Measurement and Research Department Reports 91-6

The Sample Strategy Of A Test Information Function In Computerized Test Design

H.H.F.M. Verstralen
N.D. Verhelst

Cito

The Sample Strategy Of A Test Information Function In Computerized Test Design

H.H.F.M. Verstalen
N.D. Verhelst

Cito
Arnhem, 1991

Abstract

Modern methods for optimal test construction require that the values or the form of a minimal information function be specified at a few well chosen values of the latent variable, the so called specification points. When the Rasch model is the IRT model, typically about five specification points suffice to prevent that the information of the constructed test has local minima between specification points considerably lower than the adjacent specified minima. For a typical test length this implies that the asymptotic property of the information function as the inverse of the variance-function of the maximum likelihood estimator is an adequate interpretation and justification for the specification of a minimum information function. As a consequence of the linear approach to optimization one has to restrict the specification of measurement accuracy to a few important ability classes. Therefore, it seems natural to choose the specification points at values of the latent variable that can be considered to represent these classes.

However, when the IRT model allows different discrimination parameters or indices, like the OPLM model, local minima of the information function within a class may be very much less than the information at the specification point, and, moreover, may be responsible for too large deviations from the asymptotic property to justify the above procedure. Therefore, a more general interpretation of the information function and an associated procedure for its specification in test construction is proposed that proves also adequate for OPLM or the Birnbaum models. The interpretation clearly shows that the information has to be specified at edges that separate ability classes, not at abilities that represent classes.

Keywords: IRT, Optimal test construction, Test Information.

Introduction

Theunissen (1985) was the first to use linear programming techniques for computerized test design. Since, several others have contributed to the subject (Boomsma, 1986, Razoux Schultz, 1987, Kester, 1988, Gademan, 1987,1989, Boekkooi-Timminga, 1989, Boekkooi-Timminga & Van der Linden, 1988, Adema & Van der Linden, 1989, Van der Linden & Boekkooi-Timminga, 1989). All these approaches assume a calibrated item bank, usually with the Rasch model, but in principle they readily generalize to other IRT-models.

The construction of a test with this approach starts by specifying an information function. It is essential to the linear programming approach that the information function is specified at a few well chosen values of the latent variable θ. A linear program then selects a test from an item bank with an information function that exceeds the specified information at the chosen values and optimizes some goal function, e.g., a minimum number of items in the test. Until now, it was recommended that the information function of the test did not show relatively low local minima between the specified values. That is, it was in general, not assumed that a test information function with low local minima between specified values could indicate desirable measurement properties.

In an earlier report, an argument was constructed, based on Fourier analysis of information functions and the sampling theorem (Oppenheim and Willsky, 1983), that for the Rasch model about five specifications in the interval [-2,2] are sufficient to prevent local minima. The sufficiency of only relatively few specification points in the Rasch model to prevent local minima is a consequence of the modest steepness of the item information functions. However, in more general IRT models that allow for different discrimination parameters or indices, like OPLM (Verhelst, et al. 1991), the steepness of item information functions shows greater variability and one has to be prepared for very steep information functions. The presence of items with relatively high discrimination indices in an item bank would result in the just mentioned local minima. If one aims at a kind of continuous measurement accuracy, that smoothly follows the specified information values, then the density of specification points must be adapted to this new situation. In that case an analysis along the lines followed in the mentioned report would be appropriate. However, we show that if one's aim with a test is to classify the examinees in a small number of adjacent ability classes, say up to ten, then local minima may even be advantageous.

The Problem of Local Information Minima

In OPLM (One Parameter Logistic Model) the probability that an examinee with ability θ scores in category $j>0$ of item i (the category response function) is given by:

$$P(X_i=j \mid \theta) = P_{ij}(\theta) = \frac{\exp(a_i(j\theta-\eta_{ij}))}{1+\sum_{c=1}^{m_i}\exp(a_i(c\theta-\eta_{ic}))}, \quad (j = 1,..,m_i), \tag{1}$$

where m_i is the maximum score of item i and a_i its discrimination index. In OPLM, the discrimination index is not estimated but part of the model hypothesis. It is not difficult to see from (1) that OPLM is a generalization of the Rasch model and the Partial Credit model as well. If $\forall i(a_i = 1)$ then (1) represents the Partial Credit model, and if, moreover, $\forall i(m_i = 1)$ then (1) has been reduced to the Rasch model. OPLM is also related to the Birnbaum two parameter logistic model. However, because the discrimination indices in OPLM are part of the model hypothesis, it yields sufficient statistics for the person parameters. As a result Conditional Maximum Likelihood estimates of the item parameters can be obtained in contradistinction to the Birnbaum model.

Denote a random variable with a Roman capital and its realization by the lower case equivalent. An OPLM test information function can now be derived as follows. Let $\underline{V}$ be a k dimensional response pattern vector, where $v_i = j$ if an examinee scores in category j of item i, and let Y be a weighted sum score of $\underline{V}$: $Y = \sum w_i V_i$, then in general (Birnbaum, 1968, p. 453), the information function of the test score Y has the form:

$$I_Y(\theta) = \frac{\left(\dfrac{\partial E(Y \mid \theta)}{\partial\theta}\right)^2}{\sigma^2(Y \mid \theta)}. \tag{2}$$

If, for an arbitrary θ one varies the weights w_i, the value of $I_Y(\theta)$ varies as well. Birnbaum proves that the maximum information is attained for optimally weighted scores X. The optimum score in OPLM is obtained by putting $w_i = a_i$, thus $X = \sum a_i V_i$. The information function $I(\theta)$ of the optimally weighted score is not indexed by X, and is called *the* information function of the test.

The derivative with respect to θ of $E(X|\theta)$ for a test of k items is:

$$E'(X \mid \theta) = \sum_{i=1}^{k} a_i \sum_{j=1}^{m_i} jP'_{ij}(\theta) \quad . \tag{3}$$

Let $\overline{V}_i(\theta) = E(V_i|\theta) \doteq \sum_{j=1}^{m_i} jP_{ij}(\theta)$, then it can be deduced from (1) that:

$$P'_{ij}(\theta) = \frac{\partial P_{ij}(\theta)}{\partial\theta} = a_i P_{ij}(\theta)(j-\overline{V}_i(\theta)) \quad . \tag{4}$$

Substitution of (4) in (3) gives:

$$\begin{aligned} E'(X \mid \theta) &= \sum_{i=1}^{k} a_i^2 \sum_{j=0}^{m_i} P_{ij}(\theta)(j^2 - j\overline{V}_i(\theta)) \\ &= \sum_{i=1}^{k} a_i^2 \sum_{j=0}^{m_i} P_{ij}(\theta)(j-\overline{V}_i(\theta))^2 \\ &= \sigma^2(X \mid \theta) \quad . \end{aligned} \tag{5}$$

The last inference in (5) follows from local stochastic independence, which implies that the conditional test score variance equals the sum of the conditional item score variances.

From (2) and (5) it follows that:

$$I(\theta) = \sigma^2(X \mid \theta) \tag{6}$$

For the Rasch model this amounts to:

$$I(\theta) = \sum_{i=1}^{k} \frac{\exp(\theta-\eta_i)}{(1+\exp(\theta-\eta_i))^2} = \sum_{i=1}^{k} P_i(\theta)(1-P_i(\theta)) \quad .$$

It follows from (6) that the value of the item information function grows quadratically in the discrimination index a_1. Moreover, from (1) and (6) it can be inferred that the discrimination index functions as a scaling factor. This scale dependency of the information function makes it less attractive as a starting point for test construction, unless the θ scale has been given a standard interpretation. An obvious possibility is to rescale the discrimination indices in an item bank such that their geometric mean or the variance of the latent variable in a population of interest equals one.

To illustrate the difficulties with automatic test design the optimal solutions were calculated for the following test construction problem in a series of six artificial conditions. The problem is to construct a test with the least possible number of binary items with an information function that attains values at least equal to 30 at $\theta = -1,1$. In all six conditions there can be selected from an unlimited number of items at $\eta = -1$, 0 and 1 with equal discrimination indices. The conditions differ in the values of a, the discrimination indices, which are 1, 2, 3, 4, 6, 8 resp. The minimum numbers of items needed from the artificial item banks are: 153, 57, 28, 16, 8 and 4 resp. With the exception of $a = 1$ the optimum test design procedure selects approximately one half of the items from $\eta = -1$ and one half from $\eta = 1$. For $a = 1$ the minimum number of items is obtained if they are all taken from $\eta = 0$. The information functions of the tests are shown in Figure 1. Clearly, there can be a tremendous drop in the value of the information function between specified points. The information function of the last two tests almost vanishes for $\theta = 0$. The first display in Figure 1 can also be interpreted as the item information function of a Rasch item if one takes the maximum value to be 38.25/153 = 0.25 in stead of 38.25. How the item information function changes by increasing the discrimination index can be inferred by comparing the first display with the last, where, for example, the left peak is approximately the sum of two item information functions with $\eta = -1$.

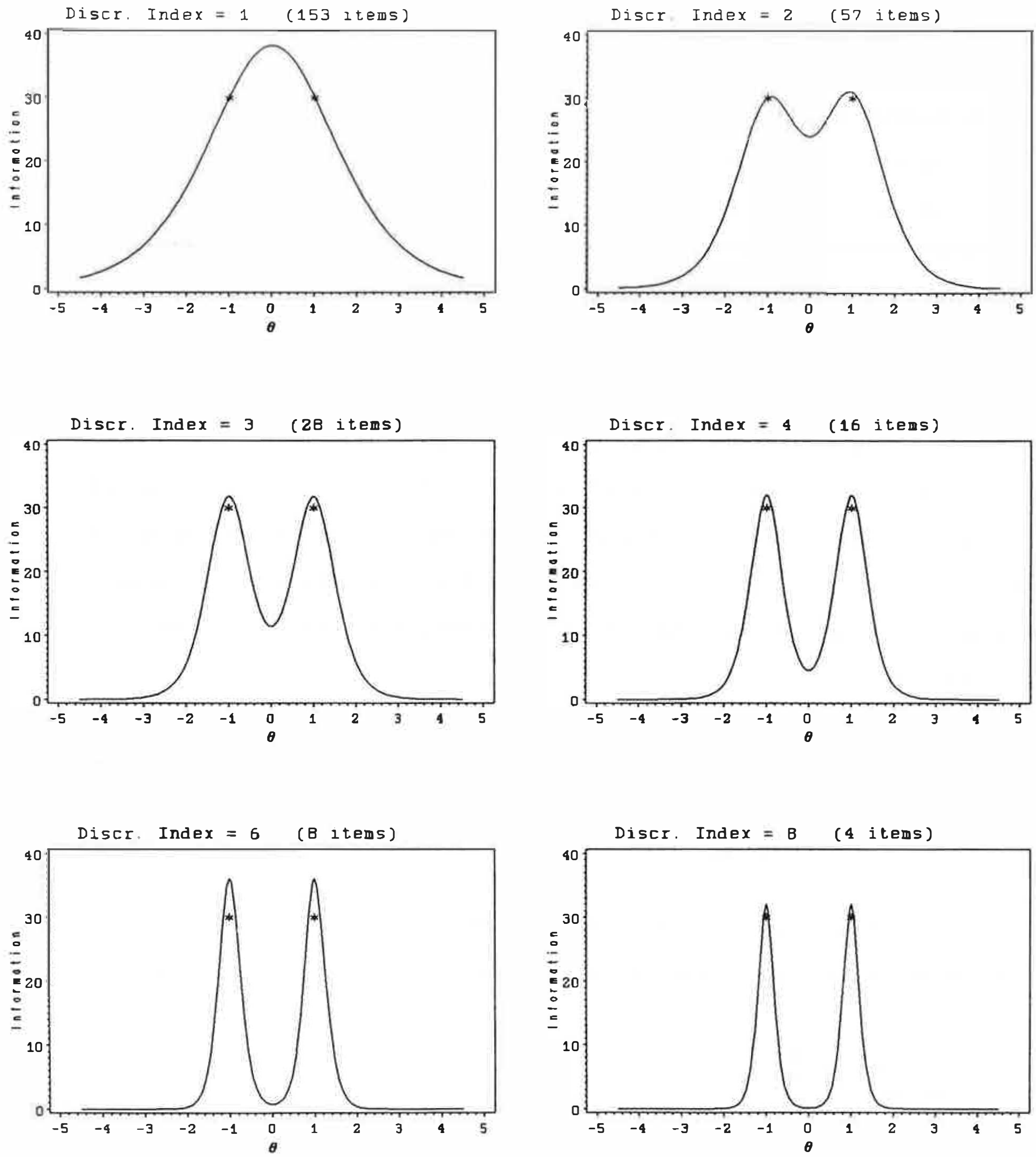

Figure 1. Information functions of optimal tests with information at least equal to 30 at $\theta = -1$ and $\theta = 1$. The tests are drawn from six artificial OPLM itembanks, all with an infinite number of binary items with difficulty parameters -1,0 and 1. The itembanks differ in the discrimination parameter of the items: 1,2,3,4,6 and 8 resp.

The Validity of the Asymptotic Property of the Information Function

In this section the consequences from a measurement point of view are investigated of having low information for values of θ in the range of interest. Birnbaum (1968, p. 457) asserts that $\hat{\theta}$, the Maximum Likelihood estimator of θ, is asymptotically $N(\theta,1/I(\theta))$ distributed. For the condition with a = 8 this would mean that the variance of $\hat{\theta}$ for θ in the interval $[-\frac{1}{2},\frac{1}{2}]$ is very large. In fact, for this test of 4 items $I(0) = 0.086$, which would mean that $\sigma^2(\hat{\theta} \mid \theta{=}0.0) = 11.65$. If the asymptotic property would apply then $\hat{\theta}$ for $\theta = 0.0$ would be almost uniformly distributed in the interval [-2,2]. In fact for $\theta{=}0.0$ the probability of being estimated between -1.0 and +1.0 is approximated by $1.0 - 2\times\Phi[-0.5/(11.65)^{1/2}] = 1.0 - 2\times0.442 = 0.116$ and it must be concluded that the test does not in any way inform us about the true θ if its value is near zero. However, this is not the case, as closer scrutiny of the conditional score distribution will show.

Define $\eta_{i0} = 0$ for all i, then it follows from (1) that the probability of a score x is:

$$P(\sum_i a_i V_i = x \mid \theta) \propto \sum_{\{\underline{v}\,:\,\sum_i a_i v_i = x\}} \prod_i \exp(a_i(v_i\theta - \eta_{iv_i})) , \qquad (7)$$

where v again represents an arbitrary response pattern with $x = \Sigma a_i v_i$ its associated weighted sum score. Define $\epsilon_{ij} \doteq \exp(a_i(j\theta-\eta_{ij}))$ for $j = 0,..,m_i$. Moreover, let

$$\gamma_x(\underline{\varepsilon},\underline{a}) \doteq \sum_{\{\underline{v}\,:\,\sum_i a_i v_i = x\}} \prod_i \varepsilon_{iv_i},$$

be called the basic combinatorial function of order x, then (7) can be expressed as

$$P\Big(\sum_i a_i V_i = x \,|\, \theta\Big) = \frac{\gamma_x(\underline{\varepsilon},\underline{a})}{\sum_z \gamma_z(\underline{\varepsilon},\underline{a})},$$

where z ranges over all possible scores in $(0,..,\sum a_i m_i)$.

As can be inferred from (7) the likelihood as a function of θ, given the item parameters, belongs to the exponential family. Therefore, the Maximum Likelihood estimator for θ is defined as $\hat{\theta}(E(X|\theta')) = \theta'$.

The four item test in the example with a = 8 yields the distribution of x and $\hat{\theta}(x)$ conditional on θ = 0.0 presented in Table 1.

TABLE 1

Distribution of x and $\hat{\theta}$ Conditional on θ = 0.0 of a Four Item Test

score	0	8	16	24	32
$\hat{\theta}(x)$	--	-1.00	0.00	1.00	--
prob	0.00000011	0.00067	0.99866	0.00067	0.00000011

$\hat{\theta}$ is obtained by Maximum Likelihood given the item parameters. Table 1 shows that the probability of an estimate for θ = 0 is almost exclusively concentrated at this same value. This shows that the information function for θ = 0 underestimates the measurement accuracy for this test. As a matter of fact in this simple example the probability of getting a right estimate for θ=0.0 is almost 1.0 and not 0.1 as expected from the asymptotic property. It must be concluded that this property of the information function as the inverse of the variance of the maximum likelihood estimator cannot always be properly applied to a finite test.

The larger the discrimination index of an OPLM item, the more it approaches a Guttman item. Therefore, this example is pushed to its very extreme if a test of 4 Guttman items is considered with difficulty parameters as in the example. For this Guttman test the variance of the conditional distribution of $\hat{\theta}$ vanishes everywhere, because the model is deterministic. The information function of the test equals zero everywhere except at the two item parameter values. From (2) and (3) it follows that, at θ = -1 and 1, the information function grows to infinity if the discrimination index is increased without bound. Everywhere else it vanishes and does not rise to infinity as expected from the asymptotic property and the fact that there the variances of the conditional distributions of $\hat{\theta}$ also vanish. But this test is generally viewed as optimal, in spite of an almost everywhere zero information function.

To advance a proper use of information functions to aid test design, there is one more consideration. An information function which is for all θ in a certain interval bounded from below at a high enough value, enables measurement on a continuous scale. However, in many cases the measurement result, especially in educational measurement, is less refined. Often at most 5 through 10 ability levels are distinguished. If one takes these levels seriously, then it only matters to minimize the probability that a pupil is incorrectly classified.

A More General Expression to Interpret the Information Function

To aid the interpretation of the information function a simple but more general expression for the relation between the measurement error of the ML estimator $\hat{\theta}(x)$ and the information function will be derived. Consider an examinee with ability θ^* and a test with information function $I(\theta)$. For this examinee the probability will be investigated that he earns a score x such that his Maximum Likelihood ability estimate $\hat{\theta}(x) < \theta'$. It was already noted that the information function of an OPLM test is equal to the weighted score variance. According to (6) a normal approximation of the distribution of the weighted score X conditional on θ^* is $N(E(X|\theta^*), I(\theta^*))$. Then the indicated probability is:

$$P(\hat{\theta}(X) < \theta' \mid \theta^*) \approx \Phi\left(\frac{E(X \mid \theta') - E(X \mid \theta^*)}{\sigma(X \mid \theta^*)}\right), \qquad (8)$$

where Φ denotes the standard normal distribution function.
By definition of the definite integral and using (2) and (5) it follows that:

$$\begin{aligned} E(X \mid \theta') - E(X \mid \theta^*) &\doteq \int_{\theta^*}^{\theta'} \frac{\partial E(X \mid \theta)}{\partial \theta} d\theta \\ &= \int_{\theta^*}^{\theta'} (I(\theta))^{1/2}\, \sigma(X \mid \theta)\, d\theta \qquad (9) \\ &= \int_{\theta^*}^{\theta'} I(\theta)\, d\theta \ . \end{aligned}$$

Substitution of (9) and (6) in (8) gives:

$$P(\hat{\theta}(X) < \theta' \mid \theta^*) \approx \Phi\left(\frac{\int_{\theta^*}^{\theta'} I(\theta)\,d\theta}{I(\theta^*)^{1/2}}\right). \tag{10}$$

The approximate conditional distribution of $\hat{\theta}$ given θ is elegantly related by (10) to the information function. To interpret (10) one could say that the larger the information 'mass' between θ^* and $\theta'<\theta^*$, an edge between ability levels, relative to the square root of the information at θ^*, the less the probability that the estimator $\hat{\theta}$ of θ^* will be less than θ'.

For applications, as in optimal test construction, often the probability of finding the estimator $\hat{\theta}$ of θ^* in each of several adjacent intervals is needed. Let, therefore,

$$y(\eta', \eta'') \doteq \int_{\eta'}^{\eta''} I(\theta)\,d\theta,$$

and let $-\infty=\theta_0<..<\theta_k=\infty$ be a partition of the real line that defines k ability classes $t_i=(\theta_{i-1},\theta_i)$ $(i=1,..,k)$. Let y_i denote $y(\theta_{i-1},\theta_i)$. Next, select some interval t_j and $\theta^* \in t_j$, and define

$$z^* = I(\theta^*)^{-1/2}.$$

To get the probability that $\hat{\theta} \in t_m$, distinguish between m<j, m>j and m=j. First, let $1\le m<j$, and

$$C = y(\theta^*, \theta_{j-1}) - \sum_{m<i<j} y_i$$

so that

$$P(\theta_{m-1} < \hat{\theta} < \theta_m) = P(\hat{\theta} \in t_m) = P(\hat{\theta} < \theta_m) - P(\hat{\theta} < \theta_{m-1})$$
$$= \Phi[z^*\{C\}] - \Phi[z^*\{C - y_m\}].$$

Second, let $j<m\le k$, and

$$D = y(\theta^*, \theta_j) + \sum_{j<i<m} y_i$$

and we have

$$P(\theta_{m-1} < \hat{\theta} < \theta_m) = P(\hat{\theta} \in t_m) = P(\hat{\theta} < \theta_m) - P(\hat{\theta} < \theta_{m-1})$$
$$= \Phi[z^*\{D + y_m\}] - \Phi[z^*\{D\}] \ .$$

If m=j then

$$P(\hat{\theta} \in t_j) = \Phi(z^*D) - \Phi(z^*C) \ .$$

Although the symbol Φ in (10) might suggest that the conditional distribution of $\hat{\theta}$ is normal, as the asymptotic property implies, this is in general not true. Note that (10) implies that the validity of the asymptotic property depends among other things on how close $I(\theta)$ in the interval $(\theta^*-cI(\theta^*)^{-1/2},\theta^*+cI(\theta^*)^{-1/2})$ can be approximated by $I(\theta^*)$, where c is, for instance, equal to 2.0. Because in that case (10) would simplify to:

$$P(\hat{\theta}(X) < \theta' \mid \theta^*) \approx \Phi[I(\theta^*)^{1/2}(\theta'-\theta^*)] \ .$$

And this means exactly that $\sigma^2(\hat{\theta}(X) \mid \theta^*) \approx 1/I(\theta^*)$. In the example of Table 1 $I(\theta)$ certainly cannot be approximated by $I(\theta^*)$ in the indicated interval, which explains the deviation from the asymptotic property.

With a finite test θ', in general, is not located exactly at the midpoint of the interval $[\hat{\theta}_r,\hat{\theta}_{r'}]$, where r is the largest possible score r with a ML estimator $\hat{\theta}_r$ smaller than or equal to θ', and r' the smallest possible score larger than r. The position of θ'in $[\hat{\theta}_r,\hat{\theta}_{r'}]$, has a sizeable influence on the probability of misclassification in the neighborhood of θ'. Therefore, the correction for discreteness $\theta'' = [\hat{\theta}(r)+\hat{\theta}(r')]/2$ will be used, where θ'' will be substituted for θ' in (10).

As an example, consider the test in Table 1, and suppose that $\theta' = -0.25$, or, indeed, any other value between -1.0 and 0.0. The largest r with a ML estimator smaller than θ' is 8, and the smallest score larger than 8 is 16. Therefore, $\theta'' = -0.5$.

Properties of the Approximation With Formula (10)

This section begins with a theoretical comparison of the probabilities of misclassification, as given by (10), for a test with constant information and a test with a single peaked information function, called a unimodal test. The result will be proved that, according to (10) the probability of misclassification for an unimodal test may be smaller than for a test with constant information as high as the maximum information of the unimodal test. The section ends with an example to illustrate this result.

Without loss of generality a unimodal test can be represented by one item with parameter $\eta = \eta_{11} = 0.0$, and $a = a_1 = 1$. The number of identical items in the unimodal test is irrelevant for the derivation of the result, therefore, we may just as well take one item. The constant information test can be represented just by an information function with constant value 0.25. Because the unimodal test has its maximum information at 0.0, the classification edge is also chosen at that value.

Misclassification results when a person with latent variable $\theta>0.0$ receives an estimate $\hat{\theta}_r \leq 0.0$, or vice versa. Denote the two probability functions of misclassification by the constant test with $P_c(\theta)$ and with the unimodal test by $P_s(\theta)$. The statement '$P_s(\theta) < P_c(\theta)$ for $\theta>0.0$' is, using (10), identical with:

$$\Phi\left(\frac{0.5 - E_s(X|\theta)}{\sqrt{I_s(\theta)}}\right) < \Phi\left(\frac{\int_\theta^0 0.25\, d\theta}{0.25^2}\right) \iff$$

$$\Phi\left(-\frac{E_s(X|\theta) - 0.5}{\sqrt{I_s(\theta)}}\right) < \Phi\left(-\frac{\theta}{2}\right)$$

where the subscript s with E and I again indicate that the unimodal test is meant. Because the normal distribution function Φ is strictly monotone increasing this is equivalent to:

$$\frac{E(X|\theta) - 0.5}{\sqrt{I(\theta)}} > \frac{\theta}{2}.$$

Using the simplification of Formula 1 for the Rasch Model this can also be written as:

$$\frac{\exp(\theta) - 0.5(1+\exp(\theta))}{\sqrt{(\exp(\theta))}} > \frac{\theta}{2} \longleftrightarrow$$

$$\exp\left(\frac{\theta}{2}\right) - \exp\left(-\frac{\theta}{2}\right) > \theta .$$

At $\theta = 0.0$ both expressions on either side are equal. Therefore, because $\theta > 0.0$, if the derivatives of both sides satisfy the inequality the result is proved. Differentiation of either side gives:

$$\frac{1}{2}\left\{ \exp\left(\frac{\theta}{2}\right) + \exp\left(-\frac{\theta}{2}\right) \right\} > 1 = \exp(0) ,$$

which is true because of the convexity of the exponential function.

So we proved that, based on approximation (10) of the conditional distribution of $\hat{\theta}_r$, that

$$P_s(\theta) < P_c(\theta) \text{ for } \theta>0. \qquad (11)$$

To test (11) against the exact distribution, an example is constructed for some illustrative calculations. In this example three measures of accuracy of a test with approximately constant information in a relevant interval on the latent trait are compared with those of two other tests. One test with a single peak, already called a unimodal test, and the other test with a doubly peaked information function, called a bimodal test. In all three cases the maximum information is identical. The three measures of accuracy are the exact probability of misclassification for these tests and the approximations of this probability as given by (10), with and without correction for continuity.

Because the exact distributions have to be calculated, it is not sufficient to specify the tests exclusively by their information functions, the underlying item parameters have to be specified as well.

For the construction of the tests some simplifications can be introduced to better highlight the essentials. Because scaling is arbitrary and variation in discrimination indices is not relevant for the subject of investigation the value of the discrimination indices is taken to be one. The same applies to the number of parameters per item. Therefore, the examples can be restricted to the Rasch model.

Let the bimodal test contain two sets of $k = 10$ items. All items within one set have the same parameter. Without loss of generality denote these two parameters by η and $-\eta$. Moreover, let this test have a minimum information value of 0.1 at the midpoint $\theta = 0.0$ between the two item parameters, approximately the local minimum of the information function at $\theta = 0$ in the example of Table 1. This minimum information value is obtained if η fulfills:

$$2kI(0;\eta) = 2kP(0;\eta)(1-P(0;\eta)) = 0.1 \quad (k = 10), \tag{12}$$

where $I(0;\eta)$ denotes the information at $\theta = 0$ of an item with parameter η.

Let $c = 0.1/2k$ then it is easy to see that for the Rasch model the solution to (12) is contained in the following equation:

$$c\exp(\eta)^2 + (2c-1)\exp(\eta) + c = 0 ,$$

which means that:

$$\exp(\eta) = \frac{(1-2c) \pm (1-4c)^{½}}{2c} , \tag{13}$$

Solving (13) it is found that $\eta' = \pm 5.288$. Therefore, two sets of 10 item parameters at +5.288 and -5.288, resp. have an information value of 0.1 at $\theta = 0.0$. Notice that $I(5.288) = 2.5$.

Next, the maximum of the unimodal test must be equated to the maximum of the bimodal test. In this case one of the two sets of the bimodal test will suffice. The information of the set with item parameter η' at the parameter value of the other set with item parameter $-\eta'$ is negligible. Therefore, the maximum information of one of the two item sets of the bimodal test almost equals the maximum information of the complete bimodal test.

The problem of constructing a test with approximately constant information is relatively simply realized by selecting enough uniformly distributed items in a wide enough interval of interest. It turns out that a test of 60 items uniformly distributed in [-11.8,11.8] yields an information function that is approximately constant and equal to the maximum information of the peaked tests in the interval [-6,6]. However, for purposes of meaningful comparison the highest $\hat{\theta}_r$ below or equal to the classification edge for all tests are made equal to -5.2280. This is the closest the three θ'' (for the correction for continuity) of the tests can get. The three next higher $\hat{\theta}_{r'}$ of the unimodal the bimodal and constant test after this equalizing correction are -4.8825, -4.8825 and -4.8876, resp. Consequently θ'' of the constant information test is somewhat closer to the classification edge than the other two.

Now that the three tests for the example are constructed, the probabilities of misclassification with classification edge at -5.2880 and the two approximations of it can be investigated in a relevant part of the interval [-5.288,5.288].

Table 2 shows the results. It follows from these results that property (11), based on approximation (10), is only partially reflected in the exact misclassification probabilities. From a more detailed extension of Table 2 it can be read that the better performance of the unimodal test holds for misclassification probabilities greater than 0.20. For lower probabilities of misclassification, that is at greater distances from θ' the constant information test performs better than the peaked test, contrary to (11). The maximum difference 0.00456 between the two probabilities is obtained at $\theta = -3.95$ where the misclassification probabilities for the constant and the unimodal tests are 0.03350 and 0.03806 resp. Similar results are found with larger numbers of items. Therefore, this finding cannot be attributed to the arbitrary numbers of items in this example. Neither can the better performance of the peaked tests in the neighborhood of θ' be attributed to a tiny difference in the information functions, because the information of the constant test was taken slightly higher in the investigated interval than the maximum of the peaked tests. It follows that (10) has to be used with some caution in comparing different tests with such very different information functions. However, where the result no longer holds, we deal with small probabilities.

TABLE 2

Probabilities of misclassification for 3 tests:
Exact : Exact probability of misclassification at θ^*
ApprC : Approximation of probability of misclassification at θ^* with formula 10 and correction for continuity
Appr : Approximation of probability of misclassification at θ^* with formula 10 without correction for continuity

	Constant Info			Bimodal			Unimodal		
θ^*	Exact	ApprC	Appr	Exact	ApprC	Appr	Exact	ApprC	Appr
-5.29	0.626	0.624	0.500	0.622	0.624	0.500	0.622	0.624	0.500
-4.94	0.403	0.405	0.289	0.399	0.405	0.288	0.399	0.405	0.288
-4.58	0.209	0.212	0.133	0.208	0.211	0.128	0.208	0.211	0.128
-3.53	0.006	0.007	0.003	0.009	0.003	0.001	0.009	0.003	0.001
-2.12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

From the above discussion and Table 2 there emerge two reassuring conclusions: First, ApprC approaches the exact probabilities very close. Second, given the first conclusion and the approximate result (11) it follows that the probabilities of misclassification for the peaked tests will very likely not be appreciably larger than those of the constant information test. In this example they differ only negligibly from those of the constant test. We may conclude that the true probabilities of misclassification are in fact comparable. The differences in misclassification probabilities between the two types of tests grow even smaller for larger numbers of items.

Conclusion

From (10) and the above analysis it may be concluded that in specifying an information function for a test, that classifies students in a few ability levels, it is not necessary to demand that it is bounded from below at a specified value everywhere in the interval of interest. It is important, though, to indicate explicitly the edges that separate the levels in which the latent trait is classified, and to specify the minimum required information at these edges. This means that the current practice of specifying an information function has to be changed. Presently, the test constructor is asked to specify the information at a θ-value that is considered representative for a class, instead of the information at the edge between classes. The evaluation of the effect of a certain specification can best be made by inspecting the conditional score distributions for a tentative test constructed according to these specifications at several values of θ. However, this procedure may be too indirect, and, therefore, inconvenient. The above analysis, however, implies a reassuring message that if the information of a test between edges would decrease, this does not deteriorate the classification qualities of this test relative to a test with constant information between edges.

References

Adema, J.J. and Van der Linden, W.J. (1989). Algorithms for computerized test construction using classical item parameters. *Journal of Educational Statistics, 14*, 279-290.

Birnbaum, A. (1968). Some latent trait models and their use in inferring anexaminee's ability. Part 5 in: F.M. Lord and M.R. Novick. *Statistical theories of mental test scores*. Reading, Mass.:Addison-Wesley.

Boekkooi-Timminga, E. (1989). *Models for computerized test construction*. Dissertation, Enschede, University of Twente.

Boekkooi-Timminga, E. and Van der Linden, W.J. (1988). Algorithms for automated test construction. In: F.J. Maarse (ed.) *Computers in psychology: Methods, instrumentation and psychodiagnostics*. Lisse: Swets & Zeitlinger.

Boomsma, Y. (1986). *Item selection by mathematical programming*. Master's thesis, Enschede: University of Twente, Department of Applied Mathematics.

Gademann, A.J.R.M. (1987). *Item selection using multiobjective programming*. Measurement and Research Department, Project 880, Report No 1. Arnhem: Cito.

Gademann, A.J.R.M. (1989). *A new heuristic to solve the item selection* problem: outline and numerical experiments. Measurement and Research Department, Project 880, Report No 5. Arnhem: Cito.

Kester, J.G. (1988). *Various mathematical programming approaches toward* item selection. Measurement and Research Department, Project 880, Report No 3. Arnhem: Cito.

Oppenheim, A.V., Willsky, A.S. and Young, I.T. (1983). *Signals and Systems*. London, Prentice-Hall.

Razoux Schultz, A.F. (1987). *Item selection using heuristics*. Measurement and Research Department, Project 880, Report No 2. Arnhem: Cito.

Theunissen, T.J.J.M. (1985). Binary programming and test design. *Psychometrika, 50*, 411-420.

Van der Linden, W.J. and Boekkooi-Timminga, E. (1989). A maximin model for test design with practical constraints. *Psychometrika, 54*, 237-247.

Verhelst, N.D., Glas, C.A.W. and Verstralen, H.H.F.M. (1991). *OPLM (The One Parameter Logistic Model)*: Arnhem: Cito, a computer program and manual.

Recent Measurement and Research Department Reports:

91-1 N.D. Verhelst & N.H. Veldhuijzen. A New Algorithm For Computing Elementary Symmetric Functions And Their First And Second Derivatives.

91-2 C.A.W. Glas. Testing Rasch Models For Polytomous Items: With An Example Concerning Detection Of Item Bias.

91-3 C.A.W. Glas & N.D. Verhelst. Using The Rasch Model For Dichotomous Data For Analyzing Polytomous Responses.

91-4 N.D. Verhelst & C.A.W. Glas. A Dynamic Generalization Of The Rasch Model.

91-5 N.D. Verhelst & H.H.F.M. Verstralen. The Partial Credit Model With Non-Sequential Solution Strategies.