DOCUMENT RESUME

ED 208 022 TM 810 711

AUTHOR Wilcox, Rand R.
TITLE On False-Positive and False-Negative Decisions with a Mastery Test.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
REPORT NO CSE-R-146
PUB DATE Oct 80
NOTE 22p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cutting Scores; *Mastery Tests; Mathematical Models; *Student Placement; *Test Reliability; True Scores
IDENTIFIERS *Binomial Error Model; *Test Length

ABSTRACT

Wilcox (1977) examines two methods of estimating the probability of a false-positive on false-negative decision with a mastery test. Both procedures make assumptions about the form of the true score distribution which might not give good results in all situations. In this paper, upper and lower bounds on the two possible error types are described which make no assumption about the form of the true score distribution. Illustrations are given on how these bounds might be used to determine the length of the test. (Author)

ON FALSE-POSITIVE AND FALSE-NEGATIVE DECISIONS WITH A MASTERY TEST

Rand R. Wilcox

CSE Report No. 146

October 1980

Test Design Project
Center for the Study of Evaluation
Graduate School of Education, UCLA
Los Angeles, California 90024

Table of Contents

PREFACE

A part of our goal at CSE has been to develop new and improved psychometric techniques to study, develop and characterize achievement tests and achievement test items. Recently our efforts have been focused on certain errors that occur when using criterion-referenced tests. In particular, we have investigated problems related to estimating and controlling the false-positive and false-negative error rates associated with a test and a population of examinees. In other words, we are concerned about passing those examinees who should pass, and retaining those examinees who need remedial work. This paper deals with one aspect of that problem.

ABSTRACT

Wilcox (1977) examines two methods of estimating the probability of a false-positive on false-negative decision with a mastery test. Both procedures make assumptions about the form of the true score distribution which might not give good results in all situations. In this paper, upper and lower bounds on the two possible error types are described which make no assumption about the form of the true score distribution. Illustrations are given on how these bounds might be used to determine the length of the test.

Introduction

Recently, Wilcox (1977) considered two methods of estimating the probability of making a false-positive or false-negative decision with a mastery test. Both of these procedures make an assumption about the form of the distribution of true scores over the population of examinees. In this ~~paper~~, upper and lower bounds to these probabilities are described which make no assumption about the true score distribution beyond that its first two moments exist. We begin by stating explicitly the model that will be used to describe a mastery test after which we consider briefly the importance of false-positive and false-negative decisions relative to the other proposed methods of characterizing such tests.

1. The Model

Consistent with Hambleton and Novick (1973), Harris (1974), Novick and Lewis (1974), Huynh (1976), Fhaner (1974), and Wilcox (1977), we may describe a mastery test as follows: An instructional program is developed with the goal of fostering certain specific skills in the students taking the course. For each skill area, a domain of test items is constructed. A total of n items is randomly sampled from this domain and administered to an examinee for the purpose of determining whether the examinee's true score, say ζ, is above or below the known criterion score ζ_0. If $\zeta \leq \zeta_0$ the examinee is a master and he/she is advanced to the next level of instruction; otherwise, the examinee is given remedial work. The decision $\zeta \leq \zeta_0$ is made if, and only if, $x \geq x_0$ where x_0 is some appropriately chosen passing score and where x is the examinee's number correct observed score. Note that the choice for the passing score x_0 may be made in accordance with the "losses" associated with the probability of a false-positive or false-negative decision (e.g., Hambleton and Novick, 1973).

For this model of a mastery test, there are two possible errors. The first is a false-positive error which occurs when $x \geq x_0$ and $\zeta < \zeta_0$. A false-negative error occurs when $x < x_0$ and $\zeta \geq \zeta_0$.

Let $\alpha = \Pr(x \geq x_0, \zeta < \zeta_0)$ and $\beta = \Pr(x < x_0, \zeta \geq \zeta_0)$. In this paper α and β are defined in terms of a group of individuals. In particular, $g(\zeta)$, the distribution of ζ, is the probability density function of true scores over a population of examinees. This is in contrast to the Bayesian approach where $g(\zeta)$ is the prior distribution for a specific examinee. (See, e.g., Novick and Lewis, 1974.)

As mentioned earlier, Wilcox (1977) describes two methods of estimating α and β both of which assume that the distribution of ζ over the population of examinees has a particular parametric form. The first estimation procedure assumes that the conditional distribution of observed scores for a single examinee is given by

$$f(x|\zeta) = \binom{n}{x} \zeta^{x} (1-\zeta)^{n-x}, \qquad (1.1)$$

the binomial probability function, and that the distribution of ζ is

$$g(\zeta) = \frac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)} \zeta^{r-1} (1-\zeta)^{s-1}, \qquad (1.2)$$

the beta distribution with parameters $r>0$ and $s>0$. For $n \geq 10$ it appears that this estimation procedure gives fairly good results even when observations are generated according to a two-term approximation to the compound bionomial distribution. The same is also true for the other estimation procedure which uses an arc-sine transformation on the observed score of an examinee and which assumes a normal prior distribution.

The assumption that ζ has a beta distribution deserves serious consideration since there is evidence that the resulting beta-binomial (or negative hypergeometric) probability model may give a good fit to data (Keats and Lord, 1962; Lord, 1965). One difficulty with this model is that, with the exception of U-shaped distributions, the distribution of true scores can have at most one mode. Thus, it is not at all clear whether the beta-binomial model will yield reasonably accurate values for α and β in every case.

One possible solution to this problem is to consider some other method of estimating the true score distribution. (See, e.g., Maritz, 1967; Lord, 1969.) However, the robustness of these alternate models in terms of estimating α and β is unknown and difficult to ascertain.

Another possibility is to use some coefficient that reflects indirectly the values of α and β but which makes no assumption about the form of $g(\zeta)$. For example, one might use the proportion of agreement (Hambleton and Novick, 1973) or Cohen's Kappa (Swaminathan, Hambleton and Algina, 1974). Several other coefficients have been proposed as well (Harris, 1974; Livingston, 1972; Brennan and Kane, 1977). In terms of α and β, all of these coefficients present at least two problems. First, the exact relationship of α and β to these other indexes is unknown. Second, none of these other indexes makes a distinction between false-positive and false-negative decisions. This latter problem is particularly troublesome since the seriousness of a false-positive decision may not be the same as the seriousness of a false-negative decision which, in turn, may have an effect on the decision rule used to determine whether ζ is above or below ζ_0. An illustration of this point arises in the situation considered by Hambleton and Novick (1973), Novick and Lewis (1974) and Huynh (1976) in which constant losses are associated with the two possible errors. Thus, we let the constants c_1 and c_2 represent the "cost" of a false-positive and false-negative decision, respectively. Within this framework a natural choice for the passing score x_0 is the one which minimizes

$$c_1\alpha + c_2\beta, \tag{1.3}$$

the Bayes risk. An index such as the proportion of agreement, is of little help in the search for an optimal passing score since we can guarantee that its maximum value of one will be attained simply by passing (or failing) every examinee. This is not to say the indexes such as the proportion of agreement or Cohen's Kappa have little or no value. Indeed, these indexes are important since, at a minimum, we want to make consistent decisions across comparable mastery tests. The advantage of α and β is that they provide a direct indication of how certain we can be that a correct decision is being made when trying to decide whether ζ is above or below ζ_0. For still more illustrations of this point, the reader is referred to Huynh (1976), Van der Linden and Mellenbergh (1977) and Wilcox (1977).

Given that it is desirable to know the values of α and β, it is natural to want to know whether their value is small regardless of the actual form of the true score distribution. With this goal in mind, we consider situations which yield upper and lower bounds for both α and β but which make no assumption about the form of $g(\zeta)$.

2. An Upper Bound as a Function of n

Before describing our main results, we note that an upper bound to α and β is readily derived when the binomial error model (Lord and Novick, 1968, Chapter 23) is assumed to hold. In other words, we are assuming that the conditional distribution of observed scores for an examinee is given by expression (1.1). From Wilcox (in press) it follows immediately that

$$\alpha \leq \sum_{x=x_0}^{n} f(x|\zeta = \zeta_0) \tag{2.1}$$

and

$$\beta \leq \sum_{x=0}^{x_0-1} f(x|\zeta = \zeta_0). \tag{2.2}$$

We observe that from a theoretical point of view, the assumption that $f(x|\zeta)$ is a binomial probability function has been criticized by several writers when an item sampling model applies (Hambleton et al., 1978; Lord and Novick, 1968, Chapter 23; Lord, 1965; Meredith and Kearns, 1973). The binomial error model would seem to deserve serious consideration in practice since even more restrictive models give a good fit to data (Keats and Lord, 1962; Lord, 1965). Nevertheless, one might prefer a more general probability function for describing the conditional distribution of observed scores. Lord (1965) as well as Lord and Novick (1968, Chapter 23) suggest that a two-term approximation to the compound bionomial be used. Results reported by Wilcox (in press) suggest that when this more general probability function is adopted, the "intuitively obvious" upper bounds to α and β given by expressions (2.1) and (2.2) will still hold. However, a rigorous proof that this is the case remains to be found.

3. Upper and Lower Bounds to α and β That are a Function of the First Two Moments of the True Score Distribution

Is it possible to improve upon the upper bounds on α and β given by expressions (2.1) and (2.2) without making any assumption about the form of $g(\zeta)$? In many cases, the answer is yes.

For notational convenience we let:

A_1 = the event $x \geq x_0$

A_1^c = the event $x < x_0$

A_2 = the event $\zeta \geq \zeta_0$

A_2^c = the event $\zeta < \zeta_0$.

The intersection of two events is denoted by the juxtaposition of the corresponding symbols. Thus, A_1A_2 represents the event $x \geq x_0$ and $\zeta \geq \zeta_0$, i.e., a correct-positive decision. We begin by deriving lower bounds to $Pr(A_1A_2)$ and $Pr(A_1^cA_2^c)$.

Let μ and σ^2 represent the mean and variance of the true scores of the examinees. In practice μ and σ^2 are unknown; however, they may be estimated as follows: Let $x_1,\ldots,x_k$ be the observed scores of k randomly selected examinees taking an n-item test. For the binomial error model (Lord and Novick, 1968, Chapter 23)

$$\hat{\mu} = (kn)^{-1}\Sigma x_i$$

$$\hat{\sigma}^2 = \hat{\mu}_1 - \hat{\mu}^2$$

may be used as an estimate of μ and σ^2 where

$$\hat{\mu}_1 = [kn(n-1)]^{-1} \Sigma(x_i^2 - x_i).$$

If a two-term approximation to the compound bionomial distribution is preferred, we still use $\hat{\mu}$ to estimate μ but we replace $\hat{\sigma}^2$ with

$$\hat{\sigma}_1^2 = \sigma_X^2 - (n-2d)\ \hat{\mu}(1-\hat{\mu})/[n(n-1)+2d]$$

where,

$$d = \frac{n^2(n-1)\sigma_\pi^2}{2[\mu_X(n-\mu_X)-\sigma^2-n\sigma_\pi^2]},$$

σ_X^2 and μ_X are the variance and mean of observed scores and where σ_π^2 is the variance of the item difficulties.

Let

$$\xi = \begin{cases} \zeta_0, & \mu < \zeta_0 \\ \mu, & \zeta_0 \leq \mu \leq 1 \end{cases}$$

and

$$U = \begin{cases} \dfrac{\sigma^2}{\sigma^2+(\xi-\mu)}, & \text{if } 0 < \sigma^2 \leq m \\ (\mu(1-\mu)-\sigma^2)/(1-\zeta_0)\zeta_0, & \text{otherwise} \end{cases}$$

where

$$m = \max\{\mu(\zeta_0-\mu), (\mu-\zeta_0)(1-\mu)\} \qquad (3.3)$$

It follows from results given by Skibinsky (1977) that

$$\Pr(A_2) \le U. \qquad (3.4)$$

From the Bonferroni inequality (see, e.g., Miller, 1966, p. 8),

$$\Pr(A_1^cA_2^c) \ge 1 - \Pr(A_1) - \Pr(A_2) \qquad (3.5)$$

which, together with (3.4) implies that

$$\Pr(A_1^cA_2^c) \ge 1 - U - \Pr(A_1). \qquad (3.6)$$

The proportion of examinees passing the test serves as an estimate of $\Pr(A_1)$. Thus, we have an estimate of a lower bound on $\Pr(A_1^cA_2^c)$.

In some cases the lower bound to $\Pr(A_1^cA_2^c)$ will be close to one which implies that both α and β are relatively small since both are less than or equal to $1 - \Pr(A_1^cA_2^c)$. In particular, (3.6) implies that

$$\alpha + \beta \le U + \Pr(A_1). \qquad (3.7)$$

A lower bound on $\Pr(A_1A_2)$ can also be derived by replacing ζ in (3.2) with

$$\xi_1 = \begin{cases} \mu, & \mu < \zeta_0 \\ \zeta_0, & \zeta_0 < \mu \le 1 \end{cases}$$

The resulting value of U, say U_1, is such that

$$\Pr(A_2^c) \le U_1.$$

Thus,

$$\Pr(A_1A_2) \ge 1 - U_1 - \Pr(A_1^c).$$

An upper bound for both α and β is readily derived as follows: Since $A_1A_2^c$ is a subset of both A_1 and A_2^c,

$$\alpha \leq Pr(A_1), \alpha \leq Pr(A_2^c), \text{ and so}$$

$$\alpha \leq \min[Pr(A_1), U_1]. \qquad (3.8)$$

Similarly,

$$\beta \leq \min[Pr(A_1^c), U] \qquad (3.9)$$

We conclude this section by describing lower bounds on both α and β. For ζ we have that

$$\begin{aligned} \alpha &= Pr(A_1A_2^c) \\ &\geq 1 - Pr(A_1^c) - Pr(A_2) \\ &\geq 1 - U - Pr(A_1^c). \qquad (3.10) \end{aligned}$$

for similar reasons

$$\beta > 1 - U_1 - Pr(A_1). \qquad (3.11)$$

An Upper Bound to α and β Assuming That The Binomial Error Model Holds

In the previous section we described upper and lower bounds to α and β which depend only on our ability to estimate the first and second moments of the true score distribution. As noted above, such estimates are readily available when the conditional distribution of observed scores for any examinee is given by a two-item approximation to a compound binomial distribution. As shown by Rutherford and Krutchkoff (1967) such estimates are also available for a wide variety of situations.

In this section we indicate how the inequalities (3.10) and (3.11) might be improved upon when the binomial error model is assumed to hold.

Since

$$\alpha = \Pr(A_1 A_2^c)$$

$$= \Pr(A_2^c)\Pr(A_1 | A_2^c)$$

it follows that

$$\alpha \le U_1 \sum_{x=x_0}^{n} f(x|\zeta<\zeta_0). \qquad (4.1)$$

From known properties about the binomial probability function (see Wilcox, in press; Fhaner, 1974) which can be derived from results given by Lehmann (1959, Chapter 3), we have that

$$\sum_{x=x_0}^{n} f(x|\zeta<\zeta_0) \le \sum_{x=x_0}^{n} \binom{n}{x} f(x|\zeta=\zeta_0).$$

Hence,

$$\alpha \le U_1 \sum_{x=x_0}^{n} \binom{n}{x} \zeta_0^x (1-\zeta_0)^{n-x} \qquad (4.2)$$

For similar reasons, it can be seen that

$$\beta \le U \sum_{x=0}^{x_0-1} \binom{n}{x} \zeta_0^x (1-\zeta_0)^{n-x} \qquad (4.3)$$

It was suggested to the author that a theorem by Markov (recently applied by Lord and Stocking, 1976) might be applied to obtain bounds on α and β. It should be pointed out, however, that the conditions of this theorem, as described by Lord and Cressie (1975), are not satisfied in general. To see this, it is sufficient to observe the first derivative of $h_1(\zeta) = \sum_{x=0}^{x_0} \binom{n}{x} \zeta^x (1-\zeta)^{n-x}$ with respect to ζ is negative.

The derivative of $h_2(\zeta) = 1-h_1(\zeta)$ is positive, but the second derivative can be negative. (See, however, Karlin and Shapley, 1953.)

5. Another Application

As an illustration and another application on how the upper bounds to α and β might be used, we consider the problem of determining how many items to include on a mastery test. For technical reasons (Fhaner, 1974; Wilcox, in press) it is necessary to formulate the above model of mastery testing in a slightly different fashion. In addition to the criterion score ζ_0, we specify the constants ζ_1 and ζ_2 where $\zeta_1<\zeta_0<\zeta_2$. If $\zeta_1<\zeta<\zeta_2$ we say that the examinee is classified correctly with probability one since there is negligible loss if a misclassification is made. However, if $\zeta\leq\zeta_1$ or $\zeta\leq\zeta_2$, we want to be reasonably certain that a correct decision is made. More specifically, we want to choose n, the test length, so that the probability of both a false-positive and false-negative decision is reasonably small. We specify this criterion by requiring

$$\alpha \leq \alpha^* \tag{5.1}$$

and

$$\beta \leq \beta^* \tag{5.2}$$

where α^* and β^* are given constants. For this model of a mastery test we now have that $\alpha = \Pr(x \geq x_0,\ \zeta \leq \zeta_1)$ and $\beta = \Pr(x < x_0,\ \zeta \geq \zeta_2)$. If $\zeta_0 = \zeta_1 = \zeta_2$, it may be impossible to choose n so that (5.1) and

(5.2) are satisfied. The solution given by Fhaner (1974) is to choose the smallest n so that simultaneously

$$\sum_{n=x_0}^{n} \binom{n}{x} \zeta_1^{x} (1-\zeta_1)^{n-x} \le \alpha^* \qquad (5.3)$$

and

$$\sum_{x=0}^{x_0-1} \binom{n}{x} \zeta_2^{x} (1-\zeta_2)^{n-x} \le \beta^* . \qquad (5.4)$$

For the sake of illustration, suppose $\alpha^*=.12$, $\beta^*=.04$, $\zeta_1=.7$, $\zeta_2=.9$, $\mu=.945$ and $\sigma^2=.003$. To determine an appropriate test length we first compute upper bounds to α and β. Since false-positive and false-negative decisions are now defined in terms of ζ_1 and ζ_2 rather than ζ_0, the expressions for ζ, ζ_1 and m are no longer appropriate. To determine an upper bound on $\Pr(\zeta \ge \zeta_2)$, we now use

$$\xi_2 = \begin{cases} \zeta_2, & \mu < \zeta_2 \\ \mu, & \zeta_2 \le \mu \le 1 \end{cases}$$

$$m = \max[\mu(\zeta_2-\mu),\ (\mu-\zeta_2)(1-\mu)]$$

and we replace U with

$$U' = \begin{cases} \dfrac{\sigma^2}{\sigma^2+(\xi_2-\mu)^2}, & \text{if } 0 < \sigma^2 \le m \\ (\mu(1-\mu)-\sigma^2)/((1-\zeta_2)\zeta_2), & \text{otherwise.} \end{cases}$$

In our example, m=.0025, ζ_2=.945 and the resulting value of U is .544. If we assume that the binomial error model holds, we may apply the arguments of the previous section which imply that

$$\beta \leq U' \mathop{}\limits_{x=0}^{x_0-1} \binom{n}{x} \zeta_2^x (1-\zeta_2)^{n-x} \tag{5.5}$$

$$= .54 \sum_{x=0}^{x_0-1} \binom{n}{x} .9^x \, .1^{n-x}.$$

As for α, we use

$$\zeta_3 = \begin{cases} \mu, & \mu < \zeta_1 \\ \zeta_1, & \zeta_1 \leq \mu \leq 1 \end{cases}$$

$$m = \max [\mu(\zeta_1-\mu), (\mu-\zeta_1)(1-\mu)]$$

and the upper bound to α is

$$U_1' = \begin{cases} \dfrac{\sigma^2}{\sigma^2+(\xi_3-\mu)^2}, & \text{if } 0 < \sigma^2 \leq m \\ (\mu(1-\mu)-\sigma^2)/((1-\zeta_1)\zeta_1), & \text{otherwise.} \end{cases}$$

For the case at hand, $m=.0135$ and $\xi_3=.7$. The resulting value of U_1' is .0476 and so

$$\alpha \leq U_1' \sum_{x=x_0}^{n} \binom{n}{x} \zeta_1^x (1-\zeta_1)^{n-x} \tag{5.6}$$

$$= .048 \sum_{x=x_0}^{n} \binom{n}{x} .7^x \, .3^{n-x}.$$

We evaluated these upper bounds for increasing values of n with x_0 chosen to be the smallest integer such that $x_0/n \geq \zeta_0$. For this particular case, the smallest value of n required so that both (5.1) and (5.2) are satisfied is $n=10$ with $x_0=8$. If inequalities (5.3) and (5.4) rather than (5.5) and (5.6) are used, we require $n=25$. Thus, we are able to justify a substantially shorter test than $n=25$ without making any assumption about the form of the true score distribution.

As a final illustration, we analyze some test data reported by Huynh (1976). A five-item arithmetic test was administered to 91 students whose test scores $x=0$, 1, 2, 3, 4 and 5 have frequencies 4, 14, 9, 17, 21, 26, respectively. The mean and variance of the true score distribution are estimated to be $\mu=.653$ and $\hat{\sigma}^2=.065$. The resulting value of U' is .52. Also, $U_1'=.77$. Thus, letting ζ_0, ζ_1, ζ_2 retain the same values as before,

$$\beta \leq .51 \sum_{x=0}^{x_0-1} \binom{n}{x} .9^x \, .1^{n-x}$$

and

$$\alpha \leq .76 \sum_{x=x_0}^{n} \binom{n}{x} .7^x \, .3^{n-x}.$$

Setting $\alpha^*=.12$ and $\beta^*=.09$, the minimum required test length is $n=19$ with $x_0=16$.

Concluding Remarks

In summary, we have indicated methods of obtaining upper and lower bounds to both α and β which make no assumptions about the form of the true score distribution. The first method depends only on our ability to determine the mean and variance of the true score distribution. As indicated above, such estimates are readily available when the binomial or compound binomial error model is assumed. The second method is based on the binomial error model which is frequently used to describe a mastery test. As was illustrated, the resulting upper bounds may be particularly useful when determining the length of the test.

ACKNOWLEDGEMENTS

The work upon which this publication is based was perfomed pursuant to a contract with the National Institute of Education, Department of Health, Education and Welfare.

Points of view or opinions stated do not necessarily represent official NIE position or policy.

REFERENCES

Brennan, R. L., & Kane, M.T. An index of dependability for mastery tests. Journal of Educational Measurement, 1977, 14, 277-289.

Fhaner, Stig. Item sampling and decision making in achievement testing. British Journal of Mathematical and Statistical Psychology, 1974, 27, 172-175.

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, J., & Coulson, D. B. Criterion-referenced testing and measurement: A review of technical issues and developments. Review of Educational Research, 1978, 48, 1-47.

Harris, C. W. Some technical characteristics of mastery tests. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.), Problems in criterion-referenced measurement (CSE Monograph Series in

Evaluation, No. 3). Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Huynh, H. On the reliability of decisions in domain-referenced testing. Journal of Educational Measurement, 1976, 13, 253-264.

Karlin, S., & Shapley, L. S. Geometry of moment spaces. Memoirs of the American Mathematical Society, 1953, 12, 1-93.

Keats, J. A., & Lord, F. M. A theoretical distribution for mental test scores. Psychometrika, 1962, 27, 59-72.

Lehmann, E. L. Testing statistical hypotheses. New York: John Wiley, 1959.

Livingston, S. A. Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 1972, 9, 13-26.

Lord, F. M. A strong true-score theory, with applications. Psychometrika, 1965, 30, 239-370.

Lord, F. M. Estimating true-score distributions in psychological testing (an empirical Bayes estimation problem). Psychometrika, 1969, 34, 259-299.

Lord, F. M., & Cressie, N. An empirical Bayes procedure for finding an interval estimate. Sankhya, 1975, 37, Series-B, 1-9.

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

Lord, F. M., & Stocking, M. L. An interval estimate for making statistical inferences about true scores. Psychometrika, 1976, 41, 79-88.

Maritiz, J. S. Smooth empirical Bayes estimation for continuous distributions. Biometrika, 1967, 54, 435-450.

Meredith, W., & Kearns, J. Empirical Bayes point estimates of latent trait scores without knowledge of the trait distribution. Psychometrika, 1973, 38, 533-554.

Miller, R. G. Simultaneous statistical inference.. New York: McGraw-Hill, 1966.

Novick, M. R., & Lewis, C. Prescribing test length for criterion-referenced measurement. In C. W. Harris, M. C. Alkin, & W. J. Popham (Eds.), Problems in criterion-referenced measurement (CSE Monograph Series in Evaluation, No. 3). Los Angeles: Center for the Study of Evaluation, University of California, 1974.

Rutherford, J. R., & Krutchkeff, R. G. The empirical Bayes approach: Estimating the prior distribution. Biometrika, 1967, 54, 326-328.

Skibinsky, M. The maximum probability of an interval when the mean and variance are known. Sankhya, 1977, 39, Series A, 144-159.

Swaminathan, H., Hambleton, R. K., & Algina, J. Reliability of criterion-referenced tests: A decision-theoretic formulation. Journal of Educational Measurement, 1974, 11, 263-267.

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

Wilcox, R. R. Estimating the likelihood of a false-positive or false-negative decision with a mastery test: An empirical Bayes appoach. Journal of Educational Statistics, 1977, 2, 289-307.

Wilcox, R. R. Applying ranking and selection techniques to determine the length of a mastery test. Educational and Psychological Measurement, in press.