DOCUMENT RESUME

ED 260 123 TM 850 450

AUTHOR van der Linden, Wim J.
TITLE Advances in the Application of Decision Theory to Test-Based Decision Making.
SPONS AGENCY Netherlands Foundation for Educational Research, The Hague.
PUB DATE Mar 85
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bayesian Statistics; Classification; *Cutting Scores; *Decision Making; Foreign Countries; Item Analysis; Literature Reviews; Mastery Tests; Personnel Selection; Secondary Education; Statistical Studies; Student Placement; *Test Results; Test Theory; *Test Use
IDENTIFIERS *Netherlands

ABSTRACT

This paper reviews recent research in the Netherlands on the application of decision theory to test-based decision making about personnel selection and student placement. The review is based on an earlier model proposed for the classification of decision problems, and emphasizes an empirical Bayesian framework. Classification decisions with threshold utility are discussed to provide an example of the application of Bayesian theory to test-based decision making. Test results from the 1981 administration of the Eindtoets Basisonderwijs are analyzed with respect to the type of secondary education chosen by Dutch students at the end of primary education: lower vocational education, lower general education, or middle general education. A 55 item bibliography is attached. (GDC)

Advances in the Application of Decision Theory to Test-based Decision Making

Wim J. van der Linden

Twente University of Technology

Enschede, The Netherlands

Paper presented in the symposium Recent developments in test theory in the Netherlands at the annual meeting of the American Educational Research Association, Chicago, Illinois, March 31 - April 4, 1985. Presentation of this paper was sponsered by the Dutch Foundation for Educational Research (SVO), The Hague, The Netherlands.

Abstract

This paper reviews recent research in the Netherlands on the application of decision theory to test-based decision making. The review is based on a classification of decision problems proposed in van der Linden (1985a) and emphasizes an empirical Bayesian framework. As a more specific example of the application of Bayesian theory to test-based decision making the problem of classification decisions with threshold utility is discussed.

Advances in the Application of Decision Theory to Test-based Decision Making

Historically, the use of psychological and educational tests has its roots in the necessity of selection and placement decisions in public domains as education, the army, and the government. This is excellently demonstrated in DuBois' (1970) historiography of such cases as Binet's early work on test development for the assignment of pupils to special education, the testing of conscripts for placement in the army during World War I, and the examination of applicants for the civil service in ancient China. It is no coincidence that in each of these domains decision making is characterized both by a high visibility and massive numbers of subjects. In such cases it seems perfectly logical to grab at tests as objective means to base decisions on. If tests had not been invented for this purpose yet, we would invent them today.

It is conspicuous that, although the practice of test use has its roots in decision making, test theory has been developed mainly as a theory of measurement. The origins of test theory are in Spearman's pioneering work on the unreliability of test scores which laid the foundations for the classical test theory as a theory of measurement error. Modern item response theory shows the same concern with measurement (parameter estimation) and was not conceived as a theory of decision making either. History of test theory shows a few exceptions, though, of which the publication of the Taylor-Russell (1939) tables, and their subsequent influence on the testing literature, and Cronbach and Gleser's (1965) well-known monograph deserve special mention. To date, the latter has been the first and only monograph attempting to provide test-based decision making with a sound theoretical basis.

Recently, however, the situation has changed somewhat and some test theorists are now seriously involved in attempts to model and to optimize the use of tests for decision making, most of them using Bayesian decision theory as their frame of reference. The major impetus for this concern has come from the introduction of modern instructional systems as individualized instruction, learning for mastery, and computer-aided instruction. In such systems there typically is much testing for instructional decision making purposes, which confronts their developers with the problem of designing and studying optimal decision procedures.

It is the goal of this paper to give a short review of recent work on the theory of test-based decision making in the Netherlands. The emphasis on Dutch contributions means that no reference is made to the mostly excellent work in this area in the U.S. as, for instance, by Huynh (1976, 1980a, 1980b), Novick and his associates (e.g., Chuang, Chen, & Novick, 1981; Novick & Lindley, 1978; Novick & Petersen, 1976), and Wilcox (1976, 1977, 1978, 1979). In the review a typology of test-based decision making given in van der Linden (1985a) is used. The paper concludes with the discussion of classification decisions as a more specific example of test-based decision making.

A Classification of Test-based Decisions

Each different type of decision making can be identified as a specific configuration of the following elements:

(1) A test providing the information the decisions are based on;

(2) One or more treatments with respect to which the decisions are made;

(3) One or more criteria by which the success of the treatments are measured.

As will be illustrated below, using these elements the following types of decisions can be distinguished:

(1) Selection;

(2) Mastery;

(3) Placement and

(4) Classification decisions.

To each of these types the following restrictions or refinements can apply:

(1) Quota restrictions. For some treatments the numbers of vacancies are constrained.

(2) Multivariate test data. The decisions are based on data from a whole test battery instead of a single test.

(3) Multivariate criteria. The success of the treatments is measured by multiple criteria.

(4) Subpopulations. The problem of culture-fair decision making arises because of the presence of subpopulations reacting differentially to the test items.

A Review of Dutch Decision Theory Research

Selection Decisions

In selection problems the decision at stake is the acceptance or rejection of individuals for a treatment. Selection decisions are characterized by the fact that the test is administered before the treatment takes place but that the criterion is measured afterwards.

Well-known examples of selection problems are the selection of personnel in industry and the admission of students to educational programs. The formal structure of a selection problem is shown in Figure 1.

Insert Figure 1 about here

Selection research in the Netherlands has a tradition as long as in any other western country, with early publications dating back to the 20's (e.g., de Quay, 1925, in van Naerssen, 1963). Most research can be considered as applied work on problems in personnel selection. Examples of popular problems are research on criterion choice and analysis, test and item selection, validity studies of test batteries, reliability of selection interviews, techniques of job sampling, and the like. Also a considerable amount of the selection literature has been devoted to ethical issues. Most applications of selection research have been in personnel psychology and not in education because Dutch education has traditionally been based on a centralized certification system and not on entrance selection. A recent exception, however, has been the selection of students for medical programs in higher education. Reviews of selection research are given in handbooks by Hofstee (1983) and Roe (1983).

As for the test theoretical framework adopted in selection research, the selection problem has generally been approached as a prediction problem in which regression lines or expectancy tables should be employed to predict whether the criterion scores of individuals exceed a certain threshold value so that their selection guarantees a success. More recent, original work along these lines has been published in which correction for the restriction of range in the validity icient of selection procedures are addressed (Brouwer & Vijn, 1978; Brouwer & Vijn, 1979, Roe, 1979). Selection decisions with quota restrictions have long been evaluated with the aid of the Taylor-Russell tables, which give success ratio's for a number of parameters characterizing the selection procedure.

A major breakthrough is selection theory in the Netherlands was offered by van Naerssen (1963, 1965a, 1970) who introduced the application of empirical Bayesian decision theory in selection research. An extensive introduction to van Naerssen's early work, which arose from a case study on the selection of drivers for the army, can be found is his addendum to Cronbach and Gleser's monograph (van Naerssen, 1965b). Among the topics dealt with in van Naerssen (1963) are the computation of optimal testing time with a fixed selection ratio, the determination of optimal selection ratio's, and two-stage selection procedures. Van Naerssen (1963) also offers some decision theory for a selection problem with two subpopulations. Apart from van Naerssen's contributions not much work on the selection problem from a decision-theoretic point of view can be found in the Netherlands. A recent exception, however, is a paper by Mellenbergh and van der Linden (1982) who give some decision theory for quota-free and fixed-quota selection from several subpopulations with a linear utility structure and illustrate their results with an application to a culture-fair testing problem.

Mastery Decisions

Unlike selection decisions, mastery decisions are made after the treatment has been administered. The decision to be made is whether the individuals who have followed the treatment meet its goals or not. A further characteristic is that in the mastery decision problem the criterion is internal to the test and not external. It is unreliability of the test as a representative of the criterion that opens the possibility of making wrong decisions and creates the mastery decision problem. Examples of mastery decisions are pass-fail and certification decisions in education, but also, e.g., decisions with respect to successfulness of therapies in clinical settings. Figure 2 displays

Insert Figure 2 about here

the formal structure of a mastery decision problem.

As opposed to the selection problem, research on the mastery decision problem in the Netherlands has been mainly test theoretic with less emphasis on applied issues. Again it was van Naerssen who took the lead and introduced the topic and its related problem of the equating of mastery standards in a series of papers (1966, 1971, 1974a). But now others have followed. The following issues have been studied more or less extensively:

1. <u>(Empirical) Bayes decision rules</u>. The problem of Bayes rules for mastery decisions with a binomial error, a beta prior and a threshold loss function has been addressed by Mellenbergh, Koppelaar, and van der Linden (1977). Mellenbergh also suggested the idea of a linear instead of a threshold loss function which has the advantage of being continuous in the true score for both the mastery and the nonmastery decision. This idea was elaborated for the classical test model with an unspecified prior in van der Linden and Mellenbergh (1977), while properties of Bayes rules for this problem were studied further in

van der Linden (1980, 1984a, 1984b). Estimation procedures and properties of Bayes rules for latent state models with threshold loss are explored in van der Linden (1978, 1981a, 1982a, 1983).

2. Decision without priors. The above decision rules are optimal in the Bayes sense for an empirical population of subjects. Van den Brink (1982) takes the position that this is not consonant with the idea of absolute measurement and gives various results for mastery testing under a binomial error model adopting a Neyman-Pearson framework of hypothesis testing. Along the same lines van den Brink and Koele (1980) and van der Linden (1982b) have studied the effect of guessing on multiple-choice items on decision rules. Minimax solutions for the binomial error model with threshold loss are discussed in Veldhuizen (1982).

3. Utility structure. Properties of Bayes rules may depend heavily on the utility structure adopted. A usual approach to the utility problem is the subjective one in which the decision theorist adopts a family of utility functions that is plausible because it meets some obvious formal conditions and the decision maker is requested to identify a member of it on intuitive grounds. The assessment of utility functions can also be based on empirical methods as lottery or scaling methods. Scaling methods for the mastery decision problem have recently been studied by Vrijhof, Mellenbergh, and van den Brink (1983).

4. Item selection. In most research reviewed earlier in this section the problem was to derive, under certain assumptions, optimal decision rules for a given test. If a domain of items from which the test has to be selected is available, another optimization problem arises, namely the optimal selection of items for decision making. Two

different lines of research can be reported. De Gruijter and Hambleton (1983) and Hambleton and de Gruijter (1983) have based their selection on the value of the item information function at the mastery standard. Simulation studies of this selection procedure against random sampling from the same domain showed a considerable improvement in terms of the percentage of misclassification of examinees for the resulting test. The same procedure, but with selection on the first derivative of the item-characteristic curve at the standard on the ability scale, was studied earlier in van Naerssen (1977a, 1977b). Mellenbergh and van der Linden (1982) have proposed a different procedure in which items are selected on the basis of their contribution to the Bayes risk. They were able to show under what conditions this criterion boils down to selection using classical item indices.

5. Evaluation of decision procedures. Measurement procedures are usually evaluated by their reliability or estimation accuracy but for decision procedures this seems less adequate. Van der Linden and Mellenbergh (1978) suggested to use the Bayes risk for this purpose and proposed to standardize this on the interval [0,1] using the risk of procedures with test scores having no and full information about the criterion as reference points. They also showed under what conditions the standardized risk is equal to classical test indices as, e.g., the reliability coefficient. In Mellenbergh and van der Linden (1979) the same procedure is outlined for test-based decision making with an external criterion (e.g. selection decisions). A different perspective on the evaluation of decision procedures is robustness analysis. A concept introduced in Vijn (1980) and explored further in Vijn and Molenaar (1981) is that of the robustness region

which is defined as the subset of the parameter space of the decision problem giving rise to the same decision rule. An application of robustness region analysis to a mastery decision problem can be found in the latter reference.

6. Standard setting. Mastery decision making supposes the presence of a threshold value or standard on the criterion separating the "masters" from the "nonmasters". A useful standard setting method is the so-called kernel-item method in which judges indicate which items present the standard best, and next the standard is computed from the statistics of these items (de Groot & van Naerssen, 1975, sect. 19.4; van Naerssen, 1974b). A proposal accounting for possible uncertainty or inaccuracy in standard setting procedures by replacing standards by distributions of possible values is elaborated in de Gruijter (1980). That there can be much inaccuracy in standard setting procedures is demonstrated in van der Linden (1982c) who used calculations under an item response model to check for specification errors in the Angoff and Nedelsky methods and found that errors larger than .20-.25 were no exception.

Placement Decisions

In placement problems several alternative treatments are available and it is the decision maker's task to assign individuals on the basis of their test scores to the most promising treatment. All individuals are administered the same test and the success of each treatment is measured by the same criterion. Unlike the selection problem, each individual is assigned to a treatment. Figure 3 shows the case of a placement decision with two treatments. Examples of placement decisions are in individualized instruction where students are assigned to different

Insert Figure 3 about here

routes through an instructional unit all leading to the same objective.

The traditional approach to the problem is that of linear regression analysis with a separate regression line for each treatment and the assignment of individuals to the treatment with the largest predicted criterion score. A Bayesian version of this approach is offered in Vijn (1980) which offers the option of incorporating previous information in placement decisions via the specification of prior distributions for the regression parameters.

Vijn's approach, although fully Bayesian, still views the placement problem as a prediction problem. A treatment of placement decisions from a decision-theoretic viewpoint is given in van der Linden (1981). This paper formalizes the placement decision as an empirical Bayes problem with different utility functions and probability models for each treatment and gives decision rules for the cases of utility functions from the threshold, linear and normal-ogive families. The paper also indicates how optimal rules for placement decisions with subpopulations can be found.

Classification Decisions

As is clear from Figure 4, the difference between classification and

Insert Figure 4 about here

placement decisions is that in the former each treatment has its own criterion. Further these two types of decisions have identical properties. Examples of classification decisions occur in vocational guidance situations when most promising schools or careers must be identified.

The most popular approach to classification decisions has again been the use of linear-regression techniques. Each criterion is then mapped on a common utility scale and the decision rule is to assign individuals to the treatment with the largest predicted utility. The classification problem has hardly been treated as a Bayesian decision problem. As a more extensive example of the application of decision theory to test-based decision making, the following section discusses the problem of classification decisions with threshold utility and illustrates the use of a Bayes rule for this case with an empirical application. A full treatment of the theory and the application is given separately in van der Linden (1985b) where further details can be obtained.

Classification Decisions with Threshold Utility

The classification problem can be formalized as follows. There is a series of individuals who can be considered to be drawn randomly from some population P and must be classified into t+1 treatments indexed by $j = 0, 1, \ldots, t$. Each treatment leads to a different distribution for P on its associated criterion which is denoted by a random variable Y_j with range R_j, which will here be considered to be continuous (although in some applications Y_j may be discrete). The test scores observed prior to the treatment are denoted by a random variable X with discrete values

$x = 0, \ldots, n$ and probability function $\lambda(x)$. It is assumed that P yields a joint distribution of test and criterion scores with probability (density) function $\eta_j(x,y_j)$.

Suppose that each treatment is followed by a mastery decision indicating whether the treatment has been successful or not. Formally, the classification problem can then be represented as a problem with threshold utility. An appropriate utility function is

$$(1) \quad u_j(y_j) = \begin{cases} w_j & \text{for } y_j \geq d_j \\ v_j & \text{for } y_j < d_j \end{cases}$$

with

$$w_j > v_j \text{ for all values of } j,$$

where d_j is the cutting score on criterion j defining the mastery decision rule for treatment j while w_j and v_j are the utilities of reaching the mastery and nonmastery status, respectively.

It is assumed that the Bayes rule for this problem has a monotone shape; i.e., takes the form of a series of cutting scores on the test

$$(2) \quad 0 = c_0 \leq c_1 \leq \ldots \leq c_j \leq \ldots \leq c_{t+1} = n \quad (t \leq n)$$

such that treatment j is assigned in the event of $c_j \leq X < c_{j+1}$ (for $j = t$ the second inequality is not strict). A necessary and sufficient condition for (2) is that

$$(3) \quad w_j - v_j \geq w_{j-1} - v_{j-1} \qquad (j = 1, \ldots, n)$$

and that

$$(4) \quad \{\Omega_j(y_j|x)\},$$

the conditional distributions of Y_j given $X = x$, are stochastically increasing. It is assumed that the treatments are in proper order reflected by their index j.

From van der Linden (1985) it follows that the expected utility of the procedure is maximal if, for each pair of treatments $(j-1, j)$, c_j is chosen as the smallest value of x for which

$$(5) \quad (w_{j-1}-v_{j-1})\,\Omega_{j-1}(d_{j-1}|x) - (w_j-v_j)\,\Omega_j(d_j|x)$$

is positive.

Since the solution of (7) only depends on the difference between w_j and v_j and not on their individual values, an interesting case arises if it can be assumed that

(6) $w_j - v_j$ = constant

for one of more pairs of adjacent treatments. Then (5) reduces to

(7) $\Omega_{j-1}(d_{j-1}|x) - \Omega_j(d_j|x)$

and it is no longer needed to specify the values of the utility parameters. A further special case is if $\Omega_j(y|x)$ can be assumed to be a location-scale family in which case analytic solutions are possible. For these and other cases, see van der Linden (1985b).

An Empirical Example

The example in this section is derived from a well-known problem in the Netherlands, namely the choice of an appropriate continuation-school at the end of primary education. Several types of secondary education are available running from lower level vocational to university track programs. A popular achievement test assisting parents and principals in making this choice is the Eindtoets basisonderwijs prepared annually by the National Institute of Educational Measurement (Cito). In the following analyses, data from the 1981 administration of the test are used, and the following types of secondary education are selected as treatments: Lower Vocational Education (LVE), Lower General Education (LGE), and Middle General Education (MGE). Success on the criterion was for each treatment defined as passing the first year of its program.

It was assumed that the probabilities of success on Y as a function of x could be modeled as a logistic distribution function. Table 1 gives the empirical proportions of successes for each treatment. As only

Insert Table 1 about here

grouped data were available logit analysis of the proportions was applied for the middles of the intervals reported in the table. The bottom line of the table shows that the data yielded a nice fit to the logit model.

Firstly, it is assumed that (6) holds for the treatments so that (7) amounts to a comparison between the logistic regression lines. The results are given in Figure 5 and show that the dominant treatment

Insert Figure 5 about here

is LGE for almost all possible test scores; only for test scores below x = 4 does the choice of another treatment (LVE) appear to be better. Secondly, the sensitivity of the solution in Figure 5 to deviation from (6) is analyzed in Table 2. As could be expected from the closeness of

Insert Table 2 about here

the logistic regression lines for LGE and MGE, the cutting score between these two treatments is most sensitive to deviations in the utility ratio from unity. The cutting score between LGE and LVE, however, appears to be quite robust to changes in the utility ratio.

Conclusion

This review of research on decision theory for test-based decision making in the Netherlands shows an early interest in the selection problem and a subsequent emphasis on the mastery decision problem. Recently offshoots to other decision problems have become visible. It is expected that the interest in decision making will continue and that more complicated types of decision making (e.g., with quota constraints, multivariate test data and criteria, and/or subpopulations) will be explored soon.

References

Brouwer, U. & Vijn, P. (1978). De empirische power van 2 procedures in "restriction of range" (case 1). Tijdschrift voor Onderwijsresearch, 3, 106-112.

Brouwer, U. & Vijn, P. (1979). Bayesiaanse schatters voor de correlatiecoefficient in "restriction of range (geval 1). Tijdschrift voor Onderwijsresearch, 4, 281-290.

Chuang, D.T., Chen, J.J. & Novick, M.R. (1981). Theory and practice for the use of cut-scores for personnel decisions. Journal of Educational Statistics, 6, 107-128.

Cronbach, L.J. & Gleser, G.L. (1965). Psychological tests and personnel decisions (2nd ed.). Urbana, Ill.: University of Illinois Press.

de Groot, A.D. & van Naerssen, R.F. (Eds.) (1973, deel I; 1975, deel II). Studietoetsen, construeren, afnemen, analyseren. Den Haag: Mouton.

de Gruijter, D.N.M. (1980). Accounting for uncertainty in performance standards. Paper presented at the Fourth International Symposium on Educational Testing, Antwerp, Belgium, June 24-27.

de Gruijter, D.N.M. & Hambleton, R.K. (1983). Using item response models in criterion-referenced test item selection. In R.K. Hambleton (Ed.), Applications of item response theory. Vancouver, British Columbia: Educational Research Institute of British Columbia.

Dubois, P.H. (1970). A history of psychological testing. Boston: Allyn & Bacon, Inc.

Hambleton, R.K. & de Gruijter, D.N.M. (1983). Applications of item response models to criterion-referenced test item selection. Journal of Educational Measurement, 20, 355-367.

Hofstee, W.K.B. (1983). Selectie. Utrecht, The Netherlands: Spectrum.

Huynh, H. (1976). Statistical considerations of mastery scores. Psychometrika, 41, 65-79.

Huynh, H. (1980). A nonrandomized minimax solution for passing scores in the binomial error model. Psychometrika, 45, 167-182. (a)

Huynh, H. (1980). Statistical inference for false positive and false negative error rates in mastery testing. Psychometrika, 45, 107-120. (b)

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. (1979). The internal and external optimality of decisions based on tests. Applied Psychological Measurement, 3, 257-273.

Mellenbergh, G.J. & van der Linden, W.J. (1981). The linear utility model for optimal selection. Psychometrika, 46, 283-293.

Mellenbergh, G.J. & van der Linden, W.J. (1982). Selecting items for criterion-referenced tests. Evaluation in Education, 5, 117-190.

Novick, M.R. & Lindley, D.V. (1978). The use of more realistic utility functions in educational applications. Journal of Educational Measurement, 15, 181-191.

Novick, M.R. & Petersen, N.S. (1976). Towards equalizing educational and employment opportunity. Journal of Educational Measurement, 13, 77-88.

Roe, R.A. (1979). The correction for restriction of range and the difference between intended and actual selection. Educational and Psychological Measurement, 39, 551-560.

Roe, R.A. (1983). Grondslagen der personeelsselectie. Assen, The Netherlands: Van Gorcum.

Taylor, H.C. & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 23, 565-578.

van den Brink, W.P. (1982). Binomiale modellen in de testleer. Proefschrift, Universiteit van Amsterdam.

van den Brink, W.P. & Koele, P. (1980). Item sampling, guessing, and decision-making in achievement testing. British Journal of Mathematical and Statistical Psychology, 33, 104-108.

van der Linden, W.J. (1978). Forgetting, guessing, and mastery: The Macready and Dayton models revisited and compared with a latent trait approach. Journal of Educational Statistics, [illegible], 305-318.

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). Estimating the parameters of Emrick's mastery testing model. Applied Psychological Measurement, 5, 517-530. (a)

van der Linden, W.J. (1981). Using aptitude measurements for the optimal assignment of subjects to treatments with and without mastery score. Psychometrika, 46, 257-274. (b)

van der Linden, W.J. (1982). Zur Schätzung der 'Proportion of Masters' in kriteriumsorientierten Tests. Zeitschrift für Empirische Pädagogik, 6, 195-201. (a)

van der Linden, W.J. (1982). Passing score and length of a mastery test. Evaluation in Education, 5, 149-165. (b)

van der Linden, W.J. (1982). A latent trait method for determining intrajudge inconsistency in the Angoff and Nedelsky techniques of standard setting. Journal of Educational Measurement, 19, 295-308. (c) van der Linden, W.J. (1983). The use of moment estimators for mixtures of two binomials with one known success parameter. Educational and Psychological Measurement, 43, 321-330.

van der Linden, W.J. (1984). Some thoughts on the use of decision theory to set cut-off scores: Comment on de Gruijter and Hambleton. Applied Psychological Measurement, 8, 9-17. (a).

van der Linden, W.J. (1984). Over absolute en nog relatievere zak-slaag beslissingen. Tijdschrift voor Onderwijsresearch, 9, 243-252. (b)

van der Linden, W.J. (1985). Decision theory in educational research and testing. In T. Husen & T.N. Postlethwaite, International encyclopedia of education: Research and studies. Oxford: Pergamon Press. (a)

van der Linden, W.J. (1985). The use of test scores for classification decisions with threshold utility. Submitted for publication. (b)

van der Linden, W.J. & Mellenbergh, G.J. (1977). Optimal cutting scores using a linear loss function. Applied Psychological Measurement, 1, 593-599.

van der Linden, W.J. & Mellenbergh, G.J. (1978). Coefficients for tests from a decision theoretic point of view. Applied Psychological Measurement, 2, 119-134.

van Naerssen R.F. (1963). Selectie van chauffeurs. Dissertatie, Wolters, Groningen.

van Naerssen, R.F. (1965). Enkele eenvoudige beslikundige toepassingen bij test en selectie. Nederlandse Tijdschrift voor de Psychologie, 20, 364-380. (a)

van Naerssen, R.F. (1965). Application of the decision theoretical approach to the selection of drivers. In L.J. Cronbach & G.L. Gleser, Psychological tests and personnel decisions (2nd ed.). Urbana, IL: University of Illinois Press. (b) van Naerssen, R.F. (1966). Het handhaven van eenmaal aangenomen normen bij opeenvolgende objectieve toetsen. Paedagogische Studiën, 43, 312-320.

van Naerssen, R.F. (1971). Een model voor tentamens. Nederlands Tijdschrift voor de Psychologie, 26, 121-132.

van Naerssen, R.F. (1974). A mathematical model for the optimal use of criterion referenced tests. Nederlands Tijdschrift voor de Psychologie, 29, 341-446. (a)

van Naerssen, R.F. (1974). Psychometrische aspecten van de kernitemmethode. Nederlands Tijdschrift voor de Psychologie, 29, 42[illegible]-430. (b)

van Naerssen, R.F. (1977). Lokale betrouwbaarheid: begrip en operationalisatie. Tijdschrift voor Onderwijsresearch, 2, 111-119. (a)

van Naerssen, R.F. (1977). Grafieken voor de schatting van de helling van itemkarakteristieken. Tijdschrift voor Onderwijsresearch, 2, 193-201. (b)

Veldhuijzen, N.H. (1982). Setting cutting scores: A minimum information approach. Evaluation in Education, 5, 141-148.

Vijn, P. (1980). Prior information in linear models. Thesis, University of Groningen.

Vijn, P. & Molenaar, I.W. (1981). Robustness regions for dichotomous decisions. Journal of Educational Statistics, 6, 205-235.

Vrijhof, B.J., Mellenbergh, G.J. & van den Brink, W.P. (1983). Assessing and studying utility functions in psychometric decision theory. Applied Psychological Measurement, 7, 341-357.

Wilcox, R.R. (1976). A note on the length and passing score of a mastery test. Journal of Educational Statistics, 1, 359-364.

Wilcox, R.R. (1977). Estimating the likelihood of false-positive and false-negative decisions in mastery testing: An empirical Bayes approach. Journal of Educational Statistics, 2, 289-307.

Wilcox, R.R. (1978). A note on decision theoretic coefficients for tests. Applied Psychological Measurement, 2, 609-613.

Wilcox, R.R. (1979). Comparing examinees to a control. Psychometrika, 44, 55-68. (c)

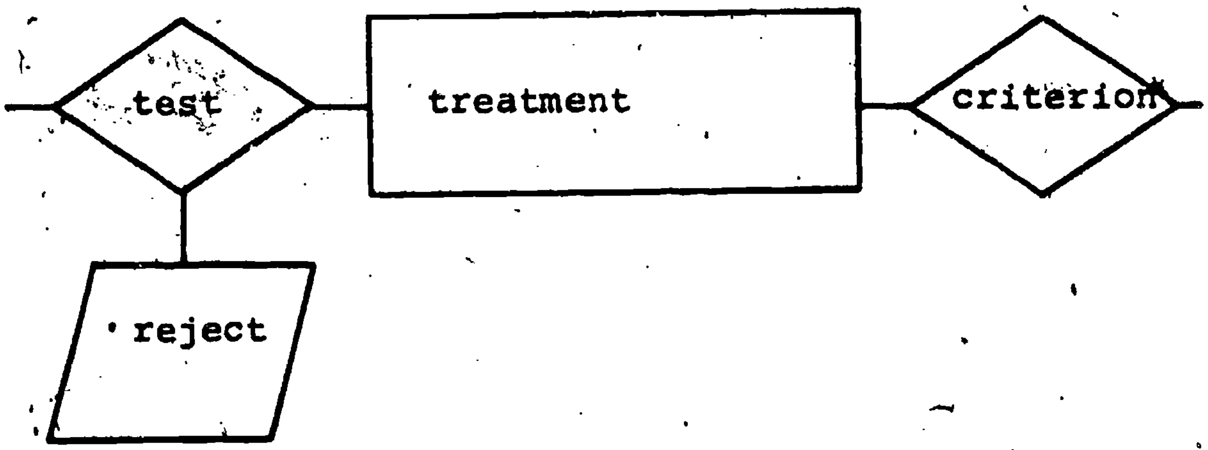

Figure 1 Flowchart of a Selection Decision

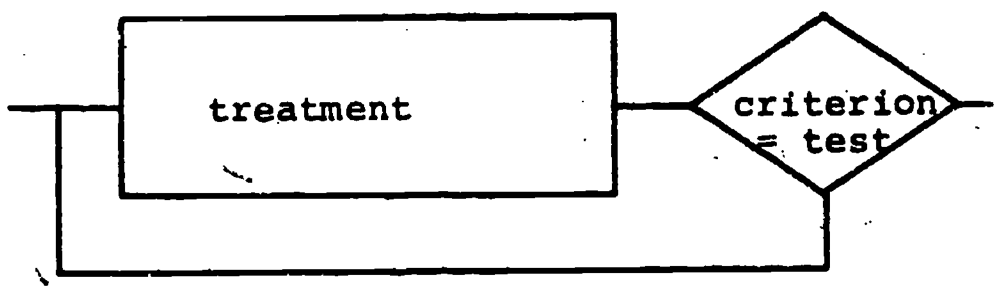

Figure 2 Flowchart of a Mastery Decision

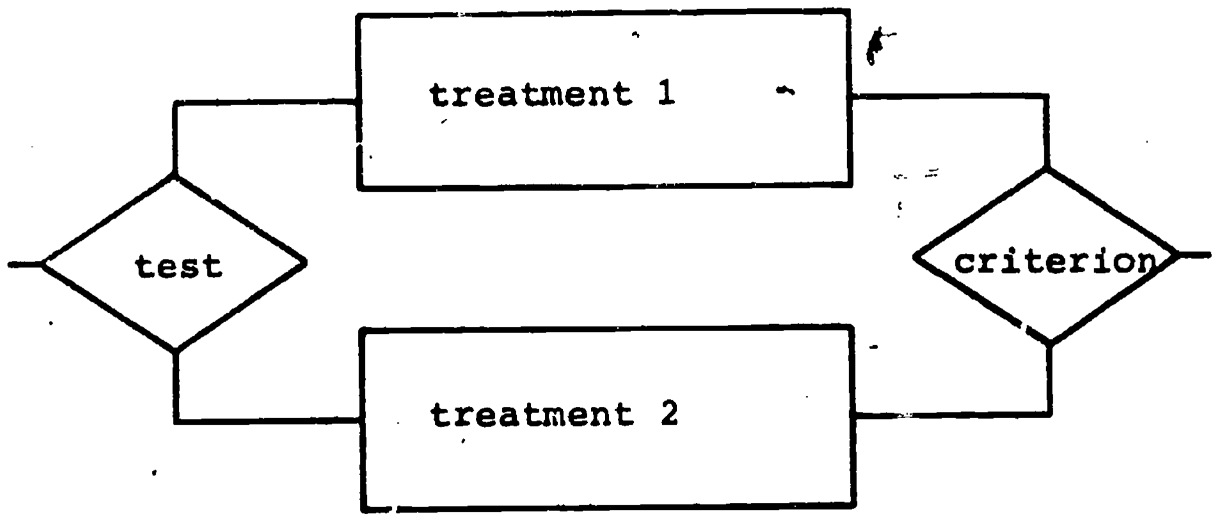

Figure 3 Flowchart of a Placement Decision
(Case of Two Treatments)

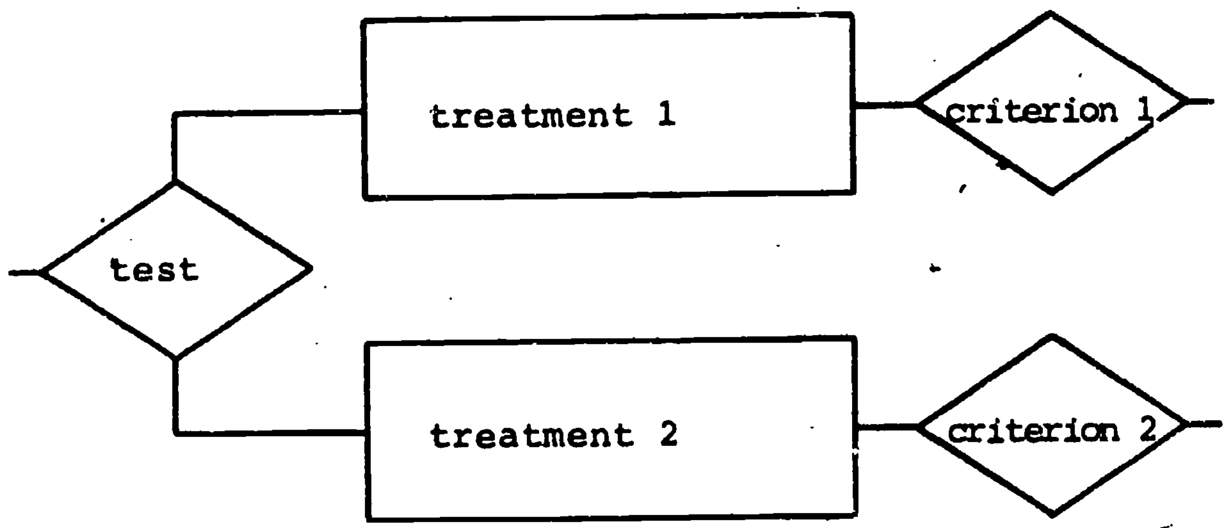

Figure 4 Flowchart of a Classification Decision
(Case of Two Treatments)

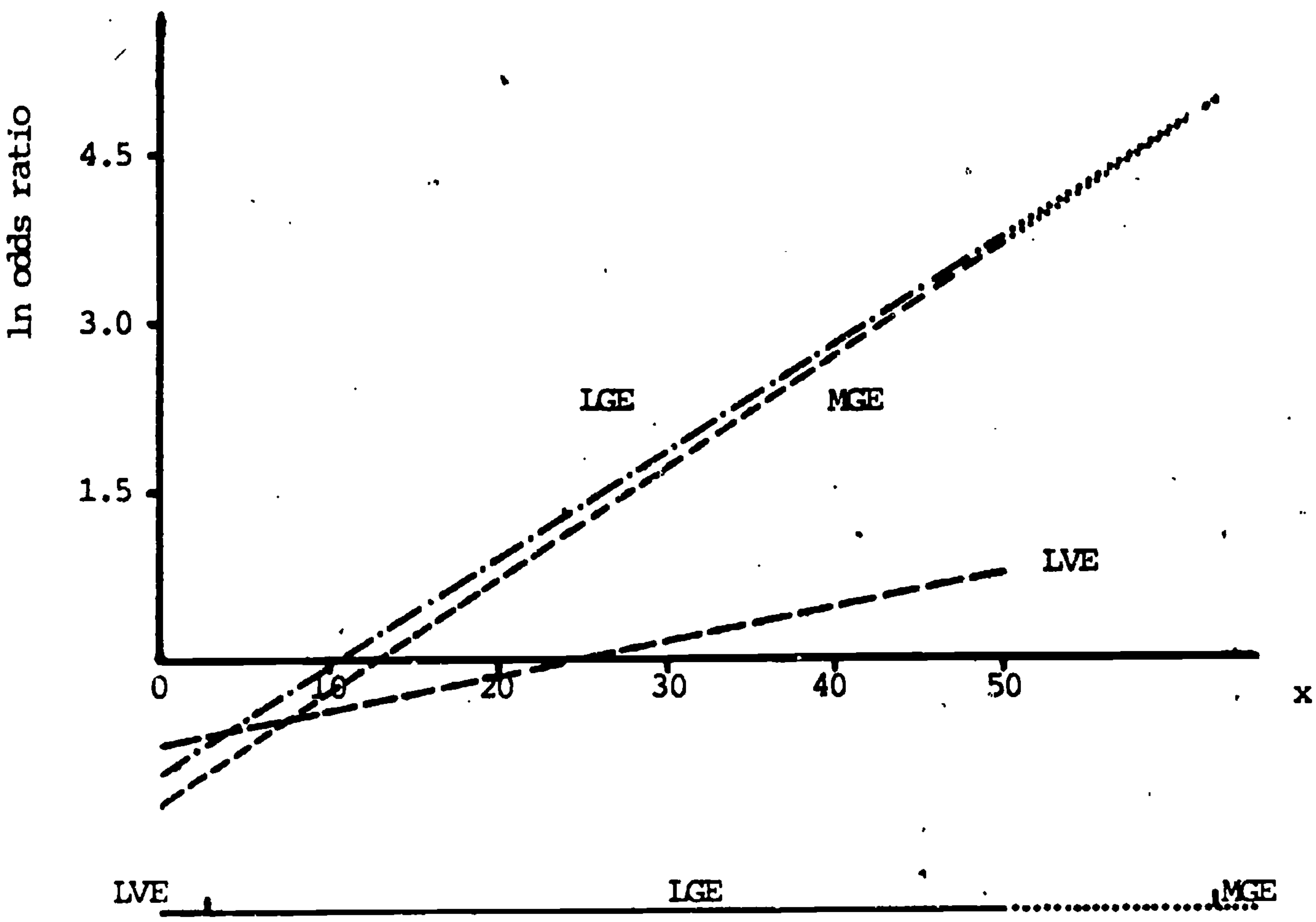

Figure 5. Logistic regression lines for the three treatments.

Table 1

Empirical Proportion of Successes as a Function of Test Scores for the Three Treatments

Test Score	Proportion of Successes After 1 Yr. LVE	LGE	MGE
0 - 5			
6 - 10	.897	.575	
11 - 15			.571
16 - 20	.929	.619	
21 - 25	.947	.760	
26 - 30	.948	.840	.788
31 - 35	.952	.890	.860
36 - 40	.959	.930	.920
41 - 45		.960	.960
46 - 50	.979	.960	.988
No. of Cases	1333	15926	2298
Slope	.031	.095	.099
Intercept	-.8	-1.0	-1.25
Model Fit	.641	.071	.105

Table 2

Optimal Cutting Scores Between the Treatments
as a Function of the Utility Ratio

Utility Ratio	Optimal Cutting Score	
	LGE/LVE	MGE/LGE
1.00	4	-
1.02	3	30
1.04	3	24
1.06	2	19
1.08	2	16
1.10	1	13
1.12	1	10
1.14	1	8
1.16	1	6
1.18	1	4
1.20	1	1

Note Utility ratio is defined as $(w_j-v_j)(w_{j-1}-v_{j-1})^{-1}$. Dash indicates cutting score outside range of test scores.

Author Notes

The author is indebted to Anita Burchartz-Huls for typing the paper.
Request of copies should be sent to Wim J. van der Linden, Afdeling Toegepaste Onderwijskunde, Technische Hogeschool Twente, Postbus 217, 7500 AE Enschede, The Netherlands.