DOCUMENT RESUME

ED 434 150		TM 030 114

AUTHOR	Veldkamp, Bernard P.; van der Linden, Wim J.
TITLE	Designing Item Pools for Computerized Adaptive Testing. Research Report 99-03.
INSTITUTION	Twente Univ., Enschede (Netherlands). Faculty of Educational Science and Technology.
PUB DATE	1999-00-00
NOTE	20p.
AVAILABLE FROM	Faculty of Educational Science and Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Ability; *Adaptive Testing; Classification; *Computer Assisted Testing; Estimation (Mathematics); Foreign Countries; *Item Banks; Simulation; Statistical Distributions; *Test Construction; Test Items
IDENTIFIERS	*Integer Programming; Test Specifications

ABSTRACT

A method of item pool design is proposed that uses an optimal blueprint for the item pool calculated from the test specifications. The blueprint is a document that specifies the attributes that the items in the computerized adaptive test (CAT) pool should have. The blueprint can be a starting point for the item writing process, and it can be used to assemble item pools in a system of rotating pools from a master pool. The blueprint is also useful for item pool maintenance. Designing the blueprint begins with analyzing the specifications for the CAT, a step amounting to the formation of a classification table involving categorization of quantitative item attributes. Using this table, an integer programming model for the assembly of the shadow tests in the CAT simulation is constructed. An estimate of the ability distribution of the identified population of examinees is obtained, and the CAT simulation is carried out using the integer programming model for the shadow tests and sampling simulees from the ability distribution. The blueprint is then calculated from the counts of the number of items from the cells in the classification table. The best way to implement the blueprint is in a sequential fashion recalculating the blueprint after a certain portion of the items has actually been written and tested so that their attribute values are known. (Contains 16 references and a list of University of Twente research reports.) (SLD)

Designing Item Pools for Computerized Adaptive Testing

Research Report
99-03

Bernard P. Veldkamp
Wim J. van der Linden

faculty of
EDUCATIONAL SCIENCE AND TECHNOLOGY

University of Twente

Department of
Educational Measurement and Data Analysis

Designing Item Pools for Computerized Adaptive Testing

Bernard P. Veldkamp
Wim J. van der Linden

Introduction

The basic assumption underlying this paper is that computerized adaptive testing (CAT) can be viewed as an instance of constrained sequential optimization. The selection of each next item in the test involves optimization of an objective function, for example, maximization of the information in the item at the current ability estimate or the fit of the information in the test to a target function. This optimization is subject to various constraints on item and test attributes dealing with, for example, their response format or content.

The idea of CAT as constrained sequential optimization is adopted in the shadow test approach to adaptive testing (van der Linden, in press; van der Linden & Reese, 1998). In this approach, at each step a full test ("shadow test") is assembled. The item to be administered is selected from this shadow test rather than the full item pool. Each shadow tests is assembled to be optimal subject to a set of constraints representing the test specifications. As a result, the CAT also meets all constraints. At the same time, it tends to have an optimal value for its objective function.

The algorithm for constrained CAT with shadow tests can be summarized as follows:

Step 1: Choose an initial value of the examinee's ability parameter θ.

Step 2: Assemble the first shadow test such that all constraints are met and the objective function is optimized.

Step 3: Administer an item from the shadow test with optimal properties at the current θ estimate.

Step 4: Update all parameters in the test assembly model.

Step 5: Assemble a new shadow test fixing the items already administered.

Step 6: Repeat Steps 3-6 until n items have been administered.

The model used to assemble the shadow tests is a 0-1 linear programming model for test assembly. An example of a test assembly model is presented later in this paper.

An important constraint in CAT is the one on the exposure rates of the items in the pool. To maintain item pool security, the items in the pool should not be administered more frequently than certain target values. Simpson and Hetter (1985) developed a probabilistic method for controlling the exposure rates of the items. After an item is selected by the CAT algorithm, a probability experiment is run to determine whether the item is or is not actually administered. By manipulating the probabilities in this experiment, the expected exposure rates of the items can be kept below their target values. Several modifications of this method have been developed (Davey & Nering, 1998; Stocking & Lewis, 1998). However, though these methods guarantee upper bounds on the exposure rates of the items, they do not guarantee substantial use of all items in the pool. Item pools can contain large sections of items items that are poor in the sense of a low contribution to the objective function optimized in the CAT algorithm or because they have attribute values that are overrepresented in the pool relative to the requirements in the constraints. Such items are seldom chosen in the CAT and point at inefficient item pool design.

To raise item pool efficiency, a method of item pool design is proposed. The main product of the method is an optimal blueprint for the item item pool calculated from the test specifications, that is, a document specifying what attributes the items in the CAT pool should have. This optimal blueprint can be used in several ways. First of all, it can be used as a starting point for the item writing process and suggest optimal division of labor among the available item writers. Second, the blueprint can be used to assemble item pools in a system of rotating pools from a master pool. Systems of rotating item pools have been proposed as an effective means of enlarging item security (Way, 1998; Way, Steffen & Anderson, 1998). Third, blueprints can be used for item pool maintenance, that is, guide periodic decisions on what items in the pool should be replaced or what types of additional items should be written.

Item Pool Design

The subject of item pool design has been addressed earlier. A general description of the process of developing item pools for CAT is presented in Flaugher (1990). This author outlines several steps in the development of an item pool and discusses current practices at these steps. A common feature of the process described in Flaugher and the method in the present paper is the use of computer simulation. However, in Flaugher's outline, computer simulation is used to evaluate the functioning of the item pool once the items have been written and field tested whereas here computer simulation is used to calculate a blueprint for the item pool to guide the item writing and pool maintenance process.

Methods of item pool design based on integer programming are presented in Boekkooi-Timminga (1991) and van der Linden, Veldkamp and Reese (in press). These methods can be used to optimize the design of item pools that have to support the assembly of a series of future linear test forms. The method in Boekkooi-Timminga follows a sequential approach calculating the numbers of items needed for these test forms maximizing their information functions. The method assumes an item pool calibrated under the one-parameter logistic (1PL) or Rasch model. The method in van der Linden, Veldkamp and Reese directly calculates a blueprint for the entire pool minimizing an estimate of the costs involved in producing the items. All other test specifications, including those related to the information functions of the test forms, are modeled as constraints in an integer programming model. This method can be used for item pools calibrated under any current IRT model. As will become clear below, the current proposal shares some of its logic with the method in van der Linden, Veldkamp and Reese. However, integer programming is used only to simulate constrained CAT with shadow tests—not to calculate numbers of items needed in the pool. Rather, these numbers are derived from computer simulation.

Both Swanson and Stocking (1998) and Way, Steffen and Anderson (1998; see also Way,1998) address the problem of designing a system of rotating item pools for CAT. This system starts with a master pool from which operational item pools are generated. A basic quantity is the number of operational pools each item should be included in, that is, the degree of item pool overlap. A heuristic based on Swanson and Stocking's (1993) weighted deviation model (WDM) is used to assemble the operational pools from the master pool such that the desired degree of overlap is realized and the pools are as similar as possible. One of the advantages of a system of rotating pools with item overlap is that the exposure rates of the items can be manipulated systematically by increasing or decreasing the number of operational pools they are included in. As will be discussed later, the method in this paper can be applied to design a system of overlapping item pools rather than assemble one from a given master pool.

Designing a Blueprint for CAT Item Pools

The process of designing on optimal blueprint for a CAT item pool goes through the following steps: First, the set of specifications for the CAT is analyzed and all item attributes figuring in the specifications are identified. As shown below, this step amounts to the formulation of a classification table involving possible categorization of quantitative item attributes. Second, using this table an integer programming model for the assembly of the shadow tests in the CAT simulation is formulated. Third, the population of examinees is identified and an estimate of its ability distribution is obtained. In principle, the true distribution is unknown but an accurate estimate may be obtained, for example, from historic data. Fourth, the CAT simulation is carried out using the integer programming model for the shadow tests and sampling simulees from the ability distribution. Counts of the number of times items from the cells in the classification table are used are collected. Fifth, the blueprint is calculated from these counts adjusting them to obtain optimal projections of the item exposure rates.

Some of these steps are now explained in more detail.

Setting Up the Classification Table

The classification table for the item pool is set up distinguishing the following three kinds constraints that can be imposed on the item selection by the CAT algorithm (van der Linden, 1998): (1) constraints on categorical item attributes, (2) constraints on quantitative attributes, and (3) constraints needed to deal with inter-item dependencies.

Categorical item attributes, such as content, format, or item author, partition an item pool into a collection of subsets. If the items are coded by multiple categorical attributes, their Cartesian product introduces a partitioning of the pool. A natural way to represent a partitioning based on categorical attributes is as a classification table. For example, let C1, C2, and C3 represent three levels of an attribute representing item content and let F1 and F2 represent two levels of an attribute representing item format. Table 1 shows the classification table for a partition which has six different cells.

Table 1

Classification table (case of two categorical attributes).

	F1	F2
C1	n_{11}	n_{21}
C2	n_{12}	n_{22}
C3	n_{13}	n_{23}

where n_{ij} represents the number of items in cell (i, j).

Classifications based on quantitative attributes are less straightforward to deal with. Examples of possible quantitative item attributes in CAT are: word counts, difficulty parameters, and discrimination indices. These attributes often have large ranges of possible values. An obvious way to overcome this obstacle is to pool adjacent values. For example, the difficulty parameter in the three parameter logistic IRT model takes real values in the interval $(-\infty, \infty)$. This interval could be partitioned, for example, into the collections of the following twelve subintervals: $((-\infty, -2.5), (-2.5, -2), \ldots, (2, 2.5), (2.5, \infty))$. After such partitioning, quantitative attributes can be used in setting up a classification tables as if they were categorical.

Inter-item dependencies deal with possible relations of exclusion and inclusion between the items in the pool. An example of an exclusion relation is the one between items in so-called enemy sets. Such items can not be included in the same test, for example, because they happen to contain clues to each others solution. However, if previous experience has shown that enemies tended to be items with certain combinations of attributes, constraints can be included in the test assembly model for the shadow tests in the CAT algorithm to prevent such combinations from happening. As a result, more realistic item exposures rates are obtained (see below). An example of an inclusion relation is the one between set-based items in a test. Such items refer to a common stimulus, for example, a common description of an experiment in a science test or a reading passage in a language test. Relations between set-based items can be dealt with by setting up a separate classification table based on the stimulus attributes. An example of this table is given in van der Linden, Veldkamp & Reese (in press). It is also possible to deal with set-based items by constraints in the model for the shadow tests. This option is elaborated in Veldkamp & van der Linden (in press) and will not be further addressed here. Both inclusion and exclusion constraints do not involve any item attribute but are logical constraints on the items themselves.

The result of this step is thus a classification table, $C \times Q$, that is the Cartesian product of table C based upon the categorical attributes and table Q based upon the quantitative attributes. Each cell of the table represents a possible subset of items in the pool that have both the same values for the categorical attributes and values for the quantitative attributes that belong to the same interval.

Constrained CAT Simulation

To find out how many items an optimal pool from each cell in table $C \times Q$ should contain, a CAT simulation study is carried out. Each cell in $C \times Q$ is represented by a decision variable in the integer programming model for the shadow test. The values of the attributes associated with the cells are thus automatically associated with their decision variables. For quantitative attributes, midpoints of the intervals can be chosen as attribute values.

The items are supposed to be calibrated by the three-parameter logistic (3PL) model:

$$P_i(\theta_j) \equiv c_i + (1 - c_i)\frac{e^{(a_i \cdot \theta_j + b_i)}}{1 + e^{(a_i \cdot \theta_j + b_i)}}, \tag{1}$$

where $P_i(\theta_j)$ is the probability that a person $j = 1 \ldots J$ with an ability parameter θ_j gives a correct response to an item $i = 1 \ldots I$, a_i is the value for the discrimination parameter, b_i for the difficulty parameter, and c_i for the guessing parameter of item i. Fisher's information in the response on item i for an examinee with ability θ_j is denoted as $I_i(\theta_j)$.

Let x_{cq} be the decision variable for an item from cell (c, q) in table $C \times Q$. Unlike the model for the shadow tests for operational CAT from an actual item pool, in item-pool design, the decision variables are no 0-1 variables but integers that determine how many items are selected from each cell (c, q). Further, let n be the length of the CAT, $\widehat{\theta}_{cq,k-1}$ be the estimate of θ_j after $k-1$ items, S_{k-1} the set of cells with nonzero decision variable after $k-1$ items have been selected. Finally, $V_g, g = 1 \ldots G$, denotes the set of cells in categorical constraint g, $V_h, h = 1 \ldots H$, the set of cells in quantitative constraint h, and $V_e, e = 1 \ldots E$, the set of cells in enemy set e.

The objective function in the model for the shadow tests is proposed to minimize an estimate of the costs involved in writing the items in the pool. Several suggestions for estimates of item writing costs are given in van der Linden, Veldkamp & Reese (in press). Generally, item writing costs can be presented as quantities k_{cq}, $(c, q) \in C \times Q$. In the empirical example below, k_{cq} is chosen to be the inverse of the numbers of items in cell (c, q) in a previous item pool, the idea being that items written more frequently are less likely to be costly. Also, if these costs are dependent on the item writer, it is recommended to adopt "item writer" as a (categorical) item attribute in the item pool blueprint. The blueprint can then also be used for optimal assignment of item-writing instructions to item writers.

The general model for the assembly of the shadow test for the selection of the kth item in the CAT is presented as:

$$\min \sum_{cq \in C \times Q} k_{cq} x_{cq} \qquad \text{(objective function)} \qquad (2)$$

subject to

$$\sum_{cq \in C \times Q} I_{cq}\left(\widehat{\theta}_{k-1}\right) x_{cq} \geq T \qquad \text{(information target)} \qquad (3)$$

$$\sum_{cq \in S_{k-1}} x_{cq} = k-1 \qquad \text{(items already selected)} \qquad (4)$$

$$\sum_{cq \in C \times Q} x_{cq} = n \qquad \text{(test length)} \qquad (5)$$

$$\sum_{cq \in V_g} x_{cq} = n_g \qquad g = 1, ..., G \qquad \text{(categorical constraints)} \qquad (6)$$

$$\sum_{cq \in V_h} f_h(x_{cq}) = n_h \qquad h = 1, ..., H \qquad \text{(quantitative constraints)} \qquad (7)$$

$$\sum_{cq \in V_e} x_{cq} \leq 1 \qquad e = 1, ..., E \qquad \text{(enemy sets)} \qquad (8)$$

$$x_{cq} \in \{0, 1, 2, ...\} \qquad (c, q) \in C \times Q \qquad (9)$$

The objective function in (2) minimizes the estimated item-writing costs. The constraint in (3) requires the information in the CAT at the simulees current ability estimate to meet a prespecified target value, T. The constraint in (4) forces the attribute values of the $k-1$ previously administered items to be part of the specifications of the shadow test for the kth item. In (5), the length of the CAT is fixed at n items. In (6) and (7), categorical and quantitative constraints are imposed on the shadow test. These constraints have been taken to be equalities here but can easily be changed for inequalities. The constraints in (8) allow the shadow test to have no more than one item with attribute values tending to results in enemies.

Alternative models are possible. For instance, in the empirical example below, no target value for the information in the CAT was available, and the objective function in (2) and the information function in the constraint in (3) were combined into a linear expression optimized in the test assembly model. Other options to deal with multi-objective decision problems are given in Veldkamp (in press).

After the shadow test for the kth item is assembled, the item with maximum information at $\widehat{\theta}_{k-1}$ among the items not yet administered is administered as the kth item.

Calculating the Blueprint

An optimal blueprint is a set of integer values for the cells in table $C \times Q$ that guarantee CATs for a prespecified number of examinees series meeting the target values for the item-exposures rates. The goal of the simulation is to produce counts of the number of times an item is administered, N_{cq}. The number of simulees should be large enough to produce stability among the relative counts.

The blueprint is calculated from the counts N_{cq} according to following formula:

$$I_{cq} = \left\lceil \frac{N_{cq}}{M} * \frac{C}{S} \right\rceil, \qquad (10)$$

where I_{cq} is the number of items in the blueprint in cell (c, q), M is the maximum number of times an item can be exposed before it is supposed to be known, S is the number of simulees in the CAT simulation, and C is the number of CATs the item pool should support.

Application of this formula is justified the following intuitive considerations. If the ability distribution in the CAT simulations is a reasonable approximation to the true ability distribution in the population, N_{cq} predicts the number of items needed in cell (c, q) rather well. Because the numbers are calculated for S simulees and the item pool should support CATs for C examinees, a correction to N_{ij} has to be made multiplying by $\frac{C}{S}$. This correction thus yields the numbers of items with attribute values corresponding to cell (c, q). However, to meet the required exposure rates, these numbers are divided by M.

The final results, rounded upwards to obtain integer values, is the optimal blueprint for the item pool looked for. The question how to realize this blueprint is postponed until the method is demonstrated by the following empirical example.

Empirical Example

As an empirical example an item pool was designed for the CAT version of the GMAT.

Five categorical item attributes were used which are labeled here as $C1, ..., C5$. Each attribute had between two and four possible values. The product of these attributes resulted in a table, C, with 96 cells.

All items were supposed to be calibrated by the 3PL model in (1). The item parameters in this model were the quantitative attributes in this example. The range of values for the discrimination parameter, a_i, is the interval $[0, \infty)$. This interval was split into nine subintervals, the ninth interval extending to infinity. The difficulty parameter, b_i, takes values in the interval $(-\infty, \infty)$. This interval was divided into fourteen subintervals. In a previous item pool, the value of the guessing parameter, c_i, was approximately the same for all items. Therefore, in the simulation, c_i was fixed at this common value. The product of the quantitative attributes resulted in a table, Q, with 124 cells. The Cartesian product of these tables, $C \times Q$, was a table with $96 \times 124 = 12096$ cells.

The integer programming model for the shadow tests in the CAT simulation had 30 constraints to deal with such attributes as test length and content. No constraints on enemy sets were introduced. Because no target for the test information function was available, the following linear combination of test information and item writing costs was optimized:

$$\max\{\lambda \sum_{cq \in C \times Q} I_{cq}\left(\widehat{\theta}_{k-1}\right) x_{cq} - (1-\lambda) \sum_{cq \in C \times Q} k_{cq} x_{cq}\} \qquad \text{(objective function)} \quad (11)$$

As estimates of item-writing costs, reciprocals of the frequencies of the items on a previous item pool for the GMAT were used, with large numbers substituted for cells with zero frequencies. Each new item administered was selected to have maximum information at the current ability estimate.

The simulees were sampled from $N(1,1)$. The initial estimate for each new simulee was set equal to $\widehat{\theta} = 0$. The CATs were simulated using software for constrained CAT with shadow tests developed at the University of Twente. The software used integer programming routines available in the linear-programming software package CPLEX (ILOG, 1998) to assemble the shadow tests. The blueprint was calculated using realistic estimates for C and M in (10). The final result was a table of item frequencies which, for security, is not revealed here.

Discussion

The method presented in this paper produces an optimal blueprint for an item pool. This blueprint serves as the best goal available to guide the item writing process. It can be used to prepare instructions for the item writers and, if the item writers were used as an attribute in the $C \times Q$ table for which empirical cost estimates were obtained, to assign these instructions to them. The best way to implement the blueprint is in a sequential fashion recalculating the blueprint after a certain portion of the items has actually been written and field tested so that their attribute values are known. Also, though an exactly realized blueprint would guarantee the exposure rates imposed on it, actual pools need estimates of additional exposure control parameters to allow for their differences from the blueprint after which they were written. In fact, the best way to view these optimal blueprints is not as a one-shot item pool design but as tools for continuous item pool management (van der Linden, Veldkamp & Reese, in press).

As already noted, the method in this paper can be adapted to design systems of rotating item pools. In this adaptation, the adjustment in (10) is replaced by an assignment problem that can be modeled as an integer programming problem again. Further details on this adaptation are provided in Veldkamp and van der Linden (in press).

References

Boekkooi-Timminga, E. (1991). *A method for designing Rasch Model-based Item Banks*. Paper presented at the annual meeting of the Psychometric Society 1991, Pinceton.

Davey, T. ,and Nering, M. (1998). *Controlling Item Exposure adn Maintaining Item Security.* Paper presented at the colloquium, Computer-Based Testing: Building the Foundation ofor Future Assessments, Philadelphia, PA, September 25-26, 1998.

Flaugher, R. (1990). Item Pools. In H. Wainer, *Computerized Adaptive Testing: A Primer (pp.41-64).* Hillsdale, NJ, Lawrence Erlbaum Associates.

CPLEX Division of ILOG, Inc. (1998). CPLEX [Computer Program and Manual]. Incline Village, NV: ILOG.

Stocking, M. and Lewis, C. (1998). Controlling Item Exposure Conditional on Ability in Computerized Adaptive Testing, *Journal of Educational and Behavioral Statistics*, 23, pp. 57-75.

Stocking, M. and Swanson, L. (1998). Optimal Design of Item Banks for Computerized Adaptive Tests, *Applied Psychological Measurement, 22*, pp. 271-280.

Swanson, L., and Stocking, M. L. (1993). A Model and Heuristic for Solving Very Large Item Selection Problems, *Applied Psychological Measurement, 17*, pp. 151-166.

Sympson, J.B., and Hetter, R.D. (1985). Controlling item-exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Military Testing Association* (pp. 973-977). San Diego, CA: Navy Personnel Research and Development Center.

van der Linden, W.J. (1998). Optimal Assembly of Psychological and Educational Tests, *Applied Psychological Measurement, 22*, pp. 195-211.

van der Linden, and Reese, L.M. (1998). A Model for Optimal Constrained Adaptive Testing, *Applied Psychological Measurement, 22*, pp. 259-270.

van der Linden, W.J. (in press). Constrained Adaptive Testing with Shadow Tests. In W.J. van der Linden and C.A.W. Glas (Eds.) *Computerized Adaptive Testing: Theory and Practice*. Bosten, MA: Kluwer Academic Publishers.

van der Linden, W.J., Veldkamp, B.P., and Reese, L.M. (in press). An Integer Programming Approach to Item Pool Design. *Applied Psychological Measurement.*

Veldkamp, B.P. (in press). Multiple Objective Test Assembly Problems. *Journal of Educational Measurement.*

Veldkamp, B.P., and van der Linden, W.J. (in press). Item Pool Desing for CAT. In W.J. van der Linden and C.A.W. Glas (Eds.) *Computerized Adaptive Testing: Theory and Practice.* Bosten, MA: Kluwer Academic Publishers.

Way, W.D. (1998). Protecting the Integrity of Computerized Testist Item Pools, *Educational Measurement: Issues and Practice, 17*, pp. 17-26.

Way, W.D., Steffen, M., and Anderson, G.S. (1998). *Developing, Maintaining, and Renewing the Item Inventory to Support Computer-Based Testing.* Paper presented at the colloquium, Computer-Based Testing: Building the Foundation ofor Future Assessments, Philadelphia, PA, September 25-26, 1998.

Send requests for information to: Bernard P. Veldkamp, Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. E-mail: veldkamp@edte.utwente.nl

Titles of Recent Research Reports from the Department of Educational Measurement and Data Analysis. University of Twente, Enschede, The Netherlands.

RR-99-03	B.P. Veldkamp & W.J. van der Linden, *Designing Item Pools for Computerized Adaptive Testing*
RR-99-02	W.J. van der Linden, *Adaptive Testing with Equated Number-Correct Scoring*
RR-99-01	R.R. Meijer & K. Sijtsma, *A Review of Methods for Evaluating the Fit of Item Score Patterns on a Test*
RR-98-16	J.P. Fox & C.A.W. Glas, *Multi-level IRT with Measurement Error in the Predictor Variables*
RR-98-15	C.A.W. Glas & H.J. Vos, *Adaptive Mastery Testing Using the Rasch Model and Bayesian Sequential Decision Theory*
RR-98-14	A.A. Béguin & C.A.W. Glas, *MCMC Estimation of Multidimensional IRT Models*
RR-98-13	E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *Person Fit based on Statistical Process Control in an AdaptiveTesting Environment*
RR-98-12	W.J. van der Linden, *Optimal Assembly of Tests with Item Sets*
RR-98-11	W.J. van der Linden, B.P. Veldkamp & L.M. Reese, *An Integer Programming Approach to Item Pool Design*
RR-98-10	W.J. van der Linden, *A Discussion of Some Methodological Issues in International Assessments*
RR-98-09	B.P. Veldkamp, *Multiple Objective Test Assembly Problems*
RR-98-08	B.P. Veldkamp, *Multidimensional Test Assembly Based on Lagrangian Relaxation Techniques*
RR-98-07	W.J. van der Linden & C.A.W. Glas, *Capitalization on Item Calibration Error in Adaptive Testing*
RR-98-06	W.J. van der Linden, D.J. Scrams & D.L.Schnipke, *Using Response-Time Constraints in Item Selection to Control for Differential Speededness in Computerized Adaptive Testing*
RR-98-05	W.J. van der Linden, *Optimal Assembly of Educational and Psychological Tests, with a Bibliography*
RR-98-04	C.A.W. Glas, *Modification Indices for the 2-PL and the Nominal Response Model*
RR-98-03	C.A.W. Glas, *Quality Control of On-line Calibration in Computerized Assessment*

RR-98-02 R.R. Meijer & E.M.L.A. van Krimpen-Stoop, *Simulating the Null Distribution of Person-Fit Statistics for Conventional and Adaptive Tests*

RR-98-01 C.A.W. Glas, R.R. Meijer, E.M.L.A. van Krimpen-Stoop, *Statistical Tests for Person Misfit in Computerized Adaptive Testing*

RR-97-07 H.J. Vos, *A Minimax Sequential Procedure in the Context of Computerized Adaptive Mastery Testing*

RR-97-06 H.J. Vos, *Applications of Bayesian Decision Theory to Sequential Mastery Testing*

RR-97-05 W.J. van der Linden & Richard M. Luecht, *Observed-Score Equating as a Test Assembly Problem*

RR-97-04 W.J. van der Linden & J.J. Adema, *Simultaneous Assembly of Multiple Test Forms*

RR-97-03 W.J. van der Linden, *Multidimensional Adaptive Testing with a Minimum Error-Variance Criterion*

RR-97-02 W.J. van der Linden, *A Procedure for Empirical Initialization of Adaptive Testing Algorithms*

RR-97-01 W.J. van der Linden & Lynda M. Reese, *A Model for Optimal Constrained Adaptive Testing*

RR-96-04 C.A.W. Glas & A.A. Béguin, *Appropriateness of IRT Observed Score Equating*

RR-96-03 C.A.W. Glas, *Testing the Generalized Partial Credit Model*

RR-96-02 C.A.W. Glas, *Detection of Differential Item Functioning using Lagrange Multiplier Tests*

RR-96-01 W.J. van der Linden, *Bayesian Item Selection Criteria for Adaptive Testing*

RR-95-03 W.J. van der Linden, *Assembling Tests for the Measurement of Multiple Abilities*

RR-95-02 W.J. van der Linden, *Stochastic Order in Dichotomous Item Response Models for Fixed Tests, Adaptive Tests, or Multiple Abilities*

RR-95-01 W.J. van der Linden, *Some decision theory for course placement*

RR-94-17 H.J. Vos, *A compensatory model for simultaneously setting cutting scores for selection-placement-mastery decisions*

RR-94-16 H.J. Vos, *Applications of Bayesian decision theory to intelligent tutoring systems*

RR-94-15 H.J. Vos, *An intelligent tutoring system for classifying students into Instructional treatments with mastery scores*

...

Research Reports can be obtained at costs, Faculty of Educational Science and Technology, University of Twente, Mr. J.M.J. Nelissen, P.O. Box 217, 7500 AE Enschede, The Netherlands.

faculty of

EDUCATIONAL SCIENCE AND TECHNOLOGY

A publication by
The Faculty of Educational Science and Technology of the University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

NOTICE

REPRODUCTION BASIS

☒ This document is covered by a signed "Reproduction Release (Blanket) form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").