Psychological Test and Assessment Modeling, Volume 53, 2011 (3), 380-393

On the impact of missing values on item fit and the model validness of the Rasch model

Christine Hohensinn[1] *& Klaus D. Kubinger*[2]

Abstract

A crucial point regarding the development and calibration of an aptitude test is the presence of missing values. In most test administrations, examinees omit individual items even in high-stakes tests. The most common procedure for treating these missing values in data analysis is to score these responses as incorrect; however, an alternative would be to consider omitted responses as if they were not administered to the examinee in question. Previous research has found that both procedures for dealing with missing values result in bias in item and person parameter estimation. Regarding test construction, not only is there an interest in item parameter estimation, but also in global and item-specific model tests as well as goodness-of-fit indices. On the basis of such statistics, it will be decided which items constitute the final item pool of a test. The present study therefore investigates the influence of two different procedures for dealing with missing values on model and item-specific tests as well as item fit indices for the Rasch model. The impact of these different treatment alternatives is shown for an empirical example and, furthermore, for simulated data. Simulations reveal that the global model test, as well as the item test, is affected by the procedures used to deal with missing values. To summarize, the results indicate that scoring omitted items as incorrect leads to seriously biased results.

Key words: Missing values; Rasch model; item fit; model test; goodness of fit statistic

[1] Correspondence concerning this article should be addressed to: Christine Hohensinn, PhD, Division of Psychological Assessment and Applied Psychometrics, Faculty of Psychology, University of Vienna, Liebiggasse 5, A-1010 Vienna, Austria, Europe; email: christine.hohensinn@univie.ac.at

[2] Faculty of Psychology, University of Vienna, Austria

Introduction

The fundamental principle of psychological assessment is to infer a non-observable latent trait variable from observed responses on test items. A crucial point in this process is the omission of items by an examinee. If there is no observable reaction to an item, it is difficult to decide what can be inferred about the latent proficiency. In practice, these omitted items are often treated as incorrect responses. The reasons for the omissions are often unclear. De Ayala, Plake, and Impara (2001) summarize three reasons for non-response of an item: firstly, for some test designs, only a subset of items are presented to examinees – this is true in adaptive testing as well as in large-scale tests using matrix sampling of items leading to multiple, different test booklets. Therefore, for items which are not administered, examinees cannot produce a response. In this case, responses are missing by design and there is no problem concerning parameter estimation. Secondly, a speed effect could occur, meaning that the examinee does not have enough time to work through the items presented at the end of a speed-limited test. The non-response to these items is not a decision of the examinee, but rather comparable to the first situation where not all items of the item pool are administered to all examinees. However, non-response here is not due to the design, but rather due to individual differences in response latency. Thirdly, items are often also omitted even if the examinees have time enough to finish the test. In this case, skipping an item appears to be a conscious decision of the examinee and it can be conjectured that there is a differential tendency to omit items, depending on certain personality traits. Matters and Burnett (2003) found that the academic self-concept, test-irrelevant thinking, achievement motivation, and self-estimate of ability discriminate between examinees without omissions and examinees with at least three omitted items in an achievement test.

The present study is concerned with the third kind of missing data – that is, the case in which the examinee omits an item although there is enough time for responding (correctly, incorrectly or even by guessing the answer).

The classification of different types of missing data, by Rubin (1976), which differs between **m**issing data (**c**ompletely) **a**t **r**andom (MCAR respectively MAR) and **m**issing data **n**ot **a**t **r**andom (MNAR), is well known. Regarding this typology, the mechanism underlying the occurrence of a missing response within the definition of MCAR means that the missings are independent of all observed variables and the variable containing the missing itself, whereas MAR only states the independence of the missing from the variable containing the missing. In contrast, an MNAR mechanism means that the occurrence of a missing value depends on the variable containing the missing (cf. Schafer & Graham, 2002). MNAR missings usually cause most of the problems in data analysis and are therefore also called "non-ignorable" missings, whereas M(C)AR are "ignorable" (Schafer & Graham, 2002).

In practice, it is hard to avoid missing values in a data set. For treatment of missing values – the most commonly applied procedure is scoring missing values as incorrect; an alternative is to treat them as not administered items. Apart from these two methods, more complex approaches for the handling of missings have been developed, including

various imputation methods (cf. Schafer & Graham, 2002) and model-based approaches (Holman, & Glas, 2005; Rose, von Davier, & Xu, 2010).

The handling of missings is difficult and rather crucial, because inappropriate handling bears the risk of a bias in parameter estimation. Within the IRT framework, a considerable bias was found in the ability parameter estimation in the presence of missings (Kubinger, 1983; De Ayala, Plake & Impara, 2001; Rose, von Davier & Xu, 2010). The studies conclude that the handling of missing data as incorrect leads to a more severe bias than just handling them as though the respective items were not administered to the examinee. Additionally, Rose, von Davier and Xu (2010) studied different procedures for dealing with missing values in item parameter estimation and found a more severe bias for the handling of missings as incorrect, than for merely treating them as not administered items.

The present study will also investigate the impact of different dealing procedures with missing data, on parameter estimation, but will focus more on the consequences of the different procedures for model and item fit. In the development of psychological and educational tests, a common practice is to eliminate poor-fitting items, in order to establish an item pool which conforms (aposteriori) to the Rasch model (cf. Kubinger, 2005). To assess the global model validness for the total item pool, Andersen's Likelihood Ratio test (Andersen, 1973) can be applied, where the likelihood of the total sample is compared to the sum of the likelihoods of two sub-samples when the total sample is split according to a relevant criterion. With regard to each item, an item-specific test with standard-normally distributed test statistics can be applied. That is, the standardized item parameter estimates $\hat{\beta}$ for an item *i,* are compared in two sub-samples (*z*-test; see Fischer, 1974):

$$z_{\mathrm{i}} = \frac{\hat{\beta}_{i1} - \hat{\beta}_{i2}}{\sqrt{\mathrm{I}(\hat{\beta}_{i1}) + \mathrm{I}(\hat{\beta}_{i2})}} \tag{1}$$

In the formula, "I" denotes Fisher's information function of the item parameter estimate. Concerning item fit statistics, there are residual-based infit and outfit statistics, based on the difference between the observed response of examinee *n* to item *i* (x_{ni}) and the expected response (e_{ni}). Infit and outfit indices are standardized residuals: $z_{ni} = (x_{ni} - e_{ni}) / \sqrt{\mathrm{Var}(x_{ni})}$. The outfit mean-squared error (outfit MSQ) equals the averaged sum of squared residuals (Wright & Masters, 1990)

$$\mathrm{o}_i = \frac{\sum_{n=1}^{N} \mathrm{z}_{ni}^2}{N} \tag{2}$$

The infit mean-squared error (infit MSQ), on the other hand, weights the sum of squared residuals according to the variance of the response (Wright & Masters, 1990):

$$i_i = \frac{\sum_{n=1}^{N} \mathrm{Var}(x_{ni}) \cdot z_{ni}^2}{\sum_{n=1}^{N} \mathrm{Var}(x_{ni})} \tag{3}$$

The expected value for infit and outfit MSQ of an item pattern which conforms to the Rasch model is 1. According to Bond and Fox (2007) fit values exceeding 1.3 indicate underfit and values less than 0.75 indicate overfit.

On the basis of the above item fit measures, poor fitting items can be identified and deleted from the item pool. With regard to missing values, the question arises as to what extent different treatments of missing data influence global model validness as well as individual item validness or fit. If there is a considerable bias introduced by certain forms of missing data treatments, this could unfortunately lead to an item of adequate psychometric quality to be deleted.

In the present study, an empirical example will illustrate whether different alternatives for dealing with missing values result in different item pools; the differences in parameter estimations with regard to the two different alternatives will be presented. Due to the fact that the "true" parameter for empirical data is never known, a simulation study is carried out which also examines the impact of adequately dealing with missing data, on parameter estimation, model validness and item goodness of fit.

IRT analyses were conducted with the *R*-package *eRm* (Mair, Hatzinger & Maier, 2010).

Empirical data

Data set and sample

Data from a mathematical competence test for 7th grade students were analyzed. This test was developed by domain experts in cooperation with psychologists. The test consisted of 30 items assigned to four different subtests. For the following analysis, data of the subtest "Calculating and Operating" were used, consisting of 9 items. This subtest measures the competence to carry out simple calculations including operations using tables and graphs. The test was administered to a total of 524 students with all examinees receiving the same items. Therefore, by design, no missings should have appeared in the data.

Missing values and different kinds of treating missing values

Although there were no missings by design, it did occur that some examinees omitted individual responses to some items of the test. As pointed out in the introduction, several reasons are hypothesized for the omission of an item. Indication of a speed effect is provided by examining the number of missing values resulting relative to the position of the

item administered in the test booklet. If a speed effect would apply, then the number of missings resulting due to item position would increase. Figure 1 shows the percentage of missing values, comparing each item from two test booklets. As the position numbers for the items inserted in the figure indicate, there may be no obvious trend for any item to have a higher percentage of missing data if it is presented later in the test booklet. Furthermore, considering that the items of this subtest were presented at the beginning or middle of the whole test, a speeding effect as a reason for omission can be excluded.

Data were analyzed in two ways: firstly, missings were dealt with as incorrect (treatment alternative 1) and secondly, missing values were not scored at all (treatment alternative 2), meaning that these missings were handled as if they were not administered to the examinee, by not considering them in data analysis.

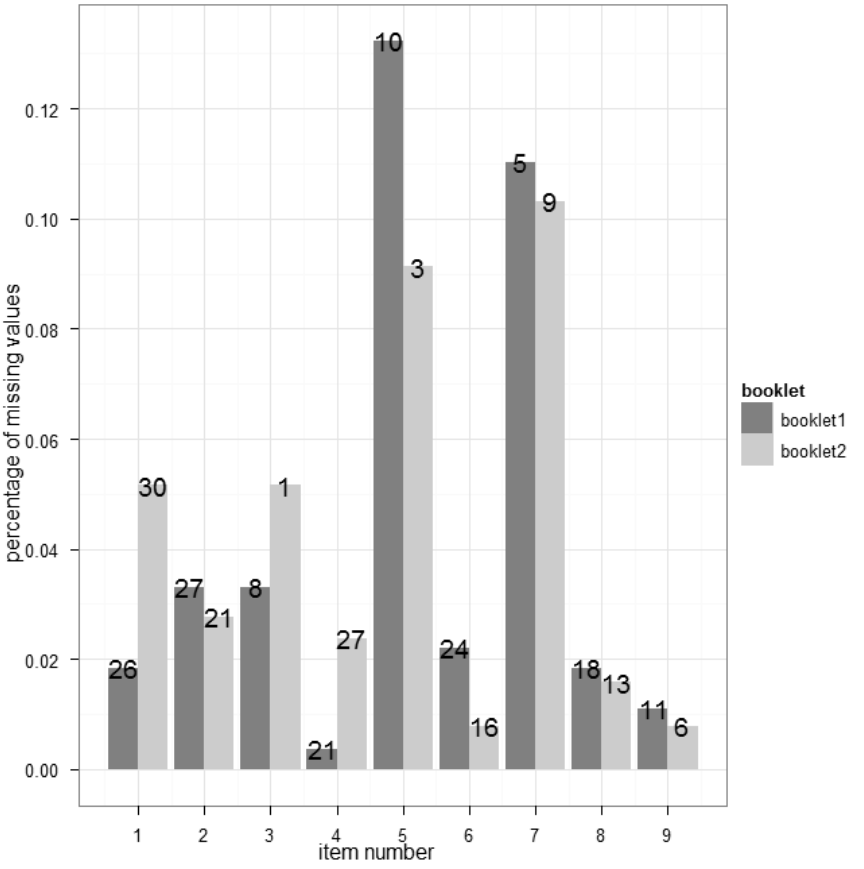

Figure 1:
Relative frequency of omitted items in each test booklet. Black numbers connected to the bars indicate the position of presentation of this item in the test form

Model validness and fit

The first step of the analysis was to test whether the mathematical competence test conforms to the Rasch model. To evaluate the test's global model validness, Andersen's Likelihood-Ratio test (Andersen, 1973) was conducted. The sample was split according to the following criteria: score, gender, native language, and regional district. If Andersen's test showed a significant result, the items were then individually tested by applying the item-specific *z*-test as given above, and their goodness of fit was examined by calculating the residual-based item fit indices. Global model validness (aposteriori) should be achieved by deleting poor-fitting items.

A comparison of the two missing data treatments resulted in a significant Likelihood-Ratio test statistic with regard to the split criterion gender (see Table 1). The *z*-test shows that item 7 is most deviant from model-conformity regardless of the kind of missing data treatment (see Table 2). In contrast, the infit statistic indicates no aberrant response patterns for any of the items. Outfit indices indicate a small overfit for item 4, which, however, cannot be explained. Hence, the results for the two kinds of treatment of missing data are very similar. Due to the severe aberrance in the *z*-test, indicating some differen-

Table 1:
Andersen's Likelihood Ratio test for both treatment alternatives. Critical χ^2-value is 20.09 (df= 8, α = 0.99)

Split Criterion	χ^2_{LRT} **(treatment alternative 1)**	χ^2_{LRT} **(treatment alternative 2)**
Score	6.76	5.80
Gender	27.26	30.87
Native language	6.68	5.64
Regional district	7.13	10.82

Table 2:
Item fit concerning the two treatment alternatives

Item number	**Treatment alternative 1**			**Treatment alternative 2**		
	***z*-test**	**Outfit MeanSQ**	**Infit MeanSQ**	***z*-test**	**Outfit MeanSQ**	**Infit MeanSQ**
1	-0.68	0.90	0.88	-0.484	0.92	0.86
2	0.78	0.81	0.87	0.81	0.81	0.87
3	2.98	1.05	1.01	2.97	1.04	0.98
4	0.95	0.55	0.84	0.61	0.58	0.84
5	-0.67	0.97	1.00	0.03	0.95	0.97
6	1.46	1.02	0.95	1.94	1.07	0.97
7	-3.67	0.93	0.90	-3.88	0.99	0.90
8	-0.51	0.97	0.91	-1.08	0.93	0.88
9	-0.71	0.93	0.97	-0.68	0.91	0.95

tial item functioning (DIF) with respect to gender for item 7, and because the model was not globally valid for the data containing this item, item 7 was deleted from the item pool. Global model validness for this reduced item pool was tested again (applying Andersen's Likelihood-Ratio test). With item 7 removed, the test showed a non-significant result. It was therefore concluded that this reduced item pool conforms to the Rasch model.

In order to observe the impact of the two missing data treatment alternatives on parameter estimation, estimated item and person parameters were compared. Figure 2 indicates that the item parameter estimates yield only small differences (situated within the range of the standard errors) between the two missing data treatments, whereas figure 3 discloses that the person parameter estimates result in clear differences for a certain subset of examinees. However, a considerable difference has been established for only a few examinees. That is, a difference in person parameter estimates of greater than 0.5 is observed for 47 examinees out of a total sample of N=524. Of course, each of these 47 examinees has omitted at least one item.

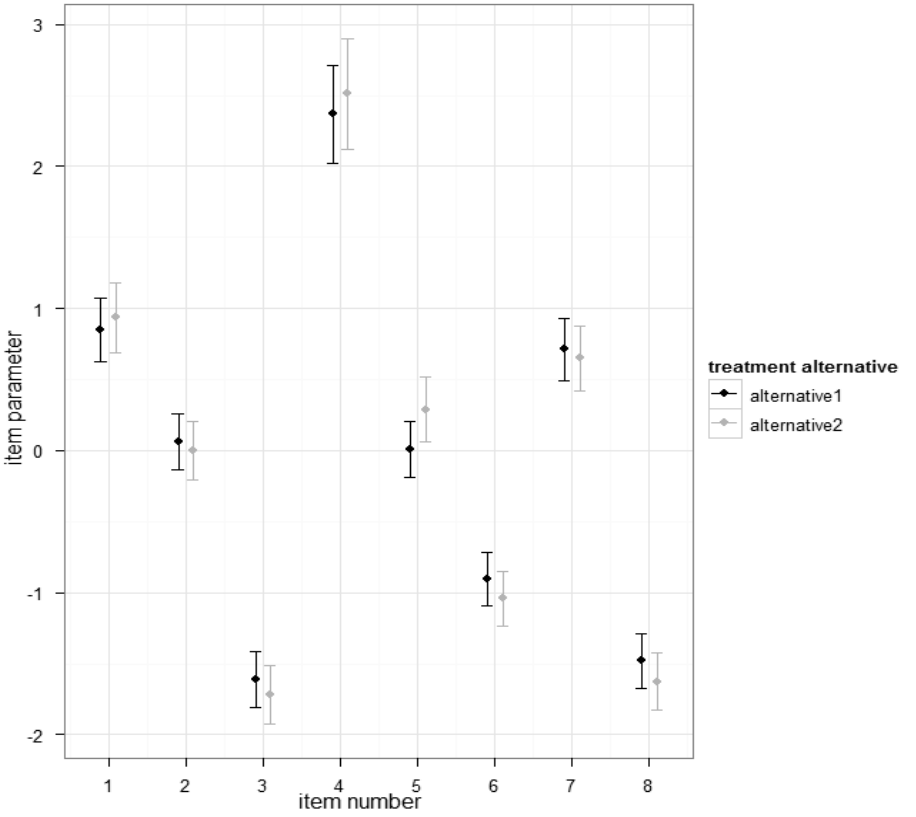

Figure 2:
Item parameter estimation applying the two treatment alternatives

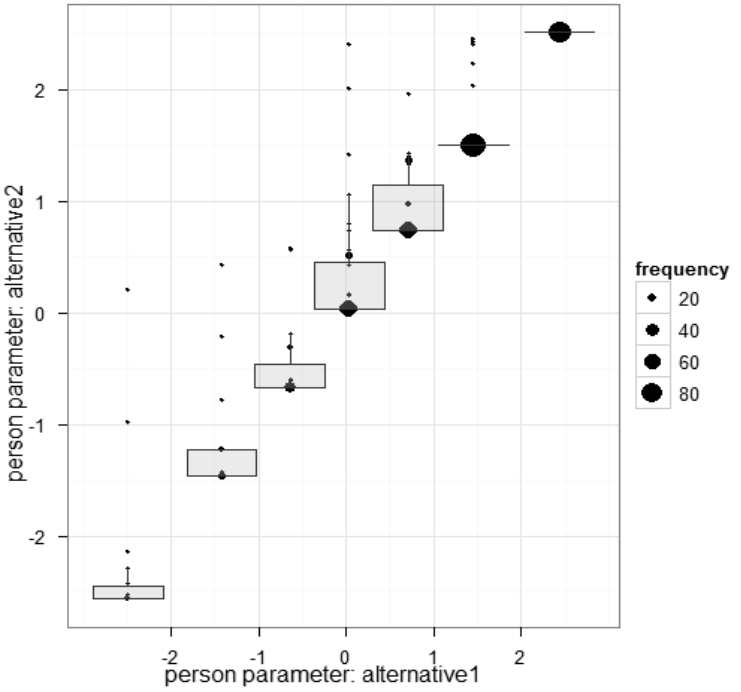

Figure 3:
Person parameter estimation applying the two treatment alternatives

Simulation

Data generation

In order to analyze the impact of the different procedures for dealing with missing values, data conforming to the Rasch model were generated. These data included no missing values and are denoted in the following as "original data". For all simulation scenarios, the number of items was fixed at $k = 15$ and the number of simulees was $N = 500$; these numbers represent commonly found test lengths for a psychological achievement (sub) test and are also seen frequently as required sample size for a psychological test calibration study with the Rasch model. Item parameters were placed equidistant in the interval [-3.5; 3.5]. In a next step, missing responses were introduced by making changes to the simulated complete data (the "original data") according to MNAR mechanism. That is, missings occur in such a way that their probability depends on the simulee's ability as well as on the item's difficulty. For the simulated data without missings, the theoretical probability p_{ij} for solving an item i was calculated for a simulee j. Subsequently, the probability for producing the case of a missing, m_{ij}, was assumed to exponentially de-

crease with a higher p_{ij} according to the following function: m_{ij} = 0.3*exp (-3*p_{ij}). This function limits the probability for missings within an interval, $0 < m_{ij} < 0.3$. The maximum of m_{ij} is reached at $p_{ij} = 0$. As mentioned in the introduction, previous research indicates that some personality variables appear to have an impact on the occurrence of omitted items. Therefore, it seems conclusive that omissions take place only for a subgroup of simulees. Three simulation scenarios were carried out, varying the number of simulees with missing values (percentages: 100, 50 and 25). For each scenario 1000 original data sets were generated and missing values were implemented subsequently. Simulation scenario S25 implemented missing data for only 25 percent of the sample, scenario S50 did this for half of the sample and S100 involved the whole sample in the missing implementation process. For each simulation, the number of missings for each simulee j was limited to a maximum of k-2. The replication number of each simulation scenario was set to $r = 1000$.

In order to compare the different missing data treatment alternatives, the implemented missing data were first scored as incorrect responses (treatment alternative 1) and were then considered as not administered (treatment alternative 2). Parameter estimation and calculations of the model's and item's validness and item fit, respectively, were conducted for the "originally" generated data, consisting of all "responses" as well as for the data with missing values and with subsequent application of the two different missing data treatments. It should be kept in mind that for each of the three simulation scenarios, separate "original data" were generated.

Results of simulations

For the simulation scenario S25, each data set had a mean of 208 missing responses (out of a total of 7500 responses), scenario S50 had a mean of 418 and scenario S100 a mean of 837 missings.

Initially, the item parameter reconstruction comparing the estimated item parameters with the given parameters for generating the data was of interest. It was analyzed whether the different methods for dealing with missing values affect the item parameter estimation and if so, to what extent. Therefore, the squared error (SQE) was calculated over all replications of each simulation scenario s:

$$\mathrm{SQE}(\hat{\beta}_s) = \sum_{i=1}^{k}\sum_{t=1}^{r}(\hat{\beta}_{it} - \beta)^2 \,. \tag{4}$$

Table 3 shows that the SQE increases with an ascending number of missing values for both kinds of missing data treatments. That is, for a higher number of missings not at random, the SQE of the item parameter recovery increases – which is an expected result. Furthermore, across all scenarios, the SQE is higher if treatment alternative 1 is applied in comparison to treatment alternative 2.

Andersen's Likelihood-Ratio test was applied to all datasets analyzed in this simulation study. The main purpose of this test is to determine whether the data conform to the

Rasch model or not. Due to the fact that data were generated according to the Rasch model, only $\alpha \cdot 100$ percent of the data matrices should yield a significant result (with α denoting the nominal type-I-error). However, we examined the number of significant results despite generating data according to the Rasch model which provides the actual type-I-error. Table 4 shows this actual type-I-error for each missing data treatment alternative, and for every simulation scenario. If no systematic bias occurs, the actual type-I-error should roughly equal the nominal α. According to Rasch and Guiard (2004) a test achieves 20%-robustness if $\alpha - 0.2 \cdot \alpha <$ actual type-I-error $< \alpha + 0.2 \cdot \alpha$, that is for the case of $\alpha = 0.05$, the actual type-I-error must lie in between the interval [0.04; 0.06]. For scenarios S25 and S50, the actual type-I-error lies within these boundaries. In scenario S100, the actual type-I-error exceeds these boundaries even for the original data set. The percentage of significant results for the data with missing responses treated as non-administered items is roughly the same as in the original data. On the contrary, for the treatment of missings as incorrect, the percentage of significant results falls below the lower boundary of the interval. The result showing that treatment alternative 1 leads to a smaller amount of significant results in S100 is also supported by the respective graph in figure 4.

In order to assess the impact of the missing treatments on the item fit measures, the squared error (SQE) was calculated again – using the known expected value of the infitMSQ and outfitMSQ – as well as the *z*-test. Regarding the latter, the calculation for some items was not possible in several replications because they were ill-conditioned (that is, in at least one of the respective sub-samples, some items have been solved either

Table 3:
SQE of item parameter reconstruction

Item parameter reconstruction	Original data (no missings included)	320.84	342.67	334.75
		S 100	S 50	S 25
	Treatment alternative 1	420.07	375.45	350.66
	Treatment alternative 2	367.03	362.09	346.57

Table 4:
Percentage of significant results for Andersen's Likelihood Ratio test (for nominal $\alpha = 0.05$)

	Simulation Scenario		
Original data (no missings included)		Treatment alternative 1	Treatment alternative 2
6.5	S 100	2.0	6.3
4.2	S 50	4.4	4.9
5.8	S 25	5.7	5.6

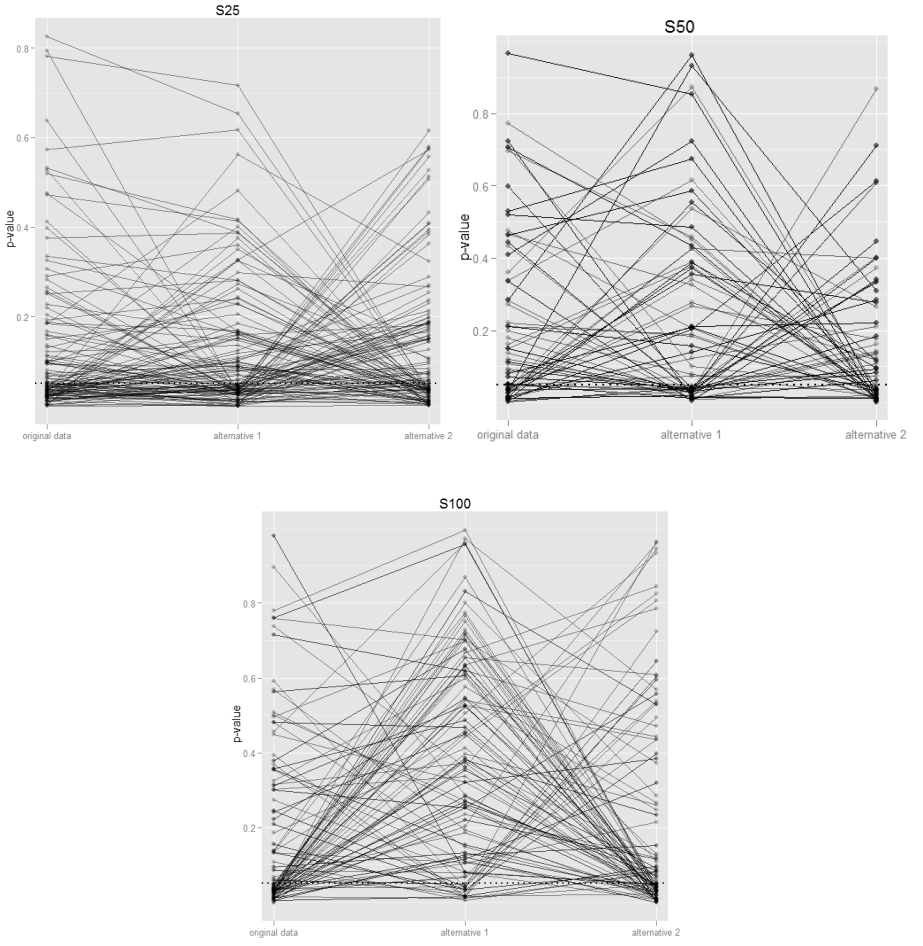

Figure 4:
p-values of Andersen's Likelihood Ratio Test for those data which yielded a significant Likelihood Ratio test in at least one of the three data conditions. ($\alpha = 0.05$)

always or never). Because the *z*-test statistic could not be computed in each replication, the SQE for the *z*-test statistic was divided by the number of estimatable items, to ensure comparability. The results are presented in Table 5. A comparison of the *z*-test resulting from the original data matrix, with those from the data matrices with missing responses, shows that the SQE is higher for treatment alternative 1 than for alternative 2. For outfit MeanSQ and infit MeanSQ, the differences are not the homogeneous and distinct. However, the SQE of the outfit MeanSQ for the two treatment alternatives is higher than the SQE of the original data. For S100 the SQE for treatment alternative 1 is higher than for

z-test	Original data (no missings included)	14.12	14.33	14.29
		S 100	**S 50**	**S 25**
	Treatment alternative 1	16.95	15.59	14.60
	Treatment alternative 2	14.14	13.96	13.92
Outfit MeanSQ	Original data (no missings included)	737.90	706.03	714.94
		S 100	**S 50**	**S 25**
	Treatment alternative 1	922.03	749.53	716.93
	Treatment alternative 2	869.75	749.99	733.93
Infit MeanSQ	Original data (no missings included)	198.40	198.77	198.81
		S 100	**S 50**	**S 25**
	Treatment alternative 1	196.71	195.71	196.45
	Treatment alternative 2	225.97	211.79	204.69

Table 5:
SQE for item fit measures summed up over all replications and items. For the z-test, the SQE was divided by the number of calculations due to the fact that, for several replications, the z-test could not be calculated for single items because this item was ill-conditioned in one subsample

treatment alternative 2. In contrast for the infit MeanSQ higher SQE's result for treatment alternative 2 in comparison to treatment alternative 1.

Furthermore the outfit MeanSQ shows generally much more deviances than the infit MeanSQ – even for the original data. That is the outfit MeanSQ seems too sensitive for the data.

Discussion

In previous studies the impact of different procedures for dealing with missing data, on parameter estimation (especially person parameter estimates), was examined. The present study aimed to examine the impact of two such methods on the fit of the Rasch model validness as well as specific item fit measures.

Similar results concerning model validness and item fit were observed when two alternatives for dealing with missing data (treating as incorrect or treating missing data as not administered) were applied for the analysis of an empirical data set of a mathematical competence test. Both treatment alternatives for missing data lead to the same (aposteriori) Rasch model conforming item pool. Furthermore, item parameter estimates do not show considerable differences between the two missing data treatments. Of course, the "true" item parameters for the empirical data were unknown.

Therefore a simulation study was conducted containing three scenarios varying the amount of missings in the data set. Treatment alternative 1 yields the highest SQE for the item parameter reconstruction. That is, if missings are treated as incorrect responses, then the item parameters are estimated with more bias than if treatment alternative 2 (or for the original data set) is used. This kind of dealing with missing values also affects the result of Andersen's Likelihood-Ratio test in such a way that the results of the global model fit differ. The global model test seems more affected if the amount of missing values is high (scenario S100). In this scenario, treating missing data as incorrect leads to a considerably lower amount of significant results than when treating them as not administered items. Regarding this result in the context of the higher item mis-reconstruction for treatment alternative 1, it is quite reasonable that results found for the model test using treatment alternative 1 show a drift. In any case, (for S100 and treatment alternative 1) Andersen's Likelihood Ratio shows too little power. We do not have a compelling explanation, but it appears as though the test's power is reduced for treatment alternative 1 because splitting the sample in high vs. low scorers (when Andersen's Likelihood Ratio test is applied) does not polarize the subsamples to such an extent as it does for the original data; and this is because high ability examinees very often belong to the low scorers due to their missing values being scored as incorrect. Of course, this problem does not arise with treatment alternative 2, when examinees are split into high vs. low scorers according only to basis of the items they responded to. Beside, in this scenario the actual type-I-risk exceeds the nominal α even for the original data set, probably due to chance.

With regard to the *z*-test, treatment alternative 1 shows the highest SQE. However, infit MeanSQ and outfit MeanSQ are less affected by the kind of dealing with missing values.

To summarize, model and item test are more adversely affected by treating missing values as incorrect and less when they are treated as not administered. This result is in concordance with previous research (De Ayala, Plake, & Impara, 2001, Rose, von Davier, & Xu, 2010) which found that treating missing values as incorrect has a stronger influence on item and person parameter estimates.

References

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38,* 123-140.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model*. New Jersey: Lawrence Erlbaum.

De Ayala, R. J., Plake, B. S., & Impara, J. C. (2001). The Impact of Omitted Responses on the Accuracy of Ability Estimation in Item Response Theory. *Journal of Educational Measurement, 38,* 213-234.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests* [Introduction to test theory]. Bern: Huber.

Holman, R., & Glas, C. A. W. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology, 58*, 1-17.

Kubinger, K. D. (1983). Anhang: Einige besondere testtheoretische Belange [Appendix: some special psychometric concerns]. In K. D. Kubinger (ed.), *Der HAWIK – Möglichkeiten und Grenzen seiner Anwendung* (pp. 229-235) [The German WISC – potentials and shortcomings of its application]. Weinheim: Beltz, 229-235.

Kubinger, K. D. (2005). Psychological Test Calibration using the Rasch Model – Some Critical Suggestions on Traditional Approaches. *International Journal of Testing, 5,* 377-394.

Mair, P., & Hatzinger, R., & Maier, M. (2010). eRm: extended Rasch modeling. R package version 0.13-0: http://cran.r-project.org/web/packages/eRm/

Matters, G., & Burnett, P. C. (2003). Psychological Predictors of the propensity to omit short-response items on a high-stakes achievement test. *Educational and Psychological Measurement, 63,* 239-256.

Rasch, D., & Guiard, V. (2004). The Robustness of Parametric Statistical Methods. *Psychology Science, 46*, 175-208.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581-592.

Rose, N., von Davier, M., & Xu, X. (2010). *Modeling Nonignorable Missing Data With Item Response Theory (IRT)*. (ETS Research Report ETS RR-10-11.) Princeton, NJ: Educational Testing Service.

Schafer, L. S., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, 7,* 142-177.

Wright, B. D., & Masters, G. N. (1990). Computation of OUTFIT and INFIT Statistics. *Rasch Measurement Transactions, 3*, 84-85.