PSYCHOMETRIKA—VOL. 46, NO. 3.
SEPTEMBER, 1981

USING APTITUDE MEASUREMENTS FOR THE OPTIMAL ASSIGNMENT OF SUBJECTS TO TREATMENTS WITH AND WITHOUT MASTERY SCORES

WIM J. VAN DER LINDEN

TWENTE UNIVERSITY OF TECHNOLOGY

For assigning subjects to treatments the point of intersection of within-group regression lines is ordinarily used as the critical point. This decision rule is critized and, for several utility functions and any number of treatments, replaced by optimal monotone, nonrandomized (Bayes) rules. Both treatments with and without mastery scores are considered. Moreover, the effect of unreliable criterion scores on the optimal decision rule is examined, and it is illustrated how qualitative information can be combined with aptitude measurements to improve treatment assignment decisions. Although the models in this paper are presented with special reference to the aptitude-treatment interaction problem in education, it is indicated that they apply to a variety of situations in which subjects are assigned to treatments on the basis of some predictor score, as long as there are no allocation quota considerations.

Key words: aptitude-treatment interaction, decision theory, criterion-referenced measurement.

Many modern instructional programs can be characterized by the fact that they allow students to reach the same learning objectives in different ways. These programs are often qualified by words like "adaptive" or "individualized". Examples are the Pittsburgh Individually Prescribed Instruction (IPI) project [Glaser, 1968], Flanagan's Program for Learning in Accordance with Needs (PLAN) [Flanagan, 1967] and Computer-Assisted Instruction (CAI) [Atkinson, 1968; Suppes, 1966; Suppes, Smith & Beard, Note 1]. Although Learning for Mastery [Block, 1971; Bloom, 1968; Bloom, Hastings & Madaus, 1971, Chap. 3] is primarily group-based, it also offers some individualization in that it provides those students who fail a mastery test with diagnostic procedures and alternative instructional methods and materials. In a typical individualized program the instruction is divided into comparatively small units or modules. In addition, all units are delimited by means of clear-cut learning objectives, each student is offered one of the different possible routes through a unit (treatments), and all students are, no matter the route they have taken, expected to attain the same minimum level of mastery at the end of each unit. At several points tests are used for monitoring, guidance, evaluation, feedback, treatment assignment, and grading purposes. Reviews of all these testing procedures can be found in Glaser and Nitko [1971], Hambleton [1974], and Nitko and Hsu [1974]. In this paper we focus exclusively on decision theoretic problems connected with treatment assignment and mastery decisions based on unit pretests and posttests, respectively. The former are as yet an underdeveloped part of the Aptitude Treatment Interaction (ATI) methodology. Decisions of this type are called placement [Cronbach & Gleser, 1965] or diagnostic decisions [Nitko & Hsu, 1974]. The latter form the object of the fast growing field of criterion-referenced measurement or mastery testing.

The author is indebted to Gideon J. Mellenbergh for his valuable comments on earlier drafts of the paper. Thanks are also due to Fred N. Kerlinger, Ivo W. Molenaar, Tjeerd Plomp, Niels Veldhuizen, Michel Zwarts, and one of the referees for their helpful comments, and to Paula Achterberg and Jolanda van Laar for typing the manuscript.

Requests for reprints should be sent to Wim J. van der Linden, Onderafdeling Toegepaste Onderwijskunde, Technische Hogeschool Twente, Postbus 217, 7500 AE Enschede, THE NETHERLANDS.

0033-3123/0600-$00.75/0

The aptitude treatment interaction hypothesis and methodology have been motivated by differential reactions of students to instructional treatments. There is no universally best treatment: students differ in their aptitudes, and a treatment that is better on the average for some students may be worse for other students and much better than expected for still others. ATI research tries to track down aptitude × treatment interactions and bridges what Cronbach [1957, 1975] has called the two disciplines of psychology—experimental and correlational research. In ATI research (for a comprehensive overview see Cronbach & Snow, 1977), a great variety of aptitudes have been used: from cognitive styles, reasoning abilities, and entering behaviors to intelligence factors and achievements in previous instructional units. Cronbach and Snow do justice to this variety and define aptitude simply as "any characteristic of a person that forecasts his probability of success under a given treatment" [1977, p. 6]. They also define treatment broadly as "any manipulable variable" [1977, p. 6]. Salomon [1962] describes a treatment classification which consists of remedial, compensatory and preferential treatments. Usually, ATI's are identified by means of ANOVA and linear regression procedures. The former is suitable when the aptitude is a natural polychotomy with a small number of classes. An ATI is supposed to exist when in a factorial analysis the combined effect of aptitude and treatment on a postunit criterion measurement shows a meaningful and significant disordinal interaction. In the case of measurements of a continuous aptitude, ATI researchers mostly use linear regression procedures and define ATI as the crossing of within-group regression lines within the relevant interval of the aptitude. Denoting the aptitude measurement by X and the criterion measurement after treatment $j (j = 0, \ldots, t)$ by Y, the within-group regression line for treatment j is

$$E_j(Y \mid x) = \alpha_j + \beta_j x, \tag{1}$$

where the index j points out that the linear conditional expectation is defined using the bivariate distribution of (X, Y) under treatment j. For $t = 1$, the hypothesis that the two regression lines cross each other within the relevant interval and thus show a disordinal interaction can be statistically tested with the Johnson–Neyman technique. The technique is also applicable when measurements of not one but two aptitudes are available, in which case (1) generalizes to a regression plane [Johnson & Neyman, 1936]. Potthoff [1964], and Erlander and Gustavson [1965] give tests which use simultaneous instead of successive confidence bounds. In all these tests the central statistic is a function of the difference between the two estimated regression lines (planes) from (1):

$$s = s(X) \equiv (\hat{\alpha}_1 - \hat{\alpha}_0) + (\hat{\beta}_1 - \hat{\beta}_0)X. \tag{2}$$

Borich and Wunderlich [1973] describe a computer program for the Johnson–Neyman technique which also tests whether the condition of homogeneity of variance across treatments—the Johnson–Neyman technique rests upon this assumption—is met. A discussion of the possibilities of using structural equation models for detecting ATI's can be found in Cronbach and Snow [1971].

The above statistical techniques deal with ATI's as scientific hypotheses in their own right. They therefore deserve to be tested. In order for ATI research to be of practical importance, however, one must be able to use the results for improving instruction and assigning students to the most promising treatments. In that case ATI data ought to be analyzed within the broader framework of decision theory and not with Neyman–Pearson rules for hypothesis testing. The decision rule for assigning students to treatments generally used in connection with one of the above statistical procedures for hypothesis testing is based on the point of intersection of the estimated regression lines. Students with aptitude scores to the right of the value of the abscissa of this point are assigned to the treatment which has the largest estimated expected criterion scores in this interval, while students on the other side of this value are assigned to the other treatment [Cronbach & Snow, 1977, p. 20]. The most serious objections against decision rules of this type are that they are based on inadequate utility considerations and that there is no explicit decision-theoretic criterion from which it follows that they are optimal. The latter objection can be countered by replacing the point of intersection with the Johnson–Neyman rule for testing the ATI hypothesis, assigning students to treatments according to the critical predictor (aptitude) values (with an appropriate random procedure for the interval for which we cannot accept the ATI hypothesis). Hypothesis-testing rules can be conceived as optimal Bayes rules [e.g. Lindgren, 1976, Sec. 8.3.3.], but then we have to adopt a utility function which in most instances will be an unrealistic representation of the existing treatment assignment problem. In a decision-theoretic approach to this problem, a utility function is chosen to represent all costs and benefits involved as realistically as possible. In addition, a probabilistic model that gives the relation between the aptitude and criterion scores for each treatment is adopted. Both are then combined into an explicit optimization criterion from which the optimal decision rule can be derived. We shall elaborate this further. Cronbach and Snow [1977, pp. 31–33] refer to this approach when they observe that taking treatment costs into account modifies their decision rule and may replace the disordinal interaction structure in the aptitude-treatment data by an ordinal utility-treatment structure. They also refer to Cronbach and Gleser [1965, Appendix 1], who set forth a model for placement decisions that can be applied to the problem of optimal treatment assignment. There may be instances in which this model is less realistic in that it assumes that for each treatment the expected payoff, which in Cronbach and Gleser's terminology equals the utility minus costs of testing, is linear in the aptitude measurement. Cronbach and Gleser also assume that the optimal placement rule is monotone [for a definition see, e.g., Ferguson, 1967, Sec. 6.1.] without looking into the consequences this has for the probability model and utility functions allowed. Finally, they assume that the aptitude measurements are normally distributed. This assumption is superfluous; we will show this in our derivations below. It also follows from the fact that in Cronbach and Gleser's model the optimal placement rule does not contain any reference to this assumption [1965, p. 311, eq. 1.22].

In programs of individualized instruction the end-of-unit test is mostly a mastery test, a test on the basis of which it is decided whether the student has mastered the instructional unit sufficiently so that he may proceed with the next unit, or has to relearn the unit and prepare himself for a new test. Procedures for analyzing mastery tests can be derived from the fast-growing methodology of criterion-referenced measurement. One of the most important problems in mastery testing is finding optimal cut-off scores on criterion-referenced tests, and to solve this problem, decision-theoretic procedures are appropriate. For a review of the criterion-referenced measurement methodology, see Hambleton, Swaminathan, Algina, and Coulson [1978]. Decision-theoretic approaches to the mastery testing problem are given and discussed in Hambleton and Novick [1973], Huynh [1976], Mellenbergh, Koppelaar, and van der Linden [1977], Swaminathan, Hambleton, and Algina [1975], van der Linden [1980], and van der Linden and Mellenbergh [1977]. That individualized instruction end-of-unit tests are usually mastery tests and thus the criterion is a dichotomy instead of a continuum has been completely neglected in ATI research. Yet, as will appear later, this distinction can have important consequences: optimal treatment assignment rules generally differ when the criterion is a dichotomy.

In this paper we elaborate the decision theoretic aspects of the ATI problem. For a number of utility functions the decision rule that optimizes the treatment assignment is given. This is done first for the case of two treatments; then the generalization to $t \geq 2$ treatments is presented. Treatments with and without mastery scores are both considered. For treatments with mastery score we assume that the observed mastery score is a given parameter already optimized using one of the mastery testing procedures referred to above.

The models presented in this paper have a larger scope than individualized instruction and educational testing. Any situation in which a certain kind of treatment, measure, or therapy has to be selected and persons, groups, or institutions can be expected to react differentially to each of these actions is a situation in which use of the optimal rules we will give can improve the decisions. Areas in which the models may be important are psychotherapy, management sciences, medicine, agricultural sciences, and so forth.

The Aptitude Treatment Decision (ATD) Problem

In the following, we shall suppose that the aptitude variable is measured by a test and assumes only discrete values. The criterion variable is also considered discrete, but sometimes we shall idealize and replace it by a continuum. Whenever this may be applicable, the results given below can be applied to all combinations of discrete and continuous variables upon replacing the summation symbols by appropriate integral signs or vice versa. For a glossary of the most important symbols used in this paper, we refer the reader to the Appendix.

We consider a hypothetical experiment consisting of a population of students being exposed to each of n possible treatments but where the students are "brain-washed" so that the effects of one treatment do not interfere with those of another. (The actual experiment needed for parameter estimation and in which different samples of students are assigned to the treatments will be described later on.) Furthermore, it is supposed that previous to the treatments an aptitude X is measured, that each of the possible treatments is followed by a measurement of some criterion Y, and that the relation between the measurement of X and the measurement of Y after treatment j can be represented by a probability (density) function $\eta_j(x, y)(j = 0, \ldots, t)$. Since the treatment is between the aptitude and the criterion measurement, it will influence the relation between X and Y, and this relation can be expected to assume a different shape for each treatment. This is indicated by the index j in $\eta_j(x, y)$. However, because the aptitude measurement takes place previous to the treatments, the distribution of aptitude scores $\lambda_j(x)$ is the same for all treatments, and

$$\lambda_j(x) = \lambda(x) \tag{3}$$

for all values of j.

The above experiment being executed, the Aptitude Treatment Decision (ATD) problem now consists of choosing a decision rule δ that assigns treatments to all possible aptitude scores such that this assignment procedure is optimal in some sense. Although there is a resemblance between the ATD problem and the standard decision problem [for the latter, see, e.g., Ferguson, 1967, DeGroot, 1970, and Lindgren, 1976] and we shall proceed in a way that can be compared with the normal form of decision analysis [DeGroot, pp. 141–142], namely by defining an optimization criterion analogous to the Bayes risk and looking for optimal decision rules, it is necessary to note an important difference between the two problems. As opposed to the standard problem, the true state (position on the criterion Y) is influenced by the chosen treatment. Consequently, the actions may entail different priors (probability distributions of Y) and conditional distributions of the aptitude X given y. This explains why in the following the two probability (density) functions $\pi_j(y)$ and $v_j(x \mid y)$ are indexed by j.

We first give a general formulation of the ATD problem and then discuss monotone decision rules, which will be further examined in this paper. The possibility of randomized procedures like behavioral and mixed decision rules [Ferguson, 1967, pp. 22–28] will be disregarded. In statistical decision theory these are included to study the full range of possibilities and to formulate results with a high level of generality. For each randomized rule, however, there exist a nonrandomized Bayes rule that is at least as good [Ferguson, 1967, p. 43]. Moreover, randomized rules, when applied in educational settings, entail the use of random devices and this will lead to acceptability problems.

AT Decision Rules

Generally, the ATD problem consists of partitioning the set of possible X values into a collection of $t + 1$ subsets $\{S_0, \dots, S_j, \dots, S_t\}$ such that treatment j is assigned to students having their X score in S_j, and this partition is optimal in some way. This partition is what we earlier referred to as a (nonrandomized) decision rule: $\delta = \{S_0, \dots, S_j, \dots, S_t\}$. To select optimal decision rules out of the many partitions that can be formed, an evaluation of the decision outcomes or utility function is needed. For the purpose of this paper, it is sufficiently general to consider the utility U as a function of the criterion Y, which is allowed to assume a different shape for each treatment, or

$$U = u_j(Y). \tag{4}$$

Using (4), the expected utility with respect to the bivariate distributions $\eta_j(x, y)$ for a decision rule $\delta = \{S_0, \dots, S_j, \dots, S_t)$ is defined as

$$B(\delta) = \sum_{j=0}^{t} \sum_{S_j} \sum_{y=0}^{n} u_j(y)\eta_j(x, y). \tag{5}$$

Decision rules are considered optimal when they maximize (5); in other words we shall be looking for optimal decision rules δ^* such that

$$B(\delta^*) = \max_{\delta} B(\delta). \tag{6}$$

The expression $B(\delta)$ can be compared with the Bayes risk in the standard decision problem [e.g., Ferguson, 1967, pp. 30–31] but differs in that it uses a different distribution of (X, Y) for each action. (Strictly speaking, the Bayes risk is not an expected utility but an expected loss, which, consequently, must be minimized instead of maximized. We have chosen the utility terminology to keep in line with ATI literature earlier referred to.) Throughout this paper, however, we shall ordinarily use the name "Bayes risk" to denote expressions like (5).

Monotone Rules

The maximization in (6) takes place over the class of all possible decision rules or partitions of the set of X values. As this may be laborious, it is better to confine ourselves to some tractable subclass of rules among which the optimal rule can be found. Besides, in educational and psychological testing one is accustomed to using cutting scores for making decisions, and decision rules having this form constitute a special subclass known as monotone rules [Ferguson, 1967, Sec. 6.1.]. A nonrandomized decision rule is monotone if there exists a series of cutting scores b_j on X such that

$$0 = b_0 \le b_1 \le \cdots \le b_j \le \cdots \le b_{t+1} = m.$$

where $t \le m$ and treatment j is assigned if $b_j \le x \le b_{j+1}$ (for $j = t$ the second inequality is not strict). The individual members of the subclass of monotone rules are obtained by varying the values of the cutting points subject to the above inequalities.

For the standard decision problem, it is known that under certain conditions the subclass of monotone rules is essentially complete, i.e., that for any nonmonotone rule there is a monotone rule that is at least as good [Ferguson, 1967, Sec. 6.1.; Karlin, 1956; Karlin & Rubin, 1956a, 1956b]. These conditions are that the probability model relating data to true state has monotone likelihood and that the actions can be ordered such that for each two adjacent actions the utility functions have a point of intersection in which the difference between the utilities changes sign. It follows that for the ATD problem with probability functions $v_j(x \mid y)$ which may vary in form across the treatments, the subclass of monotone rules is essentially complete if the following ratio of likelihoods

$$\frac{v_j(y;\, x)}{v_{j+1}(y';\, x)} \tag{7}$$

is monotone nondecreasing in x for every pair of values for y and y' with $y > y'$ and an ordered series of values, $y_0 \leq \cdots \leq y_j \leq \cdots \leq y_{t-1}$, exists for which the utility functions satisfy

$$\begin{aligned} u_j(Y) - u_{j+1}(Y) &\geq 0 \quad \text{for} \quad Y < y_j \\ u_j(Y) - u_{j+1}(Y) &\leq 0 \quad \text{for} \quad Y > y_j \end{aligned} \tag{8}$$

($j = 0, 1, \ldots, t - 1$). Thus, for distributions and utility functions obeying these two conditions, we know that the optimal nonrandomized (Bayes) rule is one of the monotone type and may ignore rules with a different shape. The condition formulated in (7) is met whenever the functions $v_j(x \mid y)$ belong to the same family having monotone likelihood (e.g., if they are—possibly different—members of the exponential family). The reason for giving (7) instead of an ordinary likelihood ratio is, however, that in principle members of *different* families could be found satisfying condition (7)—a statistical problem that, to our knowledge, has yet to be investigated. To establish if the condition formulated in (8) is met, that is, if each two adjacent utility functions have a point of intersection in which the difference between the utilities changes sign, it will usually be necessary to rearrange the treatments. Throughout this paper, we shall assume that the treatments are in the appropriate order indicated by their value of j.

Optimal Assignment with Monotone Utility Functions

We consider the general case with probability functions $v_j(x \mid y)$ and utility functions $u_j(Y)$ obeying (7) and (8), respectively. In most applications it will be realistic to assume that these utility functions are monotone nondecreasing in Y (an increase in the criterion score will never lead to a decrease in utility) and possibly contain a constant indexed by j representing the (fixed) amount of treatment costs. This assumption is not needed for the next few steps, however. For the case of two treatments ($t = 1$) we may replace equation (5) by

$$B(b) = \sum_{x=0}^{b-1} \sum_{y=0}^{n} u_0(y)\eta_0(x, y) + \sum_{x=b}^{m} \sum_{y=0}^{n} u_1(y)\eta_1(x, y), \tag{9}$$

$b \equiv b_1$ being the boundary point indicating the monotone rules or partitions to which we may confine ourselves. From (9) it follows that

$$\begin{aligned} B(b) &= \sum_{x=0}^{b-1} E_0[u_0(Y) \mid x]\lambda(x) + \sum_{x=b}^{m} E_1[u_1(Y) \mid x]\lambda(x) \\ &= \sum_{x=0}^{m} E_0[u_0(Y) \mid x)]\lambda(x) + \sum_{x=b}^{m} \{E_1[u_1(Y) \mid X] - E_0[u_0(Y) \mid x]\}\lambda(x), \end{aligned} \tag{10}$$

where E_j indicates that the expectation has been taken over a distribution indexed by j. Since the first term is a constant and $\lambda(x) \geq 0$ for all values of x, b is optimal for the smallest (integer) value of x for which

$$E_1[u_1(Y) \mid x] - E_0[u_0(Y) \mid x] \tag{11}$$

is positive. The comparison of the conditional expected utilities in (11) is a general solution which, for given utility functions and distributions, can be found numerically by choosing trial values for x, computing both conditional expectations and determining the smallest x value for which the difference between these expectations is positive. If the next trial value is always chosen to be the middle of the interval in which the optimal x value has now been determined to lie, this can be accomplished within $\log_2(m + 1)$ steps. (When this is not an integer, rounding off upwards gives the required upper bound.) The solution can also be found graphically by having a computer plot both conditional expectations. However, these procedures may be cumbersome and it seems better to introduce additional restrictions and to look for analytical solutions.

It is first supposed that, still obeying (8), $u_j(Y)$ has a linear shape:

$$u_j(Y) = f_j Y + g_j. \tag{12}$$

The parameter g_j can, for instance, represent the constant amount of costs of treatment j and will in that case have a nonpositive value. The connection between utility and criterion score is also determined by the parameter f_j. Note that this parameter is indexed by j; hence, function (12) can represent the utility structure of situations in which, next to a constant amount of treatment costs, variable amounts of treatments costs are involved depending upon the performance of the student. An example is a treatment in which remedial teaching is offered as a function of the student's level of functioning. Cronbach and Snow [1977, p. 31–33] call attention to the fact that taking (a constant amount of) treatment costs into account translates the within-group regression lines and may change a disordinal interaction into an ordinal one. Their translation amounts to applying (12) with $f_j = 1$ for all values of j to the regression lines.

Using utility function (12), expression (11) is replaced by

$$f_1 E_1(Y \mid x) - f_0 E_0(Y \mid x) + g_1 - g_0. \tag{13}$$

For the case where the regression functions can be assumed linear it follows that the optimal value of b is equal to

$$b^* = \operatorname{int}\left(\frac{g_0 - g_1 + f_0 \alpha_0 - f_1 \alpha_1}{f_1 \beta_1 - f_0 \beta_0}\right) + 1 \tag{14}$$

where α_j and β_j are the regression parameters from (1) and the function indicated by the mnemonic "int" replaces the value of its argument by the greatest integer not greater than this value.

Note that (14) is unstable for

$$f_0 \beta_0 \simeq f_1 \beta_1. \tag{15}$$

This can be interpreted for the case $f_0 = f_1 = 1$, because (14) then takes the form

$$b^* = \operatorname{int}\left(\frac{g_0 - g_1 + \alpha_0 - \alpha_1}{\beta_1 - \beta_0}\right) + 1 \tag{16}$$

and (15) reduces to $\beta_0 \simeq \beta_1$. Formula (16) yields the (Bayes, randomized) decision rules for situations that Cronbach and Snow have described by their translation of the regression lines to allow for (fixed amounts of) treatment costs. When $f_0 = f_1 = 1$, condition (15) amounts to approximately parallel regression lines and the optimal cut-off score b^* shows, dependent upon the sign of the numerator of (16), a tendency to go to its maximal or minimal possible value. If in addition $\alpha_0 = \alpha_1$, then both regression lines coincide and the optimal cut-off score takes on its minimal possible value for $g_0 < g_1$, its maximal value for $g_0 > g_1$, and it is indefinite for $g_0 = g_1$. Thus, in this case we always assign students to the cheapest treatment and cannot make up our minds if both treatments are equally expensive.

The following numerical example illustrates the procedure: Suppose that the fixed amounts of costs for both treatments are about equal but that treatment $j = 1$, unlike treatment $j = 0$, offers a quantity of remedial instruction and alternative materials that depend on the student's achievement. This might be represented by the parameter values $g_0 = g_1 = f_0 = 1$ and $f_1 = .5$. Also suppose that the distributions of (X, Y) yield regression lines with $\alpha_0 = 22.8$, $\alpha_1 = 17.1$, $\beta_0 = -0.3$, and $\beta_1 = 1.7$. Finally, assume $n = 25$. From (14) it follows that the optimal cut-off $b^* = 13$. Using the classical regression approach, on the other hand, would have resulted in $b^* = 3$. The difference arises from the fact that f_2 was chosen to be equal to .5 instead of 1. This was done, just as in the examples below, to highlight the possibly large effect on the cut-off score of utility functions different from the ones being implied by the regression approach.

Lindley [1976] and Novick and Lindley [1978], following Berhold [1973], make a plea to choose distribution functions as utility functions. There can be situations in which these functions are realistic in that they are bounded and nondecreasing, and moreover, they can sometimes be readily combined with the probability model involved. Note that the linear utility function given in (12) is proportional to the uniform c.d.f. In order to shed some more light on the use of distribution functions as utility functions, a continuous idealization of Y is considered and the normal ogive utility function

$$u_j(Y) = \Phi\left(\frac{Y - \mu_j}{\sigma_j}\right), \tag{17}$$

is chosen where Φ is the standard normal c.d.f. with parameters μ_j and σ_j. Lindley [1976] and Novick and Lindley [1978] show that

$$E_j[u_j(Y) \mid x] = \Phi\left\{\frac{E_j(Y \mid x) - \mu_j}{[\text{var}_j(Y \mid x) + \sigma_j^2]^{1/2}}\right\}, \tag{18}$$

under the assumption that the conditional p.d.f. $\omega_j(Y \mid x)$ is normal with conditional mean and variance $E_j(Y \mid x)$ and $\text{var}_j(Y \mid x)$, respectively. Expression (11) is now replaced by

$$\Phi\left\{\frac{E_1(Y \mid x) - \mu_1}{[\text{var}_1(Y \mid x) + \sigma_1^2]^{1/2}}\right\} - \Phi\left\{\frac{E_0(Y \mid x) - \mu_0}{[\text{var}_0(Y \mid x) + \sigma_0^2]^{1/2}}\right\}, \tag{19}$$

and the smallest integer for which expression (19) is positive is also the value for which

$$\frac{E_1(Y \mid x) - \mu_1}{[\text{var}_1(Y \mid x) + \sigma_1^2]^{1/2}} - \frac{E_0(Y \mid x) - \mu_0}{[\text{var}_0(Y \mid x) + \sigma_0^2]^{1/2}} \tag{20}$$

is positive for the first time. With known distributions $\omega_j(Y \mid x)$ this value can be found graphically or numerically.

For the particular case in which the regression functions are linear and homoscedasticity may be assumed, substitution of (1) into (20) yields

$$b^* = \text{int}\left[\frac{\varepsilon_0(\mu_1 - \alpha_1) - \varepsilon_1(\mu_0 - \alpha_0)}{\varepsilon_0 \beta_1 - \varepsilon_1 \beta_0}\right] + 1 \tag{21}$$

with

$$\varepsilon_j \equiv [\text{var}_j(Y \cdot X) + \sigma_j^2]^{1/2}$$

and

$$\text{var}_j(Y \cdot X) \equiv \{1 - [\text{cor}_j(X, Y)]^2\}\text{var}_j(Y), \tag{22}$$

where $\text{cor}_j(X, Y)$ and $\text{var}_j(Y)$ are the linear correlation coefficient between X and Y and the variance of Y under treatment j, respectively.

As an illustration, suppose again that the decision problem of our previous example is given, but that, on the basis of a pilot study of the utility of criterion scores for students and teachers, it seems better to use a normal ogive transformation of the criterion scores and to represent the situation by utility function (17) with $\mu_0 = \mu_1 = 0$, $\sigma_0 = 1$ and $\sigma_1 = .5$. With the additional information that $\text{cor}_0(X, Y) = .81$, $\text{cor}_1(X \cdot Y) = .70$, $\text{var}_0(Y) = 10.0$, and $\text{var}_1(Y) = 12.9$, it follows from formula (21) that now $b^* = 7$.

Optimal Assignment to Treatments with a Mastery Score

Next, we introduce a mastery score c on criterion Y so that students with $Y \geq c$ may proceed to the next instructional unit but students with $Y < c$ are retained and receive some extra learning time or remedial instruction. One way to formalize this mastery score and its consequences for students and instructional program is to introduce a threshold utility function with parameter c

$$u_j(Y) = \begin{cases} w + a_j & \text{for} \quad Y \geq c \\ v + a_j & \text{for} \quad Y < c \end{cases} \tag{23}$$

in which w and v represent the treatment-independent and a_j the treatment-dependent part (e.g. treatment costs) of the utility structure. It will generally hold that $w > v$. We shall first examine the solution for (23), and then introduce some extra conditions yielding analytical solutions for the optimal cut-off score b^*.

The Bayes risk can now be written as

$$\begin{aligned} B(b) = & \sum_{x=0}^{b-1} \sum_{y=0}^{c-1} (v + a_0)\eta_0(x, y) + \sum_{x=0}^{b-1} \sum_{y=c}^{n} (w + a_0)\eta_0(x, y) \\ & + \sum_{x=b}^{m} \sum_{y=0}^{c-1} (v + a_1)\eta_1(x, y) + \sum_{x=b}^{m} \sum_{y=c}^{n} (w + a_1)\eta_1(x, y). \end{aligned} \tag{24}$$

As the value of b that maximizes (24) also maximizes (24) for any linear transformation of (23), we can rescale (23) without any loss of generality and put $v = 0$ and $w = 1$. This gives

$$\begin{aligned} B(b) & = \sum_{x=0}^{b-1} \sum_{y=c}^{n} \eta_0(x, y) + a_0 \sum_{x=0}^{b-1} \lambda(x) + \sum_{x=b}^{m} \sum_{y=c}^{n} \eta_1(x, y) + a_1 \sum_{x=b}^{m} \lambda(x) \\ & = \sum_{x=0}^{m} \sum_{y=c}^{n} \eta_0(x, y) + a_0 + \sum_{x=b}^{m} \left\{ \sum_{y=c}^{n} [\omega_1(Y \mid x) - \omega_0(Y \mid x)] + a_1 - a_0 \right\} \lambda(x). \end{aligned} \tag{25}$$

Since the first double sum in (25) is a constant and $\lambda(x)$ is nonnegative for all values of x, b^* is that value of x for which

$$\sum_{Y=c}^{n} [\omega_1(Y \mid x) - \omega_0(Y \mid x)] + a_1 - a_0$$

or

$$\Omega_0(c \mid x) - \Omega_1(c \mid x) + a_1 - a_0, \tag{26}$$

Ω_j being the c.d.f. of $Y \mid x$, is positive for the first time.

We now examine the solution for the case in which the treatment-dependent part in (23) is approximately equal for both treatments, i.e. when (26) can be replaced by

$$\Omega_0(c \mid x) - \Omega_1(c \mid x). \tag{27}$$

If Y given x is normally distributed with (conditional) expectations $E_j(Y \mid x)$ and variances $\text{var}_j(Y \mid x)$, the solution—i.e. the smallest value of x which makes (27) positive—is equivalent to that which makes

$$\frac{c - E_0(Y \mid x)}{[\text{var}_0(Y \mid x)]^{1/2}} - \frac{c - E_1(Y \mid x)}{[\text{var}_1(Y \mid x)]^{1/2}} \tag{28}$$

positive for the first time. For linear regression functions (1) this results in

$$\frac{c - \alpha_0 - \beta_0 x}{[\text{var}_0(Y \mid x)]^{1/2}} - \frac{c - \alpha_1 - \beta_1 x}{[\text{var}_1(Y \mid x)]^{1/2}}.$$

Under homoscedasticity $\text{var}_j(Y \mid x)$ reduces to the constant $\text{var}_j(Y \cdot X)$ given in (22) and

$$b^* = \text{int}\left\{\frac{(c - \alpha_1)[\text{var}_0(Y \cdot X)]^{1/2} - (c - \alpha_0)[\text{var}_1(Y \cdot X)]^{1/2}}{\beta_1[\text{var}_0(Y \cdot X)]^{1/2} - \beta_0[\text{var}_1(Y \cdot X)]^{1/2}}\right\} + 1. \tag{29}$$

Again using the two distributions from our preceding example with $\alpha_0 = 22.8$, $\alpha_1 = 17.1$, $\beta_0 = -0.3$, $\beta_1 = 1.7$, $\text{cor}_0(X, Y) = .81$, $\text{cor}_1(X, Y) = .70$, $\text{var}_0(Y) = 10.0$, $\text{var}_1(Y) = 12.9$, and the additional information that criterion Y has a mastery score $c = 20$, it follows from (29) that the optimal cut-off $b^* = 4$.

Note how in expressions (26)–(29) the optimal decision is a function of the mastery score c. In situations in which the treatment assignment is followed by end-of-unit mastery decisions and threshold utility is a realistic assumption, this compels the involvement of the mastery score in the assignment procedure. There is a particular case for which this does not apply, namely when

$$\text{var}_0(Y \mid x) = \text{var}_1(Y \mid x) \tag{30}$$

for all values of x. Then, solving (28) reduces to a comparison between the two regression functions, and b^* is the first value of x to the right of the intersection of the two functions. With linear regression this amounts to

$$b^* = \text{int}\left(\frac{\alpha_1 - \alpha_0}{\beta_0 - \beta_1}\right) + 1. \tag{31}$$

Optimal Assignment with Utility Defined on the True Criterion

In the event of unreliable criterion scores or criterion scores that are inefficient estimators of an underlying true or latent parameter, it seems better to revise the utility function (4) and define it as a function of the true criterion score T:

$$U = u_j(\mathrm{T}). \tag{32}$$

This amounts to the notion that utilities should not be based on fallible impressions but on what the student is actually able to perform. Henceforth we shall assume that T is defined as the expected criterion score over replications, a definition that not only corresponds with classical test theory but with any test model in which Y is an unbiased estimator of some parameter of interest T. Cronbach and Snow's assertion that unreliability of the outcome measure does not bias the interpretation of ATI [1977, p. 34] is motivated by their interest in testing the disordinality of within-group regression lines and is generally not valid for the decision theoretic approach and utility functions examined in this paper. Unlike Cronbach and Snow [1977, p. 33], we are not interested in the unreliability of the aptitude measurements since we shall never be able to use true aptitude scores for treatment assignment; we always confine ourselves to the available observed or estimated aptitude scores. Nevertheless, it must be taken into account that in situations where for some treatments the true aptitude and criterion scores are stochastically dependent, a reduction in the Bayes risk may be gained by improving the reliability of the aptitude measurement.

As a consequence of (32) we have to replace (9) by

$$B(b) = \sum_{x=0}^{b-1} \sum_{y=0}^{n} \int_0^n u_0(\tau)\psi_0(x, y, \tau)\, d\tau + \sum_{x=b}^{m} \sum_{y=0}^{n} \int_0^n u_1(\tau)\psi_1(x, y, \tau)\, d\tau, \tag{33}$$

where $\psi_j(x, y, \tau)$ is the probability (density) function of (X, Y, T) under treatment j. Integrating and summing yields

$$B(b) = \sum_{x=0}^{b-1} E_0[u_0(\mathrm{T}) \mid x]\lambda(x) + \sum_{x=b}^{m} E_1[u_1(\mathrm{T}) \mid x]\lambda(x),$$

or, on completing the first sum,

$$B(b) = \sum_{x=0}^{m} E_0[u_0(\mathrm{T}) \mid x]\lambda(x) + \sum_{x=b}^{m} \{E_1[u_1(\mathrm{T}) \mid x] - E_0[u_0(\mathrm{T}) \mid x]\}\lambda(x),$$

which is optimal for the smallest value of x for which

$$E_1[u_1(\mathrm{T}) \mid x] - E_0[u_0(\mathrm{T}) \mid x] \tag{35}$$

is positive. We know however that

$$E(\mathrm{T} \mid x) = E(Y - e \mid x) = E(Y \mid x), \tag{36}$$

where e is the error of estimation in the unbiased estimator Y of T. We arrive again at the solutions given by (14), (20), and (21).

Analogous to the preceding, we now examine the consequences for treatments with a true mastery score and define the threshold loss function as

$$U = u_j(\mathrm{T}) = \begin{cases} w + a_j & \text{for} \quad \mathrm{T} \geq d \\ v + a_j & \text{for} \quad \mathrm{T} < d, \end{cases} \tag{37}$$

where d denotes the true mastery score. Note that we do not use c again, inasmuch as c may differ from d when obtained by decision theoretic optimization procedures which we mentioned in our discussion of mastery testing. Following the same derivation as before, it appears that b^* is equal to that value of x for which

$$\int_d^n [\theta_1(\mathrm{T} \mid x) - \theta_0(\mathrm{T} \mid x)]\, d\tau + a_1 - a_0 \tag{38}$$

or

$$\Theta_0(d \mid x) - \Theta_1(d \mid x) + a_1 - a_0 \tag{39}$$

is positive for the first time, with $\theta_j(\mathrm{T} \mid x)$ denoting the p.d.f. of T given x and Θ_j its c.d.f.

Unlike the monotone nondecreasing utility function given in (32), we see that for treatments with a mastery score it does matter whether we choose an observed or true criterion. This is so even if we would choose $c = d$, since the difference refers to the c.d.f.'s and not to the mastery score used. To determine the consequences for solution (29), we again assume normal conditional distributions (now of T given x) with linear regression and homoscedasticity. Equation (22) assumes the form

$$\mathrm{var}_j(\mathrm{T} \cdot X) = \{\mathrm{cor}_j(Y, Y') - [\mathrm{cor}_j(X, Y)]^2\}\mathrm{var}_j(Y), \tag{40}$$

where $\mathrm{cor}_j(Y, Y')$ is the reliability coefficient of measurement Y after treatment j. Substituting (40) instead of (22) into (29) gives the analogue of (29) for utility function (37) (with $g_0 = g_1 = g$). Both solutions coincide in the event of distributions of criterion measurements $\pi_j(y)$ having $\mathrm{cor}_j(Y, Y') = 1$ for $j = 0, 1$. However, this seems unlikely for behavioral measurements, and, moreover, in ATI research one normally looks for aptitudes which have a differential validity or predictability [Cronbach & Gleser, 1967, pp. 57–59; Lord & Novick, 1968, pp. 271–273] and this, in combination with condition (3) and for criterion measurements with a supposedly constant standard error of measurement, leads to distributions $\pi_j(y)$ with different reliabilities.

To illustrate the consequences of the attenuation correction in (40) for the solution given in (29), we again compute the optimal cut-off score b^* using both distributions of our preceding examples but now with the additional information that $d = c = 20$ and the criterion measurements have reliabilities $\mathrm{cor}_0(Y, Y') = .66$ and $\mathrm{cor}_1(Y, Y') = .85$. It appears that b^* is equal to 7 instead of 4.

Generalization to $t \geq 2$

In the foregoing we confined ourselves to assignment decisions with two treatments ($t = 1$). The theory can be readily generalized to $t \geq 2$ treatments, however. We shall show this below, and in doing so we still assume that the decision problem has a monotone character, that is, that the likelihoods $v_j(y; x)$ and $\tau_j(\tau; x)$ are monotonic [equation (7)], the utilities satisfy the inequalities given in (8), and the optimal decision rule therefore can be found in the class of rules having the form of an ordered series of cut-off scores.

In the preceding sections we obtained the solutions by an optimization method in which we arrived at a sum consisting of the partial sum to $x = b$ for some expression and from $x = b$ for another. The optimal value of b was found by completing the first partial sum and adjusting the second sum correspondingly. The generalization of this method to decision rules of the form

$$0 = b_0 \leq \cdots \leq b_j \leq \cdots \leq b_{t+1} = m, \tag{41}$$

with $2 \leq t \leq m$, where b_j is the cutting score between treatment j and $j + 1$, will be illustrated with the help of expression (27) (threshold utility defined on Y). It can easily be verified that for multiple monotone decisions the Bayes risk is the following generalization of (9):

$$B(b_1, \ldots, b_t) = \sum_{x=0}^{b_1-1} [1 - \Omega_0(c \mid x)]\lambda(x) + \sum_{x=b_1}^{b_2-1} [1 - \Omega_1(c \mid x)]\lambda(x) + \cdots + \sum_{x=b_t}^{m} [1 - \Omega_t(c \mid x)]\lambda(x). \tag{42}$$

Completing the first sum and omitting the factor $\lambda(x)$ because, as a nonnegative quantity which does not change sign, it plays no role in optimizing (42), results in

$$B_1(b_1, \ldots, b_t) = \sum_{x=0}^{m} [1 - \Omega_0(c \mid x)] + \sum_{x=b_1}^{b_2-1} [\Omega_0(c \mid x) - \Omega_1(c \mid x)] + \cdots + \sum_{x=b_t}^{m} [\Omega_0(c \mid x) - \Omega_t(c \mid x)]. \tag{43}$$

This sum is maximal if each of its terms is maximal; thus, we may conclude that b_1^* is equal to the smallest value of x making

$$\Omega_0(c \mid x) - \Omega_1(c \mid x) \tag{44}$$

positive.

Substituting b_1^* for b_1 into (42), completing the second sum from b_1^* up to and including m, adjusting all following terms, and leaving out $\lambda(x)$ and all constants yields

$$B_2(b_2, \ldots, b_t) = \sum_{x=b_1*}^{m} [1 - \Omega_1(c \mid x)] + \sum_{x=b_2}^{b_3-1} [\Omega_1(c \mid x) - \Omega_2(c \mid x)] + \cdots + \sum_{x=b_t}^{m} [\Omega_1(c \mid x) - \Omega_t(c \mid x)]. \tag{45}$$

Now b_2^* is equal to the first value of x making

$$\Omega_1(c \mid x) - \Omega_2(c \mid x)$$

positive. The values $b_3^*, \ldots, b_t^*$ are found repeating the same procedure.

The above procedure shows how rules can be found for the multiple monotone AT decision problem. *Starting from the lowest cutting score the results in the preceding sections are applied to each pair of adjacent treatments.*

Using Qualitative Information to Improve AT Decisions

Suppose that we are able to classify the students on a qualitative attribute (e.g. previous schooling) and that the resulting distribution also shows stochastic dependence with the criterion distribution for some treatments. How can this information be used to improve the treatment assignment?

We denote this classification by K, with $K = 0, \ldots, k, \ldots, q$, and define

$$p_k \equiv \text{prob}\{K = k\}.$$

If R_j is the set of ordered pairs of K and X values leading to treatment j, then the Bayes risk (for a utility function defined on T) is equal to

$$B(\delta) = \sum_{j=0}^{t} \sum_{R_j} \sum_{y=0}^{n} \int_0^n u_j(\tau)\psi_{jk}(x, y, \tau \mid k)p_k \, d\tau, \tag{46}$$

where $\psi_{jk}(x, y, \tau \mid k)$ is the p.(d.)f. of (X, Y, T) given $K = k$. For conditional distributions having monotone likelihood and utility functions satisfying the inequalities given in (8), the second coordinates in the t sets R_j show for each possible value of K the partitioning corresponding with the monotone decision rules we studied earlier, and the optimal collection of sets R_j are found by applying the preceding theory for each class or subpopulation of K. The result is thus a series of $(t + 1)(q + 1)$ cut-off scores, namely $t + 1$ scores b_{jk} for each class k.

In principle, the foregoing suggests lines along which a generalization of the theory in this paper to two and more aptitudes can be found. Aptitude measurements are always discrete and can therefore be considered a classification consisting of m possible classes. The computations involved, however, increase quickly in magnitude with the number of classes and the number of aptitudes, and the amount of data required to arrive at stable results may be impractical. It seems better to use the extensions we have given here only for natural dichotomies or trichotomies.

The foregoing also suggests an extension to situations in which the test is culturally biased against some subpopulations. For this problem there are solutions based on expected utility models as well [Gross & Su, 1975; Mellenbergh & van der Linden, 1981; Peterson, 1976; Peterson & Novick, 1976]. When there are subpopulations $L = 0, \ldots, l, \ldots, r$ against which the aptitude or the criterion test may be biased, the Bayes risk can be defined analogous to these expected utility models as

$$B(\delta) = \sum_{j=0}^{t} \sum_{Q_j} \sum_{y=0}^{n} \int_0^n u_{jl}(\tau)\psi_{jl}(x, y, \tau \mid l)p_l \, d\tau, \tag{47}$$

with

$$P_l = \text{prob}\{L = l\}$$

where $\psi_{jl}(x, y, \tau \mid l)$ is the p.(d.)f. of (X, Y, T) given $L = l$ and Q_j denotes the set of ordered pairs of L and X values for which treatment j is chosen. The difference between (46) and (47) is that in the expected utility models for culturally fair testing the utility function is defined as $u_j(T)$ and allowed to assume a different form for each subpopulation. The optimal decision rule can, however, be found similarly.

Discussion

This paper was mainly motivated by the fact that regression line intersections, although they are often recommended and used for this purpose, lead to decision rules that in many situations are far from optimal. Moreover, as we pointed out earlier, they follow from considerations in which utility functions and optimization criterion are not chosen explicitly. Looking back, we are now able to indicate some utility functions, types of distributions, and an optimization criterion together yielding the point of intersection of regression lines as the optimal cut-off point. For two regression lines with parameters as designated in (1), the point of intersection has an x coordinate with value $(\alpha_1 - \alpha_0)/(\beta_0 - \beta_1)$. This is precisely the argument of the greatest integer function in solution (14) for $g_0 = g_1$ and $f_0 = f_1$, which means that the use of the point of intersection involves a Bayes rule with linear utility under the restriction $g_0 = g_1$ and $f_0 = f_1$ and distributions with linear regresssion. It is also the argument of the greatest integer function in solution (21) for $\varepsilon_0 = \varepsilon_1$ and $\mu_0 = \mu_1$, and the one in (29) for $\text{var}_0(Y \cdot X) = \text{var}_1(Y \cdot X)$. The former implies normal ogive utility with equal location, distributions with linear regression, and a restriction on the scale parameter of the normal ogive and the conditional criterion score variances. The latter calls for distributions with linear regressions and, for example, homoscedasticity. Other combinations of utilities and distributions may also lead to this decision rule. Thus, there are some situations in which the use of the intersection is optimal; the point is, however, that the use of the intersection is not universally best as a decision rule and that in a situation with different utilities or in which the required distributional assumptions are violated much better rules can be found.

Several techniques are in use to assess utility structures. Most texts on decision theory contain sections devoted to utility theory in which lottery methods are proposed for this purpose [e.g., Luce & Raiffa, 1957, Chap. 2; for a recent modification, see Novick & Lindley, 1979]. But in principle any psychological scaling method can be used. Although helpful techniques are available, this does not mean that, for example, in programs of individualized instruction the assessment of utilities is always a simple matter. Much depends on the way the situation to which these techniques are to be applied may be reduced beforehand. When several consequences of the decision outcomes have to be taken into account, among which less tangible consequences as, for instance, psychic well-being or societal effects, the assessment of utilities may be extremely complicated and not possible without additional assumptions. On the other hand, when only a few clearly defined consequences are deemed important, the situation resembles those in business applications of which many successful illustrations are available [e.g., Keeney & Raiffa, 1976]. Elsewhere [van der Linden, 1980] we have indicated that in choosing a utility function fit to the decisionmaker's utilities should not be the only requirement. The choice of a utility function ought to be a compromise between at least three requirements: (a) fit to the decisionmaker's utilities; (b) fit to the psychometric model relating test scores to true future states; and (c) robustness of results with respect to its parameters. This last requirement has not received much attention (an exception is Vijn, Note 2), but it is very important. Especially to the decisionmaker who is uncertain about his parameter specifications, it can be a soothing thought that parameter values differing somewhat from the ones that were actually used lead to the same decision. For an example in the area of mastery testing, we refer to van der Linden [1980].

This paper was also motivated by the fact that introducing treatments with mastery scores might increase the practical value of the ATD problem. In addition, we have called attention to the fact that in case of unreliable criterion scores utility functions ought to be redefined as a function of the underlying true or latent criterion variable. It appears that this changes the optimal decision rule for treatments with a mastery score (threshold loss).

So far the distributions dealt with in our solutions have been model or population quantities. In applications these must be estimated, and it is important that the data come from the correct experiment and not, for example, from individualized instructional programs in which students are already assigned to treatments on the basis of their scores on the aptitude in question. In a proper experiment students from the same aptitude score distribution are randomly drawn and assigned to treatments, after which their criterion performances are measured. This is to guarantee that condition (3) is satisfied. The bivariate score distributions that arise in this way can be used to estimate the necessary parameters and to test the distributional assumptions and, when needed, the linearity of the regression functions.

The parameters to be estimated are all familiar parameters, such as conditional expectations (regression functions), variances, and linear regression parameters. Estimation of these, for which standard statistical theory provides estimators with favorable properties, can yield errors of estimation that propagate in computing the assignment rules of this paper. It is therefore recommended not to use too small samples in the foregoing experiment. This does not mean, however, that small samples necessarily yield inaccurate results. The way errors of estimation propagate depends on the whole structure of the rule, involving not only distributional parameters but utility parameters as well. One of the referees of an earlier draft of this paper suggested the establishment of "regions of indifference" as in the Johnson–Neyman technique, i.e., intervals on the aptitude variable for which we are indifferent to the two adjacent treatments because of sampling error. A practical way to proceed may be to establish (simultaneous) confidence intervals for the parameters, and to determine the range of cutting score values associated with these intervals. A large range implies that large errors are likely to have crept into the determinations of the cutting score. It is recommended that this be done when, for practical reasons, only samples of limited size can be used.

The solutions given in this paper apply to treatment assignment problems with a monotone character. The solutions for nonmonotone problems are still to be examined. It is however, always possible to use observed frequencies of Y given x and compute for each treatment the conditional average utility given x

$$A_j(x) \equiv \frac{\sum_{y=0}^{n} u_j(y) 0_j(y \mid x)}{\sum_{y=0}^{n} 0_j(y \mid x)},$$

with $0_j(y \mid x)$ denoting the observed frequency of Y given x for treatment j. (In this empirical approach $\lambda(x)$ plays no role.) The optimal procedure can then consist of assigning those treatments to the values of x for which $A_j(x)$ is maximal across j. It should be realized that this empirical approach can yield decision rules that are extemely unstable unless based on large numbers of cases. Moreover, as already pointed out, the acceptability of nonmonotone rules may be questionable in situations in which the parties involved are used to decision rules having the form of cut-off scores. Therefore when possible, it would be appropriate to redesign the tests to obtain decision problems that do have a monotone character.

Finally, we repeat that the ATD problem differs from the (Bayes) decision problem presented in standard texts [e.g. DeGroot, 1970; Ferguson, 1967; Lindgren, 1976, Chap. 8]. In this problem there is the same prior for each action, whereas in the ATD problem the priors are allowed to vary and assume the form which represents the effects of the action on the true state. All decision problems in which the true state is a future state influenced by the action to be taken have this property. Further examination of the formal structure and the statistical aspects of the ATD problem seems therefore a valuable line of research.

Appendix: Glossary of Symbols

The most important symbols used in this paper are defined as follows:

X	aptitude measurement; $X = 0, \ldots, m$;
Y	criterion measurement: $Y = 0, \ldots, n$;
T	true criterion; $\mathrm{T} \varepsilon [0, n]$;
K	qualitative attribute; $K = 0, \ldots, k, \ldots, q$;
L	subpopulation; $L = 0, \ldots, l, \ldots, r$;
U	utility;
j	treatment; $j = 0, \ldots, t$;
S_j	set of X values leading to treatment j;
R_j	set of (K, X) values leading to treatment j;
Q_j	set of (L, K) values leading to treatment j;
b	cut-off score on X;
c	mastery score on Y;
d	mastery score on T;
$\lambda(x)$	p.f. of X;
$\pi(y)$	p.(d.)f. of Y;
$\eta(x, y)$	p.(d.)f. of (X, Y);
$\psi(x, y, \tau)$	p.(d.)f. of (X, Y, T);
$\omega(y \mid x)$	p.(d.)f. of Y given $X = x$;
$\nu(x \mid y)$	p.f. of X given $Y = y$;
$\theta(\tau \mid x)$	p.d.f. of T given $X = x$;
α	intercept of the regression line of Y on X;
β	slope of the regression line of Y on x;
v, w, a_j	parameters in the threshold utility function for treatment j;
g_j, f_j	parameters in the linear utility function for treatment j;
μ_j, σ_j	parameters in the normal ogive utility function for treatment j.

REFERENCE NOTES

1. Suppes, P., Smith, R., & Beard, M. *University level computer-assisted instruction at Stanford: 1975* (Technical Report No. 265). Stanford, Cal.: Stanford University, Institute for Mathematical Studies in the Social Sciences, October, 1975.
2. Vijn, P. *Prior information in linear models*, Thesis, University of Groningen, The Netherlands, 1980.

REFERENCES

Atkinson, R. C. Computer-based instruction in initial reading. In *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service, 1968.

Berhold, M. H. The use of distribution functions to represent utility functions. *Management Science*, 1973, *19*, 825–829.

Block, J. H. *Mastery learning: Theory and practice*. New York: Holt, Rinehart and Winston, 1971.

Bloom, B. S. Learning for mastery. *Evaluation Comment*, 1968, *1*, (12).

Bloom, B. S., Hastings, J. T., & Madaus, G. F. *Handbook on formative and summative evaluation of student learning*. New York, N.J.: McGraw-Hill, 1971.

Borich, G. D., & Wunderlich, K. W. Johnson–Neyman revisited: Determining interactions among group regressions and plotting regions of significance in the case of two groups, two predictors and one criterion. *Educational and Psychological Measurement*, 1973, *33*, 155–159.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1967, *12*, 671–684.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, *30*, 116–127.

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

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington Publishers, Inc., 1977.

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

Erlander, S., & Gustavsson, J. Simultaneous confidence region in normal regression analysis with an application to road accidents. *Review of the International Statistical Institute*, 1965, *33*, 364–378.

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

Flanagan, J. C. Functional education for the seventies. *Phi Delta Kappan*, 1967, *49*, 27–32.

Glaser, R. Adapting the elementary school curriculum to individual performance. In *Proceedings of the 1967 Invitational Conference on Testing Problems.* Princeton, N.J.: Educational Testing Service, 1968.

Glaser, R., & Nitko, A. J. Measurement in learning and instruction. In R. L. Thorndike (Ed.), *Educational Measurement.* Washington, D.C.: American Council on Education, 1971.

Gross, A. L., & Su, W. H. Defining a "fair" or "unbiased" selection model: A question of utilities. *Journal of Applied Psychology*, 1975, *60*, 345–351.

Hambleton, R. K. Testing and decision-making procedures for selected individualized instructional programs. *Review of Educational Research*, 1974, *44*, 371–400.

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

Hambleton, R. K., & Swaminathan, H., Algina, R., & Coulson, D. B. Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research*, 1978, *48*, 1–47.

Huynh, H, Statistical considerations of mastery scores. *Psychometrika*, 1976, *41*, 65–79.

Johnson, P. D., & Neyman, J. Tests of linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1936, *1*, 57–93.

Karlin, S. Decision theory for Pólya type distributions. Case of two actions, I. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* (Vol. 1). Berkeley, Cal.: University of California Press, 1957.

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

Karlin, S., & Rubin, H. The theory of decision procedures for distributions with monotone likelihood ratio. *Annals of Mathematical Statistics*, 1956, *27*, 272–299.

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

Lindgren, B. W. *Statistical Theory* (3rd Ed.). New York: MacMillan Publishing Co., Inc., 1976.

Lindley, D. V. A class of utility functions. *The Annals of Statistics*, 1976, *4*, 1–10.

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

Luce, R. D., & Raiffa, H. *Games and decisions.* New York: John Wiley & Sons, Inc., 1957.

Mellenbergh, G. J., Koppelaar, H., & van der Linden, W. J. Dichotomous decisions based on dichotomously scored items: A case study. *Statistica Neerlandica*, 1977, *31*, 161–169.

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

Nitko, A. J., & Hsu, T. C. Using domain-referenced tests for student placement, diagnosis and attainment in a system of adaptive, individualized instruction. *Educational Technology*, 1974, *14*, 48–54.

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

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

Petersen, N. S. An expected utility model for "optimal" selection. *Journal of Educational Statistics*, 1976, *1*, 333–358.

Petersen, N. S., & Novick, M. R. An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, 1976, *13*, 3–29.

Potthoff, R. F. On the Johnson–Neyman technique and some extensions thereof, *Psychometrika*, 1964, *29*, 241–256.

Salomon, G. Heuristic models for the generation of aptitude-treatment interaction hypotheses. *Review of Educational Research*, 1972, *42*, 327–343.

Suppes, P. The use of computers in education. *Scientific American*, 1966, *215*, 206–221.

Swaminathan, H., Hambleton, R. K., & Algina, J. A Bayesian decision-theoretic procedure for use with criterion-referenced tests. *Journal of Educational Measurement*, 1975, *12*, 87–98.
van der Linden, W. J. Decision models for use with criterion-referenced tests. *Applied Psychological Measurement*, 1980, *4*, 469–492.
van der Linden, W. J., & Mellenbergh, G. J. Optimal cutting scores using a linear loss function. *Applied Psychological Measurement*, 1977, *1*, 593–599.

Manuscript received 1/3/80
First revision received 8/15/80
Final version received 3/3/81