DOCUMENT RESUME

ED 310 127 TM 013 700

AUTHOR	Vos, Hans J.
TITLE	Simultaneous Optimization of Decisions Using a Linear Utility Function. Research Report 88-15.
INSTITUTION	Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE	Nov 88
NOTE	56p.; Portions of this paper were presented at the European Meeting of the Psychometric Society (Enschede, Netherlands, 1987).
AVAILABLE FROM	Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Bayesian Statistics; College Freshmen; Computer Assisted Instruction; *Decision Making; *Expert Systems; Higher Education; *Linear Programing; *Mastery Tests; Medical Students; Science Tests; Selection
IDENTIFIERS	Decision Rules; Empirical Bayes Estimation; Free Response Test Items; Medical Knowledge Tests; *Optimization; *Utility Function

ABSTRACT

The purpose of this paper is to simultaneously optimize decision rules for combinations of elementary decisions. As a result of this approach, rules are found that make more efficient use of the data than does optimizing those decisions separately. The framework for the approach is derived from empirical Bayesian theory. To illustrate the approach, two elementary decisions--selection and mastery decisions--are combined into a simple decision network. A linear utility structure is assumed. Decision rules are derived both for quota-free and quota-restricted selection-mastery decisions for several subpopulations. An empirical example of instructional decision making in an individual study system concludes the paper. The example involves 43 freshmen medical students (27 were disadvantaged and 16 were advantaged with respect to elementary medical knowledge). Both the selection and mastery tests consisted of 17 free-response items on elementary medical knowledge with test scores ranging from 0 to 100. The treatment consisted of a computer-aided instructional program. Three data tables and three figures are provided. (Author/TJH)

Simultaneous Optimization of Decisions Using a Linear Utility Function

Research Report

88-15

Hans J. Vos

ment of
EDUCATION

University of Twente

Division of Educational Measurement
and Data Analysis

Colofon:
Typing: Lorette Padberg
Cover design: Audiovisuele Sectie TOLAB Toegepaste Onderwijskunde
Printed by: Centrale Reproductie-afdeling

Simultaneous Optimization of Decisions Using a Linear Utility Function

Hans J. Vos

Applications of decision theory to computer based adaptive instructional systems / Hans Vos - Enschede : University of Twente, Department of Education, November, 1988. - 34 pages

Abstract

The purpose of this paper is to simultaneously optimize decision rules for combinations of elementary decisions. As a result of this approach, rules are found that make more efficient use of the data than optimizing these decisions separately. The framework for the approach is derived from (empirical) Bayes theory. To illustrate the approach, two elementary decisions (viz. selection and mastery decisions) are combined into a simple decision network. A linear utility structure is assumed. Decision rules are derived both for quota-free and quota-restricted selection-mastery decisions in case of several subpopulations. An empirical example of instructional decision making in an individual study system concludes the paper.

Introduction

Decision problems in educational and psychological testing can be classified in many ways. An elegant typology of test-based decisions has been given in van der Linden (1985, 1988). Each type of decision making in this typology can be viewed as a specific configuration of three basic elements, namely a test, a treatment, and a criterion. In general, the following four different types of decision problems can be distinguished: selection, mastery, placement, and classification.

Educational applications of the four types of decision making can be found in such fields as the admission of students to schools (selection), pass-fail decisions (mastery), the aptitude-treatment-interaction paradigm in instructional psychology (placement), and vocational guidance situations where most promising schools must be identified (classification).

In Hambleton and Novick (1973), Huynh (1976, 1977), Mellenbergh and van der Linden (1981), Novick and Petersen (1976), Petersen (1976), Petersen and Novick (1976), van der Linden (1980, 1981, 1987) and Vos (1988), these elementary decision problems have been studied extensively; these authors also indicate how - analytically or numerically - optimal decision rules can be found using (empirical) Bayesian decision theory.

The four elementary decisions can be met both in their pure forms or in combinations with each other. The latter is

the case, for instance, in test-based decision making in individualized study systems (ISS's), which can be conceived of as networks consisting of these various types of decisions as nodes (Vos & van der Linden, 1987). In such systems decision making can be viewed as proceeding students through a network of several of the elementary decisions.

The purpose of this paper is the simultaneous optimization of combinations of elementary decisions using a decision-theoretic approach. Compared with separate optimization of elemer ary decisions, two main advantages can be identified. First, rules making more efficient use of the data can be found. Second, utility structures can be made more realistic. In order to illustrate the approach, in this paper a selection and a mastery decision will be combined into a simple decision network, and it will be indicated how optimal rules for guiding students through such a system can be derived. The first advantage of the simultaneous approach is illustrated using this simple system. For instance, when optimizing acceptance-rejection rules in the combined decision network, pass-fail decisions to be made later can already be taken into account (see also Figure 2). The second advantage will be explained after the utility function for the combined decision has been specified.

For each elementary decision, one or more of the following restrictions may apply (van der Linden, 1988):

(1) Multiple populations. The problem of culture-fair decision making may arise because of the presence of subpopulations reacting differently to the test items, e.g.

for populations defined by race or sex. In such a case, the test items are often be assumed to be "biased" against some of the populations.

(2) Quota restrictions. For some treatments, due to shortage of resources, the number of vacancies are constrained.

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

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

In the present paper, only restrictions will be made with respect to the presence of subpopulations and the number of students to be accepted for some treatments. First, the problem of culture-fair decision making will be considered for a quota-free selection problem. Next, optimal rules will be derived for quota-restricted selection problems using methods of constrained optimization. The final section presents some empirical examples of optimal cut-off scores for quota-free as well as quota-restricted selection-mastery decisions for two subpopulations referred to as the disadvantaged and the advantaged populations.

Statement of the Problem

As noted before, a well-known example of combinations of elementary decisions in education is an individualized instruction system. Figure 1 shows a flowchart of a system in which a selection decision is followed by a treatment, here an instructional module. Then a mastery decision follows,

after which a placement decision assigns the students to two different routes through a module both leading to the same learning objective. Real-life ISS's often have more decision points.

Insert Figure 1 about here

Selection-mastery decisions may occur in an ISS, for instance, when decisions on the admission of students to the system should be made. Then a selection test is administered before the treatment takes place and students promising satisfactory results on the criterion are accepted for the first module of the instructional program (see Figure 2). Furthermore, let us suppose that the criterion is unreliably measured, which is not uncommon in ISS's. If success on the criterion is measured by a threshold value separating "masters" from "nonmasters", then, in fact, after the treatment a mastery decision has to be taken, and the problem is a selection-mastery decision problem. Students who have reached the module objectives may proceed with the next module. However, students who failed are provided with supplemental instruction, extra learning time, corrective feedback, and the like. These students have to prepare themselves for a new mastery test.

Insert Figure 2 about here

In the following, we shall suppose that in the selection-mastery decision problem g ($g \geq 2$) subpopulations reacting differently to the test items can be distinguished. Furthermore, it is assumed that the observed selection test score variable X, the observed mastery test score variable Y, and the true score variable T underlying Y, i.e. the criterion score, assume only continuous values. Formally, the presence of populations reacting differently to test items implies different cut-off scores for each population. Therefore, let x_{ci} and y_{ci} denote the cut-off scores for subpopulation i ($i = 1,2,\ldots\ldots,g$) on the observed test score variables X and Y, respectively. However, the cut-off score t_c on the criterion score T is assumed to be equal for each population and is set in advance by the decision-maker. The combined decision problem can now be stated as choosing values of x_{ci} and y_{ci} that, given the value of t_c, are optimal in some sense.

In the present paper, linking up with common practice in criterion-referenced testing, we consider only decisions in which the decision rules δ have a monotone form: students are admitted to a treatment if their test score is above a certain cutting point and rejected otherwise. They can be defined for our example in the following way:

$$(1) \qquad \delta(X, Y) = \begin{cases} a_0 & \text{for } X < x_{ci} \\ a_1 & \text{for } X \geq x_{ci},\ Y < y_{ci} \\ a_2 & \text{for } X \geq x_{ci},\ Y \geq y_{ci}, \end{cases}$$

where a_0, a_1, and a_2 stand for the actions to reject a student, to retain an accepted student, and to advance an accepted student, respectively.

An appropriate framework for dealing with decision problems such as the above is (empirical) Bayesian decision theory (e.g., DeGroot, 1970; Ferguson, 1967; Keeney & Raiffa, 1976; Lindgren, 1976). Besides the actions, probabilities and utilities are two other fundamental elements in a Bayesian procedure. In case of an ISS, a probability model predicts the outcomes of the several possible routes for the students, and a utility structure evaluates the outcomes predicted. The optimal procedure as prescribed by Bayesian decision theory is to look for a decision rule that maximizes expected utility.

With respect to the first element, it will be assumed that for each population i, the probability function $\Omega_i(x,y,t)$ of the joint distribution (X,Y,T) is available. Note that, due to the presence of different populations reacting differently to test items, different probability functions for each population should be assumed.

Also, the decision-maker may have different utilities associated with different populations. Hence, in addition to separate probability distributions, the decision-maker has to

specify explicitly his/her utility function for each subpopulation separately.

The utility structure dealt with in this paper is a linear function of the criterion variable T, which seems to be a realistic representation of the utilities actually incurred in many decision making situations. In a recent study, for instance, it was shown by van der Gaag (1987) that many empirical utility structures could be approximated by linear functions.

Monotonicity Conditions

As mentioned before, in a decision-theoretic approach, optimal decision rules are found by optimizing expected utility. However, the restriction to monotone rules in our paper is only correct if there are no nonmonotone rules with higher expected utility. It is here that the notion of an essentially complete class of decision rules comes in handy. An essentially complete class is defined as a class of decision rules as good as rules outside this class (e.g., Ferguson, 1967, p.55).

In case of separate elementary decisions, the monotonicity conditions are known (Ferguson, 1967, sect. 6.1; Karlin & Rubin, 1956). Two conditions have to be met: First the probability model relating observed test score Z to true score T should have a monotone likelihood ratio (MLR), i.e. it is required that for any $t_2 < t_1$, the likelihood ratio $f(z|t_1)/f(z|t_2)$ is a nondecreasing function of z. Second, the utility function should be monotone; that is, the actions

should be ordered such that for each two adjacent actions the utility functions have at most one intersection point. If these conditions are met, a monotone solution is said to exist. It should be noted that for the classification problem these conditions do not hold without modifications (van der Linden, 1987).

To guarantee that the monotone rule of the combined decision problem belongs to an essentially complete class, the following extra condition (Lehmann, 1959, sect. 3.3) should hold:

(2) For any $t_2 < t_1$, the likelihood ratio $k(x,y|t_1)/k(x,y|t_2)$ is a nondecreasing function in each of its arguments; that is, for any $t_2 < t_1$ and fixed values of $Y = y_0$ and $X = x_0$, the likelihood ratios $k(x,y_0|t_1)/k(x,y_0|t_2)$ and $k(x_0,y|t_1)/k(x_0,y|t_2)$ are nondecreasing functions of x and y, respectively.

It will be shown below that, in addition to the conditions of MLR and monotone utility, condition (2) is sufficient for a monotone solution to exist for the combined decision problem. The condition of monotone utility is elaborated in the next section.

Linear Utility Function for a Selection–Mastery Decision

Generally speaking, a utility function evaluates the total consequences of all possible decision outcomes. Formally, it is a function $u_{ji}(t)$ that describes the utility incurred when action a_j ($j = 0,1,2$) is taken for the student from subpopulation i whose true score is t.

Mellenbergh and van der Linden (1981) and van der Linden and Mellenbergh (1977) use a linear utility function for determining optimal cutting scores on the separate decisions. Here, their function is restated for the combined decision problem as a linear function in T for subpopulation i (see also Figure 3):

$$(3) \qquad u_{ji}(T) = \begin{cases} b_{0i}(t_c - t) + d_{0i} & \text{for } X < x_{ci} \\ b_{1i}(t - t_c) + d_{1i} & \text{for } X \geq x_{ci},\ Y < y_{ci} \\ b_{2i}(t - t_c) + d_{2i} & \text{for } X \geq x_{ci},\ Y \geq y_{ci}, \end{cases}$$

where b_{0i}, $b_{2i} > 0$.

For each action a_j ($j = 0,1,2$), this function consists of a constant term and a term proportional to the difference between the criterion performance t of a student and the minimum level of satisfactory criterion performance t_c. The parameters d_{0i}, d_{1i}, and d_{2i} can represent, for example, the costs of testing or the cost of following an instructional module. The condition b_{0i}, $b_{2i} > 0$ is equivalent to the

statement that for the rejected students and the accepted students who passed the mastery test, utility is a strictly decreasing and increasing function of t, respectively.

It should be noticed that it cannot be said beforehand whether the utility associated with action a_1, i.e. $u_{1i}(t)$, is increasing or decreasing, because the utility of the combined decision depends on the utilities associated with the selection as well as the mastery decision. Depending on either the influence of the utility associated with the acceptance or with the fail decision is the most important, $u_{1i}(t)$ is an increasing or decreasing function of t, respectively. Figure 3 displays an example of a combined linear utility function for $b_{1i} > 0$.

Insert Figure 3 about here

In the Introduction, it was remarked that one of the main advantages of a simultaneous approach was that more realistic utility structures could be used. Formula 3 nicely demonstrates how a utility function defined on the ultimate criterion of the ISS ("master" or "nonmaster") can be properly brought into a previous decision (selection decision).

Gross and Su (1975) pointed out that 'fair' selection is a question of utilities. Whether a selection procedure is believed to be fair to the various subpopulations which can

be distinguished depends on the utilities of those involved in the selection process. From this point of view, the linear utility model can be used to allow for the fact that the students might belong to a disadvantaged or advantaged subpopulation by choosing separate parameter values for the subpopulations involved. Suppose, for example, that subpopulation h is considered more advantaged than i. In choosing values of the parameters of the linear utility function this can be taken into account by requiring that incorrect decisions are considered worse for subpopulation i than for h, while correct decisions are considered more valuable for i than for h. This amounts to choosing values of the slope parameters such that $b_{0i} > b_{0h}$ and $b_{2i} > b_{2h}$ for all t. Since $b_{1i} > 0$ implies that the influence of the utility associated with the acceptance decision is the most important, it will hold that $b_{1i} > b_{1h}$ for $b_{1i}, b_{1h} > 0$. Following the same line of reasoning, it is required that $b_{1i} < b_{1h}$ if $b_{1i}, b_{1h} < 0$.

The possible actions are supposed to be ordered as a_0, a_1, and a_2. Using the fact that, as can be seen from Figure 3, the difference between the utilities change sign precisely once, the condition of monotone utility for the utility function defined by Formula 3 can be expressed as

(4) $$u_{1i}(t)-u_{0i}(t) = (b_{1i}+b_{0i})(t-t_c) + d_{1i}-d_{0i} > 0 \qquad \text{for } t > t_{10,i}$$

$$u_{1i}(t)-u_{0i}(t) = (b_{1i}+b_{0i})(t-t_c) + d_{1i}-d_{0i} < 0 \qquad \text{for } t < t_{10,i}.$$

(5) $$u_{2i}(t)-u_{1i}(t) = (b_{2i}-b_{1i})(t-t_c) + d_{2i}-d_{1i} > 0 \qquad \text{for } t > t_{12,i}$$

$$u_{2i}(t)-u_{1i}(t) = (b_{2i}-b_{1i})(t-t_c) + d_{2i}-d_{1i} < 0 \qquad \text{for } t < t_{12,i}.$$

where $t_{10,i}$ and $t_{12,i}$ ($t_{10,i} \leq t_{12,i}$) denote the T coordinates of the intersection of utility line $u_{1i}(t)$ with $u_{0i}(t)$ and $u_{2i}(t)$, respectively. Furthermore, it is assumed that the functions $u_{1i}(t)-u_{0i}(t)$ and $u_{2i}(t)-u_{1i}(t)$ are strictly increasing functions of t, implying that the slope parameters $(b_{1i}+b_{0i})$, $(b_{2i}-b_{1i}) > 0$. Using the fact that b_{0i}, $b_{2i} > 0$, this means that the following condition should hold for the utility parameter b_{1i}:

(6) $$b_{2i} > b_{1i} \quad , \text{ if } b_{1i} > 0$$

$$b_{0i} > -b_{1i} \quad , \text{ if } b_{1i} < 0.$$

Optimal Cutting Scores for Quota-Free Selection

In this section, optimal cutting scores are derived for the combined decision problem in case of quota-free selection. That is, we are looking for pairs of cutting scores (x_{ci}, y_{ci}) such that the overall expected utility is a maximum.

Overall Expected Utility

In maximizing overall expected utility, first the expected utility of a random student from the ith subpopulation will be calculated, which, as monotone solutions are looked for, can be written as

$$(7) \qquad E[u_i(T|x_{ci}, y_{ci})] = \int_{-\infty}^{x_{ci}}\int_{-\infty}^{\infty} [b_{0i}(t_c-t)+d_{0i}]w_i(x,t)dtdx +$$

$$\int_{x_{ci}}^{\infty}\int_{-\infty}^{y_{ci}}\int_{-\infty}^{\infty} [b_{1i}(t-t_c)+d_{1i}]\Omega_i(x,y,t)dtdydx +$$

$$\int_{x_{ci}}^{\infty}\int_{y_{ci}}^{\infty}\int_{-\infty}^{\infty} [b_{2i}(t-t_c)+d_{2i}]\Omega_i(x,y,t)dtdydx,$$

where $w_i(x,t)$ is the joint probability function of X and T in subpopulation i. Let $E_i(T|x)$, $q_i(x)$, $k_i(x,y)$, and $E_i(T|x,y)$ denote the regression function of the criterion variable T on X, the probability function of X, the joint probability function of X and Y, and the regression function of the criterion variable T on X and Y in subpopulation i, respectively, then (7) can be written as

$$(8) \qquad E[u_i(T|x_{ci},y_{ci})] = \int_{-\infty}^{\infty}\{b_{0i}[t_c-E_i(T|x)]+d_{0i}\}q_i(x)dx$$

$$+ \int_{x_{ci}}^{\infty}\{[b_{0i}+b_{1i}][E_i(T|x)-t_c]$$

$$+ d_{1i}-d_{0i}\}q_i(x)dx$$

$$+ \int_{x_{ci}}^{\infty}\int_{y_{ci}}^{\infty}\{[b_{2i}-b_{1i}][E_i(T|x,y)-t_c]$$

$$+ d_{2i}-d_{1i}\}k_i(x,y)dydx,$$

Now, the decision procedure is viewed as a series of separate decisions, each of which involves one random student, and it follows that the overall expected utility is a weighted average of the expected utilities for the individual populations. Thus, overall expected utility of the combined decision problem is:

$$(9) \qquad E[u(T|x_{ci},y_{ci})] = \sum_{i=1}^{g} p_i E[u_i(T|x_{ci},y_{ci})],$$

where p_i, $\sum_{i=1}^{g} p_i = 1$, is the proportion of students from subpopulation i in the total population of students.

In quota-free selection there is no restriction as to the number of students that can be accepted for the treatment. Therefore, Formula 9 is maximized if the expected utility of a random student is maximized. This is done by maximizing Formula 9 for each subpopulation separately. The maximum of $E_i[u(T|x_{ci},y_{ci})]$ now depends only on the second and third term in the right-hand side of (8), because

the first term is independent of x_{ci} and y_{ci}. Using a result from decision theory (see e.g., Chuang, Chen, & Novick, 1981) stating that for any prior distribution of t, $E[u(T|z)]$ is a nondecreasing function of z if $f(z|t)$ has MLR and $u(t)$ is a nondecreasing function of t, and assuming monotonocity condition (2), it follows from (5) that $E_i[u_{2i}(t)-u_{1i}(t)|x,y] = [b_{2i}-b_{1i}][E_i(T|x,y)-t_c]+d_{2i}-d_{1i}$ is a nondecreasing function in each of its arguments. Since $(b_{2i}-b_{1i}) > 0$, this implies that $\frac{\partial}{\partial x} E_i(T|x,y)$ and $\frac{\partial}{\partial y} E_i(T|x,y) > 0$. Similarly, using (4) instead of (5), it follows that $E_i[u_{1i}(t)-u_{0i}(t)|x] = [b_{1i}+b_{0i}][E_i(T|x)-t_c]+d_{1i}-d_{0i}$ is a nondecreasing function of x, implying that, since $(b_{1i}+b_{0i}) > 0$, $\frac{d}{dx} E_i(T|x) > 0$. Using $q_i(x)$, $k_i(x,y) \geq 0$, it follows now that the sign of the sum of the second and third term changes only once from negative to positive, and, therefore, $E[u_i(T|x_{ci},y_{ci})]$ will reach its maximum for one pair of cutting scores (x_{ci},y_{ci}).

Maximizing Expected Utility for a Random Student

Necessary conditions for the optimal values of the cutting scores, say x'_{ci} and y'_{ci}, optimizing the expected utility for a random student from subpopulation i, $E_i[u(T|x_{ci},y_{ci})]$, can be obtained by differentiating $E_i[u(T|x_{ci},y_{ci})]$ with respect to x_{ci} and y_{ci}, setting the resulting expressions equal to zero, and solving simultaneously for x_{ci} and y_{ci}.

Using the property that for any bivariate distribution $f(x,y)$, it holds that

$$\frac{\partial}{\partial s}\int_x^{\infty}\int_{-\infty}^{s} f(x,y)dydx = -\frac{\partial}{\partial s}\int_x^{\infty}\int_s^{\infty} f(x,y)dydx = \int_x^{\infty} f(x,s)dx.$$

For the derivative of $E_i[u(T|x_{ci},y_{ci})]$ with respect to y_{ci} this results in

$$(10) \qquad \frac{\partial}{\partial y_{ci}} E_i[u(T|x_{ci},y_{ci})] =$$

$$= s_i(y_{ci}) \int_{x_{ci}}^{\infty} \{[b_{1i}-b_{2i}][E_i(T|x,y_{ci})-t_c]+d_{1i}-d_{2i}\}$$

$$z_i(x|y_{ci})dx = 0,$$

where $z_i(x|y_{ci})$ and $s_i(y)$ denote the posterior probability function of X given $Y = y_{ci}$ and the marginal probability function of Y in subpopulation i, respectively. Since $s_i(y) \geq 0$ (the possibility of $s_i(y) = 0$ will be ignored), it follows that (10) can be replaced by

$$(11) \qquad \int_{x_{ci}}^{\infty} \{[b_{2i}-b_{1i}][E_i(T|x,y_{ci})-t_c]+d_{2i}-d_{1i}\}z_i(x|y_{ci})dx = 0.$$

Similarly, differentiating $E_i[u(T|x_{ci},y_{ci})]$ with respect to x_{ci}, using $q_i(x) \geq 0$ (the possibility of $q_i(x) = 0$ will also be ignored), results in

$$(12) \qquad [b_{0i}+b_{1i}][E_i(T|x_{ci})-t_c]+d_{1i}-d_{0i} +$$

$$\int_{y_{ci}}^{\infty} \{[b_{2i}-b_{1i}][E_i(T|x_{ci},y)-t_c]+d_{2i}-d_{1i}\}m_i(y|x_{ci})dy = 0,$$

with $m_i(y|x_{ci})$ being the posterior probability function of Y given $X = x_{ci}$. Now, solving the system of Equations 11 and 12 for x_{ci} and y_{ci}, one obtains the optimal cutting scores x'_{ci} and y'_{ci}.

Linear Regression

For given regression functions and probability density functions, the optimal quota-free decision strategy is represented by the system of Equations 11 and 12. If the monotonicity conditions are not strict or it does not hold that $s_i(y)$ or $q_i(x) > 0$ in the neighborhood of the solution, the optimal decision strategy may not be unique. Throughout this paper it will be assumed that conditions like these are fulfilled.

Since the relations between the test scores and the criterion (true score) in the regression functions are not directly observable, psychometric models are needed to estimate these relations. Possible psychometric models are the linear regression functions $\theta_i+\Gamma_i x$ and $\alpha_i+\beta_i x+\tau_i y$ for $E_i(T|x)$ and $E_i(T|x,y)$, respectively. Since the probability functions of T given X=x, and T given X=x and Y=y in subpopulation i are normal (see e.g., Johnson & Kotz, 1970), it follows that they belong to the exponential family, and, hence, they do possess

the property of MLR and MLR in each of its arguments, respectively (see e.g., Chuang, Chen, & Novick, 1981). In addition to the properties $\frac{d}{dx} E_i(T|x) = \Gamma_i$, $\frac{\partial}{\partial x} E_i(T|x,y) = \beta_i$, and $\frac{\partial}{\partial y} E_i(T|x,y) = \tau_i > 0$ (see e.g., Lord & Novick, 1968), it then follows that the monotonicity conditions are fulfilled. Assuming linear regression, it can be shown from classical test theory that the linear regression of T on X is given by

$$(13) \qquad E_i(T|x) = E_i(Y|x) = \mu_{Y,i} + \rho_i(\sigma_{Y,i}/\sigma_{X,i})(x-\mu_{X,i}),$$

$\mu_{Y,i}$, $\mu_{X,i}$, ρ_i, $\sigma_{Y,i}$, and $\sigma_{X,i}$ being the population means of Y_i and X_i, the population correlation between X_i and Y_i, and the population standard deviations of Y_i and X_i, respectively. From (13), it follows that

$$(14) \qquad \Gamma_i = \rho_i(\sigma_{Y,i}/\sigma_{X,i})$$
$$\theta_i = \mu_{Y,i} - \Gamma_i\mu_{X,i}.$$

Furthermore, using results from classical test theory, it can be shown that the linear regression of T on X and Y can be written as

$$(15) \qquad E_i(T|x,y) =$$
$$\mu_{Y,i} + (\sigma_{Y,i}/\sigma_{X,i})\{(\rho_i-\rho_{YY',i}\rho_i)/(1-\rho_i^2)\}$$
$$(x-\mu_{X,i}) + \{(\rho_{YY',i}-\rho_i\sqrt{\rho_{YY',i}})/(1-\rho_i^2)\}(y-\mu_{Y,i}),$$

$\rho_{YY',i}$ being the reliability coefficient of Y_i. From (15), it follows that

$$(16) \qquad \beta_i = (\sigma_{Y,i}/\sigma_{X,i})\{(\rho_i-\rho_{YY',i}\rho_i)/(1-\rho_i^2)\}$$

$$\tau_i = (\rho_{YY',i}-\rho_i\sqrt{\rho_{YY',i}})/(1-\rho_i^2)$$

$$\alpha_i = -\mu_{X,i}\beta_i + \mu_{Y,i}(1-\tau_i).$$

All quantities appearing in (14) and (16) can be estimated straightforward; thus, estimates of the linear regression functions can be calculated.

Iterative Solution in Case of the Bivariate Normal Model

In order to solve the system of Equations 11 and 12 for x_{ci} and y_{ci}, the decision-maker must specify the joint probability function of X and Y. It is assumed that the variables X and Y have possibly different bivariate normal distributions in each subpopulation. Assuming that the X and Y scores are in their standardized form, this can be written as

$$(17) \qquad k_i(x_N, y_N) =$$

$$[2\pi\sqrt{1-\rho_i^2}]^{-1} \exp[-(x_N^2-2\rho_i x_N y_N+y_N^2)/(2(1-\rho_i^2))].$$

where x_N and y_N denote the standardized scores $(x-\mu_X)/\sigma_X$ and $(y-\mu_Y)/\sigma_Y$ of X and Y, respectively. For the standardized bivariate normal distribution in (17), the conditional distribution of X_N given $Y_N = y_n$ is normal with expected value $\rho_i y_N$ and variance $(1-\rho_i^2)$. Likewise, the distribution of Y_N given $X_N = x_N$ is normal with expected value $\rho_i x_N$ and variance $(1-\rho_i^2)$.

Substituting $\alpha_i + \beta_i x + \tau_i y_{ci}$ and $N(\rho_i y_{N,ci}, 1-\rho_i^2)$ into Equation 11 for $E_i(T|x,y_{ci})$ and $z_i(x_N|y_{N,ci})$, respectively, and using the property that the primitive function of $xe^{-\frac{1}{2}x^2}$ is equal to $-e^{-\frac{1}{2}x^2}$, it follows that Equation 11 will take the form

$$(18) \quad f(x_c, y_{ci}) = \{(b_{2i}-b_{1i})(\alpha_i+\beta_i\mu_{X,i}+\tau_i y_{ci}+\beta_i\sigma_{X,i}\rho_i y_{N,ci}-t_c)+ d_{2i}-d_{1i})\sigma_{X,i}\{1-\Phi[(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}]\}+ (b_{2i}-b_{1i})\beta_i\sigma_{X,i}^2\sqrt{(1-\rho_i^2)} \varphi[(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}] = 0,$$

where $\Phi[.]$ and $\varphi[.]$ denote the standard normal distribution function and the standard normal density, respectively.

Similarly, substituting $\theta_i + \Gamma_i x_{ci}$, $\alpha_i + \beta_i x_{ci} + \tau_i y$, and $N(\rho_i x_{N,ci}, 1-\rho_i^2)$ into Equation 12 for $E_i(T|x_{ci})$, $E_i(T|x_{ci},y)$, and $m_i(y_N|x_{N,ci})$, respectively, results in

(19) $$g(x_{ci},y_{ci}) = \{(b_{0i}+b_{1i})(\theta_i+\Gamma_i x_{ci}-t_c)+d_{1i}-d_{0i}\}+\{(b_{2i}-b_{1i})(\alpha_i+\beta_i x_{ci}+\tau_i\mu_{Y,i}+\tau_i\rho_i\sigma_{Y,i}x_{N,ci}-t_c)+d_{2i}-d_{1i}\}(1-\Phi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}])+(b_{2i}-b_{1i})\tau_i\sigma_{Y,i}\sqrt{(1-\rho_i^2)}\varphi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}] = 0.$$

The system of Equations 18 and 19 cannot be solved analytically for x_{ci} and y_{ci}, but it can be solved iteratively using Newton's method for systems of nonlinear equations (see, e.g., Ortega & Rheinholdt, 1970). Updated estimates $x'_{ci,j+1}$ and $y'_{ci,j+1}$ after iteration j+1 are obtained using the following formulas:

(20) $$x'_{ci,j+1} = x'_{ci,j} - [(f\frac{\partial}{\partial y_{ci}}g - g\frac{\partial}{\partial y_{ci}}f)/J(f,g)]$$

$$y'_{ci,j+1} = y'_{ci,j} - [(g\frac{\partial}{\partial x_{ci}}f - f\frac{\partial}{\partial x_{ci}}g)/J(f,g)],$$

where $J(f,g) = \frac{\partial}{\partial x_{ci}}f\frac{\partial}{\partial y_{ci}}g - \frac{\partial}{\partial x_{ci}}g\frac{\partial}{\partial y_{ci}}f$ represents the Jacobian of the functions $f(x_{ci},y_{ci})$ and $g(x_{ci},y_{ci})$. It is recommended that the cut-off score t_c on the true score scale T is used as a first approximation to x'_{ci} and y'_{ci}.

In order to solve the nonlinear system of Equations 18 and 19 via the iterative procedure given in (20), the partial derivatives of $f(x_{ci},y_{ci})$ and $g(x_{ci},y_{ci})$ are needed. They are given as

(21) $$\frac{\partial}{\partial x_{ci}} f(x_{ci}, y_{ci}) =$$

$$-[\sqrt{(1-\rho_i^2)}]^{-1}\varphi[(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}]$$
$$\{(b_{2i}-b_{1i})(\alpha_i+\beta_i x_{ci}+\tau_i y_{ci}-t_c)+d_{2i}-d_{1i}\}.$$

(22) $$\frac{\partial}{\partial y_{ci}} f(x_{ci}, y_{ci}) =$$

$$(b_{2i}-b_{1i})\sigma_{X,i}/\sigma_{Y,i}\{(\tau_i\sigma_{Y,i}+\beta_i\rho_i\sigma_{X,i})$$
$$(1-\Phi[(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}])+$$
$$\rho_i/[\sqrt{(1-\rho_i^2)}]\varphi[(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}]$$
$$(\alpha_i+\beta_i x_{ci}+\tau_i y_{ci}-t_c)\}+(d_{2i}-d_{1i})\rho_i/[\sqrt{(1-\rho_i^2)}]$$
$$\sigma_{X,i}\varphi[(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}]/\sigma_{Y,i}.$$

(23) $$\frac{\partial}{\partial x_{ci}} g(x_{ci}, y_{ci}) =$$

$$(b_{0i}+b_{1i})\Gamma_i+$$
$$(b_{2i}-b_{1i})/\sigma_{X,i}\{(\beta_i\sigma_{X,i}+\tau_i\sigma_{Y,i}\rho_i)$$
$$(1-\Phi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}])+$$
$$\rho_i/[\sqrt{(1-\rho_i^2)}]\varphi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}]$$
$$(\alpha_i+\beta_i x_{ci}+\tau_i y_{ci}-t_c)\}+(d_{2i}-d_{1i})\rho_i/[\sqrt{(1-\rho_i^2)}]$$
$$\varphi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}]/\sigma_{X,i}.$$

(24) $$\frac{\partial}{\partial y_{ci}} g(x_{ci}, y_{ci}) =$$

$$-[\sqrt{(1-\rho_i^2)}]^{-1}\varphi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}]$$
$$\{(b_{2i}-b_{1i})(\alpha_i+\beta_i x_{ci}+\tau_i y_{ci}-t_c)+d_{2i}-d_{1i}\}/\sigma_{Y,i}.$$

An interesting special case of the combined linear utility function arises when $d_{0i} = d_{1i} = d_{2i}$. In that case, all utility parameters d_{ji} ($j = 0,1,2$) vanish from Equations 18 and 19. In other words, if the amount of constant utility, d_{ji}, for each action is equal, then there is no need to choose values for d_{ji} in determining the optimal cutting scores x'_{ci} and y'_{ci}.

It can be shown (Lord & Novick, 1968, sect. 17.2) that the standard normal distributions appearing in (18)–(24) are almost interchangeable with logistic functions for a scale parameter equal to 1.7. The logistic model will be preferred in the iterative procedure because it is easier to work with mathematically than the standard normal model. Using this approximation, we may rewrite the standard normal distributions as follows:

$$\Phi[(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}] =$$
$$[1+\exp\{-1.7(x_{N,ci}-\rho_i y_{N,ci})/\sqrt{(1-\rho_i^2)}\}]^{-1},$$

$$\Phi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}] =$$
$$[1+\exp\{-1.7(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}\}]^{-1}.$$

The iterative procedure is implemented in a computer program called NEWTON.

Special Solutions

The optimal solution for the separate mastery and selection decision can both easily be derived from Equations 11 and 12 by imposing certain restrictions on x_{ci} and y_{ci}, respectively.

First, putting $x_{ci} = -\infty$ in Equation 11, that is accepting all students, and using $\int_{-\infty}^{\infty} z_i(x|y_{ci})dx = 1$, Equation 11 will take the form

$$(25) \qquad [b_{2i}-b_{1i}][E_i(T|y_{ci})-t_c]+d_{2i}-d_{1i} = 0.$$

Putting both utility lines $u_{1i}(t)$ and $u_{2i}(t)$ in Formula 3 equal to each other, it appears that the t coordinate of the intersection, $t_{12,i}$, is equal to $t_c+(d_{1i}-d_{2i})/(b_{2i}-b_{1i})$, which implies (25) can be replaced by $E_i(T|y_{ci}) = t_{12,i}$. This solution yields the same optimal cutting score y'_{ci} as the one given by van der Linden and Mellenbergh (1977) for the separate mastery decision. Analogous to the combined decision problem, a psychometric model is needed to specify the regression function $E_i(T|y_{ci})$. For this purpose, the classical test model with linear regression (Lord & Novick, 1968, p.55) will be assumed, which is known as Kelley's regression line:

$$(26) \qquad E_i(T|y_{ci}) = \rho_{YY',i}y_{ci} + (1-\rho_{YY',i})\mu_{Y,i}.$$

Substituting (26) into (25), gives

(27) $y'_{ci} = \mu_{Y,i} + \{t_c - \mu_{Y,i} + {}'_{1i} - d_{2i})/(b_{2i} - b_{1i})\}/\rho_{YY',i}$.

Analogous to the derivation of the optimal separate mastery decision from (11), the optimal separate selection decision can be derived from (12) by putting $y_{ci} = -\infty$, that is, advancing all accepted students. Doing so, and using Formula 3, it follows that

(28) $E_i(T|x_{ci}) = t_{02,i} = t_c + (d_{0i} - d_{2i})/(b_{0i} + b_{2i})$,

where $t_{02,i}$ denotes the t coordinate of the intersection of utility line $u_{0i}(t)$ with $u_{2i}(t)$. Also, this optimal solution is the same as the one reached by Mellenbergh and van der Linden (1981) for the separate selection decision. Adopting the linear regression function from classical test theory again, it follows from (28) that the optimal cutting score x'_{ci} can be expressed in closed form as

(29) $x'_{ci} = \mu_{X,i} + \{t_c - \mu_{X,i} + (d_{0i} - d_{2i})/(b_{0i} + b_{2i})\}/\rho_{XX',i}$,

where $\rho_{XX',i}$ denotes the reliability coefficient of X_i.

An interesting case arises when $d_{1i} = d_{2i}$ in Equation 27. Whenever this occurs, both utility lines $u_{1i}(t)$ and $u_{2i}(t)$ intersect at t_c, and thus, Equation 27 takes the form

(30) $y'_{ci} = \mu_{Y,i} + (t_c - \mu_{Y,i})/\rho_{YY',i}$.

In other words, if the amounts of constant utility associated with the actions retaining and advancing a student in the separate mastery decision are equal or there are no constant utilities at all, then there is no need to assess the parameters b_{1i} and b_{2i}. In the numerical example, this situation will be further elaborated.

Similarly, all utility function parameters vanish from (29) whenever $d_{0i} = d_{2i}$; thus, (29) can be further simplified to

$$(31) \qquad x'_{ci} = \mu_{X,i} + (t_c - \mu_{X,i})/\rho_{XX',i}.$$

Optimal Cutting Scores for Quota-Restricted Selection

In quota-restricted selection only a fixed number of students can be accepted for the instructional program. The selection constraint can be expressed as

$$(32) \qquad p_0 = \sum_{i=1}^{g} p_i [\mathrm{Prob}(X \geq x_{ci})] = \sum_{i=1}^{g} p_i [\int_{x_{ci}}^{\infty} q_i(x)dx],$$

where $0 < p_0 < \sum_{i=1}^{g} p_i = 1$ represents the fixed proportion of all students that can be accepted.

The values of x'_{ci} and y'_{ci} optimizing the overall expected utility of the combined decision problem, can now be found by introducing the selection constraint into the

function to be optimized (Equation 9) through a Lagrange multiplier λ:

$$(33) \qquad L(x_{ci}, y_{ci}, \lambda) = \sum_{i=1}^{g} p_i E[u_i(T|x_{ci}, y_{ci})]$$

$$+ \lambda\{p_0 - \sum_{i=1}^{g} p_i [\int_{x_{ci}}^{\infty} q_i(x)dx]\},$$

where λ is a constant.

Differentiating $L(x_{ci}, y_{ci}, \lambda)$ with respect to y_{ci} and x_{ci}, setting the resulting expressions equal to zero, and using $p_i > 0$, yields

$$(34) \qquad \frac{\partial}{\partial y_{ci}} E[u_i(T|x_{ci}, y_{ci})] = 0,$$

$$(35) \qquad \frac{\partial}{\partial x_{ci}} E[u_i(T|x_{ci}, y_{ci})] + \lambda q_i(x_{ci}) = 0.$$

As can be noticed from Equation 10, the solution to Equation 34 is the same as the solution for the case of quota-free selection given by Equation 18.

The first term in the left-hand side of Equation 35 represents the derivative of the expected utility of a random student with respect to x_{ci} in the unrestricted situation (Equation 12). Substituting this partial derivative into Equation 35, and using $q_i(x) \geq 0$, it follows that

$$(36) \qquad [b_{0i}+b_{1i}][E_i(T|x_{ci})-t_c]+d_{1i}-d_{0i}-\lambda + \int_{y_{ci}}^{\infty} \{[b_{2i}-b_{1i}][E_i(T|x_{ci},y)-t_c]+d_{2i}-d_{1i}\}m_i(y|x_{ci})dy = 0.$$

Inserting the expressions for the linear regression functions and $N(\rho_i x_{N,ci}, 1-\rho_i^2)$ for $m_i(y_N|x_{N,ci})$, and integrating Equation 36, results in

$$(37) \qquad h(x_{ci},y_{ci}) = \{(b_{0i}+b_{1i})(\theta_i+\Gamma_i x_{ci}-t_c)+d_{1i}-d_{0i}-\lambda\}+\{(b_{2i}-b_{1i})(\alpha_i+\beta_i x_{ci}+\tau_i\mu_{Y,i}+\tau_i\rho_i\sigma_{Y,i}x_{N,ci}-t_c)+d_{2i}-d_{1i}\}(1-\Phi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}])+(b_{2i}-b_{1i})\tau_i\sigma_{Y,i}\sqrt{(1-\rho_i^2)}\varphi[(y_{N,ci}-\rho_i x_{N,ci})/\sqrt{(1-\rho_i^2)}] = 0.$$

Since it has been assumed that the joint distribution of X and Y is a possibly different bivariate normal distribution in each population, it follows that $q_i(x)$ is a normal distribution with mean $\mu_{X,i}$ and variance $\sigma_{X,i}^2$ (see, e.g., Johnson & Kotz, 1972). Hence, the restriction of Equation 32 can be written as

$$(38) \qquad v(x_{c1},x_{c2},\ldots,x_{cg}) = \sum_{i=1}^{g} p_i\{1-\Phi[x_{N,ci}]\} - p_0 = 0.$$

Now, the solution for the quota-restricted selection model is found by solving the system of Equations 18, 37, and 38 for

the (2g + 1) unknown parameters x_{ci}, y_{ci}, and λ. Note that with quota-restricted selection, unlike quota-free selection, the optimal cutting scores x'_{ci} and y'_{ci} ($i = 1,\ldots,g$) are dependent upon each other.

In order to apply Newton's iterative method to solve the given system of nonlinear Equations, the partial derivatives are required again. From (37) and (19), it can easily be verified that $\frac{\partial}{\partial x_{ci}} h(x_{ci},y_{ci}) = \frac{\partial}{\partial x_{ci}} g(x_{ci},y_{ci})$, $\frac{\partial}{\partial y_{ci}} h(x_{ci},y_{ci}) = \frac{\partial}{\partial y_{ci}} g(x_{ci},y_{ci})$, and $\frac{\partial}{\partial \lambda} h(x_{ci},y_{ci}) = -1$. The derivatives of $f(x_{ci},y_{ci})$ and $g(x_{ci},y_{ci})$ with respect to x_{ci} and y_{ci} were given in Equations 21 until 24, respectively. Furthermore, it follows from (38) that

$$(39) \qquad \frac{\partial}{\partial x_{ci}} v(x_{c1},x_{c2},\ldots,x_{cg}) = (-p_i/\sigma_{X,i})\varphi[x_{N,ci}].$$

Analogous to the quota-free selection model, as can easily be seen from Equations 18, 37, and 38, no values for the utility function parameters d_{ji} ($j = 0,1,2$) have to be specified when the amount of constant utility for each action is equal.

A computer program called LAGRANGE has been written to obtain an optimal decision rule. In the program, the optimal solution (x'_{ci},y'_{ci}) of the quota-free selection model can be used as a first approximation in the iterative procedure. A numerical example illustrating the procedure is given in the next section.

A Numerical Example

The linear utility model for optimal selection-mastery decisions was applied to a sample of 43 freshmen in medicine. Both the selection and mastery tests consisted of 17 free-response items on elementary medical knowledge with test scores ranging from 0-100. The treatment consisted of a computer-aided instructional (CAI) program.

Due to prior knowledge, the total population of 43 students could be distinguished with respect to elementary medical knowledge into a disadvantaged and an advantaged subpopulation of 27 and 16 students, respectively. Let the disadvantaged and advantaged population be referred to as subpopulation 1 and 2, respectively. The normal models assumed for the distributions X_i and Y_i showed a satisfactory fit to the test data for a Kolmogorov-Smirnov goodness-of-fit test with p-values of 0.869, 0.934, 0.867, and 0.993 for X_1, Y_1, X_2, and Y_2, respectively. The differences between the theoretical and observed cumulative distribution functions were 0.0686, 0.1035, 0.1495, and 0.1067, respectively.

The teachers of the course considered students as having mastered the subject matter if their test scores were at least 55. Therefore, t_c was fixed at 55.

The means, standard deviations and correlation between X and Y, were computed for each subpopulation using the maximum likelihood estimates. Furthermore, the reliabilities of the test scores were computed as coefficient α (Cronbach, 1951)

for each subpopulation. The results of these computations are shown in Table 1.

Insert Table 1 about here

First, the quota-free situation is considered. Since population 2 was considered more advantaged than 1, it should hold that $b_{01} > b_{02}$, $b_{21} > b_{22}$, $b_{11} > b_{12}$ for b_{11}, $b_{12} > 0$, and $b_{11} < b_{12}$ for b_{11}, $b_{12} < 0$. Besides these conditions for the utility parameters in Formula 3, condition (6) should hold for the utility parameters b_{ji} ($j = 0,1,2$; $i = 1,2$). Substituting the values of the statistics of Table 1 into Equations 14 and 16, and using the computer program NEWTON, Equations 18 and 19 were then solved for x_{ci} and y_{ci} ($i = 1,2$) with t_c as starting values. To illustrate the dependence of the results on the utility structure, optimal cutting scores were computed for 10 different values of the utility parameters b_{ji} and d_{ji} ($j = 0,1,2$; $i = 1,2$). The absolute values of b_{ji} and d_{ji} for utility function 1 until 5 were the same as the absolute values for utility function 6 until 10. However, the sign of b_{1i} was taken negative in the last five runs, taking into account the fact that the sign of b_{1i} could not be specified beforehand. The results are reported in Table 2.

Insert Table 2 about here

As can be seen from Table 2, the consequence of increasing the parameters b_{0i} and b_{2i} ($i = 1,2$) is a decrease of the cutting scores. Furthermore, inspection of Table 2 shows that a decrease of the amount of constant utility, d_{ji} ($j = 0,1,2$; $i = 1,2$), implies that the cutting scores have to be raised.

Finally, it can be seen that for the simultaneous solution the optimal selection scores are lower for the disadvantaged than for the advantaged group. Conversely, the optimal mastery scores are lower for the advantaged group. This is an important conclusion, which can be argued by the fact that the disadvantaged students should be accepted sooner. On the other hand, however, they should stay longer in the treatment to be sure that they have mastered the instructional unit sufficiently so that they may proceed with the next unit.

Using Equations 27 and 29, the optimal cutting scores were also computed for the separate mastery and selection decisions. Since no constant amounts of utility were assumed for utility functions 2 and 7, Equations 30 and 31 were used to compute the optimal cutting scores for these two utility specifications. The results are also reported in Table 2. Table 2 shows that the optimal selection as well as the optimal mastery scores in the separate model have been raised compared to those in the combined model.

To give an impression of the gain in overall expected utility by using the simultaneous approach, the ratio of overall expected utility for the separate and simultaneous solution has been calculated. The overall expected utilities have been calculated by substituting the optimal cutting scores from Table 2 into Equation 8. The third term in the right-hand side of (8) has been computed by using numerical integration methods, while the first and second term have been integrated analytically yielding respectively

$$b_{0i}(t_c-\theta_i-\Gamma_i\mu_{X,i})+d_{0i},$$

$$\{(b_{0i}+b_{1i})(\theta_i+\Gamma_i\mu_{X,i}-t_c)+d_{1i}-d_{0i}\}$$
$$\{1-\Phi[x_{N,ci}]\}+\sigma_{X,i}\Gamma_i(b_{0i}+b_{1i})\varphi[x_{N,ci}].$$

The computer program UTILITY calculates the overall expected utility; the results are displayed in Table 2.

Finally, the quota-restricted situation is considered. The proportions p_i ($i = 1,2$) of the student population belonging to each subpopulation were set estimated as n_i/n, where n represents the total sample size and n_i represents the number of students in the sample in subpopulation i. The proportion p_0 of the total student population that could be accepted for the instructional treatment was arbitrarily set equal to 0.333. Using the computer program LAGRANGE, the system of Equations 18, 37, and 38 were then solved for x_{ci} and y_{ci} ($i = 1,2$) with the optimal solution of the quota-free

situation as starting values. The optimal cutting scores were computed again for 10 different values of b_{ji} and d_{ji}; the results are shown in Table 3.

Insert Table 3 about here

From Table 3 it can be seen that the optimal selection scores x'_{c1} and x'_{c2} in the quota-restricted model have to be raised compared to those in the quota-free model. This result is in accordance with our expectations, because fewer students can be accepted in the restricted situation. Also, it follows from Table 3 that the optimal mastery scores y'_{c1} and y'_{c2} are higher than for the quota-free model.

Finally, it can be noticed that, analogous to the quota-free situation, the optimal selection and mastery scores are lower and higher for the disadvantaged than for the advantaged group, respectively.

Discussion

In this paper an approach to instructional decision making for combinations of elementary decisions has been presented. A useful application of simultaneous decision making can be found in the area of instructional decision making in ISS's. As an example, two elementary decisions (viz. a selection and a mastery decision) were combined into a simple ISS to

indicate how by simultaneous optimization of such networks, optimal rules for proceeding students through ISS's can be designed within a Bayesian decision-theoretic framework. The utility structure adopted in this combined decision problem was a linear utility function.

Further examination of the "best" way to represent more complicated instructional networks of combinations of elementary decisions seems a valuable line of research. Such instructional networks can also be formalized with the aid of Bayesian statistics and optimal rules for these simultaneous optimization problems can be found.

Also, more efforts are needed to examine other more realistic forms utility functions might take in certain educational applications. For example, the normal ogive utility function (Novick & Lindley, 1978) which takes utility to be a nonlinear function of the true score. Such a utility structure might be adopted, for instance, when it is reasonable to assume a leveling-off effect.

Finally, an interesting line of research seems to be to design "optimal CAI-networks" using the method of simulation. On the basis of the derived theoretical optimal decision rules, then, for a simulated distribution of students, it can be determined in which instructional network the shortest time is spent to reach a certain final mastery level.

Author's Note

Portions of this paper were presented at the European Meeting of the Psychometric Society, 1987, Enschede, The Netherlands. The author is indebted to Wim J. van der Linden for his valuable comments on earlier drafts of the paper and to Jan Gulmans for providing the data for the empirical example.

The computer programs NEWTON, LAGRANGE and UTILITY are available on request from the author.

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. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

DeGroot, M.H. (1970). Optimal statistical decisions. New York: McGraw-Hill.

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. (1967). Statistical considerations of mastery scores. Psychometrika, 41, 65-79.

Huynh, H. (1976). Two simple cases of mastery scores based on the beta-binomial model. Psychometrika, 41, 65-78.

Johnson, N.L., & Kotz, S. (1970). Distributions in statistics: Continuous univariate distributions. Boston: Houghton Mifflin.

Karlin, S., & Rubin, H. (1956). Distributions possessing a monotone likelihood ratio. Journal of the American Statistical Association, 1, 637-643.

Keeney, D., & Raiffa, H. (1976). Decisions with multiple objectives: Preferences and value trade-offs. New York: John Wiley and Sons.

Lehmann, E.L. (1959). Testing statistical hypothesis. New York: John Wiley & Sons.

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

Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Reading: M.A.: Addison-Wesley.

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

Novick, M.R., & Petersen, N.S. (1976). Towards equalizing educational and employment opportunity. Journal of Educational Measurement, 13, 77-89.

Ortega, J.M., & Rheinholdt, W.C. (1970). Iterative solution to nonlinear equations in several variables. Academic Press.

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

Petersen, N.S. (1976). An expected utility model for "optimal" selection. Journal of Educational Statistics, 4, 333-358.

Petersen, N.S., & Novick, M.R. (1976). An evaluation of some models for culture-fair selection. Journal of Educational Measurement, 13, 3-31.

van der Gaag, N.L. (1987). Applications of empirical linear utility functions to mastery decisions. Paper presented at the European Meeting of the Psychometric Society, University of Twente, The Netherlands, June 1987.

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 treatments with and without mastery scores. Psychometrika, 46, 257-274.

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

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. (1988). Applications of decision theory to test-based decision making. In R.K. Hambleton & J.N. Zaal (Eds.), New developments in testing: Theory and applications. Amsterdam, The Netherlands: North-Holland.

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. (1988). The use of decision theory in the Minnesota Adaptive Instructional System. Journal of Computer-Based Instruction, Vol. 15, No. 2, 65-71.

Vos, H.J., & van der Linden, W.J. (1987). Designing optimal rules for instructional decision making in CAI systems. In J. Moonen & T. Plomp (Eds.), Developments in educational software and courseware. Oxford: Pergamon Press.

Table 1

Statistics Selection and Mastery Tests (X and Y)

Statistic	Disadvantaged		Advantaged	
	X	Y	X	Y
Mean	50.875	62.626	56.453	67.148
Standard Deviation	10.981	11.645	11.674	13.344
Reliability	0.762	0.775	0.783	0.813
Correlation	0.8564		0.8685	

Table 2
Optimal Cutting Scores Quota-Free Selection

No.	Utility Specifications: Disadv.		Utility Specifications: Adv.		Cutting Scores (x'_{ci}, Y'_{ci}): Simultaneous Disadv.	Simultaneous Adv.	Separate Disadv.	Separate Adv.	$E[u_i(T \mid x_{ci}, Y_{ci})]_{sep.} / E[u_i(T \mid x_{ci}, Y_{ci})]_{simult.}$ *100 Disadv.	Adv.
(1)	$b_{01}=3.5$	$d_{01}=-2$	$b_{02}=3$	$d_{02}=-2$						
	$b_{11}=2$	$d_{11}=-3$	$b_{12}=1.5$	$d_{12}=-3$	$x_{c1}=42.72$	$x_{c2}=44.32$	$x_{c1}=56.69$	$x_{c2}=55.06$		
	$b_{21}=3$	$d_{21}=-4$	$b_{22}=2.5$	$d_{22}=-4$	$Y_{c1}=41.99$	$Y_{c2}=39.91$	$Y_{c1}=54.07$	$Y_{c2}=53.44$	27.16	68.90
(2)	$b_{01}=3.5$	$d_{01}=0$	$b_{02}=3$	$d_{02}=0$						
	$b_{11}=2$	$d_{11}=0$	$b_{12}=1.5$	$d_{12}=0$	$x_{c1}=42.38$	$x_{c2}=43.95$	$x_{c1}=56.29$	$x_{c2}=54.60$		
	$b_{21}=3$	$d_{21}=0$	$b_{22}=2.5$	$d_{22}=0$	$Y_{c1}=40.53$	$Y_{c2}=38.41$	$Y_{c1}=52.78$	$Y_{c2}=52.21$	35.61	73.23
(3)	$b_{01}=3.5$	$d_{01}=-3$	$b_{02}=3$	$d_{02}=-3$						
	$b_{11}=2$	$d_{11}=-7$	$b_{12}=1.5$	$d_{12}=-7$	$x_{c1}=43.68$	$x_{c2}=45.51$	$x_{c1}=58.10$	$x_{c2}=56.69$		
	$b_{21}=3$	$d_{21}=-12$	$b_{22}=2.5$	$d_{22}=-12$	$Y_{c1}=54.15$	$Y_{c2}=50.54$	$Y_{c1}=59.24$	$Y_{c2}=58.36$	5.57	55.25
(4)	$b_{01}=5$	$d_{01}=-2$	$b_{02}=3.5$	$d_{02}=-2$						
	$b_{11}=3$	$d_{11}=-3$	$b_{12}=2$	$d_{12}=-3$	$x_{c1}=42.50$	$x_{c2}=43.93$	$x_{c1}=56.53$	$x_{c2}=54.90$		
	$b_{21}=6$	$d_{21}=-4$	$b_{22}=5$	$d_{22}=-4$	$Y_{c1}=40.99$	$Y_{c2}=38.90$	$Y_{c1}=53.21$	$Y_{c2}=52.62$	39.58	75.49
(5)	$b_{01}=15$	$d_{01}=-2$	$b_{02}=14$	$d_{02}=-2$						
	$b_{11}=3$	$d_{11}=-3$	$b_{12}=2$	$d_{12}=-3$	$x_{c1}=42.38$	$x_{c2}=43.85$	$x_{c1}=56.39$	$x_{c2}=54.71$		
	$b_{21}=10$	$d_{21}=-4$	$b_{22}=9$	$d_{22}=-4$	$Y_{c1}=40.73$	$Y_{c2}=38.62$	$Y_{c1}=52.96$	$Y_{c2}=52.39$	26.47	66.88
(6)	$b_{01}=3.5$	$d_{01}=-2$	$b_{02}=3$	$d_{02}=-2$						
	$b_{11}=-2$	$d_{11}=-3$	$b_{12}=-1.5$	$d_{12}=-3$	$x_{c1}=42.31$	$x_{c2}=43.47$	$x_{c1}=56.69$	$x_{c2}=55.06$		
	$b_{21}=3$	$d_{21}=-4$	$b_{22}=2.5$	$d_{22}=-4$	$Y_{c1}=40.80$	$Y_{c2}=38.84$	$Y_{c1}=53.04$	$Y_{c2}=52.52$	28.39	62.45
(7)	$b_{01}=3.5$	$d_{01}=0$	$b_{02}=3$	$d_{02}=0$						
	$b_{11}=-2$	$d_{11}=0$	$b_{12}=-1.5$	$d_{12}=0$	$x_{c1}=41.95$	$x_{c2}=43.06$	$x_{c1}=56.29$	$x_{c2}=54.60$		
	$b_{21}=3$	$d_{21}=0$	$b_{22}=2.5$	$d_{22}=0$	$Y_{c1}=40.52$	$Y_{c2}=38.51$	$Y_{c1}=52.78$	$Y_{c2}=52.21$	36.34	67.35
(8)	$b_{01}=3.5$	$d_{01}=-3$	$b_{02}=3$	$d_{02}=-3$						
	$b_{11}=-2$	$d_{11}=-7$	$b_{12}=-1.5$	$d_{12}=-7$	$x_{c1}=43.57$	$x_{c2}=44.87$	$x_{c1}=58.10$	$x_{c2}=56.69$		
	$b_{21}=3$	$d_{21}=-12$	$b_{22}=2.5$	$d_{22}=-12$	$Y_{c1}=41.97$	$Y_{c2}=40.22$	$Y_{c1}=54.07$	$Y_{c2}=53.75$	8.86	47.51
(9)	$b_{01}=5$	$d_{01}=-2$	$b_{02}=3.5$	$d_{02}=-2$						
	$b_{11}=-3$	$d_{11}=-3$	$b_{12}=-2$	$d_{12}=-3$	$x_{c1}=42.13$	$x_{c2}=43.16$	$x_{c1}=56.53$	$x_{c2}=54.90$		
	$b_{21}=6$	$d_{21}=-4$	$b_{22}=5$	$d_{22}=-4$	$Y_{c1}=40.68$	$Y_{c2}=38.73$	$Y_{c1}=52.92$	$Y_{c2}=52.39$	40.39	71.28
(10)	$b_{01}=15$	$d_{01}=-2$	$b_{02}=14$	$d_{02}=-2$						
	$b_{11}=-3$	$d_{11}=-3$	$b_{12}=-2$	$d_{12}=-3$	$x_{c1}=42.22$	$x_{c2}=42.79$	$x_{c1}=56.39$	$x_{c2}=54.71$		
	$b_{21}=10$	$d_{21}=-4$	$b_{22}=9$	$d_{22}=-4$	$Y_{c1}=40.63$	$Y_{c2}=38.79$	$Y_{c1}=52.88$	$Y_{c2}=52.32$	26.83	64.75

Table 3
Optimal Cutting Scores Quota-Restricted Selection

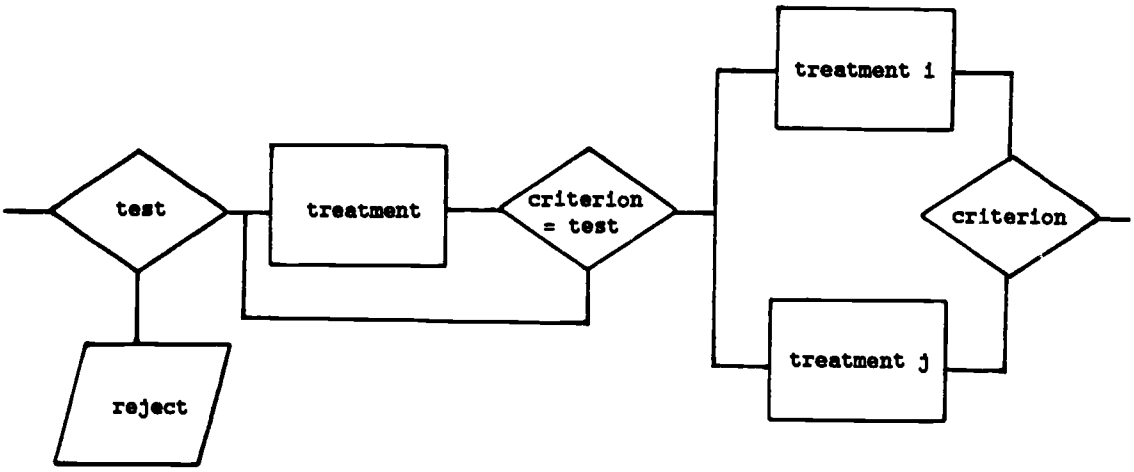
No.	Utility Specifications Disadv.		Adv.		Cutting Scores ($x'_{ci} \cdot Y'_{ci}$) Simultaneous Disadv.	Adv.
(1)	$b_{01}=3.5$	$d_{01}=-2$	$b_{02}=3$	$d_{02}=-2$		
	$b_{11}=2$	$d_{11}=-3$	$b_{12}=1.5$	$d_{12}=-3$	$x_{c1}=56.48$	$x_{c2}=59.34$
	$b_{21}=3$	$d_{21}=-4$	$b_{22}=2.5$	$d_{22}=-4$	$Y_{c1}=48.54$	$Y_{c2}=47.21$
(2)	$b_{01}=3.5$	$d_{01}=0$	$b_{02}=3$	$d_{02}=0$		
	$b_{11}=2$	$d_{11}=0$	$b_{12}=1.5$	$d_{12}=0$	$x_{c1}=56.48$	$x_{c2}=59.34$
	$b_{21}=3$	$d_{21}=0$	$b_{22}=2.5$	$d_{22}=0$	$Y_{c1}=47.76$	$Y_{c2}=46.36$
(3)	$b_{01}=3.5$	$d_{01}=-3$	$b_{02}=3$	$d_{02}=-3$		
	$b_{11}=2$	$d_{11}=-7$	$b_{12}=1.5$	$d_{12}=-7$	$x_{c1}=56.48$	$x_{c2}=59.34$
	$b_{21}=3$	$d_{21}=-12$	$b_{22}=2.5$	$d_{22}=-12$	$Y_{c1}=51.53$	$Y_{c2}=50.37$
(4)	$b_{01}=5$	$d_{01}=-2$	$b_{02}=3.5$	$d_{02}=-2$		
	$b_{11}=3$	$d_{11}=-3$	$b_{12}=2$	$d_{12}=-3$	$x_{c1}=55.95$	$x_{c2}=60.19$
	$b_{21}=6$	$d_{21}=-4$	$b_{22}=5$	$d_{22}=-4$	$Y_{c1}=47.73$	$Y_{c2}=47.09$
(5)	$b_{01}=15$	$d_{01}=-2$	$b_{02}=14$	$d_{02}=-2$		
	$b_{11}=3$	$d_{11}=-3$	$b_{12}=2$	$d_{12}=-3$	$x_{c1}=56.98$	$x_{c2}=58.58$
	$b_{21}=10$	$d_{21}=-4$	$b_{22}=9$	$d_{22}=-4$	$Y_{c1}=48.15$	$Y_{c2}=46.07$
(6)	$b_{01}=3.5$	$d_{01}=-2$	$b_{02}=3$	$d_{02}=-2$		
	$b_{11}=-2$	$d_{11}=-3$	$b_{12}=-1.5$	$d_{12}=-3$	$x_{c1}=56.52$	$x_{c2}=59.27$
	$b_{21}=3$	$d_{21}=-4$	$b_{22}=2.5$	$d_{22}=-4$	$Y_{c1}=47.94$	$Y_{c2}=46.54$
(7)	$b_{01}=3.5$	$d_{01}=0$	$b_{02}=3$	$d_{02}=0$		
	$b_{11}=-2$	$d_{11}=0$	$b_{12}=-1.5$	$d_{12}=0$	$x_{c1}=56.52$	$x_{c2}=59.27$
	$b_{21}=3$	$d_{21}=0$	$b_{22}=2.5$	$d_{22}=0$	$Y_{c1}=47.78$	$Y_{c2}=46.54$
(8)	$b_{01}=3.5$	$d_{01}=-3$	$b_{02}=3$	$d_{02}=-3$		
	$b_{11}=-2$	$d_{11}=-7$	$b_{12}=-1.5$	$d_{12}=-7$	$x_{c1}=56.52$	$x_{c2}=59.27$
	$b_{21}=3$	$d_{21}=-12$	$b_{22}=2.5$	$d_{22}=-12$	$Y_{c1}=48.57$	$Y_{c2}=47.37$
(9)	$b_{01}=5$	$d_{01}=-2$	$b_{02}=3.5$	$d_{02}=-2$		
	$b_{11}=-3$	$d_{11}=-3$	$b_{12}=-2$	$d_{12}=-3$	$x_{c1}=55.97$	$x_{c2}=60.15$
	$b_{21}=6$	$d_{21}=-4$	$b_{22}=5$	$d_{22}=-4$	$Y_{c1}=47.57$	$Y_{c2}=46.91$
(10)	$b_{01}=15$	$d_{01}=-2$	$b_{02}=14$	$d_{02}=-2$		
	$b_{11}=-3$	$d_{11}=-3$	$b_{12}=-2$	$d_{12}=-3$	$x_{c1}=57.01$	$x_{c2}=58.53$
	$b_{21}=10$	$d_{21}=-4$	$b_{22}=9$	$d_{22}=-4$	$Y_{c1}=48.11$	$Y_{c2}=46.01$

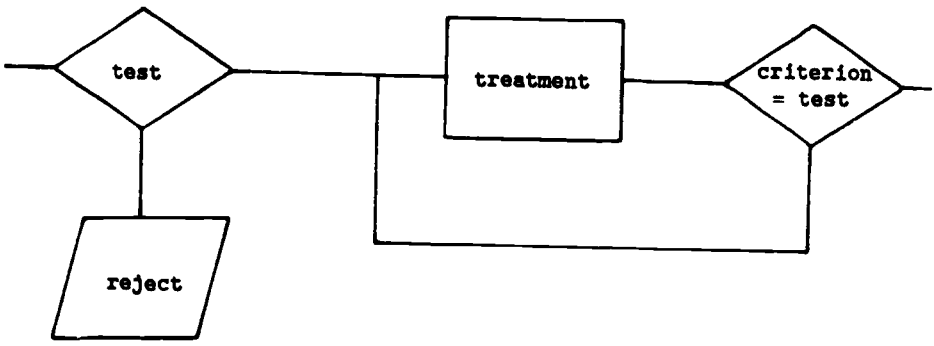
Figure Captions

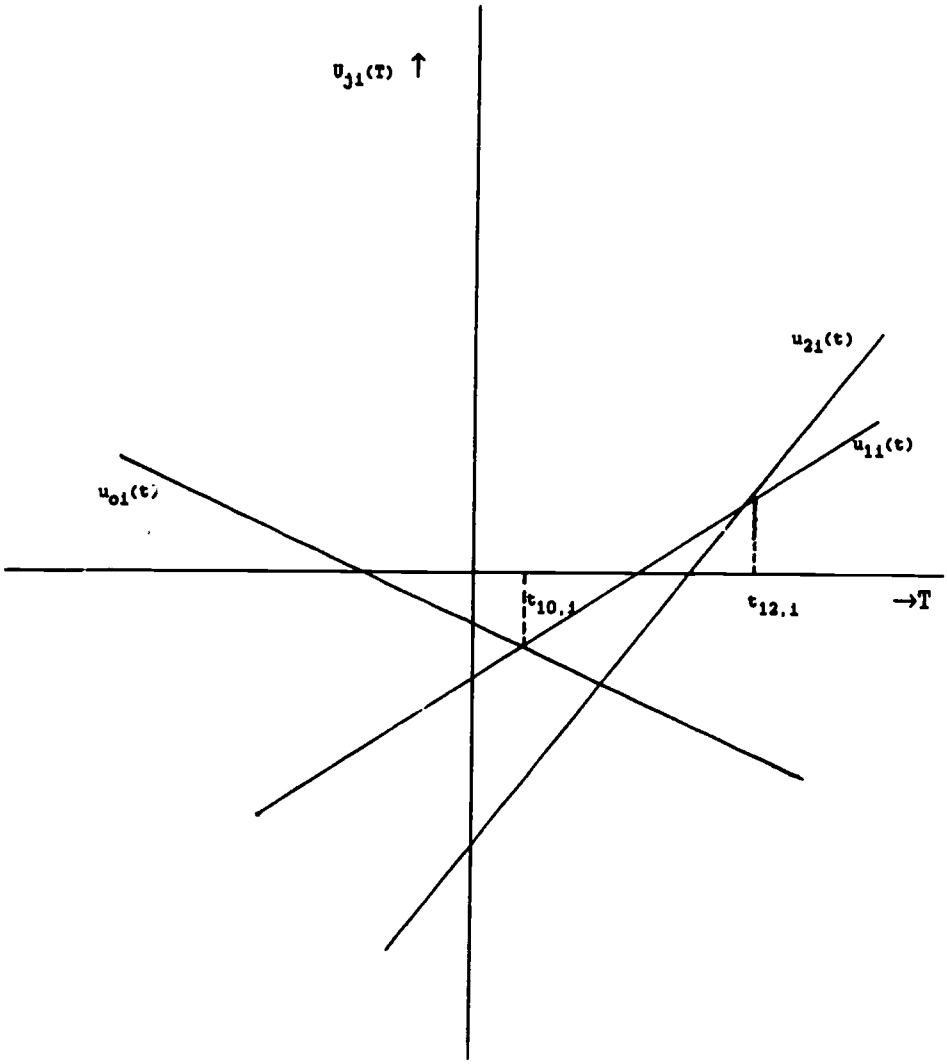
Figure 1. Example of an individualized study system

Figure 2. A system of one selection and one mastery decision

Figure 3. Example of a linear utility function for a selection–mastery decision ($b_{1i} > 0$)


test
treatment
criterion
= test
treatment i
criterion
treatment j
reject


test
treatment
criterion
= test
reject


$v_{j1}(T)$ ↑
$u_{21}(t)$
$u_{11}(t)$
$u_{01}(t)$
$t_{10,1}$
$t_{12,1}$
→T

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

RR-87-1 R. Engelen, *Semiparametric estimation in the Rasch model*

RR-87-2 W.J. van der Linden (Ed.), *IRT-based test construction*

RR-87-3 R. Engelen, P. Thommassen, & W. Vervaat, *Ignatov's theorem: A new and short proof*

RR-87-4 E. van der Burg, & J. de Leeuw, *Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis*

RR-87-5 H. Kelderman, *Estimating a quasi-loglinear models for the Rasch table if the number of items is large*

RR-87-6 R. Engelen, *A review of different estimation procedures in the Rasch model*

RR-87-7 D.L. Knol & J.M.F. ten Berge, *Least-squares approximation of an improper by a proper correlation matrix using a semi-infinite convex program*

RR-87-8 E. van der Burg & J. de Leeuw, *Nonlinear canonical correlation analysis with k sets of variables*

RR-87-9 W.J. van der Linden, *Applications of decision theory to test-based decision making*

RR-87-10 W.J. van der Linden & E. Boekkooi-Timminga, *A maximin model for test design with practical constraints*

RR-88-1 E. van der Burg & J. de Leeuw, *Nonlinear redundancy analysis*

RR-88-2 W.J. van der Linden & J.J. Adema, *Algorithmic test design using classical item parameters*

RR-88-3 E. Boekkooi-Timminga, *A cluster-based method for test construction*

RR-88-4 J.J. Adema, *A note on solving large-scale zero-one programming problems*

RR-88-5 W.J. van der Linden, *Optimizing incomplete sample designs for item response model parameters*

RR-88-6 H.J. Vos, *The use of decision theory in the Minnesota Adaptive Instructional System*

RR-88-7 J.H.A.N. Rikers, *Towards an authoring system for item construction*

RR-88-8 R.J.H. Engelen, W.J. van der Linden, & S.J. Oosterloo, *Item information in the Rasch model*

RR-88-9 W.J. van der Linden & T.J.H.M. Eggen, *The Rasch model as a model for paired comparisons with an individual tie parameter*

RR-88-10 H. Kelderman & G. Macready, *Loglinear-latent-class models for detecting item bias*

RR-88-11 D.L. Knol & M.P.F. Berger, *Empirical comparison between factor analysis and item response models*

RR-88-12 E. van der Burg & G. Dijksterhuis, *Nonlinear canonical correlation analysis of multiway data*

RR-88-13 J. Kogut, *Asymptotic distribution of an IRT person fit index*

RR-88-14 J.J. Adema, *The construction of two-stage tests*

RR-88-15 H.J. Vos, *Simultaneous optimization of decisions using a linear utility function*

Research Reports can be obtained at cost from Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

tment of

EDUCATION

A publication by
the Department of Education
of the University of Twente

ede

nds