. Too frequently, π_0 has been accepted at face value and consequently the decision problem has been incorrectly simplified.

In conclusion, we remain less optimistic than van der Linden about the merits of decision theory for setting standards. However, we share with him the wish to pursue a decision-oriented search for improved solutions to the problem of setting standards.

References

- de Gruijter, D. N. M., & Hambleton, R. K. (1984). On problems encountered using decision theory to set cutoff scores. *Applied Psychological Measurement*, 8, 1-8.
- Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 1, 159-170.
- van der Linden, W. J. (1984). Some thoughts on the use of decision theory to set cutoff scores: Comment on de Gruijter and Hambleton. *Applied Psychological Measurement*, 8, 9-17.

Authors' Addresses

Send requests for reprints or further information to Dato N. M. de Gruijter, Educational Research Center, University of Leyden, Boerhaavelaan 2, 2334 EN Leyden, The Netherlands; or Ronald K. Hambleton, University of Massachusetts, Laboratory of Psychometric and Evaluative Research, Hills South, Room 152, Amherst MA 01003, U.S.A.