Assembling Tests for the Measurement of Multiple Traits

Wim J. van der Linden

University of Twente

For the measurement of multiple traits, this paper proposes assembling tests based on the targets for the (asymptotic) variance functions of the estimators of each of the traits. A linear programming model is presented that can be used to computerize the assembly process. Several cases of test assembly dealing with multidimensional traits are distinguished, and versions of the model applicable to each of these cases are discussed. An empirical example of a test assembly problem from a two-dimensional mathematics item pool is provided. *Index terms: asymptotic variance functions, linear programming, multidimensional IRT, test assembly, test design.*

A standard procedure for assembling tests from an item pool fitting a unidimensional item response theory (IRT) model was suggested by Birnbaum (1968). The central quantity in his suggestion is the test information function (TIF), which is defined as Fisher's information about the unknown trait parameter θ in the responses to the test taken as a function over the range of possible values of the trait parameter, θ. For a one-dimensional IRT model, the TIF is given by

$$I(\theta) = -\mathrm{E}\left[\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right], \quad -\infty < \theta < \infty, \tag{1}$$

where $L(\theta)$ is the likelihood statistic associated with the responses to the test.

Birnbaum's (1968) suggestion was to first design a target for the information function of the test and then select items in the test such that the sum of their information functions matches the target. The procedure capitalizes on the fact that local independence between item responses guarantees additivity of the item information functions. If $I_i(\theta)$ is the information function of item i, defined analogously to Equation 1 for the likelihood statistic associated with the response to this item, it holds that

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta), \tag{2}$$

where n is the number of items in the test.

If $\hat{\theta}$ is the maximum likelihood estimator (MLE) of θ, it holds that

$$\mathrm{Var}(\hat{\theta}|\theta) \to 1/I(\theta) \quad \text{for } n \to \infty, \tag{3}$$

where Var is the variance operator (e.g., Kendall & Stuart, 1976, chap. 18). Note that because of this reciprocity, setting a target for the information function is equivalent to setting a target for the (asymptotic) variance function of $\hat{\theta}$.

In practice, in spite of the additivity of the item information functions, the problem of selecting n items from a pool of a realistic size, such that the sum of the functions matches the target best over the range of

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 20, No. 4, December 1996, pp. 373–388
© Copyright 1996 Applied Psychological Measurement Inc.
0146-6216/96/040373-16$2.85

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

possible θ values, is not a trivial task. The prohibitively large number of possible combinations rules out manual optimal test assembly. In fact, even for a high-speed computer explicit enumeration of all possible solutions and selecting the best solution is unrealistic. The problem becomes more difficult still if the test has to meet various constraints on the selection of the items related to the distributions of, for example, item content, item format, or the values of certain item parameters. To implement Birnbaum's procedure, efficient algorithms are needed that reduce the set of feasible solutions to a smaller set of candidate solutions and then select an optimal solution.

Application of Linear Programming

Formally, the problem of test assembly is a constrained combinatorial optimization problem that, in its mathematical generality, has been studied in such fields as applied mathematics, decision theory, and operations research (Nemhauser & Wolsey, 1988; Wagner, 1975). Therefore, attempts to implement Birnbaum's procedure in a computer algorithm have been based on techniques of combinatorial optimization, in particular on techniques of (mixed) integer programming from the field of linear programming (LP). Although suggestions to resort to LP for solving test assembly problems were made earlier (Feuerman & Weiss, 1973; Votaw, 1952; Yen, 1983), the first LP model for a variation of Birnbaum's procedure was published by Theunissen (1985). Ever since, modeling various test assembly problems as an LP problem and finding algorithms and heuristics to solve the model for an optimal solution has been a fruitful field of research [e.g., Adema (1990, 1992a, 1992b); Adema, Boekkooi-Timminga, & van der Linden (1991); Adema & van der Linden (1989); Armstrong & Jones (1992); Armstrong, Jones, & Wu (1992); Boekkooi-Timminga (1987, 1990); Timminga & Adema (1995); van der Linden (1994); van der Linden & Boekkooi-Timminga (1988, 1989); van der Linden & Luecht (1996); important heuristic approaches to the same problems have been presented by Ackerman (1989), Luecht & Hirsch (1992), and Swanson & Stocking (1993)].

The Maximin Model

The model taken as a starting point for the problem of multidimensional test assembly is the maximin model for unidimensional assembly (van der Linden & Boekkooi-Timminga, 1989). It is assumed that a test of n items must measure an interval of possible θ values with uniform accuracy, and that the test assembler wants to control this behavior at θ points θ_k, $k = 1, ..., K$. Decision variables x_i, $i = 1, ..., I$, are defined for each item in the pool, which take the value 1 if the item is included in the test and 0 otherwise. The maximin model is

maximize y, (4)

subject to

$$\sum_{i=1}^{I} I_i(\theta_k)x_i - y \geq 0, \quad k = 1, \ldots, K, \tag{5}$$

$$\sum_{i=1}^{I} x_i = n, \tag{6}$$

$$x_i \in \{0,1\}, \quad i = 1, \ldots, I, \tag{7}$$

and

$$y \geq 0. \tag{8}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

The model is based on the idea that a common lower bound y to each of the values of the TIF at θ_k, $k = 1, \ldots, K$, defined by the inequality in Equation 5, should be maximized, as is done by the objective function in Equation 4. At the same time, Equation 6 constrains the length of the test to size n. Equations 7 and 8 define the ranges of values of the decision variables in the model.

The model can be generalized to a target for the TIF of any shape by providing the variable y in Equation 5 with coefficients r_k that govern the relative height of the TIF at $\theta_1, \ldots, \theta_K$ (van der Linden & Boekkooi-Timminga, 1989). For ease of exposition, only the case of a uniform target will be considered here. Also, a catalog of additional linear constraints is available to model test specifications with respect to such categories as item content, item format, testing time, the values of classical or IRT item parameters, and interdependencies between test items (van der Linden & Boekkooi-Timminga, 1989). For an illustration of the use of some of the constraints, see the empirical example below.

The maximin model has been implemented as one of the options in the computer program CONTEST (Timminga, van der Linden, & Schweizer, 1996), which contains a large selection of algorithms and heuristics to solve the model for an optimal combination of values for its decision variables. Quick heuristics to solve certain test assembly problems have been presented in Ackerman (1989) and Luecht & Hirsch (1992). If the model has a network flow structure, computation of an optimal solution simplifies dramatically (e.g., Armstrong et al., 1992).

Purpose

This paper presents models for the optimal assembly of tests measuring more than one trait. However, unlike a unidimensional IRT model, for a model with multiple θs Fisher's information measure is no longer a scalar but a (nondiagonal) matrix. Also, the (asymptotic) variances of the MLEs of the θs are not given by the reciprocals of the diagonal elements of the information matrix; they are given by the diagonal elements of the variance-covariance matrix, which is the inverse of the information matrix. Hence, the motivation to use a target directly for Fisher's information measure fails for the case of multidimensional test assembly. To solve the problem, the use of targets for the variance functions in the model are explored. Then a generalization of the maximin model and a heuristic for the assembly of tests in the presence of multiple θs is proposed, and various cases of multidimensional test assembly are discussed.

Multidimensional Test Assembly

The multidimensional IRT model considered here is the logistic model discussed by McKinley & Reckase (1983), Reckase (1985, 1997), and Samejima (1974). The case of two θs (θ_1, θ_2) is considered. Let the response variables U_{ij} take the value 1 if the response of person $j = 1, \ldots, N$ to item $i = 1, \ldots, n$ is correct and the value 0 otherwise. The model is defined by the following logistic response function:

$$P_i(\theta_1,\theta_2) \equiv P\left(U_{ij} = 1 \middle| a_{1i}, a_{2i}, d_i, \theta_1, \theta_2\right) \equiv \frac{\exp\left(a_{1i}\theta_1 + a_{2i}\theta_2 + d_i\right)}{1 + \exp\left(a_{1i}\theta_1 + a_{2i}\theta_2 + d_i\right)}, \tag{9}$$

where (a_{1i}, a_{2i}) are the discrimination parameters of item i for θ_1 and θ_2, respectively, and d_i can be interpreted as a composite parameter representing the easiness of the item. It is assumed here that these item parameters are known and that the model is used to estimate θ_{1j} and θ_{2j} from a realization of the response variables $U_{ij} = u_{ij}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, N$.

Variance Functions

For two θs, Fisher's information matrix is defined as

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

$$I(\theta_1,\theta_2) \equiv -\mathrm{E}\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta_1^2} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \ln L}{\partial \theta_2^2} \end{bmatrix}, \tag{10}$$

where L now is the likelihood statistic associated with the data under the model in Equation 9. Following the derivation in Ackerman (1994, Appendix) and using the notation $P_i \equiv P_i(\theta_1,\theta_2)$ and $Q_i \equiv 1 - P_i(\theta_1,\theta_2)$, the following result is obtained for the model in Equation 9:

$$I(\theta_1,\theta_2) = \begin{bmatrix} \sum_{i=1}^{n} a_{1i}^2 P_i Q_i & \sum_{i=1}^{n} a_{1i} a_{2i} P_i Q_i \\ \sum_{i=1}^{n} a_{1i} a_{2i} P_i Q_i & \sum_{i=1}^{n} a_{2i}^2 P_i Q_i \end{bmatrix}. \tag{11}$$

Standard techniques for matrix inversion yield the variance-covariance matrix (V) of the MLEs of (θ_1,θ_2):

$$V(\hat{\theta}_1,\hat{\theta}_2|\theta_1,\theta_2) = \begin{bmatrix} \frac{\sum_{i=1}^{n} a_{2i}^2 P_i Q_i}{|I(\theta_1,\theta_2)|} & \frac{-\sum_{i=1}^{n} a_{1i} a_{2i} P_i Q_i}{|I(\theta_1,\theta_2)|} \\ \frac{-\sum_{i=1}^{n} a_{1i} a_{2i} P_i Q_i}{|I(\theta_1,\theta_2)|} & \frac{\sum_{i=1}^{n} a_{1i}^2 P_i Q_i}{|I(\theta_1,\theta_2)|} \end{bmatrix}, \tag{12}$$

where

$$|I(\theta_1,\theta_2)| = \left(\sum_{i=1}^{n} a_{1i}^2 P_i Q_i\right)\left(\sum_{i=1}^{n} a_{2i}^2 P_i Q_i\right) - \left(\sum_{i=1}^{n} a_{1i} a_{2i} P_i Q_i\right)^2 \tag{13}$$

is the determinant of the matrix in Equation 11, which is assumed here to be nonzero. The diagonal elements of the matrix in Equation 12 are the (asymptotic) variances of the MLEs of θ_1 and θ_2, respectively:

$$\mathrm{Var}(\hat{\theta}_1|\theta_1,\theta_2) = \left(\sum_{i=1}^{n} a_{2i}^2 P_i Q_i\right)\left[\left(\sum_{i=1}^{n} a_{1i}^2 P_i Q_i\right)\left(\sum_{i=1}^{n} a_{2i}^2 P_i Q_i\right) - \left(\sum_{i=1}^{n} a_{1i} a_{2i} P_i Q_i\right)^2\right]^{-1}, \tag{14}$$

and

$$\mathrm{Var}(\hat{\theta}_2|\theta_1,\theta_2) = \left(\sum_{i=1}^{n} a_{1i}^2 P_i Q_i\right)\left[\left(\sum_{i=1}^{n} a_{1i}^2 P_i Q_i\right)\left(\sum_{i=1}^{n} a_{2i}^2 P_i Q_i\right) - \left(\sum_{i=1}^{n} a_{1i} a_{2i} P_i Q_i\right)^2\right]^{-1}. \tag{15}$$

Equation 14 shows that the (asymptotic) variance of $\hat{\theta}_1$ for true θ is a point in a two-dimensional space. Thus, the variance of $\hat{\theta}_1$ not only depends on the true value of θ_1 but also on the value of θ_2. The same holds for the (asymptotic) variance of $\hat{\theta}_2$ in Equation 15. Also, note that the two variances differ only by the factors a_{2i}^2 and a_{1i}^2 in the two numerators.

Taking the variances in Equations 14–15 as functions over the complete two-dimensional θ space, two

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

variance functions are defined—one for $\hat{\theta}_1$ and the other for $\hat{\theta}_2$. Figure 1 shows the plots of three pairs of variance functions, each for a different test. The first test had nine items, three with larger values for the first discrimination parameter and six with the reverse pattern: a_1 = (2.0, 2.0, 2.6, 1.2, 1.5, 1.7, 1.2, .8, .9), a_2 = (.1, 1.1, 1.7, 2.4, 2.0, 3.0, 1.9, 2.1, 1.8), and d_i = 0.0 for all items. The second test had six items, with the following values for the two discrimination parameters: a_1 = (1.8, 2.6, 1.7, 1.8, 2.2, 2.0), a_2 = (2.0, 1.8, 1.9, 1.7, 1.8, 1.7), and d_i = −2.0 for all items. The six items in the third test had values for the first discrimination parameter exactly twice those for the second parameter, except for Item 6 for which the values slightly deviated from this proportion: a_1 = (2.0, 2.0, 2.6, 2.4, 2.0, 3.0), a_2 = (1.0, 1.0, 1.3, 1.2, 1.0, 1.7), and d_i = 0.0 for all items. The result is a case of *weak identifiability*, which reveals itself by a variance function for $\hat{\theta}_1$ with low values only locally along a line in the θ plane and a function for $\hat{\theta}_2$ that never takes on any small value. (Note that for readability in all three figures the surfaces are cut at a height of 100.) The figures show a large variety of possible shapes for the two variance functions. Therefore, only a carefully designed test assembly algorithm can give these functions a desired shape.

Targets for Variance Functions

Targets for the two variance functions are proposed for the multidimensional test assembly process. Graphically, this means that tests are assembled such that the plots of their variance functions meet previously defined forms. For example, if θ_1 is considered to be more important than θ_2, a target for $\mathrm{Var}(\hat{\theta}_1|\theta_1,\theta_2)$ uniformly lower than that for $\mathrm{Var}(\hat{\theta}_2|\theta_1,\theta_2)$ over the θ area of interest makes sense. The choice of targets for variance functions is in spirit with the criterion of A-optimality in optimal design theory (van der Linden, 1994).

Computational Complications

Test assembly with simultaneous targets for two distinct functions is an example of a multiobjective decision problem. Standard approaches to decision problems with two objectives are, for example, to combine the two objectives into one objective function or to focus on one as the objective function and represent the other by a constraint with an optimally selected bound. More important, however, is the fact that the two expressions in Equations 14 and 15 are nonlinear. A realistic objective function based on the difference between the two expressions and their targets will also be nonlinear. Due to this complication, algorithms allowing for optimal multidimensional test assembly that operate in polynomial time are not available. Hence, unless the problem is trivially small, the use of a heuristic that yields good, but not necessarily the best, solutions seems to be the only remaining possibility.

The Multidimensional Maximin Model

Further analysis of the variance functions in Equations 14 and 15 reveals that, although nonlinear, they consist of sums, each of which is additive in the items. The role of these sums becomes more obvious if decision variables are added to Equation 14, and the variance function of $\hat{\theta}_1$ is written as

$$\mathrm{Var}(\hat{\theta}_1|\theta_1,\theta_2) = \left[\sum_{i=1}^{I} a_{1i}^2 P_i Q_i x_i - \left(\sum_{i=1}^{I} a_{1i} a_{2i} P_i Q_i x_i\right)^2 \Big/ \left(\sum_{i=1}^{I} a_{2i}^2 P_i Q_i x_i\right)\right]^{-1}. \quad (16)$$

Thus, for a fixed value of (θ_1,θ_2), the function in Equation 16 decreases if the values of the decision variables x_i, i = 1, ..., I, are selected such that

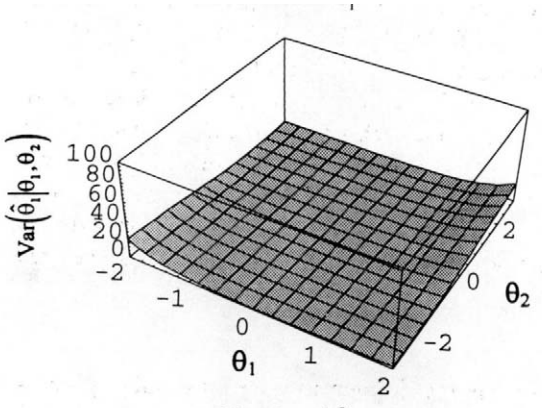
$$\sum_{i=1}^{I} a_{1i}^2 P_i Q_i x_i \text{ increases;} \quad (17)$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

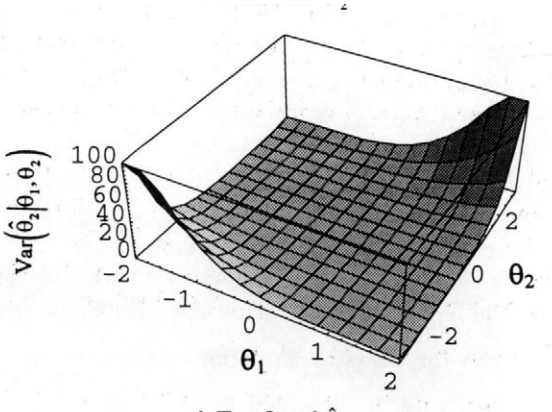
Figure 1

Variance Functions for $\hat{\theta}_1$ and $\hat{\theta}_2$ (See Text for Item Parameter Values)

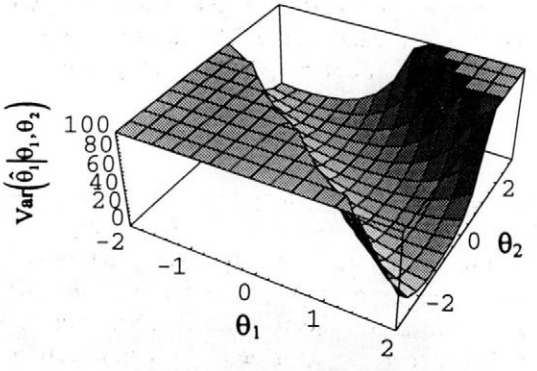
a. Test 1 and $\hat{\theta}_1$



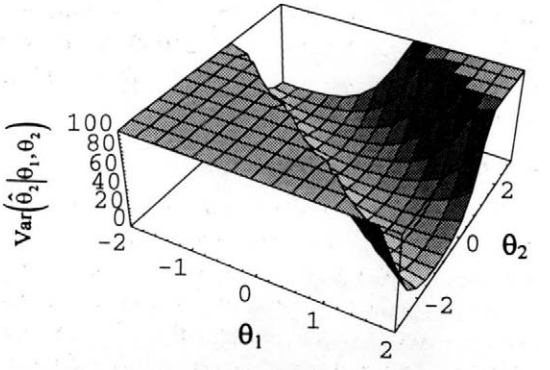
b. Test 1 and $\hat{\theta}_2$



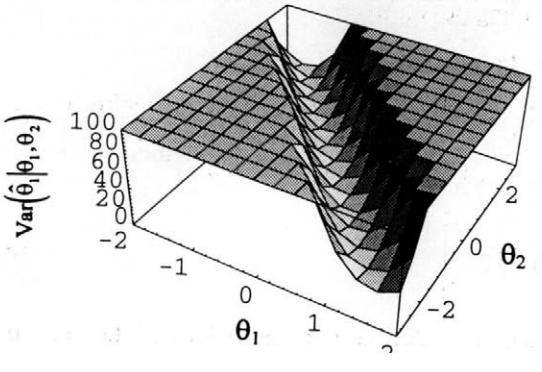
c. Test 2 and $\hat{\theta}_1$



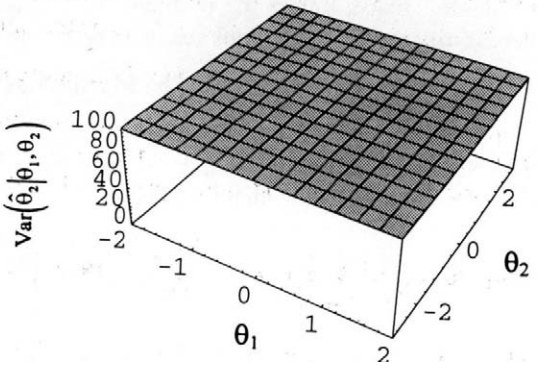
d. Test 2 and $\hat{\theta}_2$



e. Test 3 and $\hat{\theta}_1$



f. Test 3 and $\hat{\theta}_2$



Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

$$\sum_{i=1}^{I} a_{2i}^2 P_i Q_i x_i \text{ increases;} \tag{18}$$

and

$$\sum_{i=1}^{I} a_{1i} a_{2i} P_i Q_i x_i \text{ decreases.} \tag{19}$$

However, note that for a fixed set of item parameter values, the expression in Equation 19 cannot decrease independently of the expressions in Equations 17 and 18. In fact, a tradeoff exists between these two sets of expressions because any choice of values that decrease the last expression also decrease the first two expressions. The optimum value of Equation 16 thus depends on the relative rates of change of the three expressions. This fact suggests the use of a heuristic in which the expression in Equation 19 is minimized for a systematically varying series of lower bounds on the expressions in Equations 17 and 18.

Consider the following variant of the maximin model in Equations 4–7 in which, for a selection of θ points $(\theta_{1p}, \theta_{2q})$, $p = 1, ..., P$, $q = 1, ..., Q$, minimization of the expression in Equation 19 is taken as the objective function and the expressions in Equations 18 and 19 are constrained by lower bounds:

$$\text{minimize } y, \tag{20}$$

subject to

$$\sum_{i=1}^{I} a_{1i} a_{2i} P_i(\theta_{1p}, \theta_{2q}) Q_i(\theta_{1p}, \theta_{2q}) x_i - y \le 0 \quad p = 1, \ldots, P, q = 1, \ldots, Q\,, \tag{21}$$

$$\sum_{i=1}^{I} a_{1i}^2 P_i(\theta_{1p}, \theta_{2q}) Q_i(\theta_{1p}, \theta_{2q}) x_i \ge c_1 \quad p = 1, \ldots, P, q = 1, \ldots, Q, \tag{22}$$

$$\sum_{i=1}^{I} a_{2i}^2 P_i(\theta_{1p}, \theta_{2q}) Q_i(\theta_{1p}, \theta_{2q}) x_i \ge c_2 \quad p = 1, \ldots, P, q = 1, \ldots, Q\,, \tag{23}$$

$$\sum_{i=1}^{I} x_i = n\,, \tag{24}$$

$$x_i \in \{0,1\}, \quad i = 1, \ldots, I\,, \tag{25}$$

and

$$y \ge 0. \tag{26}$$

The basic idea is to systematically vary the values of c_1 and c_2 until optimal variance functions are found.

Selection of Values for c_1 and c_2

First, note that the following inequalities hold:

$$0 \le \sum_{i=1}^{I} a_{1i}^2 P_i Q_i x_i \le .25 \sum_{\max} a_{1i}^2\,, \tag{27}$$

and

$$0 \le \sum_{i=1}^{I} a_{2i}^2 P_i Q_i x_i \le .25 \sum_{\max} a_{2i}^2\,, \tag{28}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

where the right-hand sums are taken over the n items with the largest values for a_{1i} and a_{2i} in the item pool, respectively. Thus, the right-hand sides are used as upper bounds for c_1 and c_2. However, note that items with high values for a_{1i} are not necessarily those with high values for a_{2i} and vice versa; therefore, these bounds will seldom be reached in practice.

Second, if c_1 and/or c_2 are set high, overconstraining may occur and no feasible solution will be found. If infeasibility is found for certain values of c_1 and c_2, no larger values have to be tried because these will also yield infeasibility.

Third, for brevity, let

$$u \equiv \sum_{i=1}^{I} a_{1i}^2 P_i Q_i \,, \tag{29}$$

$$v \equiv \sum_{i=1}^{I} a_{2i}^2 P_i Q_i \,, \tag{30}$$

and

$$w \equiv \sum_{i=1}^{I} a_{1i} a_{2i} P_i Q_i \,. \tag{31}$$

Suppose no dependences exist between u, v, and w. The following partial derivatives then show the impact of u and v on the variance function of $\hat{\theta}_1$:

$$\frac{\partial \mathrm{Var}\left(\hat{\theta}_1 | \theta_1, \theta_2\right)}{\partial u} = \frac{-v^2}{\left(uv - w^2\right)^2} \,, \tag{32}$$

and

$$\frac{\partial \mathrm{Var}\left(\hat{\theta}_1 | \theta_1, \theta_2\right)}{\partial v} = \frac{-w^2}{\left(uv - w^2\right)^2} \,. \tag{33}$$

As $u,v,w \geq 0$, the derivatives are negative for all possible values of u, v, and w (provided that $uv \neq w$). Consequently, as already assumed in Equations 17 and 18, for a fixed value of (θ_1, θ_2), $\mathrm{Var}\left(\hat{\theta}_1 | \theta_1, \theta_2\right)$ is minimal if u is minimal and v is maximal. However, in the model, w is minimized, and the derivatives in Equations 32 and 33 show that if an optimal solution is approached, the marginal contribution of v to $\mathrm{Var}\left(\hat{\theta}_1 | \theta_1, \theta_2\right)$ is likely to be smaller than the contribution of u. If w approaches 0, the contribution of v becomes negligible. By symmetry, the reverse conclusion holds for the contributions of u and v to the variance function of $\hat{\theta}_2$. This suggests that a larger value of c_1 relative to c_2 favors minimization of $\mathrm{Var}\left(\hat{\theta}_1 | \theta_1, \theta_2\right)$, whereas a smaller value of c_1 favors minimization of $\mathrm{Var}\left(\hat{\theta}_2 | \theta_1, \theta_2\right)$. However, the actual problem is one of combinatorial optimization over a finite pool of possible values for the item parameters. Also, as already noted, these values create dependencies between the expressions in Equations 17–19. Therefore, it is recommended that this suggestion be evaluated for the actual item pool in use. For an empirical example, see the analyses presented below.

A Heuristic

The following heuristic can be used to find a (nearly) optimal solution to the test assembly problem:

1. Select a grid of values for $(\theta_{1p}, \theta_{2q})$ that covers the θ area of interest. Because the variance functions are well-behaved smooth functions, a 3×3 or 4×4 grid will generally suffice. There is no need to space the points evenly or to have the same numbers of points along both dimensions.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

2. Select a series of values for (c_1, c_2) covering the range of possible values below the upper bounds in Equations 27 and 28, taking into account the distribution of the values of the item parameters in the pool as well as the goal of the test (see below);
3. Solve the model in Equations 20–26 using, for example, standard software for LP or one of the algorithms in CONTEST (Timminga & van der Linden, 1996);
4. Calculate the two variance functions for each solution in the previous step;
5. Based on the results, repeat Steps 3 and 4 for a finer grid of values for (c_1, c_2) in the neighborhood of the value for which the best variance functions were obtained;
6. Repeat Step 5 until the fit of the variance functions to their targets cannot be improved any further.

Experience with earlier runs of the heuristic for a given item pool can be used to make the first selection of values for (c_1, c_2) more effective. For example, once infeasibility is met for certain values of (c_1, c_2), it makes no sense to use larger values for (c_1, c_2) for any later test assembled from the same item pool. This conclusion remains valid when items are removed from the pool or new constraints are added. An implementation of the heuristic for the case of two flat variance functions is described in the empirical example below.

Different Cases of Multidimensional Test Assembly

Five different cases of test assembly are considered in which multidimensionality of the item pool plays a role. For each case, a different use of the multidimensional model in Equations 20–26 is proposed, with the exception of one case that leads to the use of a modified version of the unidimensional model in Equations 4–8. The main criteria used to classify the five cases are (1) whether the traits are intentional or should be viewed as "nuisance traits," and (2) whether or not the traits underlying the test should display a "simple structure."

Case 1: Two Intentional Traits

In Case 1, test items are designed to measure two traits, and scores are reported on both traits for each examinee. Thus, for each possible (θ_1, θ_2) the test should produce variances of $\hat{\theta}_1$ and $\hat{\theta}_2$ that meet realistic targets.

The model to be used for Case 1 is the multidimensional maximin model in Equations 20–26 with linear constraints added for any remaining test specifications. As already suggested, the relative sizes of the values of c_1 and c_2 can be used to control for the importance of the two variance functions.

Case 2: One Intentional and One Nuisance Trait

The test items in the pool are designed to measure one intentional trait but are also sensitive to another trait. When scoring the test, the nuisance trait is ignored and only a score for the intentional trait is reported. An obvious example of a nuisance trait is "differential item functioning," because a focal and a reference group have different distributions on the nuisance trait. Removing the effect of the nuisance trait by fitting a two-dimensional IRT model and scoring only the intentional trait will likely yield trait estimates that are more informative than simply removing all items sensitive to the nuisance trait from the test.

The best approach in this case is to ignore the variance function for the estimator of the nuisance trait, and set a target for the intentional trait only. If θ_2 is the nuisance trait, this approach is implemented if Case 1 is applied, but with c_2 small relative to c_1. Again, additional linear constraints can be added to the model to deal with other test specifications.

Case 3: One Composite Trait

In Case 3, both traits are intentional but estimates of the linear combination $\beta_1\theta_1 + \beta_2\theta_2$, with $\beta_1, \beta_2 > 0$ (weights chosen by the test assembler), are reported. A practical motivation for Case 3 might be that the

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

construct measured by the test is truly two-dimensional but that test users want a single score equally reflecting both traits. The variance function of the estimator of the linear composite is equal to

$$\mathrm{Var}\left(\beta_1\hat{\theta}_1+\beta_2\hat{\theta}_2|\theta_1,\theta_2\right)=\beta_1^2\mathrm{Var}\left(\hat{\theta}_1|\theta_1,\theta_2\right)+\beta_2^2\mathrm{Var}\left(\hat{\theta}_2|\theta_1,\theta_2\right)+2\beta_1\beta_2\mathrm{Cov}\left(\hat{\theta}_1,\hat{\theta}_2|\theta_1,\theta_2\right), \tag{34}$$

where Cov is the covariance (Ackerman, 1994, equations 15–16). Although Equation 34 is also an expression consisting of the sums of the elements in the information matrix in Equation 11, analysis of Equation 34 shows that it misses the monotonicity that could lead to the conditions in Equations 17–19. The best solution in Case 3, therefore, is to rotate the trait space such that in the reparameterized model the composite corresponds to the first trait dimension. Then, Case 3 is identical to Case 2.

Case 4: Simple Trait Structure

The item pool is again assumed to measure two intentional traits, but the test has to be assembled such that one subtest is maximally informative on θ_1 and another subtest on θ_2. Case 4 may arise if, for diagnostic purposes, test performance must be reported at the item level and it is thus necessary to know which items best measure θ_1 and which items best measure θ_2.

Let n_1 be the number of items required to be informative on θ_1 and n_2 the number of items informative on θ_2. An obvious approach in Case 4 first applies the multidimensional model in Equations 20–26 to assemble n_1 items under the condition $c_1 > c_2$, and a second time to assemble n_2 items under the reverse condition $c_1 < c_2$ removing the items already selected from the pool. However, a clear disadvantage of a sequential approach is that items fitting the constraints of the second subtest better may already have been selected for the first subtest. Also, it is not possible to directly constrain item selection with respect to item content, format, and so forth, at the level of the complete test.

A more favorable solution, therefore, is to select the two subtests simultaneously. This choice leads to an adaptation of the multidimensional model in Equations 20–26. New decision variables x_{is} are introduced that take the value of 1 if item i is assigned to subtest s, and the value 0 otherwise (s = 1, 2). The adapted model is

$$\text{minimize } y, \tag{35}$$

subject to

$$\sum_{s=1}^{2}\sum_{i=1}^{I} a_{1i}a_{2i}P_i\left(\theta_{1p},\theta_{2q}\right)Q_i\left(\theta_{1p},\theta_{2q}\right)x_{is}-y\geq 0,\quad p=1,\ldots,P,\quad q=1,\ldots,Q, \tag{36}$$

$$\sum_{s=1}^{2}\sum_{i=1}^{I} a_{1i}^2P_i\left(\theta_{1p},\theta_{2q}\right)Q_i\left(\theta_{1p},\theta_{2q}\right)x_{is}\geq c_1\quad p=1,\ldots,P,\quad q=1,\ldots,Q, \tag{37}$$

$$\sum_{s=1}^{2}\sum_{i=1}^{I} a_{2i}^2P_i\left(\theta_{1p},\theta_{2q}\right)Q_i\left(\theta_{1p},\theta_{2q}\right)x_{is}\geq c_2\quad p=1,\ldots,P,\quad q=1,\ldots,Q, \tag{38}$$

$$\sum_{s=1}^{I} x_{is}=n_s,\quad s=1,2, \tag{39}$$

$$\sum_{s=1}^{2} x_{is}\leq 1,\quad i=1,\ldots,I, \tag{40}$$

$$x_{is}\in\{0,1\}, \tag{41}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

and

$$y \geq 0. \tag{42}$$

New constraints in the model are those in Equation 39 that define the lengths of the two subtests and those in Equation 40 that prevent the items from being assigned to both subtests. The model can be solved using the heuristic proposed here. However, doubling the number of decision variables generally has an effect on the speed of the algorithms and heuristics comparable to doubling the size of the item pool; consequently, some of the heuristics slow down considerably.

Case 5: Simple Trait Structure

For completeness, the case of two subpools of items each fitting a unidimensional IRT model but with the complete pool fitting only a two-dimensional model is mentioned. The practical motivation for assembling a test with this simple structure for its trait space is the same as that in Case 4. Again, a simple solution would assemble the two subtests sequentially, but the same objections to sequential assembly apply. A model for simultaneous assembly can be obtained using the same decision variables in Equations 4–8 as in Equations 35–42.

Discussion

Cases 1–5 demonstrate that it is never correct to assemble tests from a multidimensional pool using traditional unidimensional procedures. It is poor test construction practice to fit a unidimensional model to a multidimensional pool and to use the parameter estimates and information functions as if they were the correct quantities. Rather, the assembly procedures must also take multidimensionality into account, even if interest is in tests that are optimal for the measurement of one-dimensional traits (Cases 2–4). The only exception is Case 5 in which subtests are assembled from separate subsets of items, each of which fits a unidimensional model (Case 5).

Empirical Example

Method

Data from an ACT Assessment Program mathematics item pool were used to assemble a test. The pool consisted of 176 items to which a two-dimensional version of the model in Equation 9 showed an acceptable fit. The items in the pool were classified according to content: plane geometry (PG), prealgebra (PA), elementary algebra (EA), coordinate geometry (CG), trigonometry (TG), and intermediate algebra (IA); and according to skill: basic skill (BS), application (AP), and analysis (AN). Two tests with flat variance functions for both abilities over the complete grid of points defined by $\theta_1, \theta_2 = -2, -1, 0, 1, 2$ were assembled, and measurement of both abilities was assumed to be intentional and equally important. One test was assembled using the basic model in Equations 20–26 (Model I). The other test was assembled adding the following set of constraints to Model I to simulate the presence of content and skill specifications in the assembly program (Model II):

$$\sum_{i \in V_{PG}} x_i \geq 5, \tag{43}$$

$$\sum_{i \in V_{PA}} x_i \geq 5, \tag{44}$$

$$\sum_{i \in V_{EA}} x_i \geq 5, \tag{45}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

$$\sum_{i \in V_{CG}} x_i \geq 5, \tag{46}$$

$$\sum_{i \in V_{TG}} x_i \geq 5, \tag{47}$$

$$\sum_{i \in V_{IA}} x_i \geq 5, \tag{48}$$

$$\sum_{i \in V_{BS}} x_i \geq 15, \tag{49}$$

$$\sum_{i \in V_{AP}} x_i \geq 15, \tag{50}$$

and

$$\sum_{i \in V_{AN}} x_i \geq 5, \tag{51}$$

where, for example, V_{PG} is the set that indexes the items with the content classification plane geometry. For both Models I and II, test length was set at $n = 50$.

Models I and II were solved using the First Acceptable Integer Solution Algorithm as implemented in the ConTEST program [a detailed description of the algorithm is given in Adema (1992a) and Timminga & van der Linden (1996, sect. 6.6)]. This algorithm is based on the following principles. First, the value of the objective function for the solution to the relaxed version of the model with decision variables $x_i \in [0, 1]$ is calculated. For test assembly problems, this value usually is an excellent upper bound to the solution to the original model. Second, a branch-and-bound search is used to find a solution to the original problem. The search stops as soon as a feasible solution with a value for the objective function larger than $(1 - \alpha)\%$ of the value of the upper bound is found. Third, optimal reduced costs in the relaxed version of the model are used to fix some of the decision variables to the values 0 or 1. This reduces the number of variables in the problem, and hence the size of the search tree. In this example, the tolerance parameter α was set equal to its default value of 5%. All runs of the computer program took less than two seconds of computing time on a 486/66MHZ personal computer to reach a solution.

Results

Because the two variance functions were assumed to be equally important, the values of c_1 and c_2 in Equations 22–23 were set equal to each other. For both models, values of $c_1 = c_2$ larger than or equal to 1.4 led to overconstraining and no feasible solutions were found. Therefore, the two models were run for $c_1 = c_2 = 0.0, .1, .2, ..., 1.3$.

The results are summarized in Table 1. Because the variance functions had to be both low and flat, the mean value (μ) plus one standard deviation (σ) of the values of the two variance functions over 25 points of the grid of (θ_1, θ_2) values was used as a summary measure to be minimized. For Model I, the minimal value of 1.318 for $\mu + \sigma$ was obtained for $c_1 = c_2 = 1.0$. Plots of the variance functions $\mathrm{Var}(\hat{\theta}_1|\theta_1, \theta_2)$ and $\mathrm{Var}(\hat{\theta}_2|\theta_1, \theta_2)$ associated with the items in this solution are given in Figures 2a and 2b, respectively. Both functions show a flat surface over the trait space considered, albeit the function for $\hat{\theta}_1$ has a tendency to slightly increase for θ_1 approaching 2.0, whereas the function for $\hat{\theta}_2$ increases for θ_2 approaching –2.0. For Model II, the best solution was obtained for $c_1 = c_2 = .9$ (1.379). This solution had 11 items different from those in the solution to Model I. Nevertheless, the numerical results in Table 1 and Figures 2c and 2d of the

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 1
Values of μ and σ for Selected Values of $c_1 = c_2$ for Model I and Model II

	Model I			Model II		
$c_1 = c_2$	μ	σ	$\mu + \sigma$	μ	σ	$\mu + \sigma$
0.0	1.551	.833	2.384	1.586	.922	2.508
.1	1.551	.833	2.384	1.586	.922	2.508
.2	1.551	.833	2.384	1.586	.922	2.508
.3	1.551	.833	2.384	1.586	.922	2.508
.4	1.551	.833	2.384	1.562	.833	2.395
.5	1.386	.505	1.891	1.387	.508	1.895
.6	1.311	.373	1.684	1.335	.384	1.719
.7	1.286	.369	1.655	1.277	.363	1.640
.8	1.180	.313	1.493	1.189	.322	1.511
.9	1.057	.295	1.352	1.085	.294	1.379
1.0	1.037	.281	1.318	1.104	.294	1.398
1.1	1.169	.320	1.489	1.231	.325	1.556
1.2	1.500	.437	1.937	1.479	.410	1.889
1.3	1.826	.486	2.313	1.907	.509	2.416
1.4	*			*		

*No feasible solution.

two variance functions show that adding the extra constraints to Model I hardly deteriorated the results.

To assess the numerical effects of setting c_1 lower or higher than c_2, solutions for Model I were computed over the full range of possible values for c_2 both for $c_1 = .2$ and $c_1 = 1.2$. These two values for c_1 were near the extremes of the range of values in Table 1 for which feasible solutions were obtained. The results are presented in Table 2. The general conclusion is that the lower value for c_1 favored minimization of the variance function for $\hat{\theta}_2$, both in terms of its average value and spread, whereas the higher value of c_1 favored minimization of the function for $\hat{\theta}_1$. For example, for $c_1 = .2$ and $c_2 = 1.2$, the variance function for

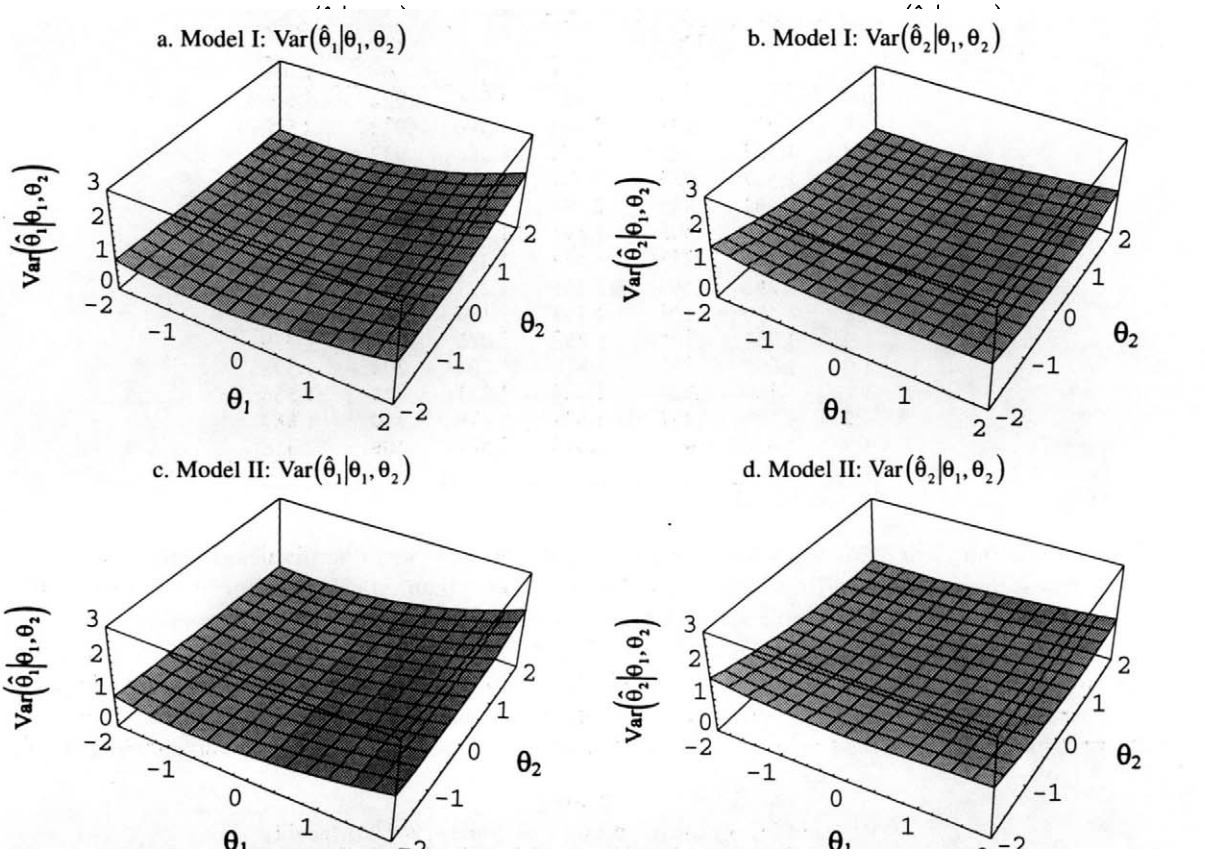
Table 2
Values of (μ_1, σ_1) and (μ_2, σ_2) for $c_1 = .2$ and $c_1 = 1.2$ (Model I)

	$c_1 = .2$				$c_1 = 1.2$			
c_2	μ_1	σ_1	μ_2	σ_2	μ_1	σ_1	μ_2	σ_2
0.0	2.029	.893	1.073	.038	1.296	.265	3.472	.516
.1	2.029	.893	1.073	.038	1.296	.265	3.472	.516
.2	2.029	.893	1.073	.038	1.296	.265	3.472	.516
.3	2.029	.893	1.073	.038	1.296	.265	3.472	.516
.4	2.029	.893	1.073	.038	1.296	.265	3.472	.516
.5	2.029	.893	1.073	.038	1.296	.265	3.472	.516
.6	2.029	.893	1.073	.038	1.235	.230	3.085	.373
.7	2.029	.893	1.073	.038	.974	.102	2.069	.123
.8	2.029	.893	1.073	.038	1.010	.110	1.698	.081
.9	2.412	1.624	1.014	.042	.923	.096	1.456	.063
1.0	2.663	2.330	.949	.042	.981	.107	1.441	.066
1.1	7.123	2.449	1.426	.062	1.212	.166	1.441	.091
1.2	12.819	1.650	2.036	.123	1.428	.232	1.572	.141
1.3	11.606	2.461	1.951	.140	1.564	.276	1.533	.146
1.4	8.520	1.769	1.661	.088	2.230	.373	1.889	.243
1.5	7.193	1.254	1.643	.107	*			
1.6	4.554	1.047	1.470	.121	*			
1.7	*				*			

*No feasible solution.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Figure 2
Variance Functions for the Tests Assembled Under Models I and II



$\hat{\theta}_1$ had a much higher mean value than the function for $\hat{\theta}_2$, whereas for $c_1 = 1.2$ and $c_2 = .2$ the opposite occurred.

Discussion

The choice to base the assembly of tests measuring multiple traits on the variance functions associated with the trait estimators seems obvious. However, as indicated above, the choice involves a multiobjective decision problem with nonlinear functions. A model and heuristic were developed here to solve the problem for the case of two traits. Implementations of the heuristic for targets other than those for the case of two intentional traits in the empirical example above still have to be examined. It is not unlikely that practical experience with the heuristic will reveal that, for some of the cases discussed above, certain patterns of item parameter values guarantee optimal variance functions. If so, this knowledge could be used to further improve the focus of the heuristic as well as future item pool design.

The general case of assembling tests from a T-dimensional item pool involves inversion of a $T \times T$ information matrix with elements analogous to those in Equation 11. The variance functions (i.e., the diagonal elements of this inverse) are generalizations of Equations 14–15 to ratios of sums of products, each of which consists of T elements from the information matrix. These elements are still linear in the

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

decision variables, but linearization of the full problem of minimizing the variance functions must deal with more complicated tradeoffs between the elements than those met in the present paper. Research on heuristics addressing this general case is in progress.

References

Ackerman, T. A. (1989, March). *An alternative methodology for creating parallel test forms using the IRT information function.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

Ackerman, T. A. (1994). Creating a test information profile for a two-dimensional latent space. *Applied Psychological Measurement, 18,* 257–275.

Adema, J. J. (1990). The construction of customized two-staged tests. *Journal of Educational Measurement, 27,* 241–253.

Adema, J. J. (1992a). Implementations of the branch-and-bound method for test construction. *Methodika, 6,* 99–117.

Adema, J. J. (1992b). Methods and models for the construction of weakly parallel tests. *Applied Psychological Measurement, 16,* 53–63.

Adema, J. J., Boekkooi-Timminga, E., & van der Linden, W. J. (1991). Achievement test construction using 0-1 linear programming. *European Journal of Operations Research, 55,* 103–111.

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

Armstrong, R. D., & Jones, D. H. (1992). Polynomial algorithms for item matching. *Applied Psychological Measurement, 16,* 365–373.

Armstrong, R. D., Jones, D. H., & Wu, I.-L. (1992). An automated test development of parallel tests from a seed test. *Psychometrika, 57,* 271–288.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 397–479). Reading MA: Addison-Wesley.

Boekkooi-Timminga, E. (1987). Simultaneous test construction by zero-one programming. *Methodika, 1,* 101–112.

Boekkooi-Timminga, E. (1990). The construction of parallel tests from IRT-based item banks. *Journal of Educational Statistics, 15,* 129–145.

Feuerman, F., & Weiss, H. (1973). A mathematical programming model for test construction and scoring. *Management Science, 19,* 961–966.

Kendall, M. G., & Stuart, A. (1976). *The advanced theory of statistics* (Vol. 2; 4th ed.). London: Griffin & Co.

Luecht, R. M., & Hirsch, T. M. (1992). Computerized test construction using average growth approximation of target information functions. *Applied Psychological Measurement, 16,* 41–52.

McKinley, R. L., & Reckase, M. N. (1983). *An extension of the two-parameter logistic model to the multidimensional latent space* (Research Rep. ONR 83-2). Iowa City IA: American College Testing.

Nemhauser, G., & Wolsey, L. (1988). *Integer and combinatorial optimization.* New York: Wiley.

Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9,* 401–412.

Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271–286). New York: Springer-Verlag.

Samejima, F. (1974). Normal ogive model for the continuous response level in the multidimensional latent space. *Psychometrika, 39,* 111–121.

Swanson, L., & Stocking, M. L. (1993). A model and heuristic for solving very large item selection problems. *Applied Psychological Measurement, 17,* 151–166.

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

Timminga, E., & Adema, J. J. (1995). Test construction from item banks. In G. H. Fischer & I. W. Molenaar (Eds.), *The Rasch model: Foundations, recent developments, and applications* (pp. 111–127). New York: Springer-Verlag.

Timminga, E., van der Linden, W. J., & Schweizer, D. A. (1996). *ConTEST* [Computer program and manual]. Groningen, The Netherlands: iec ProGAMMA.

van der Linden, W. J. (1994). Optimum design in item response theory: Applications to test assembly and item calibration. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 308–318). New York: Springer-Verlag.

van der Linden, W. J., & Boekkooi-Timminga, E. (1988). A zero-one programming approach to Gulliksen's matched random subsets method. *Applied Psychological Measurement, 12,* 201–209.

van der Linden, W. J., & Boekkooi-Timminga, E. (1989). A maximin model for test design with practical constraints. *Psychometrika, 53,* 237–247.

van der Linden, W. J., & Luecht, R. M. (1996). An optimization model for test assembly to match observed-score distributions. In G. Engelhard & M. Wilson

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

(Eds.), *Objective measurement: Theory into practice* (Vol. 3, pp. 405–418). Norwood NJ: Ablex Publishing Company.

Votaw, D. F. (1952). Methods of solving some personnel classification problems. *Psychometrika, 17,* 255–266.

Wagner, H. M. (1975). *Principles of operations research* (2nd ed.). London: Prentice/Hall.

Yen, W. M. (1983). Use of the three-parameter model in the development of standardized achievement tests. In R. K. Hambleton (Ed.), *Applications of item response theory* (pp. 123–141). Vancouver: Educational Research Institute of British Columbia.

Acknowledgments

The author is indebted to Wim M. M. Tielen for his computational support and to Terry A. Ackerman for the dataset used in the empirical example.

Author's Address

Send requests for reprints or further information to Wim J. van der Linden, Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: vanderlinden@edte.utwente.nl.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/