

# Some Thoughts on the Use of Decision Theory to Set Cutoff Scores: Comment on de Gruijter and Hambleton

Wim J. van der Linden  
Twente University of Technology  
Enschede, The Netherlands

In response to an article by de Gruijter and Hambleton (1984), some thoughts on the use of decision theory for setting cutoff scores on mastery tests are presented. This paper argues that decision theory offers much more than suggested by de Gruijter and Hambleton and that an attempt at evaluating its potentials for mastery testing should address the full scale of possibilities. As for the problems de Gruijter and Hambleton have raised, some of them disappear if proper choices from decision theory are made, while others are inherent in mastery testing and will be encountered by any method of setting cutoff scores. Further, this paper points at the development of new technology to assist the mastery tester in the application of decision theory. From this an optimistic attitude towards the potentials of decision theory for mastery testing is concluded.

In contrast with current notions, tests are more often used for decision making than for measurement purposes. In fact, one of the few situations where tests are used as pure measurement instruments is in research, for instance, when reactions of subjects to varying conditions are measured or when relationships between variables are to be determined. In all applied fields, though, tests are predominantly used to make such decisions as admission to schools, selection of personnel, treatment assignment in therapies, pass-fail decisions in instructional units, vocational guidance deci-

sions, program and product evaluations, certification, placement in the military, and so forth. In the light of this overwhelming number of applications, it seems contradictory that psychometricians have invested so much energy in the improvement of tests as measurement instruments. For example, one of the basic differences between the two types of test uses is that in measurement the concern, generally, is in uniform accuracy of the instrument along some part of the scale, whereas in decision making this is seldom a prerequisite.

A favorable exception to this somewhat one-sided interest is found in the mastery testing literature. For the most part, this is due to a seminal paper by Hambleton and Novick (1973). In their paper, one of the three "classics" in the history of criterion-referenced measurement (van der Linden, 1982a), the authors proposed the use of (Bayesian) decision theory to optimize and analyze cutoff scores on mastery tests. The proposal has stimulated others to work out decision-theoretic solutions as well, and a great variety of versatile contributions have followed. Currently, notwithstanding Glass's (1978) critical comments (see, however, Hambleton, 1978; Popham, 1978), statistical decision theory has been adopted as the basic paradigm in mastery testing research. In view of this, de Gruijter and Hambleton's (1984) paper is a timely and welcome attempt at evaluating the progress being made in this area. The paper, however, strikes the reader by the fact that, despite a basic trust in the potential of decision theory for

mastery testing, it has some pessimistic overtones. These are mainly due to five different obstacles that, according to the authors, are encountered when using decision theory to set cutoff scores on tests.

The present reaction is inspired by a somewhat more optimistic attitude towards the use of decision theory in mastery testing. In particular, it is felt that de Gruijter and Hambleton are not correct in passing their five problems fully to the account of decision theory. Also, decision theory offers many unique advantages that they do not mention but should be put forward to allow a fair comparison with alternatives. In the following, some aspects of decision theory and its potential for use in mastery testing are elucidated further. This is not only to complete the exposé in de Gruijter and Hambleton's paper but also to set the stage for a subsequent critical review of the five problems. In the final section the most important conclusions from this review are summarized.

#### Aspects and Potential of Decision Theory

In a section entitled "The Basic Paradigm for Obtaining Optimal Cut-off Scores," de Gruijter and Hambleton discuss an algorithm for obtaining an optimal cutoff score on the test adopting an (empirical) Bayesian framework with threshold utility and assuming an item sampling model. This may leave the reader who is not familiar with the subject with the impression that the contribution of decision theory to mastery testing amounts to this. Therefore, the perspective will be broadened somewhat here. Since it is not the intention to offer a coherent treatment of decision theory, the presentation will be point by point and without any emphasis on logical order.

#### The Nature of Decision Theory

A helpful distinction, which may clarify much of the discussion, is between decision theory as a *theory* versus a *collection* of decision rules. The former corresponds with decision theory as a branch of statistics concerned with the formal aspects of decision problems, the definition of classes of de-

cision rules, and the examination of their properties against certain criteria of optimality. (A standard reference here is a textbook such as Ferguson's, 1967.) The latter is a reduction of the former. Such a reduction is natural in an applied setting where the main interest is in solving a practical problem at hand; it may, however, become harmful as soon as reasonable alternatives are disregarded or what is applied is no longer understood.

The first thesis this author would defend is that, even if decision theory in the latter sense is not applied, i.e., if cutoff scores are not determined using rules from decision theory, it still is most useful for those engaged in mastery testing to be conversant with decision theory in the former sense. Decision theory allows an exact formulation of the mastery testing problem, introduces explicit utility considerations, clarifies the role to be played by test data in decision making, offers means for analyzing the behavior of any decision rule (whether optimal or not), shows under what conditions these rules take a certain shape, and so forth.

#### Impact on Mastery Testing

An example of how decision theory has clarified the mastery testing problem is the following. In much of the mastery testing literature, it has long been ignored that mastery testing involves *two* different cutoff scores. In this tradition, the setting of the mastery standard and the subsequent inference whether examinees display mastery or not, i.e., mastery testing, are confused. Decision theory, on the other hand, helps to neatly distinguish between a cutoff score on the true score variable (mastery standard) and the inference rule defined on the observed test scores, which, under conditions normally satisfied in testing (e.g., Ferguson, 1967, sect. 6.1), takes the shape of a cutoff score on the test.

It can readily be demonstrated that under their respective models, popular methods such as the Angoff and Nedelsky methods yield standards on the true score scale (van der Linden, 1982b); these methods are, as opposed to common belief, not methods for setting cutoff scores on tests. On the

other hand, for instance, Berk's (1976) contrasting groups method and Zieky and Livingston's (1977) borderline group method aim at observed test score distributions and are, again in contrast with common belief, not standard-setting methods.

Ignoring one of the two cutoff score problems, as in the aforementioned methods, amounts to assuming that the test is free of measurement error, exactly the same misunderstanding as ignoring the difference, say, between a true and an observed test score or between an attenuated and a disattenuated correlation coefficient. Even Glass (1978), in his otherwise lucid (but controversial) paper, mistakenly classifies decision theory as a standard-setting method—and subsequently blames it for the fact that it appears to rest on a mastery standard! Decision-theoretic techniques are not substitutes for standard-setting methods but must follow each time such a method is used to allow decisions to be based on fallible test data.

### Behavior of Cutoff Scores

An example of how decision-theoretic analysis can reveal the behavior of a decision rule can be given using de Gruijter and Hambleton's example of the (empirical) Bayes cutoff score under threshold loss. It is shown how the behavior of this optimal cutoff score has a noteworthy dependency on the true score distribution. Suppose that, in these authors' notation, the distribution shows many examinees with  $\pi > \pi_0$  and only a few with  $\pi < \pi_0$ . As the substitution  $a = c = 0$  yields no loss of generality, the expected utility associated with the *optimal* cutoff score may be written as

$$U = bB + dD, \quad (1)$$

whereas it further seems obvious to assume that  $b, d < 0$  (lower utility for incorrect than for correct decisions). Now, for the given population, it holds that  $B$ , the probability of a false master, and  $D$ , the probability of a false nonmaster, tend to 0 and  $\text{Prob}\{X < x_c\}$ , respectively. In such cases it thus holds approximately that

$$U = d\text{Prob}\{X < x_c\}, \quad (2)$$

and it follows that the expected utility tends to be maximal for a low value of  $x_c$ .

In educational terms this implies that hard-working populations are rewarded by *low* cutoff scores. The same analysis shows that low-performing populations are penalized by *high* cutoff scores. This behavior of the optimal value of  $x_c$  was documented for the case of threshold utility by Mellenbergh, Koppelaar, and van der Linden (1977). The same reward mechanism was observed for the case of linear utility by van der Linden and Mellenbergh (1977). As the phenomenon is due to the basic fact the Bayesian solutions use regression functions in the opposite direction, it was coined the "regression from the mean effect" in a treatment by van der Linden (1980).

### Robustness Analysis

As noted earlier, decision-theoretic analysis is most useful even if no decision-theoretic optimization takes place. The foregoing point illustrated a result of such an analysis that may be of avail when selecting cutoff scores on a more intuitive basis. Another example of an analysis useful in mastery testing is robustness analysis, also mentioned by de Gruijter and Hambleton.

As decision theory is based on a complete mapping of the mastery testing problem, it is the proper framework for studying the robustness of decision rules. A recent study relevant to mastery testing is Vijn and Molenaar (1981). Their approach is innovative in that it adopts the concept of a robustness region as the total set of all parameter specifications and assumptions for which the same decision is optimal. The identification of such regions can take place prior to the selection of a specific decision rule and provides valuable information to the decision maker that may assist him/her in the choice of a rule, even if this is not exclusively based on decision-theoretic optimization.

### Selection, Placement, and Classification

It is worthwhile to look beyond the mastery testing problem and to see how decision theory clarifies other decision problems in educational and psychological testing as well. In general, three other

basic types of decisions can be distinguished: selection, placement, and classification decisions.

Selection decisions have been treated decision theoretically by Cronbach and Gleser (1965). This monograph was one of the first attempts to base test applications on statistical decision theory. Solutions to the placement decision problem have also been given in Cronbach and Gleser, whereas the decision-theoretic approach to this type of decision was developed further in van der Linden (1981). Decision theory of classification problems still remains to be elaborated.

Elsewhere the author shows that it is possible to treat these four basic types of decisions in a unified fashion using (empirical) Bayesian decision theory (van der Linden, in press). Within this framework it is also possible to deal with conditions occasionally met in practice such as quota restrictions, multivariate test information, or the presence of differential reaction to test items by subpopulations of testees (culture-fair testing problem). Such a unified approach will enable mastery testing research to profit from solutions to other types of decision problems as well.

Moreover, modern educational systems such as computer-aided instruction or other forms of individualized instruction are usually constructed as systems with several points at which mastery, classification, or placement decisions have to be made. Though this has to be worked out further, decision theory in principle opens the possibility to optimize such systems *simultaneously*, no doubt yielding better solutions than when these decisions are optimized separately. It seems unlikely that a more intuitive approach or an approach based on a less complete mapping of decision problems will address such problems successfully.

### Branches of Decision Theory

Though decision theory gives the impression of a unified theory, nevertheless, important subdivisions are present. One of the differences concerns the way decision rules are ordered to choose the most preferred (or optimal) rule. It is here where, e.g., Bayesian decision theory and minimax theory go different ways, confronting the mastery tester

with an important choice. The advantages of minimax theory for mastery testing have been set forth in papers by Huynh (1980) and Veldhuizen (1982). Another choice is whether or not (subjective) prior knowledge as to true states is allowed (Bayesian versus empirical Bayesian theory). Also, within these frameworks further choices are necessary as to the utility structure and the test model to be adopted.

A review of utility functions and test models current in mastery testing is given in van der Linden (1980). The important thing to note here is that each of these subdivisions and specifications impose different conditions on the application of decision theory to mastery testing. Consequently, as will be shown more in detail below, de Gruijter and Hambleton's five problems cannot hold universally.

### De Gruijter and Hambleton's Five Obstacles

The foregoing has set the stage for a critical review of de Gruijter and Hambleton's five problems. The author's premises in this review are that (1) some of these problems are inherent in mastery testing and not in the application of decision theory; (2) some of these problems depend on what choices are made from decision theory, and may disappear when proper choices are made; and (3) for some of these problems new technology to aid the decision maker in his/her choices is being developed and should be developed further.

### Optimal Cutoff Scores on Various Occasions

This point illustrates that some of the problems raised by de Gruijter and Hambleton pertain to problems inherent in mastery testing and its educational context, and not to the application of decision theory. Every other conceivable method of setting cutoff scores on tests will run up against the same phenomenon—that in a series of test administrations students, when preparing for a test, tune their efforts to results on previous administrations. For example, it is a common experience among teachers that classes and exams known as easy to students are usually taken less seriously.

De Gruijter and Hambleton are correct in their observation that such interactions between cutoff score and students' learning, despite a common mastery standard, generally lead to a series of different optimal cutoff scores. Unlike these authors, however, this is not viewed as a criticism of decision theory in this paper.

As a first comment, it is noted that the suggestion that this interaction leads to a decrease in the expected utility associated with the optimal cutoff score, and thus to a loss incurred by the educational institution, is not correct. In their example, the authors compare only the two right-hand terms of their Equations 7 and 8 under the assumption  $a > c$ . A similar comparison of the two left-hand terms reveals that under the obvious assumption of unchanged distributions of  $X$  given  $\pi$ , it holds that  $B^* > B$  and  $D^* < D$  (an increase in the number of true nonmasters leads to a decrease in the number of false nonmasters and an increase in the number of false masters). Thus, if  $b > d - (a - c)$ ,

$$(d - a)D^* + (b - c)B^* > (d - a)D + (b - c)B, \quad (3)$$

which may compensate for de Gruijter and Hambleton's Equation 9. Whether a loss or a profit is incurred depends fully on the utility structure of the mastery testing problem involved. The same conclusion follows for  $a \leq c$ .

The fact that students may show the above behavior is not an argument against but just an important case for the application of decision theory to mastery testing. Only (Bayesian) decision theory does allow for changes in true score distributions. The correct conclusion from analyses such as de Gruijter and Hambleton is not that loss or profit is incurred but that *more loss or less profit than necessary will be incurred if decision-theoretic optimization is omitted*.

It is known how the optimal (or Bayes) cutoff score behaves as a function of the true score distribution. This is precisely the "regression from the mean effect" mentioned earlier. If the true score distribution moves up to the left, the optimal cutoff score rises; if it moves in the other direction, it drops. For a series of test administrations with the same mastery standard, this effect creates a perfect negative feedback mechanism. In the long

run it may force students to maintain their study efforts at an acceptable level. It is instructive to compare this with norm-referenced cutoff score setting (e.g., grading on the curve). These cutoff scores vary up and down with the performances of students. Discord with this uncritical automatism led to the mastery testing movement, which in Bayesian decision theory finds not only optimal decision rules but also a possible control mechanism for maintaining acceptable study efforts with respect to its fixed mastery standards.

### The Cutoff Score as a Target for Examinees

The requirement that students should be informed of standards beforehand to let them know what they are expected to meet seems sensible for many educational situations. Before embarking on some comments on this point raised by de Gruijter and Hambleton, two things are noted. First, ironically, it is exactly this requirement that creates the problem in the foregoing section. As soon as students are informed of standards in advance, they may use information on previous test administrations to react in the above fashion. Second, it is noted that the same situation can arise because of other reasons, for instance, item analysis or calibration previous to test scoring. The requirement is thus not exclusive to the application of decision theory.

Whether or not this requirement is problematic depends on what choices from decision theory are made. In minimax theory, for example, cutoff scores are not based on maximization of expected utility and are generally not dependent on previously unknown distribution parameters. Hence minimax cutoff scores can be set prior to test administration without any problem (for details, see Huynh, 1980; Veldhuizen, 1982).

If mastery testers prefer (empirical) Bayesian theory, several strategies to reduce the problem are available. Following are three suggestions: First, of course, it is always possible to inform students of the mastery standard and to explain to them how Bayesian optimization works. In particular, the regression from the mean effect can be explained, possibly with some numerical demonstrations, so

that students are able to assess the effects of their preparation. Second, students could be provided with the mastery standard and an *interval* of possible cutoff score values. The eventual cutoff score is then selected optimally from this interval after the administration of the test. The loss associated with this procedure can be negligible if the interval is calculated from ranges of realistic values of distribution parameters. Moreover, the procedure provides students with a meaningful target, especially if they understand how the eventual choice from the interval depends on their study efforts.

Third, the cutoff score can be set in advance using deliberate guesses of parameter values. What parameter values need to be substituted depends on the test model and the utility structure adopted. For the classical test model and linear utility, only relatively innocent parameters such as the mean of the observed score distribution and the test reliability have to be estimated (van der Linden & Mellenbergh, 1977). Estimates can be based on results from comparable test administration and/or impressions of the students' preparation. Substitution of various estimates to evaluate the behavior of the cutoff score (robustness analysis) seems valuable and suggests the development of interactive computer programs to assist the mastery tester in his/her definitive choice. Strategies such as those above circumvent the problem of cutoff scores as targets and lead, in the sense of the expected utility criterion, to better results than cutoff scores set beforehand solely on an intuitive basis.

### Choice of Subpopulations

The choice of subpopulations, put forward as a problem by de Gruijter and Hambleton, is not inherent in decision theory or in its applications. Nothing in decision theory compels the distinguishing of subpopulations or even considering the sampling of students. The use of a less than perfectly reliable test (measurement error) is already a sufficient reason to apply decision theory. In such cases, the only "population" involved in the application of empirical Bayesian theory referred to by de Gruijter and Hambleton is the empirical true score distribution of the students who took the test

and for whom a decision has to be made. There is no doubt whatsoever as to who do or do not belong to this "population." On the other hand, decision theory is quite able to deal with student sampling and subpopulation structures when confronted with the necessity to do so.

The point to be noted, however, is that the necessity to distinguish subpopulations arises from the mastery testing problem and its educational or societal context and not from the application of decision theory. Any other method of standard setting will meet the same problem as it arises.

The distinction between possible subpopulations was made first in testing for selection. There the fact that test items may be biased against minority groups led to societal protests and created the problem of culture-fair selection. As Gross and Su (1975) and Petersen and Novick (1976) have argued convincingly, fair selection is a question of utilities. A selection procedure is "fair" if the utility structure underlying its selection rule reflects the utilities of those involved in the selection process (see also Mellenbergh & van der Linden, 1981). These authors have demonstrated how decision theory can be applied to solve the culture-fair selection problem. This application supposes that the definition of group membership has been solved. If no closed definition of group membership is available, the problem as such is ill defined; and for ill-defined problems, however urgently they are felt, simply no satisfying solutions are possible.

It remains to be seen if the issue of culture fairness will become as important in mastery testing as in testing for selection. Mastery testing usually takes place in a series of small instructional units and in such situations less is at stake than in, e.g., college admission. But if the issue arises, decision theory can be used to solve well-defined problems successfully, whereas it is unclear as yet how alternative methods could address even such problems.

### Choice of Test Model and Inaccurate Parameter Estimates

These two problems are taken together here because, to some extent, they depend on each other.

What model is chosen determines the parameters to be estimated; conversely, the choice of the test model may be determined by the availability of accurate estimation methods. Like de Gruijter and Hambleton, the problems of determining mastery standards and utility parameters will be dealt with here though they do not necessarily involve statistical estimation.

The problem of parameter estimation and the propagation of its errors in solutions for practical problems is a delicate one and deserves adequate attention. This holds for decision theory as well as for any other approach involving parametric modeling. The author's point of view in the present problem is that it should be reduced as far as possible. In order to achieve this the mastery tester has a full range of options from decision theory and psychometric theory at his/her disposal and may have support from a developing new decision-making technology (see the examples below).

Whether distribution parameters have to be estimated depends on several choices. For instance, if minimax theory is chosen and the simple binomial model is assumed to hold, no estimation of distribution, test, or item parameters is required. Other solutions assume only weak models, for example, in the linear utility model where no more than the classical test model with Kelley's regression line is required (van der Linden & Mellenbergh, 1977). In this model only estimates of the mean observed score and the classical test reliability need to be provided. More involved models, such as the empirical beta-binomial model, have more parameters to be estimated and may require considerable sets of test data. Also, as noted by de Gruijter and Hambleton, there may be the risk of model misfit.

The (well-known) lesson from this is that the decision maker should aim at the weakest test model compatible with his/her choice from decision and utility theory. It is advisable to already be aware of the pros and cons of available options when setting up the testing procedure. For example, the use of item sampling and compound binomial models can be circumvented by item banking techniques. Item banking with calibrated items permits systematic item selection, and the selection of items with the same success probability at the mastery stan-

dard is a sufficient condition for applying minimax theory under the simple binomial model. Item banking also prevents the problem of the relationship between  $\pi$  and  $\zeta$  dwelt on by de Gruijter and Hambleton.

As for the determination of  $\pi_0$ , de Gruijter and Hambleton have made some useful proposals. However, the most obvious proposal, namely, the use of standard-setting methods, seems to be overlooked. (These methods must, of course, be genuine standard-setting methods, and not methods ignoring the essential difference between standards and cutoff scores; see above.) Among the most popular methods are Angoff's and Nedelsky's. Elsewhere the author has presented empirical results showing that these methods suffer from serious inaccuracy due to inconsistent specifications of the success probabilities (van der Linden, 1982b). It is, however, possible to improve the accuracy of these methods (of which the former is preferable on many grounds) to any desired level (1) by transforming them into interactive procedures confronting the standard setter with intermediate results and (2) by requesting him/her to reconsider inconsistent specifications until a satisfactory result is obtained. (For the methodology needed for these procedures, see van der Linden, 1982b. Work is in progress to develop computer dialogues to implement these interactive procedures.)

A comparable approach to the choice of utility functions has already been implemented in CADA, an interactive computer program by Novick, Isaacs, and Dekeyrel (1977). Although, in principle, any psychological scaling method could be used to assist the decision maker in specifying his/her utilities, the application of such new technology as this computer-based utility determination is especially suited to educational settings. The same holds for the development of computer-based methods of standard setting above. In many educational settings computers have come to stay, and the increasing number of applications of item banking and computer-aided testing are excellent environments in which the use of this new technology could flourish.

Another area where comparable computer programs are welcome is in the possibility of inter-

active robustness or sensitivity analysis already alluded to above. In this application the program assists the decision maker in evaluating the effect of minor deviations from his/her parameter specifications on the resulting cutoff score. A robust cutoff score in the range of parameter values investigated may be reassuring to the decision maker; a sensitive score, on the other hand, may be a reason to revise the specifications. Such analyses may cover the simultaneous variation of utility function parameters and test model parameters (e.g., within the range of their standard error of estimation) but also only the latter, for example, to evaluate the effect of a priori estimates with standard setting previous to the test administration (see above). The challenge is to develop user-oriented program packages which converse with the user in an understandable way and do not bother him/her with technicalities.

#### Summary and Conclusions

In the foregoing, attention has been paid to the fact that decision theory has to offer much to mastery testing theory and practice. In response to five problems put forward by de Gruijter and Hambleton it has been indicated that some of these problems are associated with specific choices from decision theory and disappear if other, possibly more sensible, choices are made. Other problems appeared to be universal to mastery testing and are met by any method of setting cutoff scores. These problems are thus not inherent in the application of decision theory. Finally, this paper has described a developing new technology that may aid the decision maker in his/her choice of parameter values.

All this is not to suggest that decision theory offers perfect solutions to any *practical* mastery testing problem. Using the full potential of decision theory is not a matter of routine, and in many situations requiring quick and easy solutions the application of decision theory may have to cope with suboptimal conditions. It is, however, clearly superior to other methods that have been proposed for the same problems and that are incorrect or just intuitive in character. When the choice is between the application of decision theory and one of these

alternatives, the former should be chosen. Or, in other words, with a motto that should guide any practical problem solving: As long as the perfect is impossible, the best should be preferred.

#### References

- Berk, R. A. (1976). Determination of optimal cutting scores in criterion-referenced measurement. *Journal of Experimental Education*, *45*, 4-9.
- Cronbach, L. J., & Gleser, G. L. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana IL: University of Illinois Press.
- de Gruijter, D. N. M., & Hambleton, R. K. (1984). On problems encountered using decision theory to set cut-off scores. *Applied Psychological Measurement*, *8*, 1-8.
- Ferguson, T. S. (1967). *Mathematical statistics; A decision theoretic approach*. New York: Academic Press.
- Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement*, *15*, 237-261.
- Gross, A. L., & Su, W. H. (1975). Defining a "fair" or "unbiased" selection model: A question of utilities. *Journal of Applied Psychology*, *60*, 345-351.
- Hambleton, R. K. (1978). On the use of cut-off scores with criterion-referenced tests in instructional settings. *Journal of Educational Measurement*, *15*, 277-290.
- Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, *10*, 159-170.
- Huynh, H. (1980). A nonrandomized minimax solution for passing scores in the binomial error model. *Psychometrika*, *45*, 167-182.
- Mellenbergh, G. J., Koppelaar, H., & van der Linden, W. J. (1977). Dichotomous decisions based on dichotomously scored items: A case study. *Statistica Neerlandica*, *31*, 161-169.
- Mellenbergh, G. J., & van der Linden, W. J. (1981). The linear utility model for optimal selection. *Psychometrika*, *46*, 283-293.
- Novick, M. R., Isaacs, G. L., & Dekeyrel, D. F. (1977). *Computer-assisted data analysis-1977. Manual for the computer-assisted data analysis (CADA) monitor*. Iowa City IA: Iowa Testing Programs.
- Petersen, N. S., & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, *13*, 3-29.
- Popham, J. W. (1978). As always, provocative. *Journal of Educational Measurement*, *15*, 297-300.
- van der Linden, W. J. (1980). Decision models for use with criterion-referenced tests. *Applied Psychological Measurement*, *4*, 469-492.
- van der Linden, W. J. (1981). Using aptitude measurements for the optimal assignment of subjects to treat-



- ments with and without mastery scores. *Psychometrika*, 46, 257–274.
- van der Linden, W. J. (1982a). Criterion-referenced measurements: Its main applications, problems, and findings. In W. J. van der Linden (Ed.), *Aspects of criterion-referenced measurement. Evaluation in Education: An International Review Series*, 5, 97–118.
- van der Linden, W. J. (1982b). A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard setting. *Journal of Educational Measurement*, 19, 295–308.
- van der Linden, W. J. (in press). Decision theory in educational research and testing. In T. Husen & T. H. Postlethwaite, *International encyclopedia of education: Research and studies*. Oxford: Pergamon Press.
- van der Linden, W. J., & Mellenbergh, G. J. (1977). Optimal cutting scores using a linear loss function. *Applied Psychological Measurement*, 1, 593–599.
- Veldhuijzen, N. H. (1982). Setting cutting scores: A minimum information approach. In W. J. van der Linden (Ed.), *Aspects of criterion-referenced measurement. Evaluation in Education: An International Review Series*, 5, 141–148.
- Vijn, P., & Molenaar, I. W. (1981). Robustness regions for dichotomous decisions. *Journal of Educational Statistics*, 6, 205–235.
- Zieky, M. J., & Livingston, S. A. (1977). *Manual for setting standards on the Basic Skills Assessment Tests*. Princeton NJ: Educational Testing Service.

#### Acknowledgment

*The author is indebted to Anita Burchartz for typing the manuscript.*

#### Author's Address

Send requests for reprints or further information to Wim J. van der Linden, Afdeling Toegepaste Onderwijskunde, Technische Hogeschool Twente, Postbus 217, 7500 AE Enschede, The Netherlands.