DOCUMENT RESUME

ED 310 124 TM 013 697

AUTHOR Kogut, Jan
TITLE Reduction of Bias in Rasch Estimates Due to Aberrant Patterns. Rapport 87-5.
INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE Dec 87
NOTE 42p.; Also cited as Project Psychometrische Aspeten van Item Banking No. 26.
AVAILABLE FROM Mediatheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Computer Simulation; *Estimation (Mathematics); *Goodness of Fit; *Latent Trait Theory; Mathematical Models; *Statistical Bias
IDENTIFIERS Aberrance; Item Parameters; Iterative Methods; *Rasch Model; *Response Patterns

ABSTRACT

The purpose of this study is to determine an efficient way to reduce the bias in estimates of the Rasch model parameters due to aberrant response patterns. First, the benefits of using one- or two-sided goodness-of-fit tests of patterns with the model are discussed. Then, the consequences of removing non-fitting patterns from Rasch model data are considered. Finally, an iterative procedure to reduce the bias is presented. This procedure replaces non-fitting patterns by certain patterns sampled according to the model. The effectiveness of this procedure is investigated in a simulation study using Rasch model data mixed with aberrant response data. It is also demonstrated that, for aberrant response behavior that too often results in ideal patterns, another strategy for detecting aberrant patterns is needed. Such a strategy would provide a possibility to detect more types of aberrant behavior and to reduce bias in the estimates of the model parameters. Four data tables and seven graphs are included. (Author/TJH)

Reduction of Bias in Rasch Estimates Due to Aberrant Patterns

Rapport
87-5

Jan Kogut

...gepaste
...DERWIJSKUNDE

Vakgroep
Onderwijskundige Meetmethoden en Data-analyse

Universiteit Twente

Project Psychometrische Aspecten van Item Banking No.26

Colofon:
Typewerk: Mevr. L.A.M. Padberg
Vormgeving omslag: Audiovisuele Sectie TOLAB Toegepaste Onderwijskunde
Reproductie: Centrale Reproductie-afdeling

Reduction of Bias in Rasch Estimates Due to Aberrant Patterns

Jan Kogut

Reduction of Bias in Rasch Estimates Due to Aberrant Patterns / Jan Kogut - Enschede : Universiteit Twente, Faculteit Toegepaste Onderwijskunde, december, 1987. - p.

Abstract

The purpose of this study is to find out an efficient way to reduce the bias in estimates of the Rasch model parameters due to aberrant response patterns. First, the benefits of using one- or two-sided goodness-of-fit test of patterns to the model are discussed. Then, the consequences of removing nonfitting patterns from Rasch model data are considered. Finally, an iterative procedure to reduce the bias is presented. This procedure replaces nonfitting patterns by certain patterns sampled according to the model. The effectiveness of this procedure is investigated in a simulation study using Rasch model data mixed with aberrant response data.

Introduction

It is known that IRT models often have to be applied to data containing aberrant response patterns; that is, patterns of persons whose response behavior deviates from the model. Such patterns have to be detected, because otherwise predictions from the model may no longer be valid. For this purpose, many person fit indices have been proposed, and still new indices are being developed (e.g. Drasgow, Levine & McLaughlin, 1987; Molenaar & Hoijtink, 1987). Even a small number of aberrant patterns in the data may affect inferences about the model. In particular, the model might be rejected, whereas it is appropriate for the majority of persons, or biased estimates of the model parameters may be obtained. For these reasons, a method to handle aberrant patterns is needed.

A common strategy to deal with aberrant, or more precise, nonfitting patterns is simply to remove them from the data. Recently, several studies of this strategy in the Rasch model (RM) were reported. In an attempt to construct a RM scale, Hoijtink (1986) removed nonfitting patterns from the data iteratively. Nonfitting patterns were excluded in the first run and nonfitting items in the second run. Roger and Hattie (1987) investigated the usefulness of the removal strategy for several popular person and item fit statistics to obtain an overall fit to the RM. These authors showed, by means of simulated data, that the removal strategy did not guarantee fit of the remaining data to the model. Kogut (1987) applied the strategy of iterative removal of nonfitting patterns to obtain a better detection of aberrant patterns. However, he showed that this strategy in general reduces the bias in the estimates of the RM parameters due to aberrant patterns, but at the same time introduces a new bias due to the exclusion of RM patterns misclassified as aberrant.

In this paper, an attempt is made to avoid the introduction of a new bias when reducing the bias on account of aberrant patterns for the RM. In the following section, it is argued that a better detection of aberrant patterns can be obtained if, depending on the expected type of aberrance in the data, a one- or two-sided test of the fit of patterns to the model is used. Next, the presence of bias in the estimates introduced by removal of nonfitting patterns from truly RM data, as well as methods to prevent such a kind of bias are discussed. Finally, a modified procedure to reduce the bias for data including aberrant patterns is presented and verified in a simulation study.

The author uses the same data sets and the same method of approximation of the null distribution of the person fit index as in his previous study (Kogut, 1987). So, in the following sections, these data and the method of approximation are used without any further comments.

Aberrance Resulting in Too Many Ideal Patterns

In this paper, the modified version of the Molenaar index, $M(\underline{X})$ (Molenaar & Hoijtink, 1987), and the above mentioned approximation of its exact null distribution are applied conditionally on the total test score. The definition of the index is as follow :

$$M(\underline{X}) = \sum_{i=1}^{k} - b_i X_i \ ,$$

where $\underline{X} = (X_1, X_2, \ldots, X_k)$ is an individual zero-one response pattern, b_i is the difficulty of item i, and k is the number of items.

From the definition of $M(\underline{X})$, some useful conclusions can be drawn. For each total test score, the associated Guttman pattern (in which only the easiest items are answered correctly and the rest incorrectly) has the largest $M(\underline{X})$ value, whereas the reversed Guttman pattern has the smallest value. In addition, ideal patterns (those for which the easiest and the majority of the relatively easy items are answered correctly) are identified by high $M(\underline{X})$ values. Likewise, rare patterns are identified by low $M(\underline{X})$ values. As is well known, the conditional probability of response pattern $\underline{X}$ given the total test score $r = \sum_{i=1}^{k} X_i$ is,

$$P(\underset{\sim}{X}=\underset{\sim}{x}|r) = 1/\tau_r(\underset{\sim}{b}) \prod_{i=1}^{k} \exp(-b_i x_i) \ .$$

where $\underset{\sim}{b} = (b_1, b_2, \ldots, b_k)$, and $\tau_r(\underset{\sim}{b})$ is the elementary symmetric function of order r (Fischer, 1974). It is easily seen that $M(\underset{\sim}{X})$ is a strictly increasing function of the likelihood $P(\underset{\sim}{X}|r)$, namely,

$$M(\underset{\sim}{X}) = \ln P(\underset{\sim}{X}|r) + \ln \tau_r(\underset{\sim}{b}) \ .$$

Thus, also for the $M(\underset{\sim}{X})$ index, rare patterns have low probabilities and ideal patterns have high probabilities. Finally, $M(\underset{\sim}{X})$ and $P(\underset{\sim}{X}|r)$ give the same order of all possible patterns associated with a fixed total score. Hence, they are equally good in detecting aberrant patterns when using the same percentile of their own null distributions.

In the literature on person fit, only patterns unlikely under the model are defined as aberrant. Therefore, for indices which order patterns in the same way as the $M(\underset{\sim}{X})$ index, only the left-tail (1%- or 5%-th) percentile of the index null distribution is used (e.g. Drasgow, Levine & Williams, 1985; Molenaar & Hoijtink, 1987). This approach is justified, if rare patterns are occurred in the data more often than predicted by the RM, for instance, in the case of guessing. Typical distributions of $M(\underset{\sim}{X})$ for RM-behaving persons and guessers are given in Figure 1(a) by means of a continuous

Insert Figure 1 about here

approximation. So, if patterns are defined as aberrant by an index value lower than the left-tail $100*\alpha\%$-th percentile, then a higher than $100*\alpha$ percentage of aberrant patterns will be classified correctly and about $100*\alpha\%$ of the RM patterns will be misclassified. On the other hand, in the case of aberrant persons who respond with a lower ability on the more difficult items than on the rest of the items, or in the case of aberrants who answer more difficult items according to the two-parameter logistic model with discrimination parameters larger than the ones of the rest of items, this approach is not appropriate. In both cases, as can be concluded from the item characteristic curves (ICC's), rare patterns are produced less often and ideal patterns more often than predicted by the RM. Therefore, in such cases, the percentage of correctly classified aberrants will, as Table 1 indicates, generally be lower than $100*\alpha$.

Insert Table 1 about here

Nevertheless, if patterns are defined as aberrant by an index value higher than the right-tail percentile associated with 100*(1–α)%, then many more aberrant patterns will be classified correctly (see Table 1). The approximate distributions of M(X) for RM-behaving persons, and persons displaying the aberrance in question are illustrated in Figure 1(b). Note that the probability of misclassifying a RM pattern as aberrant, the Type I error, is α again. This is why, in order to effectively detect the aberrance of this kind, a right-tail rather than a left-tail test has to be used.

A definition of ideal patterns as aberrant may sound very strange because these patterns are just the most probable ones under the model. Yet, for the purpose of an effective detection of deviations from the model this possibility should be taken into account.

To conclude, an efficient detection of aberrant patterns, can be only achieved if the direction of the person fit test is in agreement with the expected type of aberrance in the data. If there is no information at all about the type of aberrance, then the use of a two-sided test is the most appropriate strategy (e.g., using the 100*(α/2)% and 100*(1–α/2)%-th percentiles).

Removal and Replacement of Nonfitting Patterns from RM Data

The removal of aberrant patterns from the data will reduce the bias in the estima... of the RM parameters. Unfortunately, we can only remove patterns which are *classified* as aberrant, and we are forced to produce classification errors. If this strategy is carried out iteratively, then generally the number of misclassified RM patterns removed will increase with the number of iterations. From a previous study (Kogut, 1987), it may be concluded that this approac' will introduce a new bias in the parameter estimates in the opposite direction.

To investigate the seriousness of such a bias, let us consider the consequences of the removal strategy if it is applied to truly *RM data*. The bias in the estimator $\hat{p}_i$ for the difficulty or ability parameter p_i, can be defined as follow :

$$\text{bias} = p_i - E(\hat{p}_i) \ .$$

where $E(\hat{p}_i)$ denotes the expected value of $\hat{p}_i$. The relevant quantity to measure bias over parameters is the root mean square error (RMSE):

$$\text{RMSE} = \{1/m \sum_i (p_i - \hat{p}_i)^2\}^{1/2} \ .$$

where m is the number of the parameters.

Let us now remove from the RM data all patterns for which the value of the Molenaar index is lower than the left-tail 100*α% : 1%, 5%, 10%, 25% and 50% -th percentile of the index distribution. The results of these five analyses are presented in Table 2 and Figure 2. Table 2

Insert Table 2 about here

shows that RMSE for the difficulty estimates increases significantly with α. For the ability estimates, the same phenomenon is also observed, yet, not in the significant degree. The estimated bias in the difficulty estimates is shown in Figure 2. Figure 2 shows that the difficulties of

Insert Figure 2 about here

easy items are underestimated (a positive bias), but these of hard items are overestimated (a negative bias). In other words, the estimates are shifted to the ends of the difficulty scale. Even at as low as the 5%-th percentile, the estimated bias for a few items on both extremes of the difficulty scale is larger than the standard error of estimate. Furthermore, it can be seen that the estimated bias increases systematically with the percentile used, and that it is more pronounced for the difficulties at the extreme parts of scale. These results could be expected, as the use of the Molenaar index in the removal strategy leads to rejection of rare patterns. Hence, it must lead to an increase of the proportions of correct answers on the easiest items and a decrease of these ones on the most difficult items. Therefore, the estimates of the difficulties will show a tendency towards dispersion.

Now, let us examine the results of the removal strategy if it is carried out iteratively. The results for four subsequent iterations, using the 5%-th percentile are presented in Table 3 and Figure 3. They show that RMSE and

---

Insert Table 3 about here

---

the estimated bias increase significantly over iterations,

Insert Figure 3 about here

---

until a certain limit is reached.

It is remarkable that the bias due to the removal of misclassified rare patterns from truly <u>RM data</u>, in comparison with the one due to truly rare aberrant patterns in <u>mixed data</u> has an opposite direction (Kogut, 1987). Thus, in the first few iterations in the removal strategy, these two types of bias can be expected to cancel each other out.

As was already noted, some types of aberrance may lead to a lot more ideal patterns than predicted by the RM. Therefore it may be useful to verify what the consequences of removing ideal patterns from RM data will be. Patterns removed are the ones for which values of the index are higher than the right-tail $100*(1-\alpha)\%$: 99%, 95%, 90%, 75% and 50% -th percentile of the index distribution. The results from these investigations are reported in Table 2 (RMSE) and Figure 4 (estimated bias). Table 2 shows that the bias for the difficulty estimates is introduced and increases with α again. In addition, Figure 4 shows that the estimated bias has an opposite

Insert Figure 4 about here

direction and a smaller size in comparison with the one arising when removing rare patterns. The reversed direction of the bias in question indicates that the difficulties of easy items are overestimated, and the difficulties of hard items underestimated. From the definition of ideal patterns it is obvious that the removal of these patterns must result in decreased proportions of correct answers on easy items and increased proportions on difficult items. Hence, the estimates of the difficulties will shift to the middle of the scale. A smaller size of the bias implies an apparently less sensitivity of the difficulty estimates to the lack of ideal patterns than to the lack of rare ones. Obviously, if rare and ideal patterns are removed simultaneously (e.g., using the $100*(\alpha/2)\%$ and $100*(1-\alpha/2)\%$-th percentiles), the resulting bias will be the sum of these two effects.

To conclude, if one removes patterns from the data that are classified as aberrant, one should take into account that, even with a small Type I error and few iterations used, a significant bias may be introduced in estimates of the difficulties due to the removal of RM patterns misclassified as aberrant.

For the RM data, the introduction of bias related to the removal of misclassified RM patterns may be simply omitted if these patterns are substituted by similar patterns sampled according to the RM with parameters equal to their estimates. More precisely, when using the $100*\alpha\%$-th percentile, patterns serving as a substitute are the first sampled ones for which the person fit index value is lower than this percentile. When using another of the proposed one- or two-sided tests of person fit, the appropriate substitution may be equally easy to realize. Because in such modified data, rare and ideal patterns are present in about the same proportions as in the RM data, the estimates of the difficulty parameters can be expected to be unbiased.

Note that the replacement of misclassified RM patterns by simply at random sampled RM patterns recovers an insufficient number of rare and aberrant patterns to prevent the introduction of new bias, as can be seen in Figure 5 for the left-tail percentiles. In particular,

Insert Figure 5 about here

when a small percentile is used, the estimates obtained in the removal strategy and in the simple replacing procedure remain almost equally biased.

So, it is clear that if proportions of rare and ideal patterns in the sample do not conform to the RM, then estimates of the RM difficulty parameters will be systematically biased. A relatively lack of rare or ideal patterns will systematically increase or decrease the dispersion of the difficulty estimates (the estimates tend to shift to the ends or to the middle of the scale), respectively. Likewise, if rare or ideal patterns occur relatively too often in the data, then the dispersion in question will decrease or increase, respectively (Kogut,1987).

A Procedure to Reduce Bias
in the RM Parameters Estimates

Let us return to the main subject of this paper, i.e., to the efficient reduction of bias inherent in estimates of the RM parameters if response data contains aberrant response patterns. Suppose that for some mixed data, the RM difficulties and the number of aberrant patterns are known. If the detection of aberrant patterns is carried out using a fixed percentile of the person fit index, then the number of misclassified RM patterns, say x, is approximately known. Now, let us replace x patterns classified as aberrant by patterns sampled from the RM in accordance with the percentile used (see the previous section), and the rest of patterns found to be aberrant by RM patterns sampled at random. Thanks to such an operation, the modified data will be purified from the aberrant patterns, but will still contain the proper proportions of rare and ideal patterns. Hence, the reduction of the original bias can be realized without any introduction of new bias. Further, the higher the percentile used, the more aberrant patterns can be detected and replaced; and finally a better reduction of bias in the estimates can be achieved. In particular, if the data contains only aberrant patterns with index values within the range of percentile used, then utilizing a higher percentile will enable us to detect all aberrant patterns, and to eliminate all bias.

Of course, in practice the item difficulties and the number of aberrant patterns are unknown, but they can be estimated from the data. So, an iterative procedure including the estimation of the RM parameters and the number of aberrant patterns can be expected to converge, and the final difficulty estimates can be expected to be less biased. This idea is realized in the following procedure.

An algorithm for the procedure is sketched below. The user of the procedure can supplement the algorithm with specific subprograms (according to software available, the type of aberrance expected in the data, the person fit index used, and the degree of reduction of bias needed). The algorithm is as follows :

(1) estimate the item and person parameters from the data set in question;

(2) calculate the $100*\alpha\%$-, $100*(1-\alpha)\%$-, or both, e.g. $100*(\alpha/2)\%$- and $100*(1-\alpha/2)\%$-th percentile(s) for the person fit index, per test score, using the estimates from step (1);

(3) estimate the number of aberrant patterns in the data, $n^j{}_A$, as follow,

$$n^j{}_A = n^{j-1}{}_A - \mathrm{int}\{(n - n^{j-1}{}_A)*\alpha\}$$

(where $n^{j-1}{}_A$ is the observed number of aberrant patterns from the previous iteration, and n is the number of nonzero or non-perfect patterns from the original data);

(4) calculate the values of the person fit index for the patterns in the original data set and determine which patterns are aberrant for each test score, using the percentiles from (2);

(5) replace the patterns classified as aberrant in (4) by RM patterns sampled from the RM with parameter values equal to their estimates from (1). More in particular:

(a) replace $\text{int}\{(n-n^j{}_A)*\alpha\}$ aberrant patterns by the first sampled RM patterns for which the index value is lower than $100*\alpha$%-, higher than $100*(1-\alpha)$%-, or one of both lower than $100*(\alpha/2)$%- or higher than $100*(1-\alpha/2)$%-th percentile(s), respectively to the chosen percentile in (2);

(b) replace the rest of aberrant patterns by the first randomly sampled RM patterns;

(6) return to step (1) and repeat the procedure on the data set created in step (5), until convergence of some reasonable criterion, say CRIT, is obtained, e.g.,

$$| b^j{}_i - b^{j-1}{}_i | < CRIT_i \;, \quad i = 1, 2, \ldots, k;$$

where $CRIT_i$=estimated SE_i (standard error) of i-th difficulty parameter in j-th iteration, $b^j{}_i$, from (1).

In step (1), the CML estimation procedure (Fischer, 1974) should be used, as only such a procedure assures unbiased estimates of the difficulties. If a test consists of a small number of items, say some 15, the exact percentiles of the person fit index should be calculated through complete enumeration of all index values and their probabilities. This method is recommended because any approximation might fail due to the high sensitivity of the index distribution to the difficulty parameters (Molenaar & Hoijtink, 1987; Snijders, 1987). Nevertheless, if a test consists of more than 15 items, then the computations get involved and an approximation has to be used. Two approximations can be recommended : a chi-square approximation (Molenaar and Hoijtink, 1987), and an approximation through sampling RM patterns (Kogut, 1987). The benefits of using higher percentiles have already been discussed, however, then more iterations will be needed to satisfy the criterion of convergence. In order to achieve the convergence criterion quicker, in step (3) $n^1{}_A$ in the first iteration can be initialized roughly as the expected number of aberrant patterns in the original data set. Finally, in step (6) another convergence criterion than SE_i may be used, as well. Yet, it should not be too hard to be fulfilled, because then the procedure might not converge.

Now the efficiency of the proposed procedure will be considered for the case of RM data mixed with aberrant response patterns. In our simulation study, the used aberrant patterns are subject to guessing to complete on the five most difficult items, and possess M(X) values lower than the 5%-th percentile. The results from this study, by the use of the CML estimation procedure (Fischer, 1974), an approximation of the 10%-th percentile of the Molenaar index through sampling RM patterns, and an initialization of n^1_A equal to 0, are presented in Table 4 and Figure 6. The procedure was run on an Olivetti M-24 microcomputer. In Table 4 can be seen that RMSE, for the

Insert Table 4 about here

difficulty estimates decreases with the number of iterations, until RMSE for the RM data is approached. The difficulty estimates from the eight and seventh iterations differed from each other no more than the specified criterion, CRIT, and therefore the procedure was stopped. The estimated bias over subsequent iterations is shown in Figure 6. The figure shows that generally each next

Insert Figure 6 about here

iteration results in a monotonical reduction of bias. So, the results are superior to those for the removal strategy in Kogut (1987), in which bias is reduced only in the first iterations, but increases in next ones. Note that bias in the ability estimates is slightly increased (see Table 4), however, it can be ignored as compared with RMSE for the RM data used.

Discussion

As is known, estimates of the RM parameters may be biased due to aberrant patterns in the data. This paper shows that the removal strategy fails to reduce the bias satisfactorily because of the fact of removing RM patterns misclassified as aberrant that introduces a new bias. The new bias may be significant even if the percentile(s) of the person fit index related to a small Type I error, and few iterations of the removal strategy are used.

To obtain estimates of the RM parameters with as little bias as possible, a modified procedure was proposed. In the procedure, nonfitting patterns are replaced by patterns sampled from the RM. Some number of the replaced patterns have to satisfy an additional criterion. A sufficient level of bias reduction can be achieved safely by the use of the percentile(s) related to a larger Type I error. In particular, if the data contains only aberrant patterns with index values within the range of percentile(s) used, then all the bias can be eliminated. With the help of the removal strategy such a reduction of bias is generally impossible.

In addition, it was shown that for aberrant response behavior that too often results in ideal patterns, another strategy for detecting aberrant patterns is needed. Such a strategy would provide us with the possibility to detect more types of aberrant behavior, and to better reduce bias in the estimates of the model parameters.

References

Drasgow, F., Levine, M.V. (1986). Optimal detection of certain forms of inappropriate test scores. Applied Psychological Measurement, 10, 59-67.

Drasgow, F., Levine, M.V., & McLaughlin, M.E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. Applied Psychological Measurement, 11, 59-79.

Drasgow, F., Levine, M.V., & Williams, E.A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. British Journal of Mathematical and Statistical Psychology, 38, 67-86.

Fischer, G.H. (1974). Einführung in die Theorie psychologischer Tests. Bern:Huber.

Hoijtink, H. (1987). Rasch schaal constructie met behulp van een passingsindex voor personen. Kwantitatieve Methoden, 25, 101-110.

Kogut, J. (1987). Detecting Aberrant Patterns in the Rasch Model, R-87-3. The Netherlands: Enschede, University of Twente, Department of Education.

Molenaar, I.W & Hoijtink, H. (1987). The many null distributions of person fit indices. Heymans Bulletin, HB-87-846-EX. Groningen: R.U.Groningen.

Phillips, S.E. (1986). The effects of the deletion of misfitting persons on vertical equating via the Rasch model. Journal of Educational Measurement, 23, 107-118.

Rogers, H.J. & Hattie, J.A. (1987). A Monte Carlo investigation of several person and item fit statistics for item response models. Applied Psychological Measurement, 11, 47-57.

Smith, R.M. (1986). Person fit in the Rasch model. Educational and Psychological Measurement, 45, 433-444.

Snijders, T. (1987). Statistical aspects of person fit assessment. Paper presented at the Fourth Workshop on IRT, University of Twente, November, 1987.

Vijver, F.J.R. (1986). The robustness of Rasch estimates. Applied Psychological Measurement, 10, 45-57.

Author's Note

The author is indebted to Martijn P.F. Berger, Ron J.H. Engelen and Wim J. van der Linden for their comments on an earlier version of the paper. The content of the paper is, however, fully his own responsibility.

Table 1

Percentages of correctly classified aberrants responding with two different abilities, using 5%- and 95%-th percentiles of M(X)

Difference in Abilities	Percentage 5% Percentile	Percentage 95% Percentile
-1.0	1.8	9.3
-2.0	1.0	16.4

Note. Ability on the five most difficult items is lower than ability on the rest of items; ability of aberrants is N(0.0, 1.53).

Table 2

RMSE of the RM parameter estimates after removal of nonfitting patterns from RM data using left- and right-tail percentiles of M(X)

Percentile	RMSE	
	Difficulty	Ability
	Left-Tail Test	
1%	0.052	0.624
5%	0.106	0.640
10%	0.206	0.666
25%	0.423	0.736
50%	0.977	0.979
	Right-Tail Test	
99%	0.061	0.610
95%	0.100	0.605
90%	0.150	0.595
75%	0.299	0.578
50%	0.513	0.560

Note. For the RM data used, RMSE is 0.055 for the difficulty and 0.616 for the ability estimates.

Table 3

RMSE of the RM parameter estimates after iterative removal of nonfitting patterns from RM data using 5%-th percentile of M(X)

Iteration	RMSE	
	Difficulty	Ability
1	0.106	0.640
2	0.146	0.652
3	0.165	0.657
4	0.166	0.657

Note. For the RM data used, RMSE is 0.055 for the difficulty and 0.616 for the ability estimates.

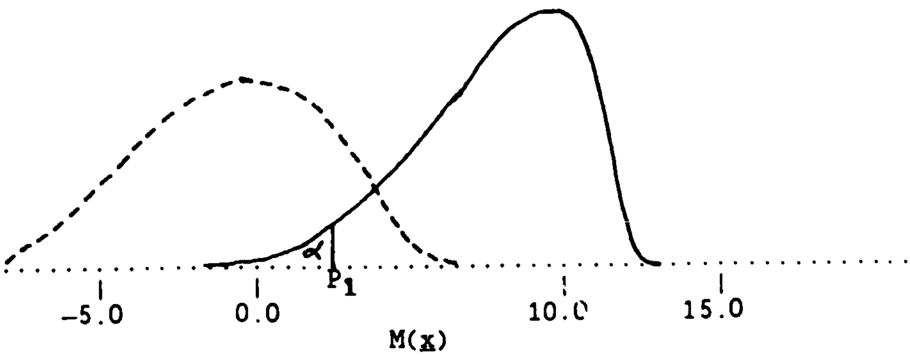
Table 4

RMSE of the RM parameter estimates after iterative replacement of nonfitting patterns from mixed data using 10%-th percentile of M(X)

Iteration	RMSE	
	Difficulty	Ability
1	0.499	0.723
2	0.300	0.732
3	0.188	0.741
4	0.126	0.746
5	0.081	0.751
6	0.069	0.755
7	0.063	0.756
8	0.070	0.757

Note. For the RM data used, RMSE is 0.055 for the difficulty and 0.616 for the ability estimates.

(a) Guessing on Easy Items



(b) Responding with Lower Ability on Difficult Items

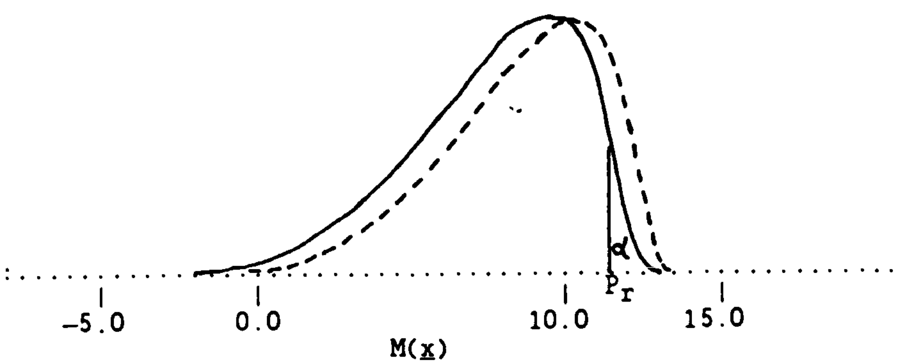


Figure 1. Approximate distributions of M(X) for RM behavior (solid line) and two types of aberrant response behavior (dashed line).

Note. P_l and P_r indicate the left- and right-tail percentiles, respectively.

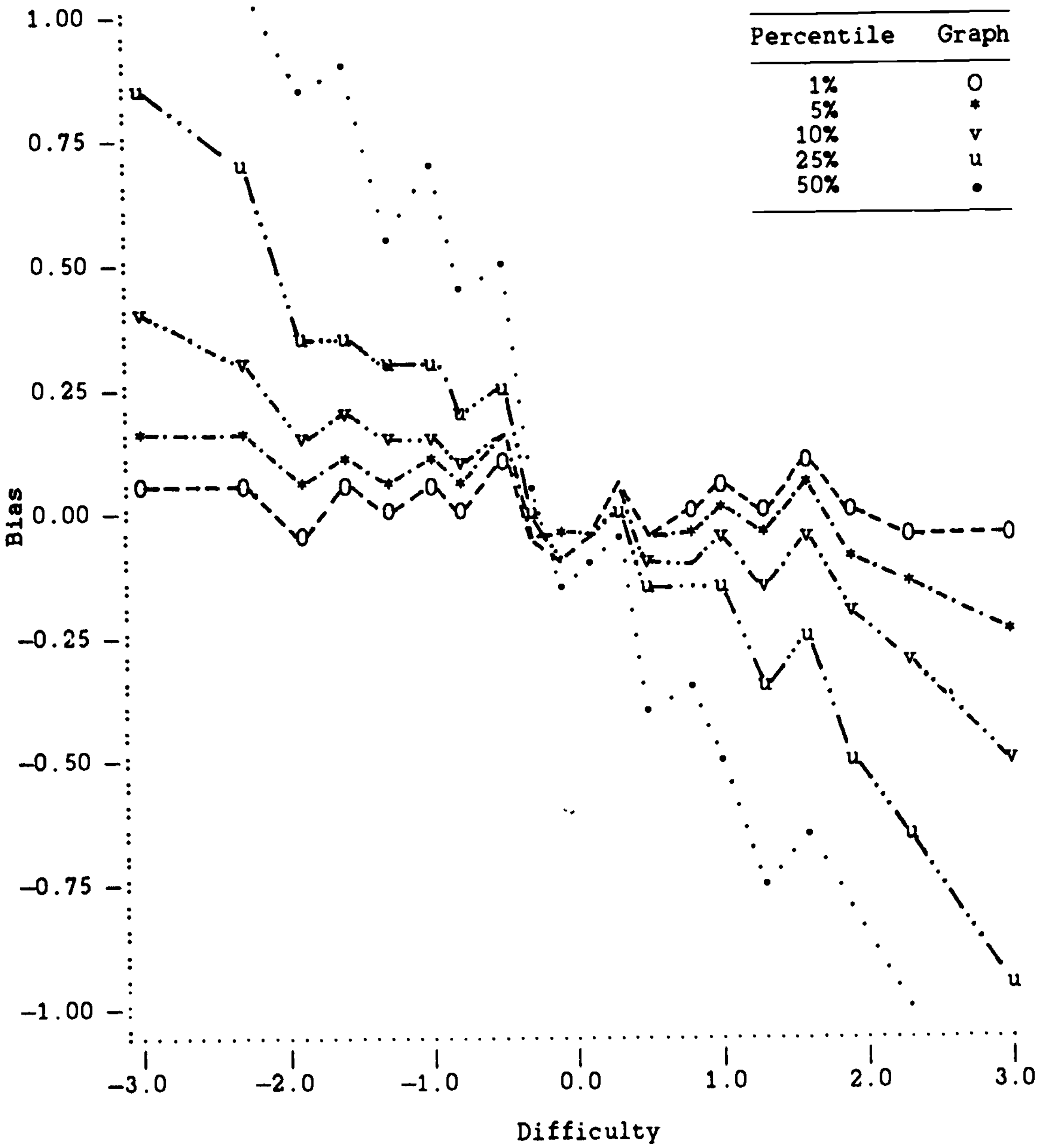


Figure 2. Estimated bias in the difficulty estimates after removal of nonfitting patterns from RM data using left-tail percentiles of M(X).

Note. Standard errors are about 0.045 for the middle and about 0.080 for the extreme difficulties.

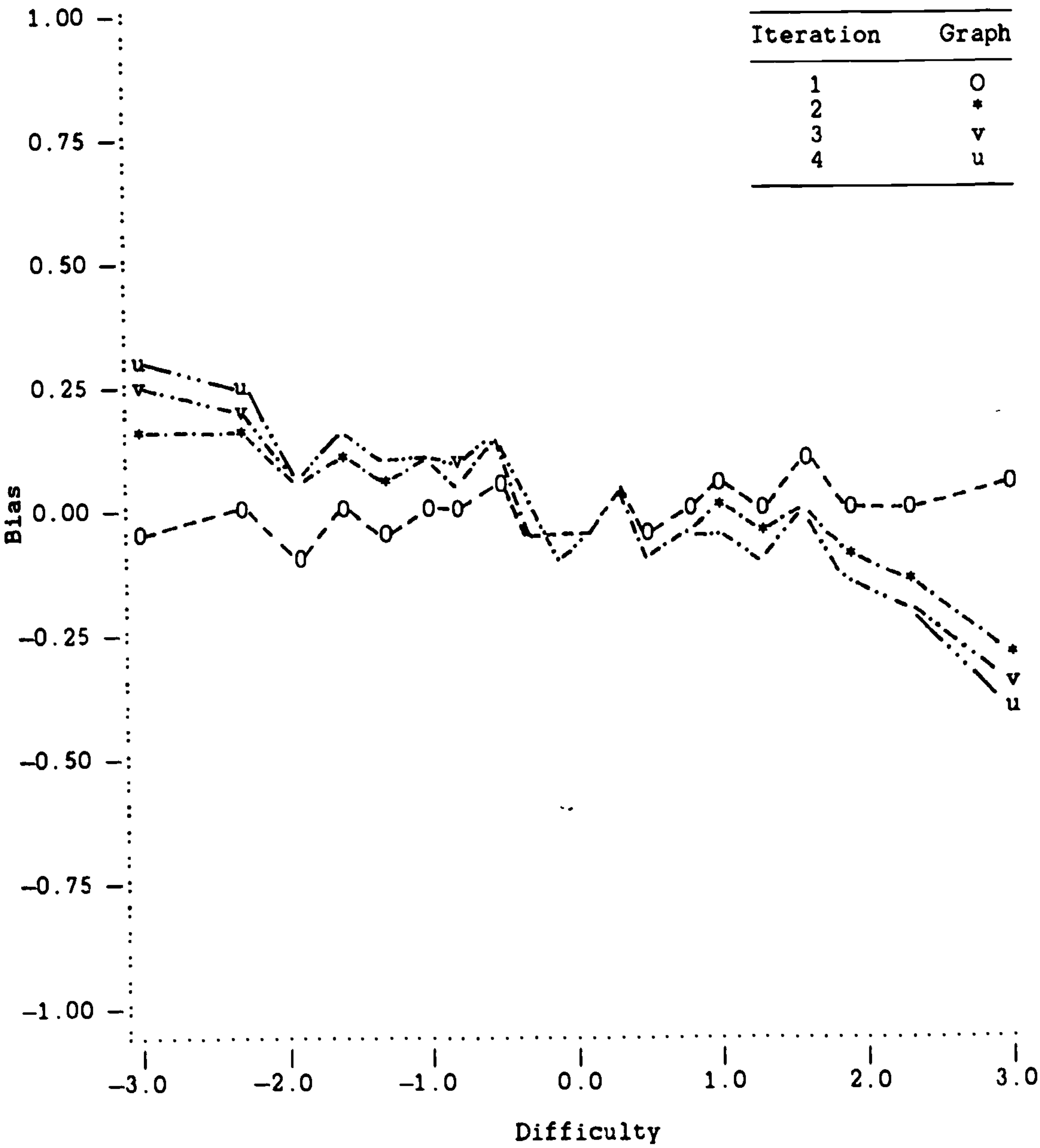
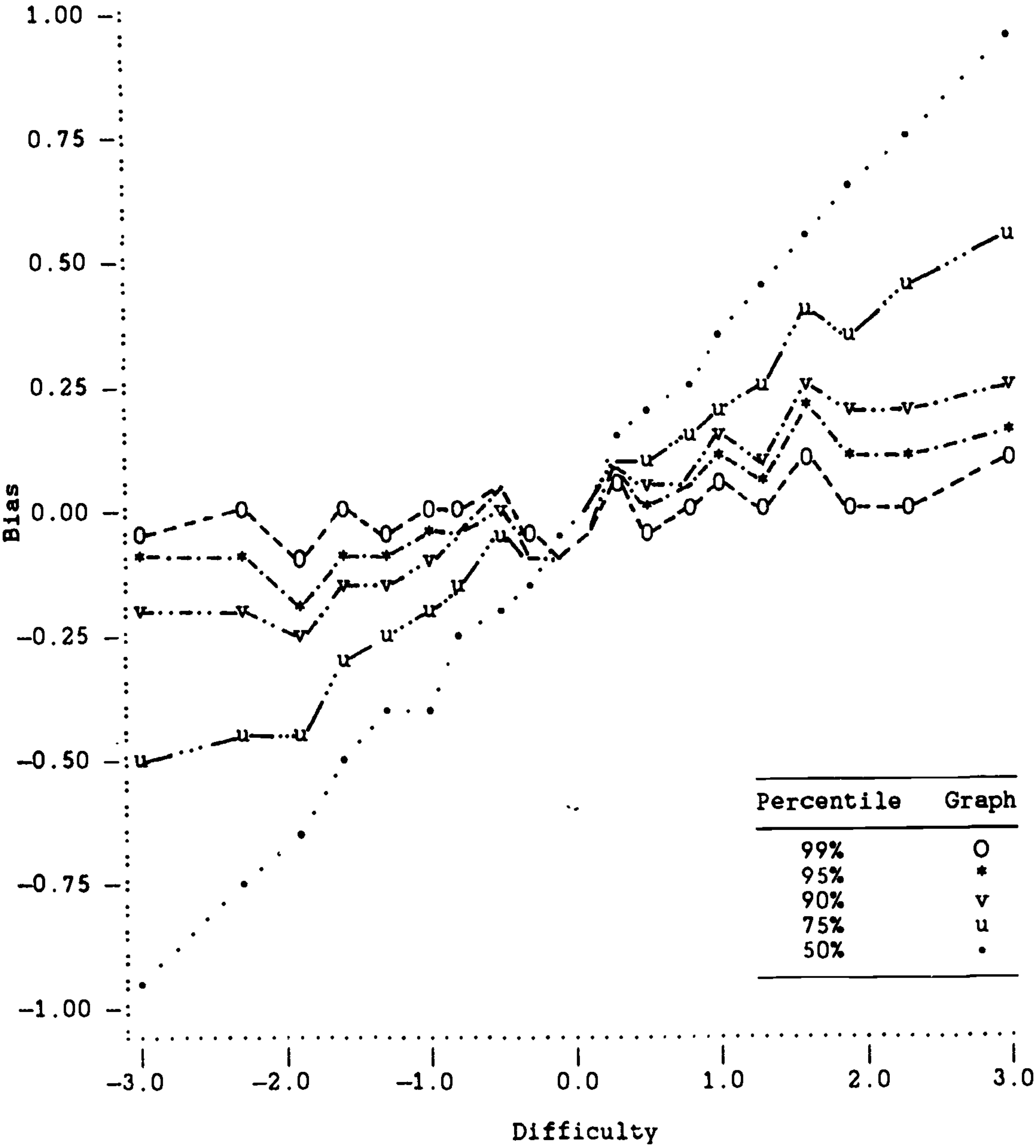


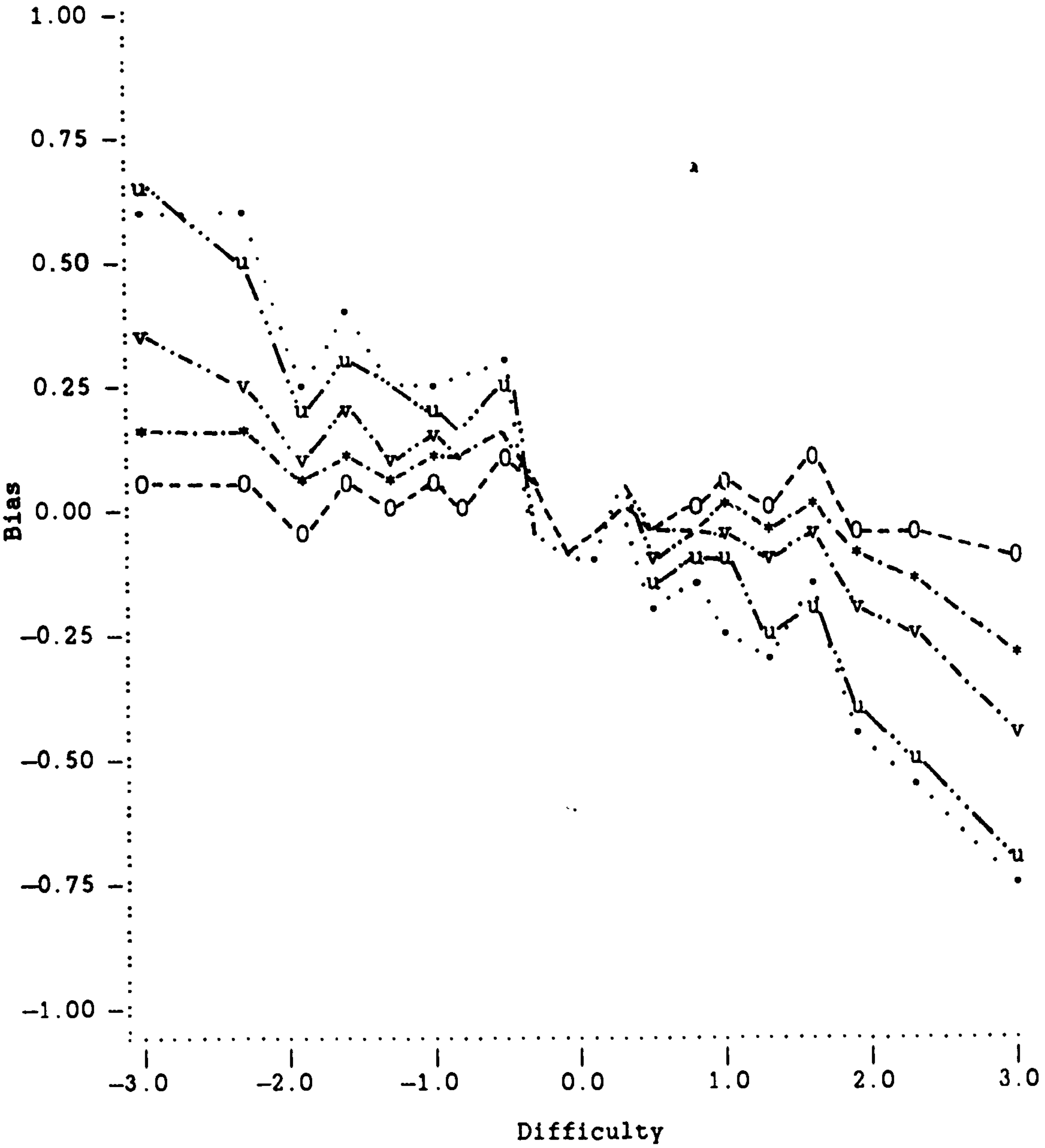
Figure 3. Estimated bias in the difficulty estimates after iterative removal of nonfitting patterns from RM data using 5%-th percentile of M(X).

Note. Standard errors are about 0.045 for the middle and about 0.080 for the extreme difficulties.



<u>Figure 4.</u> Estimated bias in the difficulty estimates after removal of nonfitting patterns from RM data using right-tail percentiles of M(X).

<u>Note.</u> Standard errors are about 0.045 for the middle and about 0.080 for the extreme difficulties.



<u>Figure 5.</u> Estimated bias in the difficulty estimates after simple replacement of nonfitting patterns from RM data using left–tailpercentiles of M(X). (See Figure 2 for an explanation of symbols).

<u>Note.</u> Standard errors are about 0.045 for the middle and about 0.080 for the extreme difficulties.

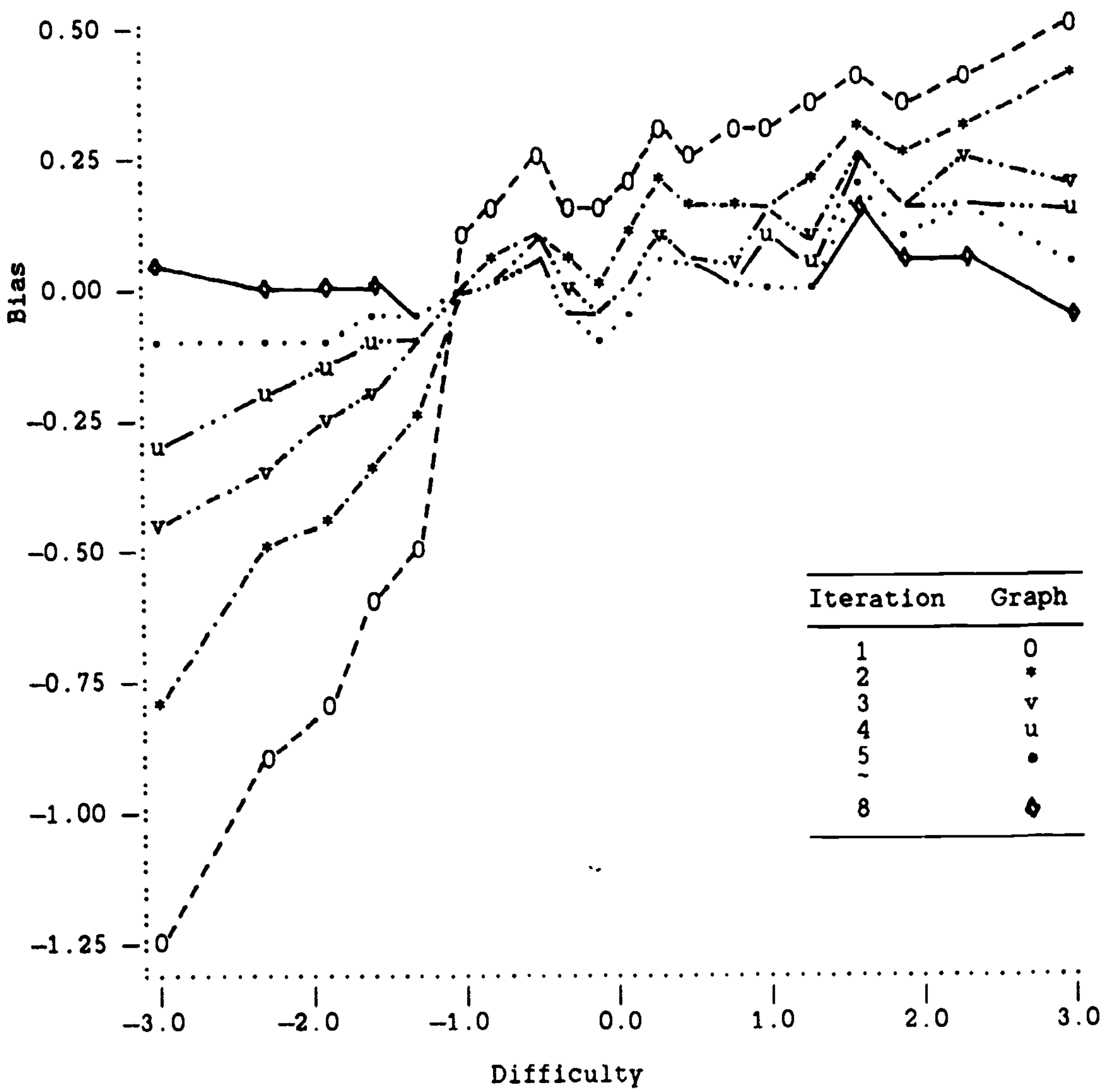


Figure 6. Estimated bias in the difficulty estimates after iterative replacement of nonfitting patterns from mixed data using 10%-th percentile of M(X).

Note. Standard errors are about 0.045 for the middle and about 0.080 for the extreme difficulties.

Titels van Recente Rapporten

R-86-1 C. Sluijter, Het bijwerken van een itembank

R-86-2 D.L. Knol, Inventarisatie van automatische itemselectie procedures voor Rasch schalen

R-86-3 G.R. Buning, T.J.H.M. Eggen, H. Kelderman, & W.J. van der Linden (red.), Het gebruik van het Raschmodel voor een decentraal toetsservicesysteem

R-86-4 J. Kogut, Review of IRT-based indices for detecting and diagnosing aberrant response patterns

R-86-5 D.L. Knol, Een overzicht van meer-dimensionale itemresponsmodellen

R-87-1 J. Adema, Testconstructie met klassieke item- en testparameters

R-87-2 D.L. Knol, Het verband tussen item respons theorie en factoranalyse voor dichotome items.

R-87-3 J. Kogut, Detecting aberrant response patterns in the Rasch model

R-87-4 J. Adema. Simultane toetsconstructie en andere doelen dan maximalisatie van α

R-87-5 J. Kogut, Reduction of bias in Rasch estimates due to aberrant patterns

Rapporten in deze reeks kunnen tegen kostprijs besteld worden bij: Mediatheek, Faculteit der Toegepaste Onderwijskunde, Universiteit Twente, Postbus 217, 7500 AE Enschede. Tel. 053-893588.

Titles of Recent Research Reports

RR-86-1 W.J. van der Linden, The use of test scores for classification decisions with threshold utility

RR-86-2 H. Kelderman, Item bias detection using the loglinear Rasch model: Observed and unobserved subgroups

RR-86-3 E. Boekkooi-Timminga, Simultaneous test construction by zero-one programming

RR-86-4 W.J. van der Linden, & E. Boekkooi-Timminga, A zero-one programming approach to Gulliksen's matched random subtests method

RR-86-5 E. van der Burg, J. de Leeuw, & R. Verdegaal, Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features

RR-86-6 W.J. van der Linden, & T.J.H.M. Eggen, An empirical Bayes approach to item banking

RR-86-7 E. Boekkooi-Timminga, Algorithms for the construction of parallel tests by zero-one programming

RR-86-8 T.J.H.M. Eggen, & W.J. van der Linden, The use of models for paired comparisons with ties

RR-86-9 H. Kelderman, Common item equation using the loglinear Rasch model

RR-86-10 W.J. van der Linden, & M.A. Zwarts, Some procedures for computerized ability testing

RR-87-1 R. Engelen, Semiparametric estimation in the Rasch model

RR-87-2 W.J. van der Linden (Ed.), IRT-based test construction

RR-87-3 R. Engelen, P. Thommassen, & W. Vervaat, Ignatov's theorem: A new and short proof

RR-87-4 E. van der Burg, & J. de Leeuw, Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis

RR-87-5 H. Kelderman, Estimating a quasi-loglinear models for the Rasch table if the number of items is large

RR-87-6 R. Engelen, A review of different estimation procedures in the Rasch model

RR-87-7 D.L. Knol & J.M.F. ten Berge, Least-squares approximation of an improper by a proper correlation matrix using a semi-infinite convex program

RR-87-8 E. van der Burg & J. de Leeuw, Nonlinear canonical correlation analysis with k sets of variables

RR-87-9 W.J. van der Linden, Applications of decision theory to test-based decision making

RR-87-10 W.J. van der Linden & E. Boekkooi-Timminga, A maximin model for test design with practical constraints

Research Reports can be obtained at costs from Mediatheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

gepaste
ONDERWIJSKUNDE

Een publicatie
van de faculteit der
Toegepaste Onderwijskunde
Twente
schede

42