DOCUMENT RESUME

ED 310 129 TM 013 702

AUTHOR	van der Linden, Wim J.
TITLE	Passing Score and Length of a Mastery Test: An Old Problem Appraoched Anew. Twente Educational Report Number 11.
INSTITUTION	Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE	Sep 80
NOTE	28p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Classification; *Cutting Scores; Error Patterns; Guessing (Tests); *Latent Trait Theory; *Mastery Tests; Mathematical Models; Testing Problems; *Test Length
IDENTIFIERS	Binomial Error Model; *Decision Theory; Optimization; Parametric Analysis; Sequential Testing

ABSTRACT

A classical problem in mastery testing is the choice of passing score and test length so that the mastery decisions are optimal. This problem has been addressed several times from a variety of viewpoints. In this paper, the usual indifference zone approach is adopted, with a new criterion for optimizing the passing score. Specifically, manipulation of probabilities of misclassification of masters versus non-masters is not incorporated into the scheme. Rather, explicit parameters are introduced to account for differences in loss between misclassifying a true master and a non-master. It appears that, under the assumption of the binomial error model, this approach yields a linear relationship between the optimal passing score and test length. The means by which different losses for both decision errors and a known base rate can be incorporated in the procedure are outlined, and the means by which a correction for guessing can be applied are described. Results are related to findings obtained in sequential probability ratio testing for binomial populations and in the latent class approach to mastery testing. (TJH)

TWENTE EDUCATIONAL REPORT NUMBER 11

PASSING SCORE AND LENGTH OF A MASTERY TEST: AN OLD PROBLEM APPROACHED ANEW

W. J. van der Linden

TWENTE UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF APPLIED EDUCATION
P.O. BOX 217
7500 AE ENSCHEDE THE NETHERLANDS

COLOFON:

Publisher : Department of Applied Education
Twente University of Technology

Typist : Paula Achterberg

Reproduction: Central Reproduction Department of the University

Edition : 60

Passing Score and Length of a Mastery Test:

An Old Problem Approached Anew

Wim J. van der Linden

Twente University of Technology

Enschede, The Netherlands

Twente Educational Report No. 11. Enschede, The Netherlands:

Twente University of Technology, Department of Applied Education,

September, 1980.

Abstract

A classical problem in mastery testing is the choice of passing score and test length so that the mastery decisions are optimal. This problem has been addressed several times from a variety of view-points. In this paper the usual indifference zone approach is adopted with a new criterion for optimizing the passing score. It appears that, under the assumption of the binomial error model, this yields a linear relationship between optimal passing score and test length, which subsequently can be used in a simple procedure for optimizing the test length. It is indicated how different losses for both decision errors and a known base rate can be incorporated in the procedure, and how a correction for guessing can be applied. Finally, the results in this paper are related to results obtained in sequential testing and in the latent class approach to mastery testing.

Key-words: Mastery Testing, Passing Score, Test Length, Domain-Referenced Testing, Decision Theory.

Passing Score and Length of a Mastery Test:

An Old Problem Approached Anew

The notion of a mastery test has arisen in the context of modern learning strategies such as learning for mastery and individualized instruction, where at several points in the instructional process decisions have to be made whether students have reached certain learning objectives or not. In most instances, this involves the administration of criterion-referenced tests and the use of decision rules assuming the form of passing scores on the test. Students with test scores exceeding the passing score are considered having reached the learning objectives (the "masters"); they are allowed to proceed with the unit or to take up a subsequent course. Students below the passing score (the "nonmasters") have to relearn the unit and to prepare for a new test.

A usual conceptualization in the area of mastery testing is that of tests as samples randomly drawn from a domain of tasks covering a well-defined learning objective. Mostly, the concern is then with the proportion of correct item responses, Π, say, to be expected when the entire domain would have been administered. Let π_m denote the passing score on this domain score variable ("mastery score"), X the number of items correct, and c the passing score on the test. A student is a true master if $\Pi > \pi_m$ and a nonmaster otherwise, but mastery is declared if $X \geq c$ and nonmastery if $X < c$. A classical problem in mastery testing is to choose a value n^* for the test length n and a value c^* for the passing score on the test c such that, for a given value of π_m, the mastery decisions are optimal.

Several authors have addressed the above problem, all using one of the binomial models for relating test scores, X, to domain scores, Π. Millman (1972, 1973), for example, assumes that the simple binomial model

$$(1) \qquad p(x) = \binom{n}{x}\pi^{x}(1-\pi)^{n-x}$$

can be used for this purpose and provides tables which for a chosen value of π_m, test length, and passing score, display the probability that a person with a given domain score is classified correctly or incorrectly. Using these tables, it is possible to optimize passing score and test length simultaneously for a selected domain score. Comparable approaches have been followed by Klauer (1972) and Kriewall (1972).

Fhanér (1974) introduced the notion of an indifference zone in the present problem. An indifference zone arises when the mastery score, π_m, is replaced by an interval, (π_0, π_1), so that examinees with $\Pi \geq \pi_1$ are considered a master, those with $\Pi \leq \pi_0$ a nonmaster, and we are indifferent with respect to examinees with $\pi_0 < \Pi < \pi_1$. The interval may be taken symmetric about π_m, but this is not necessary. For true masters and nonmasters the probability of a misclassification is largest for domain scores π_1 and π_0, respectively. Fhanér proposed as a solution to choose the minimum value of n and a value of c for which the probabilities of misclassification

$$(2) \qquad \alpha = \sum_{x=c}^{n} \binom{n}{x}\pi_0^{x}(1-\pi_0)^{n-x}$$

and

(3) $$\beta = \sum_{x=0}^{c-1} \binom{n}{x} \pi_1^{\,x} (1-\pi_1)^{n-x}$$

are not larger than preassigned values α^* and β^*. This is no closed-form solution, and binomial tables must be entered to find the optimal values of n and c. It is possible to use a normal approximation, however, and in that case a closed-form solution is obtained (Fhanér, 1974). Wilcox (1976) has adopted the same approach and has suggested computer search routines using the incomplete beta function to find the solution for the binomial case.

As van den Brink and Koele (1980) have pointed out, it is possible to correct the above solution for the possibility of guessing on multiple-choice or ture-false items. To perform this correction, they adopt the knowledge or random guessing model and simply replace the parameter π in the binomial model by

(4) $$\pi_g + g(1 - \pi_g),$$

π_g being the domain score corrected for guessing and g the guessing parameter. For the latter they suggest the use of the reciprocal of the number of item alternatives, q^{-1}.

Novick and Lewis (1974) and de Gruijter (1979) present a Bayesian approach to the present problem extending the model with the beta distribution as a prior for the binomial parameter π.

As several authors (e.g., Wilcox, 1976) have noted, different preassigned values α^* and β^* can be selected to allow for differences in loss between misclassifying a true master and a nonmaster. In this paper, we will take a somewhat different approach and represent possible differences in loss not indirectly -- by manipulating both probabilities of misclassification -- but via the introduction of explicit parameters. But before doing so, we will prepare this approach and consider the case of a decision-maker who is indifferent to both classification errors. It appears that in this case there is a simple linear relation between the optimal passing score and test length. This can be utilized to find the solution in this particular case but also plays an important part in the more general case of different losses.

Indifference to Both Classification Errors

In the Wilcox solution a number P^*, $1/2 < P^* < 1$, is chosen, and next values for n and c are determined so that the value for n is minimal and both α and β are not larger than P^*. The fact that the same restriction P^* is imposed on α and β reflects that both misclassification errors are considered equally serious. But this can also be expressed in another way; If there is equal loss in misclassifying a true master and a nonmaster, only the size of the probability of misclassification and not the type of misclassification concerns us. If so, it seems natural not to look for values of n and c for which (2) and (3) are both below the same predetermined number P^* but for values for which their average

$$(5) \qquad \left[\sum_{x=0}^{c-1} \binom{n}{x}\pi_1^{\,x}(1-\pi_1)^{n-x} + \sum_{x=c}^{n} \binom{n}{x}\pi_0^{\,x}(1-\pi_0)^{n-x}\right] /2$$

meets such a requirement.

We first assume n to be fixed and look for the value of c minimizing the new target function defined in (5). The constant factor 1/2 may thereby be ignored. Adding terms to the first sum in (5) and subtracting these from the second yields

$$(6) \qquad \sum_{x=0}^{n} \binom{n}{x} \pi_1^{\,x} (1-\pi_1)^{n-x} - \sum_{x=c}^{n} \binom{n}{x} \left[\pi_1^{\,x}(1-\pi_1)^{n-x} - \pi_0^{\,x}(1-\pi_0)^{n-x} \right].$$

Since the first sum is equal to 1, the value of c for which (6) is minimal depends only on the bracketed factor in the second sum. We know that the binomial probability function has monotone likelihood in x (Ferguson, 1967, sect. 5.2), which implies that the ratio $\pi_1^{\,x}(1-\pi_1)^{n-x}/\pi_0^{\,x}(1-\pi_0)^{n-x}$ is monotone increasing in x. So there is a value of x for which the sign of this factor changes from negative to positive. If we set c equal to this value, the second sum in (6) contains all positive terms and (6) is minimal. Thus, (6) is minimal for the value c^* obeying

$$\pi_0^{c^*}(1-\pi_0)^{n-c^*} = \pi_1^{c^*}(1-\pi_1)^{n-c^*}.$$

Logarithmizing both sides and simplifying, it appears that

$$(7) \qquad \frac{c^*}{n} = \frac{\ln \dfrac{1-\pi_0}{1-\pi_1}}{\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}} \, .$$

This result is most interesting: The left-hand side is the optimal value of c expressed as a relative score. The right-hand side is a constant which is independent of test length and only a function of the boundary values of the indifference zone. Thus, whenever an indifference zone is established, we can easily compute (7) from its boundary values and immediately know the optimal passing-score for any test length.

We now use the linear relation between c^* and n to find an optimal value, n^*, for the latter and thereby follow a simple procedure analogous to the one in the Wilcox solution. First, a number P^*, $1/2 < P^* < 1$, is selected which serves as an upper bound to (5). The value n^* is determined as the smallest value of n for which (5) is not larger than P^*. Second, the ratio c^*/n is computed from the indifference zone via (7). Third, a trial value for n^* is chosen, and (de)cumulative binomial tables are entered with this value and the implied value of c^* to compute (5). Fourth, if this computation yields a value smaller than P^*, a lower trial value for n^* is selected and step three is carried out again. For values of (5) larger than P^*, a larger trial value is selected. This proces is repeated until the smallest value of n is found for which (5) is not larger than P^*. This is n^*.

In the above procedure, trial values for n^* may be chosen not involving an integer value for c^*. As follows from (6), in that case the first integer value above c^* must be used. (The choice of integer value below c^* would imply adding negative terms to the second sum in (6) making this suboptimal.)

Although binomial tables for values of n up to 150 are available (Ordnance Corps, 1952), tables in most text-books do not go further than n = 20. It is known, however, that indifference zone methods are rather conservative and that, for strong requirements on (5) or narrow indifference zones, values of n^* larger than 20 can be expected. (For an impression, see Table 1 in Fhanér, 1974). When longer tests are needed and no special tables are available, one has to resort to a computer or a calculator for the above procedure. The programming involved is compartively simple, though, and some calculators posses even hardware facilities for binomial probabilities.

Another possibility is to use an approximation to the binomial distribution function which is simple enough for hand calculation. A straight-forward approximation, based on the central limit theorem, is to replace (5) by

$$(8) \qquad \left[\Phi\left\{\frac{c/n - \pi_1}{[\pi_1(1-\pi_1)/n]^{1/2}}\right\} + \Phi\left\{\frac{\pi_0 - c/n}{[\pi_0(1-\pi_0)/n]^{1/2}}\right\}\right]/2,$$

Φ denoting the standard normal distribution function. Using this normal approximation, we need not compute (8) completely for each trial value for n^*. Substituting (7) into (8), it appears that this can be written as

$$(9) \qquad \left[\Phi(an^{1/2}) + \Phi(bn^{1/2})\right]/2,$$

with

$$(10) \qquad a \equiv \left[\pi_1(1-\pi_1)\right]^{-1/2} \left[\ln \frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}\right]^{-1} \ln\left\{\left(\frac{\pi_0}{\pi_1}\right)^{\pi_1} \left(\frac{1-\pi_0}{1-\pi_1}\right)^{1-\pi_1}\right\}.$$

and

$$(11) \qquad b \equiv \left[\pi_0(1-\pi_0)\right]^{-1/2} \left[\ln \frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}\right]^{-1} \ln\left\{\left(\frac{\pi_1}{\pi_0}\right)^{\pi_0} \left(\frac{1-\pi_1}{1-\pi_0}\right)^{1-\pi_0}\right\}.$$

The last two expressions depend only upon the indifference zone boundaries, π_0 and π_1. Once a and b are calculated from these boundaries, the iterative procedure can be applied directly to (9). The reader who is familiar with the cumulative normal distribution can use well-known reference values as, for example, $\Phi(-1) = .1587$, $\Phi(0) = .5000$, and $\Phi(1) = .8413$, to quickly establish whether trial values for n^* meet the restriction P^* imposed on (9).

It is known that the normal approximation in (8) can be less accurate, notably when it is used for approximating tail probabilities of skew binomial distributions. This situation arises when both indifference zone boundaries are larger than .70, say, and strong requirements are imposed on (5). A variety of better approximations are given in Molenaar (1973). When choosing one, we are, however, faced with a dilemma. Generally, the more accurate the approximation, the more cumbersome its calculation. Most approximations can be used in combination with the iterative procedure for test length determination only if one has access to a computer, but in that case the procedure can as well be carried out directly with (5). A reasonable accurate approximation to (5), which is not too complex, is

$$(12)\qquad \left[\Phi\left\{(4c+3)^{1/2}(1-\pi_1)^{1/2} - (4n-4c-1)^{1/2}\pi_1^{1/2}\right\} + \Phi\left\{(4n-4c-1)^{1/2}\pi_0^{1/2} - (4c+3)^{1/2}(1-\pi_0)^{1/2}\right\}\right]/2$$

(for a discussion and some numerical results, see Molenaar, 1973, eq. 3.20, pp. 111 - 114). It is recommended that this approximation be used when strong requirements are imposed on (5) and the choice of the indifference zone entails skew binomial distributions. In order to reduce the calculations, a good strategy is to use (8) until it gives a solution and next to use (12) to find out whether it can be improved.

As noted before, the worth of the procedure proposed in this paper lies in the ease with which binomial tables can be consulted. It utilizes a simple linear relation between optimal passing score and test length so that for each trial value for n^* results for only one passing score need to be obtained. In the other indifference zone methods several trial values for c^* must be tested for each trial value for n^* until the combination (c^*, n^*) meeting the requirements is found. The fact that c^* has a simple relation to n gives (7) value in its own right. It can be used, for example, to find optimal passing scores on new tests when test length has to be fixed for some practical reason, or to establish whether passing scores that have already been used in practice satisfy the optimality condition considered in this paper. As will be shown in the next section, another advantage of the present procedure is the possibility of incorporating different losses for false positive and false negative decisions.

Different Losses for Both Decision Errors

So far it has been assumed that the loss incurred for a false positive decision (granting the mastery status to a nonmaster) is equal to the loss for a false negative decision (granting the nonmastery status to a master). We now assume that both losses take on different values and incorporate this in the procedure by replacing the average in (5) by the weighted average

$$(13) \qquad \left[\ell_1 \sum_{x=0}^{c-1} \binom{n}{x}\pi_1^{\,x}(1-\pi_1)^{n-x} + \ell_0 \sum_{x=c}^{n} \binom{n}{x}\pi_0(1-\pi_0)^{n-x}\right](\ell_0 + \ell_1)^{-1},$$

where ℓ_0 is the loss of misclassifying a nonmaster and ℓ_1 of a master.

Following the same derivation as before, (13) is minimal for the value $c°$ given by

$$(14) \qquad \frac{c°}{n} = \frac{\ln \dfrac{1-\pi_0}{1-\pi_1}}{\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}} + \frac{\ln \lambda}{n\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}},$$

λ denoting the loss ratio ℓ_0/ℓ_1.

Comparing (14) with (7), several things can be noted: The right-hand side of (14) displays an additive structure consisting of two different parts. The first part is equal to (7), and thus again a constant dependent only upon the indifference zone boundaries; the second part represents the influence of the loss ratio on the optimal passing score and is, as opposed to the first part, dependent on test length. When different losses for both misclassifications have to be reckoned with, the optimal passing score (expressed as a relative score) is thus equal to the one for the case of equal losses plus a test length dependent correction. For loss ratios larger than one this correction is positive, while it is negative for ratios smaller than one.

It should be noted that in the second term of (14) test length figures only in the denominator. This implies that the longer the test is the smaller the absolute size of the influence of the loss ratio on the optimal relative passing score will be. Table 1 shows this for loss ratio values

Insert Table 1 about here

from 1:4 to 4:1. For example, the relative passing score on a 10-item test must be raised by .173 to account for a loss ratio $\lambda = 3$, while this is only .035 for a 50-item test.

In view of the determination of optimal test length, it is helpful to rewrite (14) into

$$(15) \qquad c^{\circ} = \frac{\ln \dfrac{(1-\pi_0)}{(1-\pi_1)}}{\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}} n + \frac{\ln \lambda}{\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}} .$$

This expression again shows a linear relation between c° and n. It has (7) as slope and this time a non-zero intercept which is a function of the loss ratio. Table 2 shows values of this intercept for loss ratio values from

Insert Table 2 about here

1:4 to 4:1 and indifference zones which can often be encountered in the practice of mastery testing. The entries in this table are thus to be added to the optimal passing score for the equal loss case, c^*, when loss ratios unequal to one are used.

The determination of optimal test length proceeds along the same lines as in the previous section. First, the number $P°$ is selected as the upper bound to (13). Its minimum value is no longer equal to 1/2. (In the previous section, this value could always be realized by randomly assigning the examinees to the mastery and nonmastery state.) Now it is equal to

$$\max\left\{\ell_0/(\ell_0+\ell_1),\ \ell_1/(\ell_0+\ell_1)\right\}, \tag{16}$$

these two values being obtained by always assigning the examinees to the mastery and nonmastery state, respectively. Second, the slope and intercept in (15) are computed. (For the latter Table 2 can be used.) Third, a trial value for $n°$, the optimal test length, is chosen, and the associated value of $c°$ is computed from (15). Binomial tables are entered with the value of $c°$ to obtain the (de)cumulative probabilities in (13), and once these are found (13) is computed. Fourth, the value computed for (13) is compared with $P°$. If it is smaller (larger) than $P°$, a lower (larger) trial value for $n°$ is selected, and the previous step is repeated. The process is stopped when the smallest value of n is met for which (13) is not larger than $P°$. This is $n°$.

When a normal approximation is needed, we replace (8), analogous to (13), by a weighted average of (de)cumulative normal probabilities. Substitution of (14) in this new target function results in

$$(17) \qquad \left[\ell_1\Phi(an^{1/2} + cn^{-1/2}) + \ell_0\Phi(bn^{1/2} - dn^{-1/2})\right](\ell_0 + \ell_1)^{-1}$$

with a and b given by (10) and (11), respectively, and c and d by

$$(18) \qquad c \equiv \left[\pi_1(1-\pi_1)\right]^{-1/2}\left[\ln\frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}\right]^{-1}\ln\lambda$$

and

$$(19) \qquad d \equiv \left[\pi_0(1-\pi_0)\right]^{-1/2}\left[\ln\frac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}\right]^{-1}\ln\lambda$$

In the above procedure, we first compute a, b, c, and d from π_0, π_1, and λ, and next substitute our trial values for n° directly into (17). If necessary, we can use the approximation in (12) to find out whether the solution thus obtained can be made more accurate.

Incorporating a Known Base Rate

If a priori knowledge about the proportion of masters is available, for instance, from previous testing programs or experiences with comparable groups of students, it may be prudent to incorporate this in the decision procedure as well.

Ignoring the examinees in the indifference zone for a while, let μ denote the proportion of masters so that $1 - \mu$ equals the proportion of nonmasters. We now use μ and $1 - \mu$ as weights in our target function and replace (13) by

$$(20)\qquad \left[\mu\ell_1 \sum_{x=0}^{c-1} \binom{n}{x}\pi_1{}^x(1-\pi_1)^{n-x} + (1-\mu)\ell_0 \sum_{x=c}^{n} \binom{n}{x}\pi_0{}^x(1-\pi_0)^{n-x}\right]\left[\mu\ell_1 + (1-\mu)\ell_0\right]^{-1}.$$

Following the same derivation as before, the optimal passing score c' proves to be given by

$$(21)\qquad \frac{c'}{n} = \frac{\ln \dfrac{1-\pi_0}{1-\pi_1}}{\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}} + \frac{\ln \lambda}{n\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}} + \frac{\ln \dfrac{1-\mu}{\mu}}{n\ln \dfrac{\pi_1(1-\pi_0)}{\pi_0(1-\pi_1)}}$$

This result is equal to (14) extended with a term containing the base rate, $(1 - \mu)/\mu$. For base rate values larger than one, this term is positive, while is is negative for values smaller than one.

The roles played by $(1-\mu)/\mu$ and λ in (21) are fully identical. For a quantitative impression of the last term in (21), Tables 1 -2 can be consulted with $(1-\mu)/\mu$ substituted for λ.

To find the optimal test length in the present case with an explicit base rate, the same procedure as in the prévious section can be followed. Even the same formulae (and Table 2) may be used. This stems from the fact that the last two terms in (21) can be reduced to the same denominator, whereupon (21) has the same structure as (14). The only modifications needed are the substitution of $\lambda(1-\mu)/\mu$ for λ and the replacement of (16) by

$$(22)\qquad \max\left\{(1-\mu)\ell_0/[(1-\mu)\ell_0+\mu\ell_1],\ \mu\ell_1[(1-\mu)\ell_0+\mu\ell_1]\right\},$$

which now is the minimum value of the upper bound P' to (20).

Guessing

As noted earlier, van den Brink and Koele (1980) proposed to use the knowledge or random guessing model to correct Fhanér's (1974) approach for the possibility of guessing on multiple-choice or true-false items. The same can be done in the approach given in this paper. We then first establish the indifference zone on the ability scale corrected for guessing, i.e., as (π_{g0}, π_{g1}), and next apply transformation (4) to obtain the values (π_0, π_1) with which we enter the formulae given in this paper.

It should be noticed, however, that experience with the knowledge or random guessing model in item response theory shows guessing parameter values somewhat less than the reciprocal of the number of alternatives, q^{-1} (Lord, in press). For example, items with four alternatives typically result in values of .22 or .23 rather than .25. It is recommended that this be taken into account when setting the guessing parameter value.

Discussion

The results presented in this paper relate to results obtained in two other areas.

The first is the tent class approach to mastery testing. In this approach it is assumed that mastery and nonmastery are two latent states underlying the test score, each entailing different probabilities of a successful reply to the items. Im Emrick's latent class model (Davis, Hickman, and Novick, 1973, pp. 32-47; Emrick, 1971; Emrick and Adams, 1969; Fricke, 1974; Macready and Dayton, 1977) two latent success probabilities are assumed, one representing the mastery and the other the nonmastery state. Emrick and Adams (1969) give an optimal passing score which, although derived and presented in a different way, is quickly seen to be equivalent to c' given in (21).

This equivalence is only formal, however. In Emrick's model the latent success probabilities, which correspond with π_0 and π_1 in (21), must be estimated from the test data. (For a review of available estimation procedures, see van der Linden, 1980.). In this paper, π_0 and π_1 represent no latent classes and need not be estimated; they are boundary values of an indifference zone on the domain-score continuum which are set on educational grounds.

Fricke (1974) has given proofs that the correction of Emrick's passing score needed for loss ratio and base rates unequal to one are independent of the base rate and the loss ratio, respectively, and of the test length. However, this follows immediately from inspecting the structure of (21) which can be viewed as a linear decomposition of c'. Van der Linden (1978) has proposed a correction for guessing for Macready and Dayton's (1977) version of Emrick's model which corresponds with the correction for guessing proposed in the previous section.

The formal correspondence between Emrick's passing score and (21) suggests the use for Emrick's model of the procedure for test length optimization developed in this paper. The only difference is then, of course, that the success parameters π_0 and π_1, in (20) - (21) must be estimated before the procedure can be applied and that, consequently, no exact but estimated results are obtained.

The second area to which the results in this paper relate is Wald's sequential probability ratio test for binomial populations. Several expressions in Wald (1947) are reminiscent of the formulae given in this paper. For example, formula (15) is equivalent to the critical numbers in the test of $\pi \leq \pi_0$ against $\pi \geq \pi_1$ (Wald, 1974, eqs. 5.1.--5.2). The only exception is that the loss ratio λ is replaced by a ratio based on the probabilities of errors of type I and II. It must be borne in mind, however, that, just as in the previous case, this equivalance is only formal and that different interpretations are involved. In sequential testing test length, or, generally, the number of observations, is a random variable, and sampling is not stopped until one of the critical numbers is exceeded. The purpose of this paper was to find an optimal test length which is fixed prior to the test administration. It should be realized, however, that when sequential testing strategies are possible this is certainly worth considering, since substantial savings in the number of test items needed can be expected (Wald, 1947, sect. 3.6).

As a final comment, we recall that the use of the binomial model imposes certain restrictions on the test items. These are, however dependent on whether the item responses can be viewed as deterministic or stochastic responses (van der Linden, 1979). Either conception of item responses involves equal item difficulties. But, when the former conception is adopted, this condition can be avoided by giving separate samples of items to each student, whereas this cannot be done for the latter. When these conditions are not met, the simple binomial model used in this paper must be replaced by the compound binomial model. It is known, however, that this entails a smaller variance of test scores, so that the procedure of this paper will result in conservative estimates of optimal test length.

References

de Gruijter, D.N.M. On the minimum number of items for pass/fail decisions in criterion-referenced testing (Memorandum 514-79). Leiden, The Netherlands: Rijksuniversiteit Leiden, Bureau Onderzoek van Onderwijs, 1979.

Emrick, J.A. An evaluation model for mastery testing. Journal of Educational Measurement, 1971, 8, 321-326.

Emrick, J.A., & Adams, E.N. An evaluation model for individualized instruction (Report RC 2674). Yorktown Hts, N.Y.: IBM Thomas J. Watson Research Center, October 1969.

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

Fhanér, S. Item sampling and decision-making in educational testing. British Journal of Mathematical and Statistical Psychology, 1974, 27, 172-175.

Fricke, R. Zum Problem von Cut-off Formeln bei lehrzielorientierten Tests. Unterrichtswissenschaft, 1974, 3, 43-56.

Klauer, K.J. Zur Theorie und Praxis des binomialen Modells lehrzielorientierter Tests. In K.J. Klauer, R. Fricke, M. Herbig, H. Rupprecht, & F. Schott, Lehrzielorientierte Tests. Düsseldorf: Schwann, 1972.

Kriewall, T.E. Aspects of applications of criterion-referenced tests. Illinois School Research, 1972, 9, 5-18.

Lord, F.M. Applications of item response theory to practical testing problems. In press.

Macready, G.R., & Dayton, C.M. The use of probabilistic models in the assessment of mastery. Journal of Educational Statistics, 1977, 2, 99-120.

Millman, J. Tables for determining number of items needed on domain-referenced tests and numbers of students to be tested (Technical Paper No. 6). Los Angeles, California: Instructional Objectives Exchange, April 1972.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 1973, 43, 205-216.

Molenaar, W. Approximations to the Poisson, binomial and hypergeometric distribution functions (Mathematical Centre Tract No. 31). Amsterdam, The Netherlands: Mathematisch Centrum, 1973.

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

Ordnance Corps. Tables of the cumulative binomial probabilities, Document ORDP 20-1 of the Ordonance Corps of the U.S. Army.

van den Brink, W.P., & Koele, P. Item sampling, guessing, and decision-making in achievement testing. British Journal of Mathematical and Statistical Psychology, 1980, 33, 104-108.

van der Linden, W.J. Forgetting, guessing, and mastery: The Macready and Dayton model revisited and compared with a latent trait approach. Journal of Educational Statistics, 1978, 3, 305-318.

van der Linden, W.J. Binomial test models and item difficulty. Applied Psychological Measurement, 1979, 3, 401-411.

van der Linden, W.J. Estimating the parameters of Emrick's mastery testing model. Submitted for publication, 1980.

Wald, A. Seq ıtial analysis. New York: John Wiley & Sons, 1947.

Wilcox, R.R. A note on the length and passing score of a mastery test. Journal of Educational Statistics, 1976, 1, 359-364.

TABLE 1

Increase of Optimal Relative Passing Score Produced by Loss Ratios Unequal to One for Some Values of n and $(\pi_0, \pi_1) = (.75, .85)$

	$\lambda=\ell_0/\ell_1$						
n	.25	.33	.50	1	2	3	4
10	-.218	-.173	-.109	.000	.109	.173	.218
20	-.109	-.086	-.054	.000	.054	.086	.109
30	-.073	-.056	-.036	.000	.036	.056	.073
40	-.054	-.043	-.027	.000	.027	.043	.054
50	-.044	-.035	-.022	.000	.022	.035	.044
60	-.036	-.029	-.018	.000	.018	.029	.036

TABLE 2

Increase of Optimal Passing Score Produced by Loss Ratios Unequal to One for Some Indifference Zones

	$\lambda=\ell_0/\ell_1$						
	.25	.33	.50	1	2	3	4
(.60, .65)	-6.491	-5.144	-3.245	.000	3.245	5.144	6.491
(.60, .70)	-3.138	-2.486	-1.569	.000	1.569	2.486	3.138
(.60, .75)	-2.000	-1.585	-1.000	.000	1.000	1.585	2.000
(.65, .70)	-6.073	-4.813	-3.037	.000	3.037	4.813	6.073
(.65, .75)	-2.891	-2.291	-1.445	.000	1.445	2.291	2.891
(.65, .80)	-1.807	-1.432	-0.903	.000	0.903	1.432	1.807
(.70, .75)	-5.516	-4.371	-2.758	.000	2.758	4.371	5.516
(.70, .80)	-2.572	-2.038	-1.286	.000	1.286	2.038	2.572
(.70, .85)	-1.562	-1.238	-0.781	.000	0.781	1.238	1.562
(.75, .80)	-4.819	-3.819	-2.409	.000	2.409	3.819	4.819
(.75, .85)	-2.180	-1.727	-1.090	.000	1.090	1.727	2.180
(.75, .90)	-1.262	-1.000	-0.631	.000	0.631	1.000	1.262
(.80, .85)	-3.980	-3.154	-1.990	.000	1.990	3.154	3.980
(.80, .90)	-1.710	-1.355	-0.855	.000	0.855	1.355	1.710
(.80, .95)	-0.890	-0.705	-0.442	.000	0.442	0.705	0.890

Acknowledgement

The author is indebted to Dato N.M. de Gruijter, Gideon J. Mellenbergh, and Ivo W. Molenaar for their comments on the first draft of this paper, and to Paula Achterberg for typing the manuscript.

Author's address

Wim J. van der Linden, Onderafdeling Toegepaste Onderwijskunde, Technische Hogeschool Twente, Postbus 217, 7500 AE Enschede, The Netherlands.