ABSTRACT

        A statistical test for the detection of answer copying on
multiple-choice tests is presented. The test is based on the idea that
answers of examinees to test items may be the results of three possible
processes: (1) knowing; (2) guessing; and (3) copying. Examinees who do not
have access to the answers of other examinees can arrive at their answers
only through the first two processes. This assumption leads to a distribution
for the number of matched incorrect alternatives between the examinee
suspected of copying and the examinee believed to be the source that belongs
to a family of "shifted binomials." Power functions for the tests for several
sets of parameter values are analyzed. It is shown that an extension of the
test to include matched numbers of correct alternatives would lead to
improper statistical hypotheses. (Author/SLD)

ED 473 531

A Statistical Test for Detecting
Answer Copying on
Multiple-Choice Tests

Leonardo S. Sotaridona
Wim J. van der Linden

TM034739

faculty of
EDUCATIONAL SCIENCE
AND TECHNOLOGY
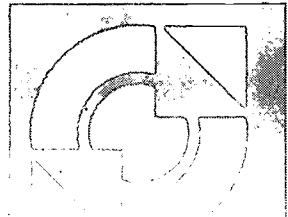
University of Twente

Department of
Educational Measurement and Data Analysis

ERIC

A Statistical Test for Detecting Answer
Copying on Multiple-Choice Tests

Wim J. van der Linden

Leonardo S. Sotaridona

Abstract

A statistical tests for the detection of answer copying on multiple-choice tests is presented. The test is based on the idea that the answers of examinees to test items may be the result of three possible processes: (1) knowing, (2) guessing and (3) copying, but that examinees who do not have access to the answers of other examinees can arrive at their answers only through the first two processes. This assumption leads to a distribution for the number of matched incorrect alternatives between the examinee suspected of copying and the examinee believed to be the source that belongs to a family of "shifted binomials". Power functions for the tests for several sets of parameter values are analyzed. It is shown that an extension of the test to include matched numbers of correct alternatives would lead to improper statistical hypotheses.

Key words: Answer Copying; Cheating; Hypothesis Testing; Multiple-Choice Testing; Shifted Binomial.

A Statistical Test for Detecting Answer Copying

on Multiple-Choice Tests

One of the first to derive a statistical test to detect answer copying on multiple-choice tests were Frary, Tideman, and Watts (1977). Their g_2 index is an attempt to evaluate the number of matching alternatives between an examinee suspected to be a copier and another examinee believed to be the source against the expected number of matching alternatives. (For convenience, we will refer to these examinees just as "copier" and "source".) The problem inherent in a test of this nature is how to obtain the distribution of the index under the null hypothesis of no copying. Frary et al. attempted to solve this problem by establishing a null model that assumes that the probability of selecting an alternative on an item is a certain function of the copier's number-correct score, the average number-correct score in the population, and the proportion of examinees in the population that selected the same alternative. Note that the first two quantities correct the probability of selecting an alternative for the examinee's relative ability in the population.

The K-index (Holland, 1996; Lewis & Thayer, 1998) is the result of an attempt to correct more explicitly for the examinee's ability. The index focuses only on the number of matching alternatives on the items that were answered incorrectly by the source. The null model is a binomial with a success parameter that is obtained by piecewise linear regression of the proportion of matching incorrect alternatives on the proportion of incorrect alternatives in each number-incorrect score group in a population of examinees. An alternative with quadratic regression is given in Sotaridona and Meijer (2002).

The most elaborate null model for a test to detect copying is the one on which Wollack's ω index is based (Wollack, 1997; Wollack & Cohen, 1998). The probability of selecting an alternative on an item for an examinee that does not copy is assumed to follow the nominal response model (Bock, 1997). The ω index has the same shape as the g_2 index but compares the observed number of matching alternatives against the (estimated) expected number under the nominal response model for all items in the test. Note that the use of the nominal response model automatically involves conditioning of the probabilities of choosing an alternative on the examinee's ability.

In spite of the attempts to condition on the examinee's ability, a fundamental feature of all three tests is their dependency on the distribution of the item scores in the population of examinees. If the population changes, the results from these tests also change: the g_2 index has to be calculated from a different proportion of times an alternative is chosen and a different average number-correct score in the population, the K index has to be based on a different regression equation, and the nominal response model has to be refitted with possible different parameter estimates or a less satisfactory fit. The fact that these three statistics are population dependent thus implies that the same pair of examinees may be tested to have been involved in answer copying in one population but not in another.

The purpose of this paper is to present a statistical test to detect answer copying on multiple-choice tests that can be used when any reference to a population of examinees is undesirable. Obviously, we need a set of assumptions to derive a statistical test, but in the current case the assumptions are only about the response behavior of the individual examinee suspected of copying. In essence, the assumptions are based on the idea that an examinee who has access to the answers of a source can arrive at his/her own answer through three different processes: (1) knowing, (2) guessing or (3) copying. Examinees who do not have access to a source can only produce answers through the first two processes. No other assumptions are made. In particular, nothing is assumed about or inferred from the distribution of item scores in a population of examinees. Also, no assumption is made about the behavior of the examinee who may have served as a source to the copier.

Derivation of the Test

Like the K-index, the test focuses on the items for which the source has an incorrect answer. We will motivate this choice later by showing that an extension of the test to include items with correct answers by the source will lead to improper statistical hypotheses.

Assumptions

The assumptions on the behavior of the copier on the items the source has answered incorrectly from which the test is derived are the following: First, if an examinee knows an item, he/she gives a correct answer. This assumption implies that if an examinee has access to the source but discovers his/her answer is incorrect, he/she does not copy but gives his/her own answer. Second, if an examinee does not know an item but has access to the source, he/she accepts the answer by the source and copies. Third, if an examinee does not know an item and does not have access to a source, he/she guesses blindly among the response alternatives. Thus, for each item incorrect by the source, we have three possible true states in which the copier can be, each characterized by a different probability of choosing the same alternative the source has chosen.

We use the following notation to present these probabilities. Let $i = 1, ..., I$ denote the items in the test and $a = 1, ..., k$ the response alternatives for these items. In addition, index s and j are used for the source and the copier, respectively. The alternatives chosen by these two examinees on item i are denoted by random variables U_{si} and U_{ji}. The set of items for which s has chosen an incorrect alternative is denoted as W_s. The size of this is denoted as w_s. Finally, a (random) indicator variable I_{sji} is used to identify the items for which examinees s and j have chosen the same alternative. That is,

$$I_{sji} = \begin{cases} 1 & \text{if} \quad U_{si} = U_{sj} \\ 0 & \text{if} \quad U_{si} \neq U_{sj}. \end{cases}$$

The three possible probabilities for examinee j to choose the same alternative on the items in W_s as s are the following:

$$\Pr(I_{sji} = 1) = \begin{cases} 0 & \text{if} \quad j \text{ knows the answer on } i \in W_s, \\ k^{-1} & \text{if} \quad j \text{ guesses blindly on } i \in W_s, \\ 1 & \text{if} \quad j \text{ copies from } s \text{ on } i \in W. \end{cases} \quad (1)$$

Hypotheses

The hypothesis to be tested is that j did not copy any of the items in W_s. We suggest to test this hypothesis against the alternative that j copied the answers for some of the items in W_s which he/she did not know. Observe that this alternative is less extreme than the hypothesis of j copying all items in W_s. Under the current alternative hypothesis, it is still possible that j actually knows some of the items in W_s and for this reason did not copy them or that he/she did not have access to the answers by s for all of the items in W_s.

Let κ_j be the number of items in the set W_s examinee j knows and γ_j the number in this set examinee j copied from s. More formally, at the level of the set of items W_s, the hypothesis to be tested is:

$$H_0 : \Pr(I_{sji} = 1) = \begin{cases} 0 & \text{for } \kappa_j \text{ items in } W_s, \\ k^{-1} & \text{for } \omega_s - \kappa_j \text{ items in } W_s, \end{cases} \tag{2}$$

against

$$H_1 : \Pr(I_{sji} = 1) = \begin{cases} 0 & \text{for } \kappa_j \text{ items in } W_s, \\ k^{-1} & \text{for } \omega_s - \kappa_j - \gamma_j \text{ items in } W_s, \\ 1 & \text{for } \gamma_j \text{ items in } W_s, \end{cases} \tag{3}$$

with $\kappa_j \geq 0$, $\gamma_j > 0$, and $\kappa_j + \gamma_j \leq w_s$.

Distribution of Matching Incorrect Alternatives

The proposed test statistic is the number of matching incorrect alternatives between j and s on the items in set W_s:

$$Z_{js} = \sum_{i \in W_s} I_{sji}. \tag{4}$$

Both hypotheses imply distributions of Z_{js} belonging to a family with probability function

$$p(z; w_s, \gamma_j, \kappa_j, k) = \begin{cases} 0 & \text{for } z < \gamma_j, \\ \binom{w_s - \kappa_j - \gamma_j}{z - \gamma_j} k^{-(z-\gamma_j)}((k-1)k^{-1})^{w_s - \kappa_j - z} & \text{for } \gamma_j \leq z \leq w_s - \kappa_j, \\ 0 & \text{for } w_s - \kappa_j < z, \end{cases} \tag{5}$$

with $\kappa_j \geq 0$, $\gamma_j \geq 0$, and $\kappa_j + \gamma_j \leq w_s$. The definition of this family follows from the fact that if j copies γ_j answers from W_s, the probabilities of observing numbers of matches smaller than γ_j are each equal to zero. Likewise, if j knows κ_j items in W_s, the probabilities of observing number of matches larger than $w_s - \kappa_j$ are each equal to zero. However, for the subset of $w_s - \kappa_j - \gamma_j$ items that j does not know and for which (s)he has not copied any answer, the number of matches follows a binomial distribution with success parameter k^{-1}. Observe that the probability of $Z_{js} = \gamma_j$ belongs to the compound event of j copying γ_j items and guessing none of the alternatives the source has chosen. Likewise, the probability of $Z_{js} = w_s - \kappa_j$ belongs to the compound event of j coping γ_j items, knowing κ_j items, and guessing the alternatives the source has chosen on all remaining items.

The function in (5) can be presented more compactly as

$$p(z; w_s, \gamma_j, \kappa_j, k) = \binom{w_s - \kappa_j - \gamma_j}{z - \gamma_j} k^{-(z-\gamma_j)} ((k-1)k^{-1})^{w_s - \kappa_j - z} I_{\{\gamma_j, w_s - \kappa_j\}}(z), \quad (6)$$

where $I_{\{\gamma_j, w_s - \kappa_j\}}(z)$ is an indicator function that is equal to 1 if $z \in \{\gamma_j, w_s - \kappa_j\}$ and equal to 0 otherwise. Because $p(z; w_s, \gamma_j, \kappa_j, k)$ is nonzero for $z \in \{\gamma_j, w_s - \kappa_j\}$, this function indicates the support of the family of distributions in (6). In spite of the presence of the binomial expression in the definition of (6), the family is *not* the binomial over the range of possible values of Z_{js}. We will refer to this family as the "shifted binomial", because it can be viewed as a binomial with its support shifted from $\{0, w_s\}$ to $\{\gamma_j, w_s - \kappa_j\}$. The size of the shift is a critical quantity because it depends both on the (unknown) number of items j knows as well as the number j has copied.

Statistical Test

Under the distribution in (6), the two hypotheses in (2)-(3) simplify to

$$H_0 : \gamma_j = 0 \quad (7)$$

and

$$H_1 : \gamma_j > 0. \tag{8}$$

However, the null distribution under which the (right-sided) test of the hypothesis in (7) has to be conducted still depends on the unknown parameter κ_j. We propose to conduct the test under the auxiliary assumption that j did not know any of the items in W_s, that is, $\kappa_j = 0$. This assumption gives us a test that tends to be more conservative than the one actually needed: From (6) it follows that the upper tail of the distribution for $\kappa_j = 0$ is further to the right than the upper tails of the distributions for $\kappa_j > 0$. As a result, setting $\kappa_j = 0$ results in a critical value for the test larger than the one needed for the (unknown) true value of κ_j at the nominal level of significance.

We feel the auxiliary assumption is permitted because it does thus not harm the copier in any way. The one who may have to pay a price for this assumption is the testing agent because of a loss of power of the test to detect answer copying. We will quantify the extent to which the critical value of the test is larger than actually needed as well as the differences in power resulting from this increase later in this paper.

A (nonrandomized) test of the hypothesis of j not having copied any answer against the alternative of j having copied the answers of some of the items in W_s with nominal significance level not larger than α has as critical value for the test statistic Z_{js} in (4) the smallest value of z^* for which the distribution in (6) yields

$$\Pr(Z_{js} \geq z^*) \leq \alpha. \tag{9}$$

Uniform Most Powerful Test

For a statistical test it is desirable to be uniformly most powerful (UMP) at the level of significance chosen. From the Karlin-Rubin theorem (e.g., Casella & Berger, 1990, sect. 8.3.2) it follows that the above test is a UMP test with level associated with the critical value in (9) provided the family in (6) has a monotone likelihood ratio in Z_{js} and

Z_{js} is sufficient statistic for the number of items copied, γ_j. It is easy to show that (6) has both properties for the case $\kappa_j = 0$.

As for the property of a monotone likelihood ratio, for the test of (7) against (8) it is sufficient to show that the ratio of (6) for $\gamma_j = \gamma > 0$ and $\gamma_j = 0$ is nondecreasing in z. Simplifying, omitting constants, and cancelling factors, the ratio can be show to be equal to

$$\frac{z!}{(z - \gamma)!},$$

which is increasing in z.

The fact that Z_{js} is a sufficient statistic for γ_j follows from the well-known factorization criterion. The factor

$$((k - 1)k^{-1})^{w_s - \kappa_j - z}$$

in (6) is independent of γ_j whereas its remaining part is dependent on γ_j and z.

It is instructive to compare this result with those for a test of a point hypothesis for the success parameter in a regular binomial family, which also is UMP. In the current case, (6) is not the regular binomial and the parameter of interest is not a success parameter but a parameter that defines both the support of the distribution and the number of Bernoulli trials on which it is based. Observe also that (6) is not UMP with nominal level α but with the actual level of significance associated with (9). An exact level α test is only possible for a randomization version of (9).

Finally, it is emphasized that the above result holds for the test in (9) which is based on the assumption $\kappa_j = 0$, but that it has not been shown that the test of (7) is UMP for an unknown value of k_j. The impact of this parameter on the power of the test will be evaluated empirically in the next section.

Power of the Test

The actual power of the above test is a function of the unknown number of items j has copied from s, γ_j. The shape of the power function depends on (1) the number of alternatives per item, k, (2) the number of items s has incorrect, w_s, and (3) the significance level chosen for the test, α, and (4) the number of items the examinee knows, κ_j.

We first present a set of power functions for the case $\kappa_j = 0$ for $k = 2, ..., 5$, $w_s = 20, 30, 40, 50$, and significance level $\alpha = .05$. The functions were calculated by first finding the critical value z^* for $\alpha = .05$ in (9) under the distribution given by (6) with $\gamma_j = 0$ and then calculating the probabilities $\Pr\{Z_{js} \geq z^*\}$ under the same distribution for $\gamma_j = 0, ..., w_s$. The power functions are presented in Figure 1. From these functions it is clear that the test has considerable power to detect copying on multiple-choice tests, particularly if the number of response alternatives per item, k, goes up. But even for a test with three-choice items the power to detect copying is already perfect if the examinee has copied approximately half of the items in W_s for $w_s = 20$ or one third for $w_s = 50$.

[Insert Figure 1 about here]

It was noted earlier that the auxiliary assumption of $\kappa_j = 0$ leads to a test that tends to be conservative. Figure 2 illustrates the effect of the assumption of $\kappa_j = 0$ on the critical value of the test for the same sets of parameter values as in Figure 1 ($k = 2, ..., 5$; $w_s = 20, 30, 40, 50$; $\alpha = .05$). The curves show the critical values as a function of the number of items j knows, κ_j. For example, the lower-left plot shows that for a test with four-choice items ($k = 4$), the assumption $\kappa_j = 0$ leads to a critical value for the test equal to $z^* = 17$. However, if the examinee actually knows $k_j = 10$ of the 40 items in the set W_s, the critical value could have been lowered to $z^* = 14$ to realize the nominal significance level of $\alpha = .05$. Except for small horizontal pieces, which are due to the discreteness of test statistic Z_{js}, all curves increase with κ_j. This feature reflects the fact that the proposed statistical test is generally conservative, unless the assumption $\kappa_j = 0$ happens to be true.

[Insert Figure 2 about here]

It was also noted that the price for using a conservative test is not paid by the examinee but by the testing agency in the form of less than optimal power. Figures 3-6 show how much larger the power of the test would have been if we had known the true value of κ_j. The curves in this figure show the increase in power relative to the power functions in Figure 1. That is, the increase in power was calculated as the difference between the power of the test for the true value of κ_j and $\kappa_j = 0$ divided by the power for $\kappa_j = 0$, and the result was plotted as a function of γ_j. Figures 3-6 show these functions for the same sets of parameter values as in the previous two figures ($k = 2, ..., 5$; $w_s = 20, 30, 40, 50$; $\alpha = .05$) and a selected set of values κ_j. Each panel in these figures shows approximately the same pattern, which can be summarized as follows. First, knowing the true value of κ_j would only lead to an increase of power for small values of γ_j. Second, the increase would be larger, the smaller the number of alternatives per item, k. These two findings are consistent with the results in Figure 1 which show that the power curves are nearly equal to 1.0 for larger values of γ_j but approach this state at a somewhat lower rate for items with fewer alternatives. Once the power is close to one, there is hardly any space for improvement left. Third, the increase in power is generally larger for large values of κ_j, with an exception at the smallest values of γ_j, where for some of the larger values of κ_j the assumption $\kappa_j = 0$ actually appears to result in a small increase in power. These exceptions are due to discrete nature of the null distribution of the test and the definition of the critical value in (9). For larger values of κ_j the actual level of Type 1 error can become smaller than α, and hence, for these values of κ_j, the power of the test can become smaller than for $\kappa_j = 0$. A randomized version of the test would not suffer from this problem, but the use of randomization in a statistical test to detect cheating on multiple-choice tests does not seem feasible. Fourth, for smaller values of κ_j the power of the test appears to be remarkably robust and the assumption of $\kappa_j = 0$ does not involve much loss of power over the entire range of values of γ_j.

[Insert Figures 3-6 about here]

The proper way to use Figures 1-6 in an application is (1) to identify the size of the set of items the source has wrong, w_s, (2) to inspect the power function for the number of alternatives per item, k, in the panel for this value of w_s in Figure 1, (3) to find out in Figure 2 how much too large the critical value is for the various possible numbers of items the examinee knows, κ_j, and (4) to use the plot for the actual value of k in Figure 3-6 to determine how much larger the power would have been if we had known κ_j.

Discussion

The question can be raised if the test could not be improved by having the statistic in (4) also include the items on which the source chooses the correct alternative. Just like the g_2 and w index, this option would allow the statistical test to derives its power from all items in the tests rather than only those that s happens to answer correctly.

However, let R_s be the subset of items the source has correct. Analogous to (1), for the items in this set a copier can be in three possible true states, with the probabilities of a matching correct alternative given by:

$$\Pr(Iji = 1) = \begin{cases} 1 & \text{if} \quad j \text{ knows the answer on } i \in R_s \\ k^{-1} & \text{if} \quad j \text{ guesses blindly on } i \in R_s \\ 1 & \text{if} \quad j \text{ copies from } s \text{ on } i \in R_s. \end{cases} \quad (10)$$

Unfortunately, the probabilities for the events of j knowing and copying the answer are equal. In the current framework, it is thus impossible to extend the test with the items in R_s, because the result would be a test confounding the difference between the events of j copying the answers s has correct and j knowing them.

The only possibility to further improve the power of the proposed test seems to get more information about the number of items the examinee actually knows, κ_j. The preceding analysis shows that this information can not come from the responses of the

examinee to the items in set R_s. However, it can come from another source. For example, in a setting where an examinee retakes a test and shows an unusual increase in test scores, it may be possible to infer a lower bound to κ_j from the first test. Figure 2 and 3 show that, particularly for items with few alternatives or sources that have only a small number of items incorrect, conducting the test not at $\kappa_j=0$ but at a lower bound to κ_j deliberatively chosen to be conservative is likely to result in an increase in power that should not be disregarded.

The third assumption on which the proposed statistical test rests is known as the "model of knowing or blind guessing" in test theory. This assumption underlies the 3-parameter logistic model in item response theory (Birnbaum, 1968) and has been used to derive the correction for guessing on multiple-choice items widely known as "formula scoring". In spite of its popularity, the assumption has been criticized because it ignores the fact that examinees may have partial knowledge. For example, they may be able to recognize some of the incorrect alternatives as wrong and guess blindly among the remaining alternatives or they may have information that helps them to guess the correct alternative with a probability larger than k^{-1}.

For an individual examinee responding to an individual item, it may be hard to identify what actual process occurs if the examinee does not know the item. We doubt if it will ever be possible to open this black box and formulate a statistical model with satisfactory validity. However, for the family of distributions in (6) it is possible to evaluate the effect of partial knowledge of some of the items in W_s on the proposed statistical test. Figure 2 shows for all parameter values that the critical value of the test never decreases but nearly always increases if (1) the number of alternatives per item decreases and/or (2) the number of items an examinees knows increases. If the examinee is thus able to exclude some of the alternatives as incorrect, the effect is a decrease in the effective number of alternatives for some of the items. Likewise, if (s)he has knowledge that leads to an increase of the probability of success on some of the items, the effect is an increase in the expected number of items the examinee has correct relative to an examinee who guesses blindly. This effect can be viewed as an increase in the number of items the

examinee knows. For both types of partial knowledge, the actual critical value for the test is higher than required, and again it is the testing agency and not the examinee who incurs the loss due to ignoring the possible presence of partial knowledge.

References

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

Bock, R. D. (1997). The nominal categories model. In W. J. van der Linden and R. K. Hambleton (Eds.), Handbook of modern item response theory (pp. 33-49). New York: Springer-Verlag.

Casella, G., Berger, R. L. (1990). *Statistical inference.* Pacific Grove, CA: Brooks/Cole.

Frary, R. B., Tideman, T. N., & Watts, T. M. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics, 2,* 235-256.

Holland, P. W. (1996). *Assessing unusual agreement between incorrect answers of two examinees using the K-index: Statistical theory and empirical support* (Technical Report No. 96-4). Princeton, NJ: Educational Testing Service.

Lewis, C., & Thayer, D. T. (1998). *The power of the K-index (or PMIR) to detect copying* (Research Report RR-98-49). Princeton, NJ: Educational Testing Service.

Sotaridona, L. S., & Meijer, R. R. (2002) Statistical properties of the K-index for detecting answer copying. *Journal of Educational Measurement, 39,* 115-132.

Wollack, J. A. (1997). A nominal response model approach to detect answer copying. *Applied Psychological Measurement, 21,* 307-320.

Wollack, J. A., & Cohen, A. S. (1998). Detection of answer copying with unknown item and trait parameters. *Applied Psychological Measurement, 22,* 144-152.

Figure Captions

Figure 1. Power functions for $k = 2, ..., 5$ and $w_s = 20, 30, 40, 50$ at significance level $\alpha = .05$.

Figure 2. Critical values as a function of κ_j for $k = 2, ..., 5$ and $w_s = 20, 30, 40, 50$ at significance level $\alpha = .05$.

Figure 3. Relative loss of power due to the number of items known by the examinee, κ_j, for $w_s = 20, 30, 40, 50$ at significance level $\alpha = .05$ (and $k = 2$).
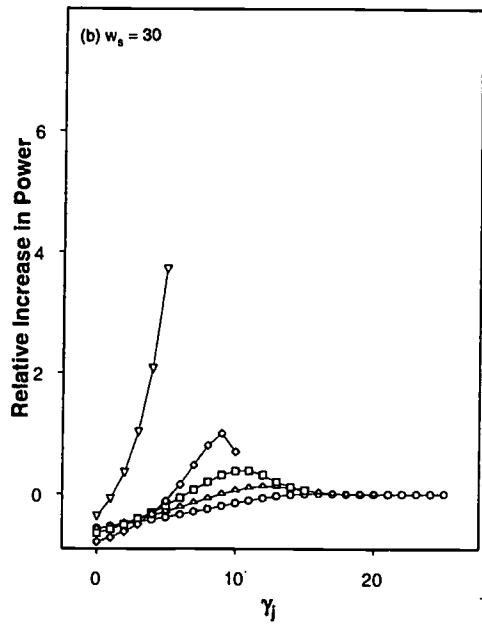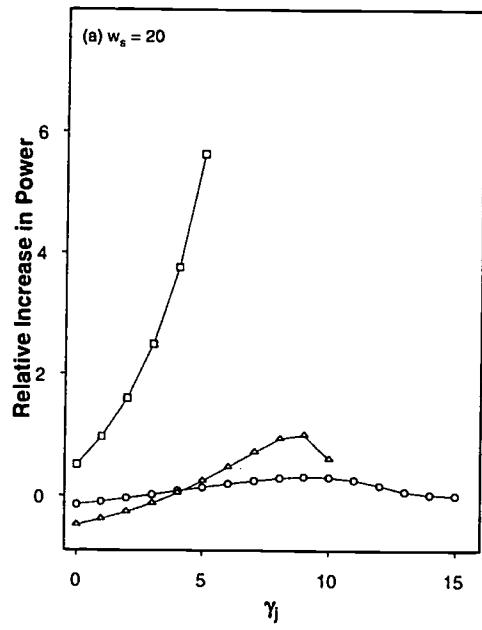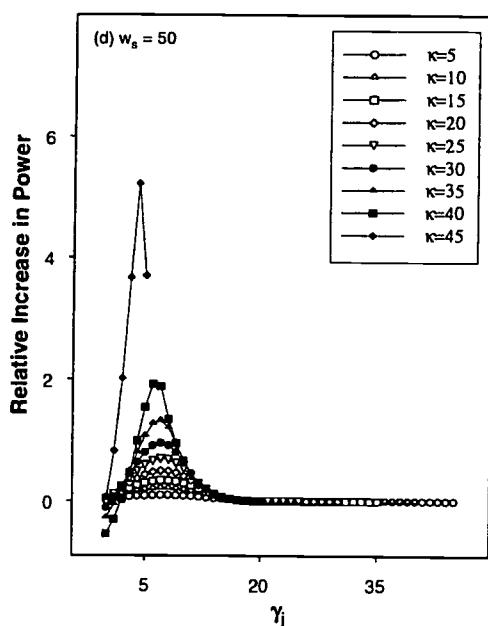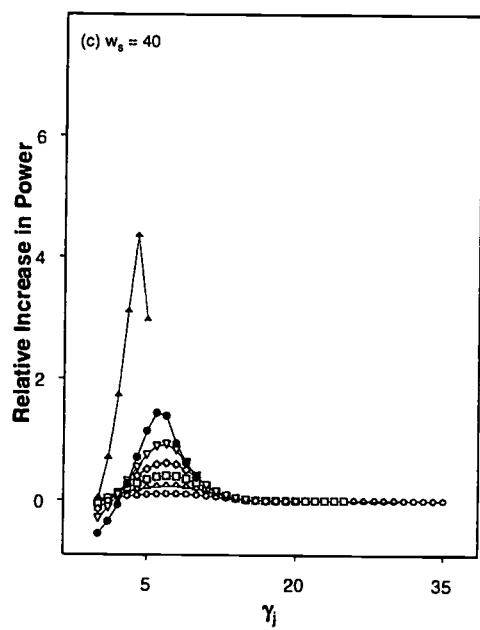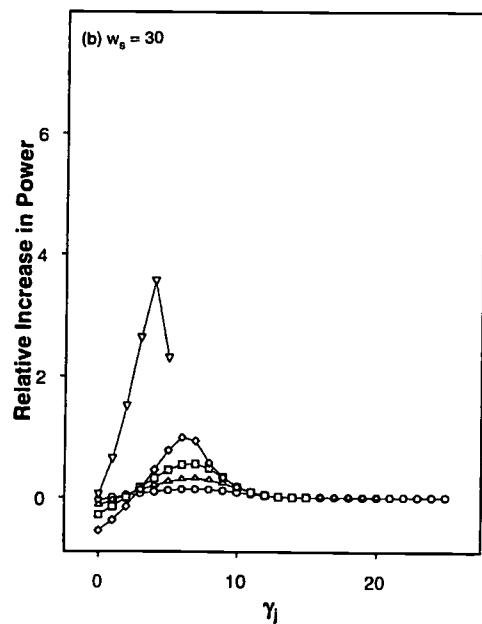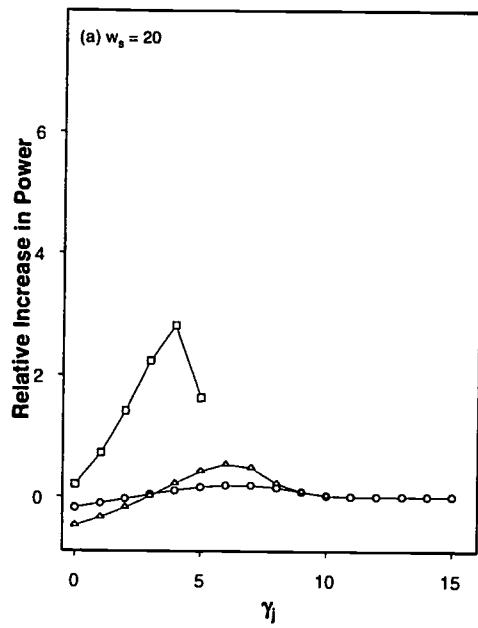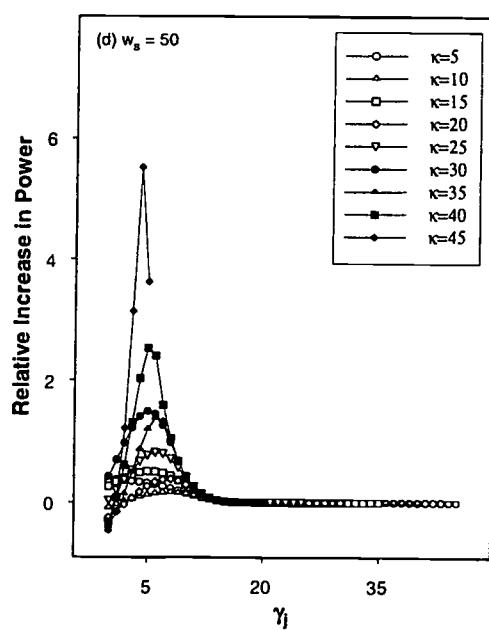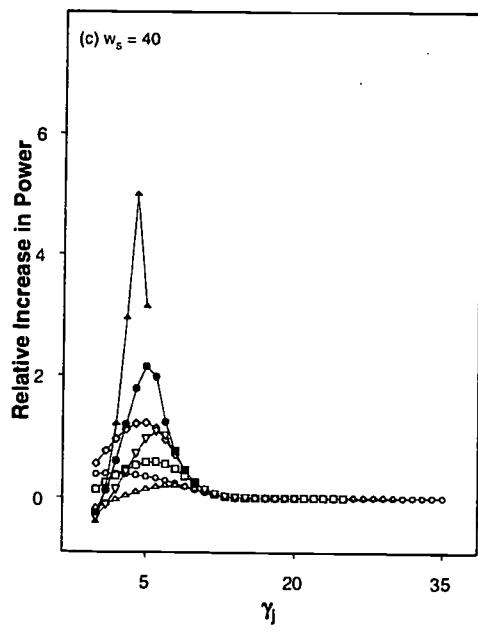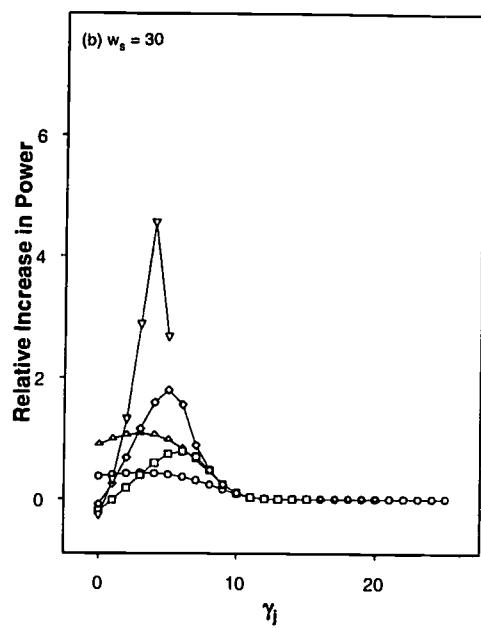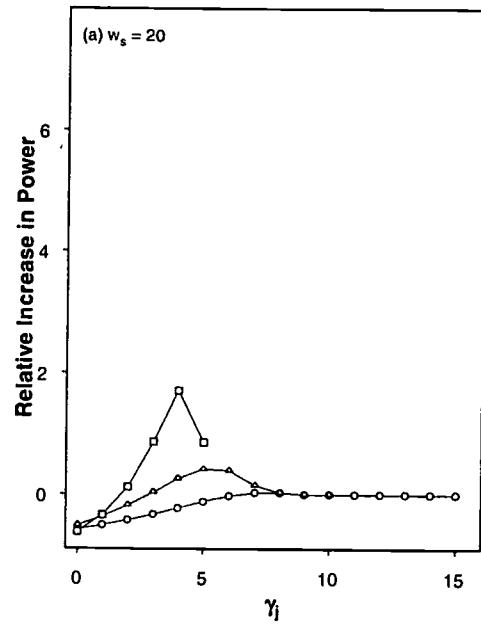
Figure 4. Relative loss of power due to the number of items known by the examinee, κ_j, for $w_s = 20, 30, 40, 50$ at significance level $\alpha = .05$ (and $k = 3$).

Figure 5. Relative loss of power due to the number of items known by the examinee, κ_j, for $w_s = 20, 30, 40, 50$ at significance level $\alpha = .05$ (and $k = 4$).

Figure 6. Relative loss of power due to the number of items known by the examinee, κ_j, for $w_s = 20, 30, 40, 50$ at significance level $\alpha = .05$ (and $k = 5$).

(a) $w_s = 20$

(b) $w_s = 30$

(c) $w_s = 40$

(d) $w_s = 50$

Relative Increase in Power

γ_j

$\kappa=5$
$\kappa=10$
$\kappa=15$
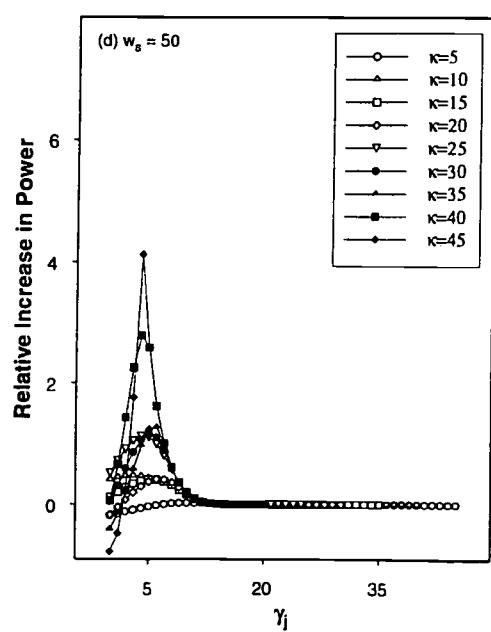$\kappa=20$
$\kappa=25$
$\kappa=30$
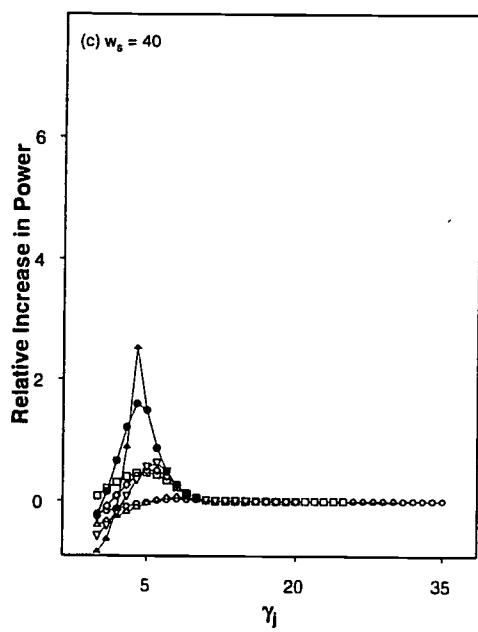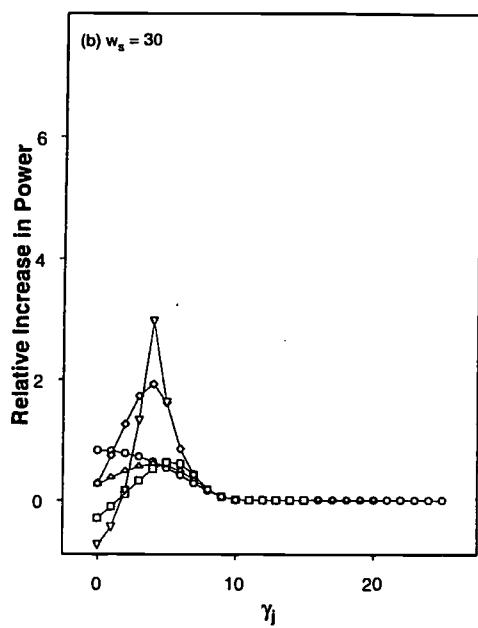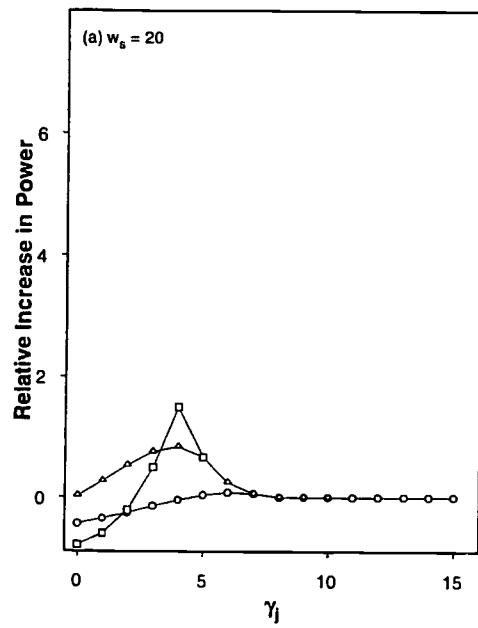$\kappa=35$
$\kappa=40$
$\kappa=45$
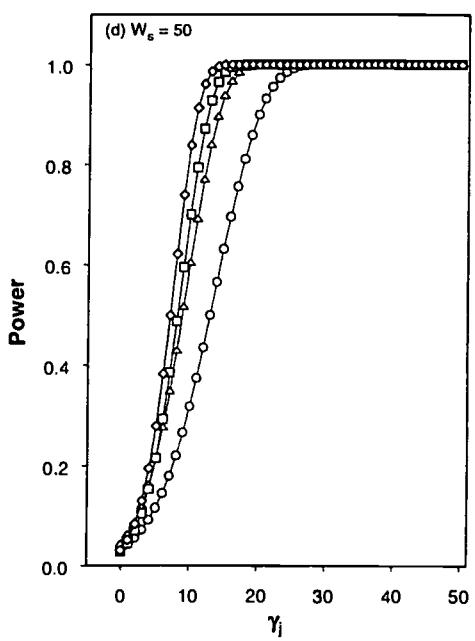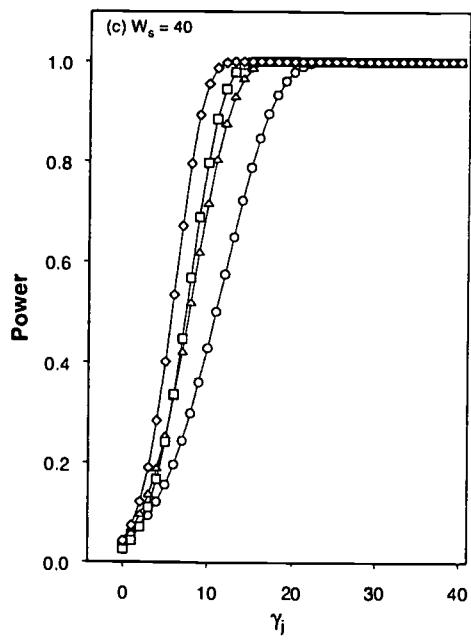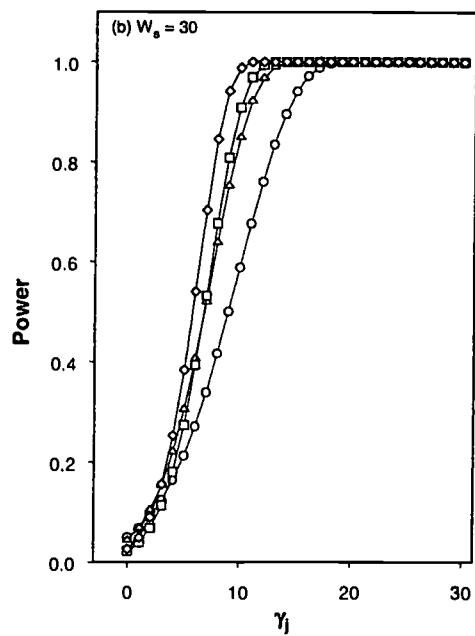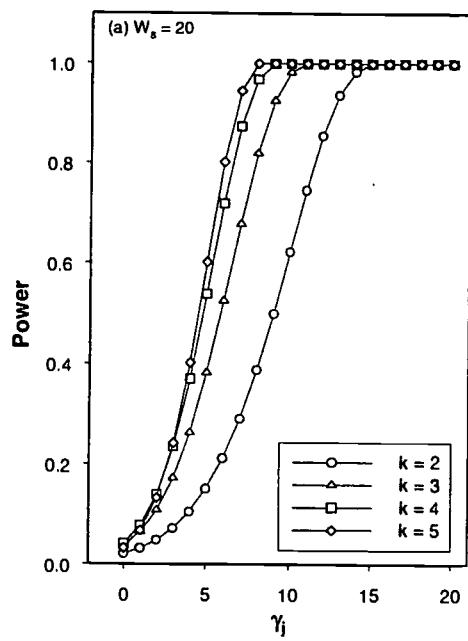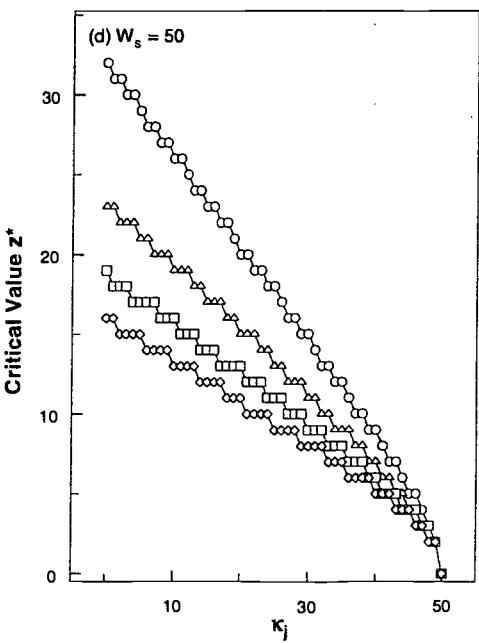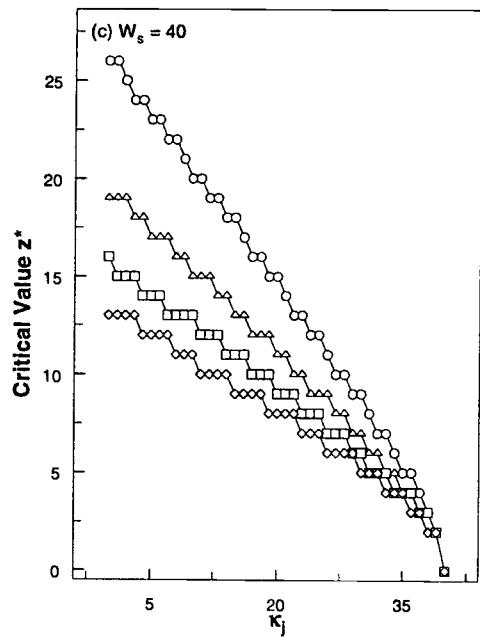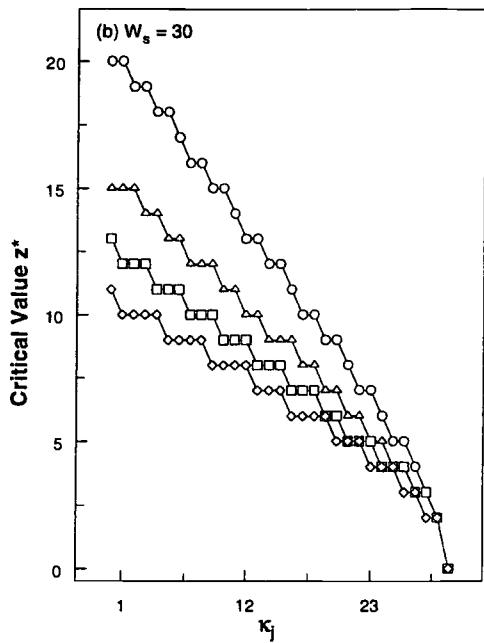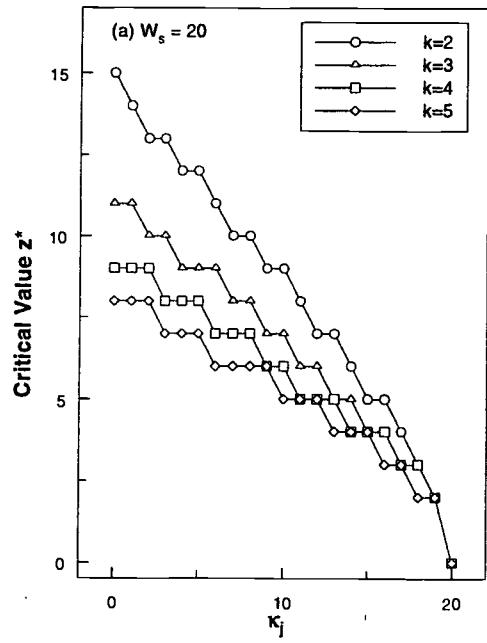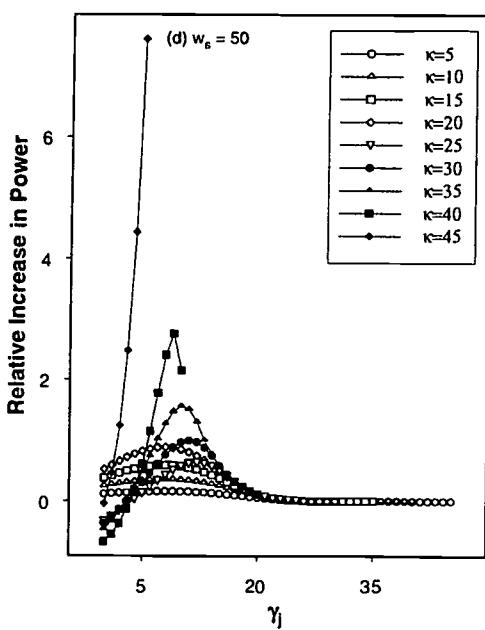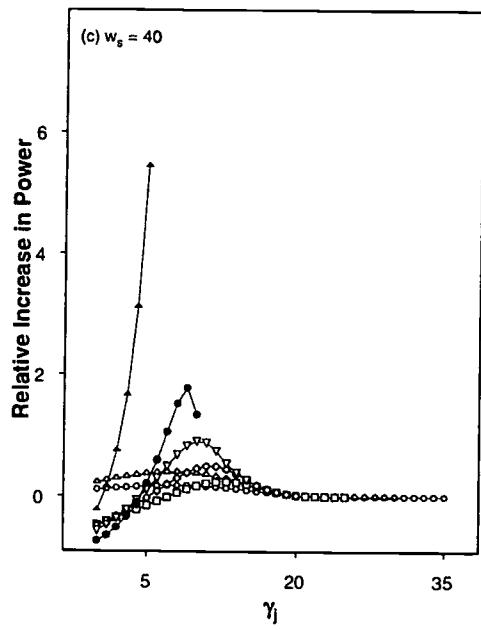
**Titles of Recent Research Reports from the Department of
Educational Measurement and Data Analysis.
University of Twente, Enschede, The Netherlands.**

RR-02-04   W.J. van der Linden & L.S. Sotaridona, *A Statistical Test for Detecting Answer Copying on Multiple-Choice Tests*

RR-02-03   W.J. van der Linden, *Estimating Equating Error in Observed-Score Equating*

RR-02-02   W.J. van der Linden, *Some Alternatives to Sympson-Hetter Item-Exposure Control in Computerized Adaptive Testing*

RR-02-01   W.J. van der Linden, H.J. Vos, & L. Chang, *Detecting Intrajudge Inconsistency in Standard Setting using Test Items with a Selected-Response Format*

RR-01-11   C.A.W. Glas & W.J. van der Linden, *Modeling Variability in Item Parameters in Item Response Models*

RR-01-10   C.A.W. Glas & W.J. van der Linden, *Computerized Adaptive Testing with Item Clones*

RR-01-09   C.A.W. Glas & R.R. Meijer, *A Bayesian Approach to Person Fit Analysis in Item Response Theory Models*

RR-01-08   W.J. van der Linden, *Computerized Test Construction*

RR-01-07   R.R. Meijer & L.S. Sotaridona, *Two New Statistics to Detect Answer Copying*

RR-01-06   R.R. Meijer & L.S. Sotaridona, *Statistical Properties of the K-index for Detecting Answer Copying*

RR-01-05   C.A.W. Glas, I. Hendrawan & R.R. Meijer, *The Effect of Person Misfit on Classification Decisions*

RR-01-04   R. Ben-Yashar, S. Nitzan & H.J. Vos, *Optimal Cutoff Points in Single and Multiple Tests for Psychological and Educational Decision Making*

RR-01-03   R.R. Meijer, *Outlier Detection in High-Stakes Certification Testing*

RR-01-02   R.R. Meijer, *Diagnosing Item Score Patterns using IRT Based Person-Fit Statistics*

RR-01-01   H. Chang & W.J. van der Linden, *Implementing Content Constraints in Alpha-Stratified Adaptive Testing Using a Shadow Test Approach*

RR-00-11   B.P. Veldkamp & W.J. van der Linden, *Multidimensional Adaptive Testing with Constraints on Test Content*

RR-00-10   W.J. van der Linden, *A Test-Theoretic Approach to Observed-Score Equating*

RR-00-09   W.J. van der Linden & E.M.L.A. van Krimpen-Stoop, *Using Response Times to Detect Aberrant Responses in Computerized Adaptive Testing*

RR-00-08   L. Chang & W.J. van der Linden & H.J. Vos, *A New Test-Centered Standard-Setting Method Based on Interdependent Evaluation of Item Alternatives*

RR-00-07   W.J. van der linden, *Optimal Stratification of Item Pools in a-Stratified Computerized Adaptive Testing*

RR-00-06   C.A.W. Glas & H.J. Vos, *Adaptive Mastery Testing Using a Multidimensional IRT Model and Bayesian Sequential Decision Theory*

RR-00-05   B.P. Veldkamp, *Modifications of the Branch-and-Bound Algorithm for Application in Constrained Adaptive Testing*

RR-00-04   B.P. Veldkamp, *Constrained Multidimensional Test Assembly*

RR-00-03   J.P. Fox & C.A.W. Glas, *Bayesian Modeling of Measurement Error in Predictor Variables using Item Response Theory*

RR-00-02   J.P. Fox, *Stochastic EM for Estimating the Parameters of a Multilevel IRT Model*

RR-00-01   E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *Detection of Person Misfit in Computerized Adaptive Tests with Polytomous Items*

RR-99-08   W.J. van der Linden & J.E. Carlson, *Calculating Balanced Incomplete Block Designs for Educational Assessments*

RR-99-07   N.D. Verhelst & F. Kaftandjieva, *A Rational Method to Determine Cutoff Scores*

RR-99-06   G. van Engelenburg, *Statistical Analysis for the Solomon Four-Group Design*

RR-99-05   E.M.L.A. van Krimpen-Stoop & R.R. Meijer, *CUSUM-Based Person-Fit Statistics for Adaptive Testing*
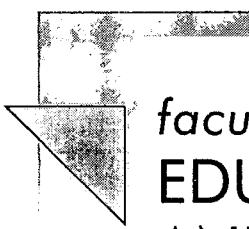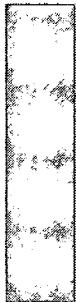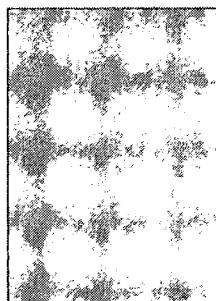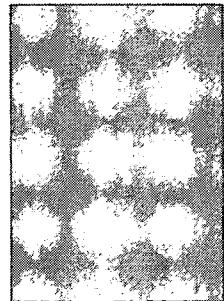
RR-99-04   H.J. Vos, *A Minimax Procedure in the Context of Sequential Mastery Testing*

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*

faculty of
EDUCATIONAL SCIENCE
AND TECHNOLOGY

ERIC®

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").

EFF-089 (3/2000)