DOCUMENT RESUME

ED 389 757 TM 024 390

AUTHOR van der Linden, Wim J.; Vos, Hans J.
TITLE A Compensatory Approach to Optimal Selection with Mastery Scores. Research Report 94-2.
INSTITUTION Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
PUB DATE Nov 94
NOTE 38p.
AVAILABLE FROM Bibliotheek, Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Decision Making; Foreign Countries; *Scores; *Scoring; *Selection; Test Items
IDENTIFIERS Compensatory Models; Decision Rules; *Mastery Model; Simultaneous Processing; Utility Functions

ABSTRACT

This paper presents some Bayesian theories of simultaneous optimization of decision rules for test-based decisions. Simultaneous decision making arises when an institution has to make a series of selection, placement, or mastery decisions with respect to subjects from a population. An obvious example is the use of individualized instruction in education. Compared with separate optimization, a simultaneous approach has two advantages. First, test scores used in previous decisions can be used as "prior" data in later decisions, and the efficiency of the decisions can be increased. Second, more realistic utility structures can be obtained defining utility functions for earlier decisions on later criteria. An important distinction is made between weak and strong decision rules. As opposed to strong rules, weak rules are allowed to be a function of prior test scores. Conditions for monotonicity of optimal weak and strong rules are presented. Also, it is shown that under mild conditions on the test score distributions and utility functions, weak rules are always compensatory by nature. To illustrate this approach, a common decision problem in education and psychology, consisting of a selection decision for treatment followed by a mastery decision, is analyzed. (Contains 1 figure, 2 tables, and 23 references.) (Author)

7

ED 389 757

A Compensatory Approach to Optimal Selection with Mastery Scores

Research Report 94-2

Wim J. van der Linden

Hans J. Vos

BEST COPY AVAILABLE

Faculty of
EDUCATIONAL SCIENCE AND TECHNOLOGY

Department of
Educational Measurement and Data Analysis

University of Twente

A Compensatory Approach to Optimal Selection
with Mastery Scores

Wim J. van der Linden
Hans J. Vos

The names of the authors are alphabetical; they are equally responsible for the contents of this paper.

A compensatory approach to optimal selection with mastery scores, Wim J. van der Linden and Hans J. Vos - Enschede: University of Twente, Faculty of Educational Science and Technology, November 1994. - 34 pages.

Abstract

This paper presents some Bayesian theory for simultaneous optimization of decision rules for test-based decisions. Simultaneous decision making arises when an institution has to make a series of selection, placements, or mastery decisions with respect to subjects from a population. An obvious example is the use of individualized instruction in education. Compared with separate optimization, a simultaneous approach has two advantages. First, test scores used in previous decisions can be used as "prior data" in later decisions, and the efficiency of the decisions can be increased. Second, more realistic utility structures can be obtained defining utility functions for earlier decisions on later criteria. An important distinction is made between weak and strong decision rules. As opposed to strong rules, weak rules are allowed to be a function of prior test scores. Conditions for monotonicity of optimal weak and strong rules will be presented. Also, it will be shown that under mild conditions on the test score distributions and utility functions, weak rules are always compensatory by nature. To illustrate the approach, a common decision problem in education and psychology, consisting of a selection decision for a treatment followed by a mastery decision, is analyzed.

Introduction

Over the past two decades, Bayesian decision theory has proven to be very useful in solving problems of test-based decision making. Historically, the first decision making problem to draw the interest of psychometricians was the selection problem in education and personnel management. Important milestones in the history of the treatment of selection decisions were the publication of the Taylor-Russell (1939) tables and Gronbach and Gleser's (1956) Psychological tests and personnel decisions. However, in spite of some of the theoretical notions in the latter, it was not after an extensive discussion on "culture-fair" selection (Gross & Su, 1975) that selection decisions were fully treated as an instance of Bayesian decision theory (Novick & Petersen, 1976).

With the advance of such modern instructional systems as individualized study systems, mastery learning, and computer-aided instruction (CAI), interest was generated in the possibility to put the problem of mastery testing on sound decision-theoretic footing. In mastery testing, the intent is to classify examinees as "masters" or "nonmasters" on the basis of their test scores, using some standard of mastery set on the true-score scale underlying the test scores. Hambleton and Novick (1973) were the first to point at the possibility of applying Bayesian decision theory to mastery testing. Optimal mastery rules for various utility or loss functions are derived in Davis, Hickman and Novick (1973), Huynh (1976, 1977, 1980) and van der Linden and Mellenbergh (1977).

Interest in decision making problems in modern instructional systems has also led to the consideration of two other types of decision making: placement and

classification decisions. In either type of decision making, test scores are used to assign examinees to one of the instructional treatments available. However, with placement decisions the success of each of the treatments is measured by the *same* criterion whereas in classification decisions each treatment involves a *different* criterion. The paradigm underlying placement decisions is the Aptitude-Treatment Interaction (ATI) hypothesis, which assumes that students may react differentially to instructional treatments, and, therefore, that different treatments may be best for different students. Classifications decisions are made if an instructional program has different tracks each characterized by different instructional objectives. Such tracking can be found in systems of comprehensive secondary education or vocational education. Bayesian decision theory for placement and classification decisions is given in Saywer (1993) and van der Linden (1981, 1987).

Typically, instructional systems as CAI do not involve one single decision but can be conceived of as networks of nodes at which one of the types of decisions above has to be made (van der Linden, 1990; Vos, 1990, 1991, 1993). An example is an instructional network starting with a selection decision, followed by several alternative instructional modules through which students are guided making placement and mastery decisions, and which ends with a summative mastery test. Decisions in CAI networks are usually based on small tests (which often consist of only a few multiple-choice items).

The question is raised how such networks of decisions should be optimized. An obvious approach is to address each decision separately, optimizing its decision rule on the basis of test data exclusively gathered for this individual decision. This approach is common in current design of instructional systems. The purpose of the present paper is to show that multiple decisions in networks can also be optimized *simultaneously*. The advantages of a simultaneous approach are twofold. First, data

gathered earlier in the network can be used to optimize later decisions. The use of such prior information can be expected to enhance the quality of the decisions--in particular if only small tests or sets of multiple-choice items are administered at the individual decision points. Second, a more realistic definition of utility or loss functions is possible, since these functions can now be defined on the ultimate success criterion in the complete network instead of on intermediate criteria measuring the success on individual treatments. In this paper, a simple decision network of a selection decision followed by one treatment and a mastery decision will be used to make our point. First the selection-mastery problem will be formalized. Then important distinctions will be made between weak and strong as well as monotone and nonmonotone decision rules. Next, a theorem will be given showing under what conditions optimal rules will be monotone. Finally, results from an empirical example will be presented to illustrate the differences between a simultaneous and a separate approach.

The Selection-Mastery Problem

A flowchart of the selection-mastery problem is given in Figure 1. An example of the problem is an instructional module with a pretest and a posttest.

Figure 1 about here

The pretest is administered to select students for the module. It is assumed that the possible actions are to admit or to reject the student for the module. The posttest is

used to decide whether or not the students have mastered the objectives of the module. Typically, the posttest is an unreliable representation of the objectives, and the criterion is supposed to be a threshold on the true score underlying the test. The possible actions are to classify a student as a master or a nonmaster.

For a randomly sampled student, let the observed scores on the selection and mastery tests be continuous random variables denoted by X and Y, with realizations x and y, respectively. Also, it is assumed that, due to measurement error in the mastery test, the criterion to be considered is the classical test theory true score underlying the mastery test. Let the true score for a randomly sampled individual be denoted by a continuous random variable T with realization t.

Further, it will be assumed that the relation between X, Y, and T can be represented by a joint density function f(x,y,t). It is important to note that the best experiment to estimate the parameters in this density function is the one in which a sample of examinees from the *full* marginal distribution of X is admitted to the treatment and the performances of these students on the mastery test Y are measured. Though it is possible to estimate the parameters from a distribution of X truncated by the fact that low performing students are not admitted to the treatment, such estimates need a parametric model for the density, which might be wrong and/or poorly estimated.

Finally, it is assumed that the standard denoting true mastery is a threshold value t_c on T.

Simultaneous Decision Rules

Let each of the possible actions be denoted by a_{ij} (i,j=0,1), where i=0,1 stand for the actions of rejecting and accepting a student and j=0,1 for the actions of retaining and advancing an accepted student. Since for a rejected student no further mastery decisions are made, the index j will be dropped for i=0.

Generally, a decision rule specifies for each possible realization (x,y) of (X,Y) which action a_{ij} is to be taken.

Weak and Strong Rules

The decision rule for the mastery decision may or may not depend on the score X on the selection test. Intuitively, one can imagine that the fact that a student has delivered a high performance on the selection test leads to a more lenient rule for the mastery decision because this prior information implies that a possible low score on the mastery test is more likely due to measurement error than to a true low performance. Simultaneous rules in which decisions are a function both of the current test score and previous test scores test will be called weak rules in this paper. As a general result, it will be proven that under obvious conditions weak rules will necessarily have a compensatory nature. The title of the paper already alludes to this result.

If decisions are only a function of current test scores, the rules will be called strong (simultaneous) rules.

For the decision network of Figure 1 a weak simultaneous rule δ can be defined as:

$$\{(x,y) : \delta(x,y) = a_0\} = A \times R$$
$$\{(x,y) : \delta(x,y) = a_{10}\} = A^c \times B(x) \qquad (1)$$
$$\{(x,y) : \delta(x,y) = a_{11}\} = A^c \times B^c(x),$$

where A, A^c, B(x), and $B^c(x)$ stand for, respectively, the sets of x and y values for which a random student is rejected or admitted for a treatment and failed or passed the mastery test. R represents the set of real numbers.

With strong rules, the sets B(x) and $B^c(x)$ are independent of x. Strong simultaneous rules can only be optimal if certain conditions are met. These conditions will be given below.

Monotone and Nonmonotone Rules

Decision rules can take a monotone or a nonmonotone form. A decision rule is monotone if cutting scores are used to partition the sample space into regions for which different actions are taken. For example, a (separate) rule for the selection decision is monotone if there exists a cutting score x_c such that all examinees with $X \geq x_c$ are admitted and those with $X < x_c$ are rejected. All other possible rules are nonmonotone.

For our decision problem, a weak monotone rule δ can be defined as:

$$\delta(X,Y) = \begin{cases} a_0 & \text{for } X < x_c \\ a_{10} & \text{for } X \geq x_c,\ Y < y_c(x) \\ a_{11} & \text{for } X \geq x_c,\ Y \geq y_c(x), \end{cases} \qquad (2)$$

with $y_c(x)$ being the cutting score on Y. The fact that this cutting score is written as a mathematical function of x will be justified below proving that $y_c(x)$ is unique

for each value of x under reasonable assumptions.

In this paper, the interest will mainly be in monotone rules. The reason for this choice is the fact that the use of cutting scores is common practice in educational and psychological testing, and that rules with a different form are frequently not acceptable. However, the restriction to monotone rules is correct only if it can be proven that for any nonmonotone rule for the problem at hand there is a monotone rule with at least the same value on the criterion of optimality used; that is, if the subclass of monotone rules is essentially complete (Ferguson, 1967, p. 55). Conditions under which the subclass of monotone (simultaneous) rules is essentially complete for the present problem will also be given below.

Strong Monotone Rules with Maximum Expected Utility (SMMEU)

To evaluate the use of cutting scores even if conditions for monotonicity are not know to hold, the case of Strong Monotone Rules with Maximum Expected Utility (SMMEU rules) is also considered. A SMMEU rule is a rule with maximum expected utility in the subclass of strong monotone rules. The attention for SMMEU ru s is motivated by the fact that educators are familiar with cutting scores as decision rules and do not have a tradition of bothering about their justification.

Thus, if the sets of conditions for both strong and monotone rules to be optimal are satisfied, the subclasses of SMMEU and strong monotone Bayes rules are identical. Otherwise, they differ.

Utility Structure

Generally, a utility function describes the utility of each possible action for the possible true states of nature. Here, the utilities involved in the combined decision problem are defined as the following additive structure

$$u_{ij}(t) = w_1 u_i^{(s)}(t) + w_2 u_j^{(m)}(t), \tag{3}$$

where $u_i^{(s)}(t)$ and $u_j^{(m)}(t)$ represent the utility functions for the separate selection and mastery decisions and w_1 and w_2 represent nonnegative weights, respectively. Since utility is supposed to be measured on an interval scale, the weights of (3) can always be rescaled as follows:

$$u_{ij}(t) = wu_i^{(s)}(t) + (1-w)u_j^{(m)}(t), \tag{4}$$

where $0 \leq w \leq 1$. For a rejected student, zero contributions to the utility for the separate mastery decision are assumed. Hence, it follows from (4) that $u_{0j}(t)$ is equal to $wu_0^{(s)}(t)$ for all j.

It should be noted that the first term of (3) and (4) is a function of t and not, for example, of a true score underlying X. This fact illustrates one of the advantages of a simultaneous approach to decision making, namely, that there is no need to resort to intermediate criteria of success but that for all decisions utility can be defined as a function of the ultimate criterion in the network.

Below more specific functions $u_i^{(s)}(t)$ and $u_j^{(m)}(t)$ will be adopted. Obviously, these functions will be chosen such that utility will be an increasing function of t for the admittance and mastery decision but decreasing functions for the rejectance and nonmastery decision. First, however, more general results will be presented.

<u>Expected Utility in the Simultaneous Approach</u>

For the decision rules in (1) and the utility structure in (4), the expected utility for the two decision rules is equal to,

$$E[U_{sim}(A^c, B^c(x))] \equiv \int_A \int_R \int_R w u_0^{(s)}(t) f(x,y,t)\,dt\,dy\,dx + \int_{A^c} \int_{B(x)} \int_R u_{10}(t) f(x,y,t)\,dt\,dy\,dx + \int_{A^c} \int_{B^c(x)} \int_R u_{11}(t) f(x,y,t)\,dt\,dy\,dx. \tag{5}$$

In a Bayesian fashion, the expected utility in (5) will be taken as the criterion of optimality in this paper.

Taking expectations, completing integrals, and rearranging terms, (5) can be written as

$$E[U_{sim}(A^c, B^c(x))] = wE[u_0^{(s)}(T)] + \int_{A^c} \{E[u_{10}(T) - w u_0^{(s)}(T) \mid x] + \int_{B^c(x)} E[u_{11}(T) - u_{10}(T) \mid x,y] h(y \mid x) dy\} q(x) dx, \tag{6}$$

where $q(x)$ and $h(y \mid x)$ denote the p.d.f.'s of X and Y given X = x.

It is interesting to note that the critical quantities in (6) are the posterior expected utilities given X=x and (X=x,Y=y). It is through these quantities that information from prior tests will play a role in later decisions in the network.

Sufficient Conditions for Monotone Rules

In this section, monotonicity conditions for the simultaneous rules are derived. First, sufficient and necessary conditions for monotone solutions for the separate selection and mastery decisions will be given. Next, sufficient conditions for weak monotone solutions will be derived. Finally, monotonicity conditions for strong simultaneous rules will be derived from the previous case by imposing additional restrictions on the test-score distributions.

Conditions for Separate Selection and Mastery Decisions

Conditions necessary and sufficient for selection and mastery rules to be (strictly) monotone are given in Chuang, Chen and Novick (1981). Two sets of conditions must be met. First, the families of distributions of the true scores T given X=x and T given Y=y must be stochastic increasing; that is, their cumulative distribution functions (c.d.f.'s) must be decreasing in x and y for all t. Second, the utility functions must be monotone. This condition requires the difference between the utility function for the rejection (nonmastery) and admittance (mastery) decision to change sign at most once.

Both conditions immediately follow from the standard decision problem addressed in statistical decision theory (e.g., Ferguson, 1967; Lindgren, 1976).

Conditions for Weak Simultaneous Rules

Let V(t|x,y) denote the c.d.f. of T given (X=x,Y=x) and H(y|x) the c.d.f. of Y given X = x. The following theorem gives a set of conditions sufficient for a weak monotone solution:

Theorem. An optimal simultaneous decision rule for the selection-mastery problem is (weak) monotone if:

$u_1^{(m)}(t) - u_0^{(m)}(t)$ is strictly increasing in t, (7)

$u_{10}(t) - wu_0^{(s)}(t)$ is strictly increasing in t, (8)

$V(t \mid x,y)$ is strictly decreasing in x and y for all t, (9)

$H(y \mid x)$ is stricly decreasing in x for all y. (10)

The first condition guarantees monotone utility for the mastery decisions.

The second condition stipulates that the difference between the utility functions for the actions a_{10} (acceptance, nonmastery) and a_0 (rejection) be an increasing function of t.

The third condition requires double (strict) stochastic increasingness for V(t|x,y). Loosely speaking, this condition is met if high true scores on the mastery test coincide with high observed scores on both the selection and mastery tests.

The last condition also requires (strict) stochastic increasingness, and thus that high scores on the mastery and selection test tend to coincide.

Not all conditions in this set are straightforward generalizations of the conditions for the separate decision problems. In particular, the conditions in (8) and (10) are new; they are needed to link the two separate decision problems.

It should be noted that there is no condition analogous to (7) for the selection problem. This is due to the fact that the utility component for this problem is defined on the true score variable for the mastery test.

In the proof of the theorem, the following lemma's are needed:

Lemma 1: Let $f(x)$ be an arbitrary function with $\int |f(x)| dx < \infty$, then for any set S of x values it holds that $\int_S f(x)dx \leq \int_{S'} f(x)dx$ with $S' = \{x: f(x) \geq 0\}$ (e.g., Ferguson, 1967, p. 201).

Lemma 2: For any increasing function $k(t)$, the expectation $E[k(T)|z]$ is an increasing function of z if and only if the c.d.f. of T given Z=z is stochastic increasing (e.g., Lehmann, 1959, p. 74).

Observe that if $k(t)$ is a constant, $E[k(T)|z)]$ is a constant too. Hence, the nondecreasing version of the lemma also holds.

Lemma 3: If (9) and (10) hold, then the marginal c.d.f. $P(t|x)$ associated with $V(t|x,y)$ is stochastic increasing in x.

This lemma is proven as follows: Let $v(t|x,y)$ be the p.d.f. of T given X =x and Y = y. By definition, $1-P(t|x) = \int_t^{\infty}\int_{-\infty}^{\infty} v(z|x,y)h(y|x)dydz = \int_{-\infty}^{\infty} [1-V(t|x,y)]h(y|x)dy$.

From (9), (10) and Lemma 2, it follows that $1\text{-}P(t|x)$ increases in x for all t, i.e., that $P(t|x)$ is stochastic increasing in x. ■

For completeness' sake, it is observed that the c.d.f. of T given Y = y is also stochastic increasing if (10) is replaced by the stronger condition of monotone likelihood ratio. However, this result is not needed in the remainder of this paper.

Lemma 4. If a function $\kappa(x,y)$ is (strictly) increasing in x and y, then the relation defined by $C = \{(x,y)|\kappa(x,y)=c,\ c \in R\}$ is a decreasing function in x.

To proof this lemma, assume that there are two pairs $(x_1,y_1) \in C$ and $(x_2,y_2) \in C$ with $x_2 > x_1$, for which $y_2 \geq y_1$. Then, by hypothesis, $\kappa(x_2,y_2) > \kappa(x_1,y_1)$, which contradicts the assumption. ■

Proof of Theorem

Applying Lemma 1 to the second term in the integral in (6), and using $h(y \mid x) \geq 0$, it follows that for all $B^c(x)$ and an arbitrary but fixed A^c:

$$E[U_{sim}(A^{c},B^{c}(x))] \leq wE[u_0^{(s)}(T) + \int_{A^c} \{E[u_{10}(T)-wu_0^{(s)}(T)|x] + \int_{B_0^c(x)} E[u_{11}(T)-u_{10}(T)|x,y]h(y|x)dy\}\, q(x)dx, \tag{11}$$

with

$$B_0^c(x) = \{y\colon E[u_{11}(T)\text{-}u_{10}(T) \mid x,y] \geq 0\}. \tag{12}$$

Again, applying the theorem to the second term in the right-hand side of (11), and using $q(x) \geq 0$, it follows that for all A^c

$$E[U_{sim}(A^c, B_0^c(x))] \leq wE[u_0^{(s)}(T)] + \int_{A_0^c} \{E[u_{10}(T) - wu_0^{(s)}(T)|x] + \int_{B_0^c(x)} E[u_{11}(T) - u_{10}(T)|x,y]h(y|x)dy\}q(x)dx, \tag{13}$$

with

$$A_0^c = \{ x: E[u_{10}(T) - wu_0^{(s)}(T)|x] + \int_{B_0^c(x)} E[u_{11}(T) - u_{10}(T)|x,y]h(y|x)dy \geq 0 \}. \tag{14}$$

It is now proven that the left-hand sides of the inequalities in (12) and (14) increase in y for all x and in x for all y, respectively. If these features hold, then (6) is maximal for the sets $A_0^c = [x_c, \infty)$ and $B_0^c = [y_c(x), \infty)$, where x_c and $y_c(x)$ are the values of x and y for which the inequalities in (12) and (14) become equalities. (The numbers x_c and $y_c(x)$ may be infinitely small or large implying that the same decisions have to be made for all examinees.)

(i) Since $u_{11}(t) - u_{10}(t) = (1-w)[u_1^{(m)}(t) - u_0^{(m)}(t)]$ and $1-w \geq 0$, it follows from the condition in (7) that the difference between these two utilities is increasing too. Therefore, (9) and Lemma 2 together imply that

$E[u_{11}(T)-u_{10}(T)|x,y]$ is increasing in y for all x and in x for all y, (15)

and thus that the sets $B_0{}^c(x)$ take the required form $[y_c(x),\infty)$ for all values of x. This result will be used in the following part of the proof.

(ii) From (8)-(10), Lemma 2 and Lemma 3, it follows immediately that the first term in the left-hand side of (14) is increasing in x.

For notational convenience, the term $E[u_{11}(T)-u_{10}(T)|x,y]$ is denoted as $\tau(x,y)$. Note that $\tau(x,y)$ is an increasing function of y which is nonnegative for $y \geq y_c(x)$ for all values of x. Now for any $x_2 > x_1$, it follows from Lemma 4 that

$$\int_{y_c(x_2)} \tau(x_2,y)h(y|x_2)dy - \int_{y_c(x_1)} \tau(x_1,y)h(y|x_1)dy > \tag{16}$$

$$\int_{y_c(x_1)} \tau(x_2,y)h(y|x_2)dy - \int_{y_c(x_1)} \tau(x_1,y)h(y|x_1)dy >$$

$$\int_{y_c(x_1)} \tau(x_1,y)[h(y|x_2) - h(y|x_1)]dy =$$

$$\int_{-\infty}^{\infty} \varphi(y)[h(y|x_2)-h(y|x_1)]dy,$$

where $\varphi(y) = I_{[y_c(x_1),\infty)}(y)\tau(x_1,y)$. By definition, $\varphi(y)$ is a nondecreasing function of y, and it follows from (10) and Lemma 2 that (16) is positive. Hence, it can be

concluded that the second left-hand term in (14) is increasing in x, and thus that the set A_0^c takes the required form $[x_c,\infty)$. ■

<u>Monotonicity Conditions for Strong Simultaneous Rules</u>

For strong simultaneous rules, $B_0^c(x)$ is not allowed to depend on x. Therefore, as an additional condition, it must hold for v(t|x,y) and the p.d.f. of T given Y=y that

$$v(t|x,y)=g(t|y). \tag{17}$$

This condition, which immediately follows from (12), implies that all information on T relevant for the decision is contained in Y=y, and that, once Y=y is given, the observation X=x does not add any information. If the condition holds, then, obviously, the use of simultaneous rules will not add any efficiency to the decision making procedure.

Calculation of Simultaneous Rules

From the theorem it follows that the optimal weak and strong simultaneous rules can be calculated as the points at which the inequalities in (12) and (14) turn into equalities. For the weak rules, it should be noted that the sets $B_0^c(x)$ and A_0^c are sequentially defined. First, for all values of x, the sets $B_0^c(x)$ are defined by (12). Only then the set A_0^c is defined by (14). Hence, optimal weak rules have to be calculated in this sequence.

SMMEU rules can be calculated solving the system of equations consisting of the partial derivatives of (6) w.r.t. x_c and y_c equated to zero.

In the empirical example below, for the calculation of all cutting scores Newton's method for solving nonlinear systems was used. The method was implemented in a computer program called NEWTON. Another program, UTILITY, was written to analyze differences in expected utility for the various rules. Copies of the programs are available from the authors of the paper upon request.

Optimal Separate Rules

It is observed that optimal rules for the separate decisions can easily be found by imposing certain restrictions on $E[U_{sim}(A^c, B^c(x))]$.

First, substituting $w = 1$ into (6), the expected utility for the separate selection decision $E[U^{(s)}(A^c)]$, can be written as

$$E[U^{(s)}(A^c)] = E[u_0^{(s)}(T)] + \int_{A^c} E[u_1^{(s)}(T) - u_0^{(s)}(T) | x]q(x)dx. \qquad (18)$$

Next, substituting $w = 0$, $A^c = R$ (i.e., accepting all students for the instructional treatment), and $B^c(x) = B^c$ into (6) gives the following result for the expected utility of the separate mastery decision:

$$E[U^{(m)}(B^c)] = E[u_0^{(m)}(T)] + \int_{B^c} E[u_1^{(m)}(T) - u_0^{(m)}(T) | y]s(y)dy, \qquad (19)$$

where s(y) denotes the p.d.f. of Y.

Analogous to the simultaneous approach, it can easily be verified that upper bounds to $E[U^{(s)}(A^c)]$ and $E[U^{(m)}(B^c)]$ are obtained for the sets of x and y values for which $E[u_1^{(s)}(T)-u_0^{(s)}(T) \mid x]$ and $E[u_1^{(m)}(T)-u_0^{(m)}(T) \mid y]$ are nonnegative, respectively. Assuming that the monotonicity conditions for the separate decisions are satisfied, the optimal cutting scores for the separate selection and mastery decisions, say $\tilde{x}_c$ and $\tilde{y}_c$, can be obtained by solving $E[u_1^{(s)}(T)-u_0^{(s)}(T) \mid x]$ and $E[u_1^{(m)}(T)-u_0^{(m)}(T) \mid y]$ for x_c and y_c, respectively. For further details, see Mellenbergh and van der Linden (1981) and van der Linden and Mellenbergh (1977).

An Empirical Example

Optimal rules were calculated for a selection-mastery decision problem consisting of a CAI module on elementary medical knowledge preceded and followed by a selection and mastery test, respectively. Both tests consisted of 21 items and had possible test scores ranging from 0-100. Data were available for a sample of 76 freshmen in a medical program. The instructors in the program considered students as having mastered the module if their true scores were larger than 55. Therefore, t_c was fixed at this value.

Score Distributions

It was assumed that (X,Y,T) followed a trivariate normal distribution. Under this assumption, the bivariate distribution of (X,Y) is also normal. Further, the regression function $E[Y|x]$ is linear.

These two observable consequences were tested against the data using a chi-square and a t-test. The probabilities of exceedance were 0.219 and 0.034, showing a satisfactory fit which confirmed our visual inspection of various plots of the distributions.

Some descriptive statistics for the two tests are given in Table 1.

Table 1 about here

Utility Structure

The following choice was made for the functions $u_i^{(s)}(t)$ and $u_j^{(m)}(t)$ in (4):

$$u_i^{(s)}(t) = \begin{cases} b_0^{(s)}(t_c - t) + d_0^{(s)} & \text{for } i = 0 \\ b_1^{(s)}(t - t_c) + d_1^{(s)} & \text{for } i = 1 \end{cases} \tag{20}$$

$$u_j^{(m)}(t) = \begin{cases} b_0^{(m)}(t_c - t) + d_0^{(m)} & \text{for } j = 0 \\ b_1^{(m)}(t - t_c) + d_1^{(m)} & \text{for } j = 1 \end{cases} \tag{21}$$

where $b_0^{(s)}, b_j^{(m)} > 0$ $(i,j = 0,1)$. The parameters $d_i^{(s)}$ and $d_j^{(m)}$ can represent, for example, the fixed amount of costs involved in following an instructional module and testing the examinees. The condition $b_0^{(s)}$, $b_1^{(s)} > 0$ states that utility be a decreasing function for the rejection decision, but an increasing function for the

acceptance decision. Similarly, the condition $b_0^{(m)}$, $b_1^{(m)} > 0$ expresses that the utilities associated with failing and passing the mastery test be decreasing and increasing functions in t, respectively.

The same utility functions were used in an analysis of separate selection and mastery decisions in Mellenbergh and van der Linden (1981) and van der Linden and Mellenbergh (1977). For other possible utility functions, see Novick and Lindley (1979).

Monotonicity Conditions

The condition in (7) is met since $b_j^{(m)} > 0$, j=0,1.

It can easily be verified that the condition in (8) is satisfied if the weight w and the parameters $b_0^{(s)}$, $b_1^{(s)}$, and $b_0^{(m)}$ are chosen such that

$$w > b_0^{(m)}/(b_0^{(s)}+b_1^{(s)}+b_0^{(m)}). \qquad (22)$$

All numerical values for the utility parameters in the example were chosen to meet these two requirements.

Under the model of a trivariate normal distribution for (X,Y,T) in this example, the conditions in (9)-(10) were met by the positive slopes of the regression lines and planes in this distribution.

Finally, the additional condition for solutions to be strong monotone in (17) was tested comparing the two regression lines E[T|x,y] and E[T|y] using an F-test. The probability of exceedance was 0.038, indicating that the result was just significant for α=.05. Therefore, only SMMEU rules and no optimal strong rules were considered.

Results for the Simultaneous Rules

For several values of the utility parameters, weak monotone and SMMEU rules were calculated. The results are reported in Table 2, where the cutting scores for the SMMEU rules are denoted as x^*_c and y^*_c.

Table 2 about here

As is clear from the results, the consequences of increasing the values of the parameters $b_i^{(s)}$ and $b_j^{(m)}$ were decreases of the optimal weak and SMMEU cutting scores on the selection test. On the other hand, a decrease of the amount of constant utility, $d_i^{(s)}$ and $d_j^{(m)}$, resulted in increases of the optimal weak and SMMEU cutting scores on the selection test. Furthermore, Table 2 indicates that the optimal weak and SMMEU cutting scores on the selection test increase in w for utility structures (1)-(3) and (4)-(6) in Table 2, whereas the opposite holds for utility structures (7)-(9) in the table.

Results for the Separate Approach

The optimal cutting scores $\tilde{x}_c$ and $\tilde{y}_c$ for the separate selection and mastery decisions are also reported in Table 2. In particular for w = 0.3, the weak cutting scores $y_c(x_c)$ on the mastery test generally were high compared with $\tilde{y}_c$.

The results did not differ much from those obtained for the weak monotone rules. This fact can be explained as follows: Students who were just accepted in the case of a weak monotone rule had to compensate their rather low cutting scores on the selection test with relatively high scores on the mastery test compared with students accepted in the case of separate rules. However, the decreasing character

of $y_c(x)$ in x implied that only students accepted with selection scores equal to or just above x_c did need these rather high scores on the mastery test to reach the mastery status.

Comparison of the Expected Utilities

For the simultaneous approach a gain in expected utility relative to the separate approach was expected. To see whether this expectation could be confirmed, the weighted sum of the expected utilities for the optimal separate rules was compared with the expected utilities for the optimal weak monotone rules. The results are also displayed in Table 2.

It can be seen that the expected utilities for the optimal weak monotone rules yielded the largest values for all utility structures. This result was in accordance with our expectations. Furthermore, Table 2 indicates that the expected utilities for the optimal weak monotone rules were only slightly larger than for the SMMEU rules. Finally, the table shows that for all three approaches, the expected utility yielded the largest value for w = 0.9. In other words, the utility for the selection decision contributed most to the expected utility for the optimal simultaneous rules in this study.

Concluding Remarks

For a monotone utility structure, Lemma 4 shows that under the natural condition of the selection and mastery test scores being stochastic increasing in the true score on the mastery test, weak cutting scores for mastery decisions are a decreasing function of the scores on the selection test. As already explained, this

feature introduces an element of compensation in the decision procedure: It is possible to compensate low scores on the mastery test by high scores on the selection test. A quantitative estimate of this effect can be calculated for the data set in the empirical example above. Substituting the estimated regression plane $E[T|x,y] = \alpha + \beta x + \gamma y$ into the left-hand-side in (12) and solving for $y_c(x)$ yields

$$y_c(x) = [(d_o^{(m)} - d_1^{(m)})/(b_o^{(m)} + b_1^{(m)}) + t_c - \alpha - \beta x]/\gamma.$$

The derivative of this equation w.r.t. x is equal to $-\beta/\gamma$, which for the data set was estimated as -.675. It follows for all utility structures in this example that the cutting score $y_c(x)$ on the mastery test has to be lowered by .675 for each score point above x_c on the selection test.

Although the area of individualized instruction is a useful application of simultaneous decision making, it should be emphasized that the optimization models advocated in this paper have a larger scope of application. For any situation in which subjects are accepted for a certain treatment on the basis of their scores on a selection test with attainments evaluated by a mastery test, the optimal rules presented in this paper can improve the decisions. An example is psychotherapy where clients accepted have to pass a success criterion before being dismissed from the therapy.

References

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*, 129-152.

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

Davis, C.E., Hickman, J., & Novick, M.R. (1973). *A primer on decision analysis for individually prescribed instruction* (ACT Technical Bulletin No. 17). Iowa City, IA: American Testing College Program, 1973.

Ferguson, T.S. (1967). *Mathematical statistics: A decision theoretic approach*. New York: Academic Press.

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., & Novick, M.R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*, 159-170.

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

Huynh, H. (1977). Two simple classes of mastery scores based on the beta-binomial model. *Psychometrika, 42*, 601-608.

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

Lehmann, E.L. (1959). *Testing statistical hypotheses*. New York: John Wiley and Sons.

Lindgren, B.W. (1976). *Statistical theory* (3rd ed.). New York: Macmillan.

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

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

Novick, M.R., & Lindley, D.V. (1979). Fixed-state assessment of utility functions. *Journal of the American Stastistical Association, 74*, 306-311.

Saywer, R. (1993, April). *Decision theoretic models for validating course placement systems*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

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

van der Linden, W.J. (1987). The use of test scores for classification decisions with threshold utility. *Journal of Educational Statistics, 12*, 62-75.

van der Linden, W.J. (1990). Applications of decision theory to test-based decision making. In R.K. Hambleton & J.N. Zaal (Eds.), *New developments in testing: Theory and applications* (pp. 129-155). Boston: Kluwer.

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

Vos, H.J. (1990). Simultaneous optimization of decisions using a linear utility function. *Journal of Educational Statistics, 15,* 309-340.

Vos, H.J. (1991). Simultaneou› optimization of the aptitude treatment interaction decision problem w ith mastery scores. In M. Wilson (Ed.), *Objective measurement: Theo y into practice* (pp. 313-331). Norwood, New Yersey: Ablex Publishing Corporation.

Vos, H.J. (1993). Simultaneous optimization of selection-placement decisions followed by a mastery decision. In F.J. Maarse et al. (Eds.), *Computers in psychology: Tools for experimental and applied psychology* (Volume 4, pp. 235-244), *4,* 235-244. Lisse: Swets & Zeitlinger.

Acknowledgment

The authors are indebted to Wilbert Kallenberg for his valuable comments and to Jan Gulmans for providing the data for the empirical example.

Table 1

Statistics Selection and Mastery Tests (X and Y)

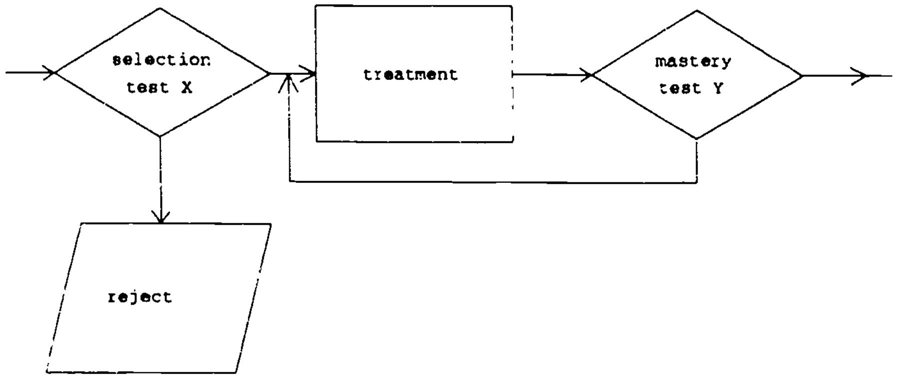
Statistics	X		Y
Mean	50.679		62.436
Standard Deviation	8.781		9.456
Reliability	0.773		0.802
Correlation		0.751	

[illegible]

[illegible]	Simultaneous Weak Monotone	Simultaneous SMMEU	Separate	Expected Utility Weak Monotone	Expected Utility SMMEU	Expected Utility Separate
[illegible]	[illegible]	[illegible]	X=41.64 [illegible]	[illegible]	[illegible]	[illegible]
[illegible]	X_c=41.41 [illegible]	[illegible]	X=41.64 / 53.41	32.64	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	X=41.49 / [illegible]	[illegible]	31.98	[illegible]
[illegible]	[illegible]	[illegible]	X=41.4[illegible]	36.38	36.35	35.32
[illegible]	[illegible]	[illegible]	X=41.4[illegible]	40.6[illegible]	40.6[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	19.05	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure Caption

Figure 1. A system of one selection and one mastery decision.


selection
test X
treatment
mastery
test Y
reject

Titles of recent Research Reports from the Department of
Educational Measurement and Data Analysis.
University of Twente, Enschede,
The Netherlands.

RR-94-2 W.J. van der Linden & H.J. Vos, *A compensatory approach to optimal selection with mastery scores*

RR-94-1 R.R. Meijer, *The influence of the presence of deviant item score patterns on the power of a person-fit statistic*

RR-93-1 P. Westers & H. Kelderman, *Generalizations of the Solution-Error Response-Error Model*

RR-91-1 H. Kelderman, *Computing Maximum Likelihood Estimates of Loglinear Models from Marginal Sums with Special Attention to Loglinear Item Response Theory*

RR-90-8 M.P.F. Berger & D.L. Knol, *On the Assessment of Dimensionality in Multidimensional Item Response Theory Models*

RR-90-7 E. Boekkooi-Timminga, *A Method for Designing IRT-based Item Banks*

RR-90-6 J.J. Adema, *The Construction of Weakly Parallel Tests by Mathematical Programming*

RR-90-5 J.J. Adema, *A Revised Simplex Method for Test Construction Problems*

RR-90-4 J.J. Adema, *Methods and Models for the Construction of Weakly Parallel Tests*

RR-90-2 H. Tobi, *Item Response Theory at subject- and group-level*

RR-90-1 P. Westers & H. Kelderman, *Differential item functioning in multiple choice items*

Research Reports can be obtained at costs from Bibliotheek, Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

Faculty of
EDUCATIONAL SCIENCE
AND TECHNOLOGY

A publication by
The Faculty of Educational Science and Technology of the University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands