A Decision Theory Model for Course Placement

Wim J. van der Linden


Stable URL:
http://links.jstor.org/sici?sici=1076-9986%281998%2923%3A1%3C18%3AADTMFC%3E2.0.CO%3B2-V


Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/aera.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
A Decision Theory Model for Course Placement

Wim J. van der Linden
University of Twente

Keywords: decision theory, mastery decisions, monotone decision rules, placement decisions, test theory

The problem of how to place students in a sequence of hierarchically related courses is addressed from a decision theory point of view. Based on a minimal set of assumptions, it is shown that optimal mastery rules for the courses are always monotone and a nonincreasing function of the scores on the placement test. On the other hand, placement rules are not generally monotone but have a form depending on the specific shape of the probability distributions and utility functions in force. The results are further explored for a class of linear utility functions.

A course placement problem is met when students with different levels of preparation enter an educational program offering classes in the same domain at several entrance levels. The typical solution of the problem is to administer a placement test to the students to decide at which level they should start. The purpose of this article is to demonstrate how (Bayesian) decision theory can be used to optimize such placement decisions.

For simplicity, the case of two courses is considered where the first course is a lower-level remedial course and the second course is the standard course. It is assumed that a placement test is administered to decide whether the students have to take the remedial course or can go straight to the standard course. In addition, it is assumed that the remedial course ends with a mastery test to decide whether those students who take the course have learned its content and can be admitted to the standard course. The mastery test could be an alternate form of the placement test; however, there is no formal psychometric requirement that the two tests be homogeneous to each other or that each test be unidimensional. A flowchart of the problem displaying the temporal relationship between the tests and courses is given in Figure 1.

To derive optimal decision rules for the placement and mastery decisions, (Bayesian) decision theory is a natural framework. Examples of a decision-theoretic treatment of the mastery problem are found in Hambleton and Novick (1973), Huynh (1976, 1977, 1982), Huynh and Perney (1979), van der Linden (1980, 1990), and van der Linden and Mellenbergh (1977). The placement problem has been addressed by Sawyer (1996) and van der Linden (1981). The novelty in this article is that optimal rules for the placement and mastery decision problems are derived simultaneously. The advantages of a simultaneous
approach have been spelled out in Vos (1994, 1995). For the present problem, they will be discussed later in this article.

**Notation**

Sampling of students from a population \( P \) is assumed. The observed scores on the placement test, administered before the first course, will be denoted by a random variable \( X \). Likewise, the observed score on the mastery test is a random variable \( Y \), for which it holds that \( Y = T + E \), with \( T \) being the classical test theory true score on the test and \( E \) the error variable. For ease of exposition, the variables are assumed to be continuous.

For each possible value of the observed placement score, the decision to assign the students either to Course 1 or to Course 2 has to be made. A placement rule can therefore be denoted either by the set of \( x \) values \( A_1 = \{x: \text{Course 1}\} \) or its complement \( A_2 = \{x: \text{Course 2}\} \). Likewise, a mastery decision rule can be represented by the set \( B_{\text{Fail}}(x) = \{y: x, \text{Fail}\} \) or its complement \( B_{\text{Pass}}(x) = \{y: x, \text{Pass}\} \). Note that the mastery rule is allowed to depend on the score on the placement test, \( X = x \). This type of mastery rule allows for the fact that, for example, a low observed score on the mastery test more likely points to low mastery than to measurement error for students with lower placement scores than for students with higher scores. The idea to formulate mastery decision rules dependent on the score on a previous placement test is also consistent with the idea developed in Bayesian statistics that using collateral information generally improves the quality of decision making. That is, there is information in both the placement and the mastery test about competence as assessed by \( t \) to enter the second course. On the other hand, as will be shown below, constraining optimal sets \( B_{\text{Fail}}(x) \) or \( B_{\text{Pass}}(x) \) to be the same set of \( y \) values for each possible value \( x \) requires additional conditions on the test score distributions.

**Distributional Assumptions**

It is assumed that if all students in \( P \) were assigned to Course 1, then the trivariate distribution of their scores \((X, Y, T)\) could be represented by a density function \( f_1(x, y, t) \). If, on the other hand, all students were assigned directly to
Course 2, then the distribution is assumed to be represented by a density function $f_2(x, y, t)$. This experiment differs from the one of an operational placement system, where students assigned directly to Course 2 would not have test scores $Y$. However, the two functions can easily be estimated from an experiment in which one sample of students is assigned to Course 1 and then to Course 2 and another sample only takes the mastery tests and then proceeds with Course 2. (Statistics needed to estimate the density functions are not reviewed here. However, it is observed that, although the decision model is Bayesian, it is not necessary to estimate the functions in a Bayesian fashion, that is, using priors for their parameters.)

For future reference, it is understood that the conditional distributions of $T$ given $(X = x, Y = y)$, $T$ given $X = x$, $Y$ given $X = x$, and $(Y, T)$ given $X = x$ are denoted by $p_i(t|x, y)$, $q_i(t|x)$, $h_i(y|x)$, and $r_i(y, t|x)$, $i = 1, 2$, respectively. The marginal distribution of $X$ is denoted by $g_i(x)$. As usual, cumulative distribution functions will be denoted by capitals.

The following assumptions about $f_i(x, y, t)$ seem plausible:

$$g_i(x) = g(x), i = 1, 2;$$

(1)

$$P_i(t|x, y)$$ is decreasing in $x$ and $y$ for all values of $t$, $i = 1, 2$; (2)

$$H_i(y|x)$$ is decreasing in $x$ for all values of $y$, $i = 1, 2$; (3)

$$Q_i(t|x)$$ is decreasing in $x$ for all values of $t$, $i = 1, 2$. (4)

The first assumption is realized in the experiment considered here because the placement test is administered to all examinees before any decision is made. The last three assumptions require the conditional distributions of the observed and true mastery test scores to be stochastically increasing in their conditioning variables. These conditions can be expected to be met if the placement and mastery tests are well designed, that is, if they are constructed such that high scores on the placement test tend to coincide with high observed and true scores on the mastery test. It can be proved that conditions as in (4) follow from the assumptions in (2) and (3) (van der Linden & Vos, 1996). For a more general treatment of the properties of stochastic order in multivariate distributions of test scores, see van der Linden (in press).

Several other assumptions could be specified to define a course placement problem with well designed tests and courses. For example, for the conditional cumulative distribution functions of $(Y, T)$ given $X = x$, $R_i(y, t|x)$, the following assumption models the effectiveness of the remedial instruction in Course 1:

$$R_i(y, t|x)$$ increases in $i$ for all values of $x, y$, and $t$. (5)

This condition states that the probability of high observed and true scores on the mastery test given a score on the placement test is larger for students assigned to Course 1 than for those assigned directly to Course 2. In other words, no matter his or her score on the placement test, a student is expected to learn something while following Course 1. However, this and other obvious properties are not needed to derive the results presented in this article.
From a distributional point of view, the decisions should maximize the stochastic largeness of the probability of Course 1 mastery, that is, make $T$ stochastically as large as possible prior to entering Course 2. However, at the same time the costs involved in following Course 1 should be avoided. The notion of costs leads to the specification of a utility structure for the problem.

**Utility Structure**

The usual approach is to define the utilities involved in mastery decisions as a function of the true score underlying the mastery test. More specifically, since mastery decisions are made only for students assigned to Course 1, it seems logical also to view the utilities involved in the outcomes of this decision as a function of the true mastery score. As for Course 2, the position is taken that the decision to assign students directly to this course can be viewed as a *mastery decision in advance* based on the placement test as a predictor of the students’ mastery scores after Course 1. It follows that the true mastery score can be used as the common criterion to measure the utilities involved in all decisions in the simultaneous placement-mastery problem addressed here. More specifically, the following utility functions are defined:

\[
\begin{align*}
\text{Assignment to Course 1 (} x \in A_1 \text{) and Fail (} y \in B_{\text{Fail}}(x)\text{):} & \quad u_{\text{Fail}}(t); \\
\text{Assignment to Course 1 (} x \in A_1 \text{) and Pass (} y \in B_{\text{Pass}}(x)\text{):} & \quad u_{\text{Pass}}(t); \\
\text{Assignment to Course 2 (} x \in A_2 \text{):} & \quad u_2(t).
\end{align*}
\]

As for the shape of the utility functions, it is assumed that

\[
\begin{align*}
u_{\text{Fail}}(t) \text{ is monotonically decreasing in } t; & \quad (7) \\
u_{\text{Pass}}(t) \text{ and } u_2(t) \text{ are monotonically increasing in } t. & \quad (8)
\end{align*}
\]

In (7) it is assumed that the loss society incurs if a student with a high level of mastery fails the test is larger than for a student with a lower level of mastery. Therefore, utility is assumed to be decreasing in true score $t$. The reverse is assumed to hold for the other two possible decisions; both for the actual mastery decision and the mastery decision in advance, the outcome has larger utility to society, the higher the true mastery level of the student. However, the actual mastery decision involves the additional costs of teaching Course 1, which are avoided for the mastery decision in advance. For traditional group-based instruction, these costs are assumed to be a constant independent of $t$. Hence,

\[
u_2(t) \geq u_{\text{Pass}}(t) \text{ for all values of } t. \quad (9)
\]

As yet, no parametric form will be assumed for the utility functions. The intention is first to explore further the general structure of problem and the optimal decision rules. Then, a parametric form for the utility functions will be introduced. In some of the examples later in the article, the difference $u_2(t) - u_{\text{Pass}}(t)$ is chosen to be a constant, whereas in others it is not a constant.
Formalization of the Problem

The condition on the utility functions in (9) suggests that if utilities were the only concern, the best decision would be to always assign the students to Course 2. This action represents the largest utility since the constant costs involved in teaching the first course are missed. On the other hand, if the decisions had to be based only on the test score distributions, the best decision would always be to assign students to Course 1, because, as the condition in (5) shows, no matter the score of the students on the placement test, the probability of having high observed or true mastery of the subject matter taught in this course increases when it is attended.

Obviously, a criterion is needed to solve this dilemma between utility and probability. In this article, the criterion of maximum expected utility, typical of the Bayesian decision model, is used. The utility function for the simultaneous problem is defined as

\[
U(A_1, B_{\text{Fail}}(x)) = \begin{cases} 
  u_{\text{Fail}}(t), & \text{if } x \in A_1 \text{ and } y \in B_1(x); \\
  u_{\text{Pass}}(t), & \text{if } x \in A_2 \text{ and } y \in B_2(x); \\
  u_2(t), & \text{if } x \in A_2. 
\end{cases}
\]  

(10)

The expected utility associated with the problem can be written as

\[
E[U(A_1, B_{\text{Fail}}(X))] = \int_{A_1} \int_{B_{\text{Fail}}(x)} \int_R u_{\text{Fail}}(t) f_1(x, y, t) \, dt \, dy 
+ \int_{B_{\text{Pass}}(x)} \int_{R} \int_R u_{\text{Pass}}(t) f_1(x, y, t) \, dt \, dy \, dx 
+ \int_{A_1} \int_{R} \int_R u_2(t) f_2(x, y, t) \, dt \, dy \, dx.
\]

(11)

It is assumed that the integrals in (11) converge. Formally, the problem is now to find placement and mastery rules, \( A_1 \) and \( B_{\text{Fail}}(x) \), for which the expression in (11) is maximal.

Moving to conditional densities, taking expectations, and noting that \( A_1 \cup A_2 = R \), \( B_{\text{Fail}}(x) \cup B_{\text{Pass}}(x) = R \), and \( E[Z] = E[E[Z \mid W]] \), (11) can be written as

\[
E[U(A_1, B_2(X))] = \int_{A_1} \left\{ E_1[u_{\text{Fail}}(T) \mid x] - E_2[u_2(T) \mid x] \right\} 
+ \int_{B_{\text{Pass}}(x)} \left\{ E_1[u_{\text{Pass}}(T) - u_{\text{Fail}}(T) \mid x, y] h_1(y(x) \mid dy \right\} g(x) \, dx 
+ E_2[u_2(T)],
\]

(12)

where the subscripts on the expectation signs index the distributions with respect to which the expectations are taken. The problem is now to find sets \( A_1 \) and \( B_{\text{Pass}}(x) \) for which (12) is maximal. Note that the last term of (12) is a constant independent of \( A_1 \) and \( B_{\text{Pass}}(x) \) and may be ignored. Since the sample spaces of
$X$ and $Y$ are not constrained in any way, the optimal sets can be found by (a) maximizing the second term in (12) over all possible sets $B_{\text{pass}}(x)$ for each value of $x$ and (b) maximizing the sum of the first two terms over all possible sets of $x$ values, $A_1$.

**General Solution to the Problem**

The general results in this paper are presented in the form of two propositions. The first proposition states that, under the conditions given above, the optimal mastery rule, $B_{\text{pass}}(x)$, is defined by a cutoff score on the mastery test score $Y$ which is a decreasing function of the placement test score $X = x$. The second proposition states that the form of the optimal placement rule, $A_1$, does not necessarily have the form of a cutoff score on the placement test.

**Proposition 1.** For conditions (1)–(3) on the conditional distributions of $T$ given $(X = x, Y = y)$ and $Y$ given $X = x$, and conditions (7) and (8) on the utility functions, the optimal sets $B_{\text{pass}}(x)$ are intervals $[y_c(x), \infty)$, where $y_c(x)$ is a decreasing function of $x$.

**Proof.** From (7) and (8) it follows that $u_{\text{pass}}(t) - u_{\text{fail}}(t)$ is increasing in $t$, whereas the condition in (2) states that the distribution of $T$ given $(X = x, Y = y)$ is stochastic increasing in the two conditioning variables. From a well-known theorem for such distributions (Lehmann, 1986, p. 116) it follows that $E_1[u_{\text{pass}}(T) - u_{\text{fail}}(T)|x, y]$ increases in $x$ and $y$. Let $\tau(x, y)$ denote this expectation, and consider the relation $\{(x, y)\mid \tau(x, y) = 0\}$. This relation defines a function $y = y_c(x)$ which is decreasing in $y$ (van der Linden & Vos, 1996, Lemma 4). Hence, for each $x$ there exists a value $y_c(x)$ such that $\tau(x, y)$ is nonnegative for $y \geq y_c(x)$ but negative for $y < y_c(x)$. Since the integral of $\tau(x, y)$ over $h_1(y|x)$ is maximal if it is taken over all values of $x$ for which the integrand is nonnegative, it follows that the optimal sets $B_{\text{pass}}(x)$ are the intervals $(y_c(x), \infty]$.

**Proposition 2.** Under the conditions stated above it does not hold generally that the optimal set $A_1$ is of the form $(-\infty, x_c]$ for some cutoff score $x_c$.

**Proof.** From (4) and (7) it follows that $E_1[u_{\text{fail}}(T)|x]$ decreases in $x$, whereas (4) and (8) imply that $E_2[u_{\text{pass}}(T)|x]$ increases in $x$. Therefore, the integrand in the first term in (12) decreases in $x$. Following an argument in van der Linden and Vos (1996), it is now proved that the integrand in the second term in (12) is nondecreasing in $x$. Substituting $B_{\text{pass}}(x) = [y_c(x), \infty)$ into (12), since $\tau(x, y)$ is nonnegative for $y \geq y_c(x)$ and $y_c(x)$ is decreasing in $x$, it holds for any $x_2 > x_1$ that

\[
\int_{y_c(x_2)}^{y_c(x_1)} \tau(x_2, y) h_1(y|x_2) dy - \int_{y_c(x_1)}^{y_c(x_2)} \tau(x_1, y) h_1(y|x_1) dy >
\]

\[
\int_{y_c(x_1)}^{y_c(x_2)} \tau(x_2, y) h_1(y|x_2) dy - \int_{y_c(x_2)}^{y_c(x_1)} \tau(x_1, y) h_1(y|x_1) dy >
\]
\[
\int_{y(x)}^{\infty} \tau(x, y) \left[ h_1(y|x_2) - h_1(y|x_1) \right] dy = \\
\int_{-\infty}^{\infty} \varphi(y) \left[ h_1(y|x_2) - h_1(y|x_1) \right] dy,
\]

(13)

where \( \varphi(y) \equiv I_{[y(x), \infty)}(y) \tau(x, y) \), and \( I_{[y(x), \infty)}(y) \) is an indicator function which takes the value 1 if \( y \in [y(x), \infty) \) and the value 0 otherwise. By definition, \( \varphi(y) \) is a nondecreasing function of \( y \), and from (3) it follows that the second term in (12) is nondecreasing indeed. Now it depends on the rate of change of the first and second terms in (12), and hence on the specific form of the utility functions and score distributions whether the integrand in the integral over \( x \) in (12) is a monotonically decreasing function of \( x \). Therefore, the optimal set \( A_1 \) does not generally take a monotone form.

Proposition 2 implies that for some utility functions and distributions, an optimal set \( A_1 \) may consist of two or more disjoint intervals of placement test scores. Generally, it does not hold that if a student with a certain achievement level is assigned to Course 1, the same should apply for all students with lower levels. The result goes against the prevailing practice of using placement rules in the form of a single cutoff score on a test. The explanation of this “anomaly” is the dilemma between utility and probability referred to in the previous section. Utilities and probabilities are different quantities. If for a given achievement level it is known that the expected increase in mastery due to attending Course 1 offsets the increase in utility involved in directly attending Course 2, no inferences as to other levels are possible without knowing the specific forms of the utility functions and probability distributions.

**Calculation of Optimal Rules**

From the proof of Proposition 1 it follows that the optimal mastery function is the decreasing function \( y = y_c(x) \) defined by

\[
E_1[u_{\text{Pass}}(T) - u_{\text{Fail}}(T)|x, y] = 0.
\]

(14)

Note that the left-hand side of (14) is the only part of (12) dependent on \( y \). Substituting \( y_c(x) \) into (12), maximal expected utility is obtained if \( A_1 \) is taken to be the set of \( x \) values for which the integrand of the outer integral is nonnegative. That is, the optimal set \( A_1 \) is given by

\[
A_1 = \left\{ x: E_1[u_{\text{Pass}}(T)|x] - E_2[u_2(T)|x] \right. \\
+ \left. \int_{y(x)}^{\infty} E_1[u_{\text{Pass}}(T) - u_{\text{Fail}}(T)|x, y]h_1(y|x)dy \geq 0 \right\}
\]

(15)

The optimal set in (15) can be found using a suitable method of numerical integration. Because in practice observed test scores are always discrete, a simple approach is to substitute all possible test scores into the expression in
(15) to find the subset for which it is nonnegative. However, to be able to do so, the regression functions in (15) must be estimated. The following section shows how to deal with the problem if linear utilities and regression functions can be assumed to hold.

**Linear Utility Structure**

For the utility functions in (6), the following linear structure is adopted:

\[
\begin{align*}
    u_{\text{Fail}}(t) &= b_{\text{Fail}}(t_c - t) + a_{\text{Fail}}, \quad b_{\text{Fail}} > 0; \\
    u_{\text{Pass}}(t) &= b_{\text{Pass}}(t - t_c) + a_{\text{Pass}}, \quad b_{\text{Pass}} > 0; \\
    u_2(t) &= b_2(t - t_c) + a_2, \quad b_2 > 0,
\end{align*}
\]

where \( t_c \) is assumed to be a threshold value on the true mastery scale which separates true masters from true nonmasters. If such a threshold does not exist, the above functions should be reparameterized absorbing \( t_c \) into the intercept parameter. Note that the slope parameters are required to be larger than zero to meet the conditions in (7)–(8). Also, generally \( a_{\text{Fail}}, a_{\text{Pass}}, a_2 < 0 \) because these utility constants represent the constant parts of the costs involved in teaching Course 1 and/or Course 2. As noted earlier, this article views the decision to assign students to Course 2 as a mastery decision in advance. Hence, utility functions (17) and (18) have the same parameter structure but may have different values for their parameters to deal with the specifics of the application. An example of the use of these utility functions will be given later in this article. Techniques to elicit utility functions for placement decisions have been addressed in Sawyer (1994).

Substituting utility functions (16)–(18) into the expression for the expected utility in (12) and using the result in Proposition 1 yields

\[
E[U(A_1, y_c(x))] = \int_{A_1} \left\{ (b_{\text{Fail}} + b_{\text{Pass}})t_c + (a_{\text{Fail}} - a_{\text{Pass}}) - b_{\text{Fail}}E_1(T|x) - b_{\text{Pass}}E_2(T|x) \\
+ \int_{y_c(x)} \left[ -(b_{\text{Fail}} + b_{\text{Pass}})t_c + (a_{\text{Pass}} - a_{\text{Fail}}) + (b_{\text{Fail}} + b_{\text{Pass}})E_1(T|x, y)h_1(y|x) \right] dy \right\} g(x) dx.
\]

From the general solution in (14) it follows that the optimal mastery function is defined by

\[
-(b_{\text{Fail}} + b_{\text{Pass}})t_c + (a_{\text{Pass}} - a_{\text{Fail}}) + (b_{\text{Fail}} + b_{\text{Pass}})E_1(T|x, y) = 0,
\]

whereas (15) implies that the following set of \( x \) values is optimal:

\[
A_1 = \{ x: (b_{\text{Fail}} + b_2)t_c + (a_{\text{Fail}} - a_2) - b_{\text{Fail}}E_1(T|x) - b_2E_2(T|x) \\
+ \int_{y_c(x)} \left[ -(b_{\text{Fail}} + b_{\text{Pass}})t_c + (a_{\text{Pass}} - a_{\text{Fail}}) + (b_{\text{Fail}} + b_{\text{Pass}})E_1(T|x, y)h_1(y|x) \right] dy \geq 0 \}.
\]
Linear Regression

If linear regression of $T$ on $X$ and $Y$ can be assumed, that is, if it holds that

$$E_1(T|x, y) = \alpha_{TXY1} + \beta_{TXY1}x + \gamma_{TXY1}y,$$

(22)

then the optimal mastery function is

$$y_c(x) = \left[ t_c - (a_{\text{Pass}} - a_{\text{Fail}})/(b_{\text{Fail}} + b_{\text{Pass}}) - \alpha_{TXY1}\gamma_{TXY1}^{-1} - (\beta_{TXY1}\gamma_{TXY1}^{-1})x \right].$$

(23)

Note that this function is linear in $x$ with a slope parameter equal to $-\beta_{TXY1}\gamma_{TXY1}^{-1}$. Since families of conditional distributions with the property in (2) have nonnegative covariances between the random variable and the conditioning variables (van der Linden, in press), it follows that $\beta_{TXY}$ and $\gamma_{TXY}$ are nonnegative, and thus that the function in (23) is nondecreasing indeed.

As classical test theory shows,

$$E_i(T|x) = E_i(Y|x)$$

(Lord & Novick, 1968, sect. 2.7). Thus,

$$E_i(T|x) = \alpha_{YX_i} + \beta_{YX_i}x.$$

(25)

Substituting (22) and (25) into (21), we find that the optimal set $A_1$ is

$$A_1 = \{x: K_1 - (b_{\text{Fail}}\beta_{YX1} + b_2\beta_{YX2})x + \int_{y_c(x)} [K_2 + (b_{\text{Fail}} + b_{\text{Pass}})(\beta_{TXY1}x + \gamma_{TXY1}y)h_i(y|x)]dy \geq 0\},$$

(26)

with

$$K_1 = b_{\text{Fail}}(t_c - \alpha_{YX1}) + b_2(t_c - \alpha_{YX2}) + (a_{\text{Fail}} - a_2),$$

(27)

$$K_2 = (b_{\text{Fail}} + b_{\text{Pass}})(\alpha_{TXY1} - t_c) + (a_{\text{Pass}} - a_{\text{Fail}}),$$

(28)

and where $y_c(x)$ is given by (23).

Classical test theory shows that the regression parameters in (23) and (26)–(28) are simple functions of the mean of $Y$, the variances of $X$ and $Y$, the correlation between $X$ and $Y$, and the reliability of $Y$. If $X$ and $Y$ are standardized, the relevant expressions are $\alpha_{YX1} = 0, \beta_{YX1} = \rho_{XY}, \alpha_{TXY1} = 0, \beta_{TXY1} = \rho_{XY}(1-\rho_{YI})/(1-\rho_{XY}^2), \text{ and } \gamma_{TXY1} = (\rho_{YY} - \rho_{XY}^2)/(1-\rho_{XY}^2)$ (Lord & Novick, 1968, chaps. 2 and 12).

Substituting the above results into (23), we obtain the following form for the optimal mastery function:

$$y_c(x) = (t_c - (a_{\text{Pass}} - a_{\text{Fail}})/(b_{\text{Fail}} + b_{\text{Pass}})) \left( \frac{1-\rho_{XY1}^2}{\rho_{YY1} - \rho_{XY1}^2} \right) - \left( \frac{\rho_{XY1}(1-\rho_{YY1})}{\rho_{YY1} - \rho_{XY1}^2} \right)x.$$
The same substitution into (26) gives

\[ A_1 = \{x: K_1 - K_3 x + \int_{y(x)} (K_2 + K_4 x + K_5 y) h_1(y|x) dy \geq 0\}, \quad (30) \]

with

\[ K_1 = (b_{\text{Fail}} + b_2) t_c + a_{\text{Fail}} - a_2, \quad (31) \]
\[ K_2 = -(b_{\text{Fail}} + b_{\text{Pass}}) t_c + a_{\text{Pass}} - a_{\text{Fail}}, \quad (32) \]
\[ K_3 = b_{\text{Fail}} \rho_{XY} + b_2 \rho_{XY2}, \quad (33) \]
\[ K_4 = (b_{\text{Fail}} + b_{\text{Pass}}) \rho_{XY1} (1 - \rho_{YY})/(1 - \rho_{XY1}^2), \quad (34) \]
\[ K_5 = (b_{\text{Fail}} + b_{\text{Pass}}) (\rho_{YY} - \rho_{XY1}^2)/(1 - \rho_{XY1}^2). \quad (35) \]

To use the above decision rules, the conditional density functions \( h_1(y|x) \) have to be estimated from the data. Fitting a well-chosen parametric form for these densities is an obvious approach to the estimation problem. If \( h_1(y|x) \) is chosen to be the normal density function, then from an appropriate change of variable in the integral, and from the fact that for the c.d.f., \( \Phi(u) \), it holds that \( |\Phi(u) - [1 + \exp(1.7u)]^{-1}| < 0.01 \) uniformly in \( u \) (Lord & Novick, 1968, sect. 17.2), it follows that for all practical purposes (30) can be replaced by

\[ A_1 \approx \left\{ x: K_1 - K_3 x + (K_2 + K_6) x \left[ 1 + \exp \left( 1.7 \frac{y_c(x) - \rho_{XY1} x}{1 - \rho_{XY1}^2} \right) \right]^{-1} + K_7 \exp \left( -\frac{1}{2} \left( \frac{y_c(x) - \rho_{XY1} x}{1 - \rho_{XY1}^2} \right)^2 \right) \geq 0 \right\}, \quad (36) \]

with

\[ K_6 = K_4 + \rho_{XY1} K_5, \quad (37) \]
\[ K_7 = (2\pi)^{-1/2}(1 - \rho_{XY1}^2) K_5. \quad (38) \]

Obviously, the second and fourth terms in the left-hand side of the inequality in (36) are monotonically decreasing in \( x \). Since the slope of \( y_c(x) \) in (29) can never be positive, the second factor of the third term is also monotonically decreasing in \( x \). However, the first factor of this term is increasing in \( x \). Thus it depends on the relative acceleration of these two factors whether or not the left-hand side of the inequality is monotonically decreasing in \( x \) and allows for an optimal placement rule of the form \( A_1 = (-\infty, x_c] \).

**Numerical Examples**

A few numerical examples were calculated to shed some light on the behavior of the optimal placement rule and mastery function in (36) and (29) for several values of their parameters. The following six cases were addressed:
Case 1: Standard parameter values. Correlations between cognitive predictor variables and college grades typically range from 0.50 to 0.60 (Etaugh, Etaugh, & Hurd, 1972; see also Schoenfeldt & Brush, 1975). A choice of 0.55 for \( \rho_{XYi} \) therefore seems realistic. Etaugh et al. found an average reliability for the grade-point average over a large variety of programs equal to 0.689. Since the average number of courses per program was 5.23, the reliability of a single grade is generally lower. On the other hand, tests for remedial courses typically are more carefully constructed than tests for regular courses. Based on these considerations, a choice of \( \rho_{YY1} = 0.60 \) was made. The cutoff score \( t_c \) was set equal to 0.5, which is a value 0.5/\( \sqrt{0.60} \) standard deviations larger than the average score on the mastery test. The values of the utility parameters \( a_{\text{Fail}} \), \( a_{\text{Pass}} \), and \( a_2 \) in (16)–(18) had to be selected relatively to the scale of \( t \). It was assumed that the costs involved in additionally teaching Course 1 were equal for the students who failed and those who passed the mastery test, that is, \( a_{\text{Fail}} = a_{\text{Pass}} \). On the other hand, sending students directly to Course 2 does involve only the costs of teaching Course 2. Therefore, utility parameter \( a_2 \) was set larger than the common value of \( a_{\text{Fail}} \) and \( a_{\text{Pass}} \). As for the values of the slope parameters \( b_{\text{Fail}} \) and \( b_{\text{Pass}} \) in (16)–(17), failing the mastery test for high values of \( t \) was assumed to be less serious than passing the test for low values. This choice is justified if the program is structured such that a retest is offered to those students who fail the test, but those who pass are allowed to continue without a further check of their true mastery levels. Hence, \( b_{\text{Fail}} < b_{\text{Pass}} \). As the placement decision is considered as a mastery decision in advance in this paper, the value of its slope parameter was set equal to value for the actual mastery decision, that is, \( b_{\text{Pass}} = b_2 \). In summary, the following set of standard values for the parameters in (29) and (36) was assumed:

\[
\begin{align*}
\rho_{XY1} &= \rho_{XY2} = 0.55, \\
\rho_{YY1} &= 0.60, \\
t_c &= 0.50, \\
a_{\text{Fail}} &= a_{\text{Pass}} = -1.0, \\
a_2 &= -0.50, \\
b_{\text{Fail}} &= 1.0, \\
b_{\text{Pass}} &= b_2 = 1.2.
\end{align*}
\]

Case 2: Higher costs of teaching Course 1. It was assumed that the constant costs involved in the teaching of Course 1 were higher than the standard value in the previous case. Therefore, the common value of the parameters \( a_{\text{Fail}} \) and \( a_{\text{Pass}} \) was lowered to −1.50. All other parameters were kept at their standard values.

Case 3: Utility less sensitive to true mastery level. To simulate the case of the utilities associated with the various decision outcomes being less sensitive to differences between the true mastery scores, the values of the three slope
parameters were halved: $b_{\text{Fail}} = 0.5$ and $b_{\text{Pass}} = b_2 = 0.6$. All other parameters were kept at their standard values.

**Case 4: Adaptive instruction.** If Course 1 is an individualized course in which the intensity of instruction is adapted to the current achievement level of the students, the savings in instructional costs involved in assigning students directly to Course 2 decrease as a function of $t$. To simulate this case, the values of $b_{\text{Fail}}$ and $b_{\text{Pass}}$ were raised to 1.8 and 2.0, respectively, but all other parameters were kept at their standard values.

**Case 5: Higher validity of placement test.** Next, it was supposed that a better placement test was available than the test studied in Case 1. Therefore, $\rho_{XY_1}$ was raised to 0.70. The height of this value must be judged relatively to the reliability of the mastery scores, which was kept at its standard value (as were the values of all other parameters).

**Case 6: Higher reliability of mastery test.** To assess the effect of a higher reliability of the mastery scores on the optimal placement and mastery rules, $\rho_{YY_1}$ was raised to 0.75, while all other parameters were left at their standard values.

For the standard parameter values (Case 1), the plot of the optimal mastery function $y_c(x)$ as a function of $x$ is given in the first part of Figure 2. It appears that the function decreases by an amount of 0.74 for an increase of the placement score $X$ equal to one standard deviation. The optimal placement decision is defined by the set $x$ values, $A_1$, for which the expression in (36) is positive. In the second part of Figure 2, $\lambda(x)$ is defined as the expression in (36) plotted as a function of $x$. To find the roots of $\lambda(x) = 0$, Newton’s method, as implemented in Mathematica’s FindRoot command (Wolfram, 1993) was used. Table 1 lists the results for the present and the following cases. Only one root was found, at $x = 0.592$. Thus, a monotone placement rule exists for this case.

The plots for all other cases are given in Figures 3–7. For higher costs of teaching Course 1 (Case 2), the root of $\lambda(x)$ moved down to $x = 0.094$, which, as expected, implies that fewer students are assigned to Course 1. For utilities less sensitive to changes in the true mastery level $t$ (Case 3), $\lambda(x)$ was again found to cross the horizontal axis only once, this time at $x = 0.092$. Making the instruction in Course 1 adaptive to the mastery levels of the students (Case 4) resulted in the existence of two cutoff scores on the placement, one at $x = 0.836$ and

**TABLE 1**

<table>
<thead>
<tr>
<th>Case</th>
<th>Roots of $\lambda(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.592</td>
</tr>
<tr>
<td>2</td>
<td>0.094</td>
</tr>
<tr>
<td>3</td>
<td>0.092</td>
</tr>
<tr>
<td>4</td>
<td>0.836</td>
</tr>
<tr>
<td>5</td>
<td>0.442</td>
</tr>
<tr>
<td>6</td>
<td>0.746</td>
</tr>
</tbody>
</table>
FIGURE 2. Optimal placement rule and mastery function for standard set of parameter values (Case 1)

FIGURE 3. Optimal placement rule and mastery function for higher costs of teaching Course 1 (Case 2)

FIGURE 4. Optimal placement rule and mastery function for utility less sensitive to true mastery level (Case 3)
FIGURE 5. Optimal placement rule and mastery function for adaptive instruction (Case 4)

FIGURE 6. Optimal placement rule and mastery function for higher validity of placement test (Case 5)

FIGURE 7. Optimal placement rule and mastery function for higher reliability of mastery test (Case 6)
van der Linden

another at \( x = 2.084 \). A higher validity of the placement test for the examinees assigned to Course 1 (Case 5) resulted in an optimal mastery function running much steeper, because the prediction of the mastery based on the scores could now be based on more valid placement scores. The placement rule was non-monotone again; two cutoff scores were found, one at \( x = 0.442 \) and the other at \( x = 2.778 \), implying that examinees with lower and very high placement scores had to be assigned to Course 1 and the others to Course 2. Finally, for a higher reliability of the mastery test for the students assigned to Course 1 (Case 6), the relative importance of the placement test as a predictor of the true mastery level went down substantially, and the optimal mastery function became a more horizontal function of the placement score. \( \lambda(x) \) now crossed the horizontal axis only at \( x = 0.746 \).

Concluding Remarks

One of the advantages of optimizing placement and mastery decisions simultaneously is that more realistic utility structures are possible. For example, the point of view taken in this article that placement decisions are in fact mastery decisions in advance could never have been translated into a utility structure if the two decisions were described by separate decision theory models. Another advantage is that the placement scores can be used as collateral information to improve the mastery decisions. The example showed that for a linear utility structure the behavior of the optimal placement rule and mastery function was realistic for various changes in the sensitivity of utility to true mastery level, the adaptivity of the instruction, and the validity and reliability of the placement and mastery tests, respectively.

A natural extension to the problem of course placement addressed here would be to include a mastery decision after Course 2. For this case, the ultimate goal would be to find a placement rule and a mastery decision rule for Course 1 that maximize the probability of mastery after Course 2 at low cost. An obvious approach would then be to define the utilities of all decision outcomes as a function of the true achievement level after Course 2 and to choose rules maximizing the expected utility. It is a topic of future research to find out how this change of problem would affect the form of the decision rules derived in this article.

The question can be raised under what conditions the optimal cutoff score on the mastery test is a constant independent of the score on the placement test rather than the decreasing function defined by (14). From (12) it immediately follows that if

\[
h_1(y|x) = h_1(y), \quad \text{for all } x,
\]

then the inner integral is only a function of \( y \), and the mastery rule becomes independent of \( X = x \). If (39) is satisfied, it holds that \( \beta_{TXY} = 0 \), and the optimal mastery function under linear regression in (29) reduces to its first term, which is indeed a constant. However, (39) represents the case in which the placement
test has no predictive validity whatsoever as to the scores on the mastery test. It is doubted whether such tests could ever play a significant role in placement decisions. In a sense, it has thus been demonstrated in this article that placement decisions can be based on valid placement tests only if they involve optimal mastery rules that are decreasing functions of the score on the placement test.

References


van der Linden, W. J. (in press). Stochastic order in dichotomous item response models for fixed tests, adaptive tests, or multiple abilities. *Psychometrika*.


van der Linden


Author

WIM J. VAN DER LINDEN is Professor, Department of Educational Measurement and Data Analysis, University of Twente, P. O. Box 217, 7500 AE Enschede, The Netherlands; vanderlinden@edte.utwente.nl. He specializes in test theory, decision theory, and applied statistics.

Received January 1, 1996
Revision received August 15, 1996
Accepted September 16, 1996
You have printed the following article:

**A Decision Theory Model for Course Placement**

Wim J. van der Linden


Stable URL:

http://links.jstor.org/sici?sici=1076-9986%28199821%2923%3A1%3C18%3AADTMFC%3E2.0.CO%3B2-V

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

**References**

**Toward an Integration of Theory and Method for Criterion-Referenced Tests**

Ronald K. Hambleton; Melvin R. Novick


Stable URL:

http://links.jstor.org/sici?sici=0022-0655%28197323%2910%3A3%3C159%3ATAIOTA%3E2.0.CO%3B2-Y

**Decision Theory Models for Validating Course Placement Tests**

Richard Sawyer


Stable URL:

http://links.jstor.org/sici?sici=0022-0655%28199623%2933%3A3%3C271%3ADTMFVC%3E2.0.CO%3B2-Q

**Patterns of College Grades across Curricular Areas: Some Implications for GPA as a Criterion**

Lyle F. Schoenfeldt; Donald H. Brush


Stable URL:

http://links.jstor.org/sici?sici=0002-8312%28197522%2912%3A3%3C313%3APCGAC%3E2.0.CO%3B2-9