# NEW MODEL-DATA FIT INDICES FOR ITEM RESPONSE THEORY (IRT): AN EVALUATION AND APPLICATION

BY

LIWEN LIU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Fritz Drasgow, Chair
Professor Hua-hua Chang
Professor Brent Roberts
Assistant Professor Nichelle Carpenter
Associate Professor Daniel Newman

# ABSTRACT

I reviewed the recently developed limited-information model fit statistics by Maydeu-Olivares and colleagues (e.g., Maydeu-Olivares & Joe, 2005; Maydeu-Olivares & Liu, 2012; Liu & Maydeu-Olivares, 2014) and conducted a simulation study to explore the properties of these new statistics under conditions often seen in practice. The results showed that the overall and piecewise fit statistics were to some extent sensitive to misfit caused by multidimensionality, although the limited-information fit statistics tended to flag more item pairs as misfit than the heuristic fit statistics. I also applied the fit statistics to three AP® exams, one personality inventory, and a rating scale used in organizational settings. Although a unidimensional IRT model was expected to fit the Physics B Exam better than the English Literature Exam, the average piecewise fit statistics showed no such difference. The fit statistics also suggested that a more advanced IRT model should be fitted to the self-rated personality inventory. Finally, the fit statistics seemed to be effective in detecting misfit caused by data skewness.

## TABLE OF CONTENTS

INTRODUCTION ………………………………………………………………………………. 1

METHOD ……………………………………………………………………………….... 24

RESULTS ……………………………………………………………………………… 30

DISCUSSION ………………………………………………………………………….. 53

CONCLUSION ……………………………………………………………………… 63

REFERENCES …………………………………………………………………...……. 64

TABLES ………………………………………………………………………………..… 68

FIGURES ……………………………………………………………………………... 110

APPENDIX A: INTERNATIONAL PERSONALITY ITEM POOL (IPIP) ………………… 128

APPENDIX B: COUNTER-PRODUCTIVE WORK BEHAVIOR (CWB) SCALE ……...… 130

# INTRODUCTION

Item response theory (IRT) is becoming the psychometric model of choice for analyzing large-scale assessments due to its statistical strengths. For example, classical test theory statistics are sample dependent, which means their values are determined by both the specific items included in a test and the specific group of examinees who take the test. By contrast, IRT item and examinee parameters are invariant across subpopulations (Embretson & Reise, 2000). That is, item parameters do not depend on the specific group tested, and ability score estimates can be computed from different sets of items for which properties are known. This allows researchers to conduct rigorous tests of item bias across groups and to compute test scores for computerized adaptive tests. Therefore, IRT can be a very useful approach for test analysis, especially on large-scale assessments.

The statistical strength of IRT is however based on important mathematical assumptions, and these assumptions must be rigorously examined. Moreover, even if the usual assumption of unidimensionality is met, model-data fit still needs to be evaluated to see whether the IRT model used to fit the data can describe the data well. Therefore, checking model assumptions and assessing model-data fit are important procedures to justify the use of IRT models.

There are numerous approaches to assessing model fit, and they are generally categorized into two groups: (a) directly checking the fundamental assumptions of IRT such as unidimensionality and local independence, and (b) examining the fit between observed scores and model predicted scores, which indicates whether the unidimensionality and local independence assumptions are violated (Swaminathan, Hambleton, & Rogers, 2007). In this paper I focused on the model-data fit approaches in the second group. To be specific, I reviewed the literature on the recently developed limited-information fit statistics and the traditional heuristic fit statistics. I also conducted several studies to compare the performance of these fit statistics.

## IRT Assumptions and Models

There are two fundamental assumptions of the most common IRT models: unidimensionality and local independence (Embretson & Reise, 2000). The IRT model is said to be unidimensional if the minimum dimension of the latent trait is one. One should always evaluate the dimensionality of the data before proceeding to perform IRT analyses. Local independence means that conditioned on an examinee's ability, item responses should be statistically independent. That is, controlling for an examinee's ability, the probability that the examinee answers one item correctly should not correlate with the probability of a correct response to another item.

There are numerous IRT models for binary responses (i.e., dichotomous) or responses with multiple categories (i.e., polytomous). For multiple choice (MC) items with either correct or incorrect answers, the three-parameter logistic model (3PLM; Birnbaum, 1968) is usually used:

$$P(u_i = 1|\theta = t) = c_i + \frac{1-c_i}{1+\exp[-1.7a_i(t-b_i)],} \qquad (1)$$

where $u_i$ is the response of the examinee with ability level $\theta$ to item *i*, $a_i$ is the item discrimination parameter, $b_i$ is the item difficulty parameter, $c_i$ is the "pseudo-guessing" parameter, and *1.7* is a scaling constant. The 3PLM is appropriate for cognitive ability tests when examinees with low ability levels can occasionally respond correctly to difficult items by guessing. When guessing is not a concern, such as for personality data, a two-parameter logistic model (2PLM) can be used due to its simplicity and some evidence of model fit:

$$P(u_i = 1|\theta = t) = \frac{1}{1+\exp[-1.7a_i(t-b_i)],} \tag{2}$$

which is the 3PLM when the pseudo-guessing parameter is zero.

For items with multiple ordered categories, Samejima's Graded Response Model (SGRM; Samejima, 1969) is a popular choice. It uses two-parameter logistic response functions to model the probability of selecting option *k* on item *i*

$$P(v_i = k|\theta = t) = \frac{1}{1+\exp\left[-1.7a_i\left(t-b_{i,k}\right)\right]} - \frac{1}{1+\exp\left[-1.7a_i\left(t-b_{i,k+1}\right)\right]}, \tag{3}$$

where $v_i$ is the examinee's response to the polytomous item *i*, *k* is the response option of item *i* selected by the examinee, $a_i$ is the item discrimination parameter, $b_{ik}$ is threshold parameter for option *k*, and *1.7* is a scaling constant.

## Traditional Approaches to Assessing the Model-Data Fit of IRT Models

Model-data fit can be evaluated by various goodness-of-fit indices, which all focus on the agreement between observed and predicted responses. Historically, the chi-square test of goodness of fit is probably the most frequently used index for such comparisons. The Pearson $\chi^2$ statistics can be written as

$$X^2 = \sum_{k=1}^{s} \frac{[O_i(k) - E_i(k)]^2}{E_i(k)}, \tag{4}$$

where *s* is the number of options for an item, $O_i\,(k)$ is the observed frequency of endorsing option *k*, and $E_i\,(k)$ is the expected frequency of option *k* under the dichotomous or polytomous IRT model. The expected frequency of a correct response to an individual item can be written as

$$E_i(k) = N \int P(v_i = k \mid \theta = t)\, f(t)dt\,, \tag{5}$$

where $f(t)$ is the probability density function of the latent trait.

Although the chi-square goodness of fit seems to be the most natural method to assess the agreement between observed and expected responses, it has some important limitations. For example, the chi-square statistic is sensitive to sample size and the test at the individual item level is insensitive to certain types of model misfits (Van den Wollenberg, 1982). To address these problems, Drasgow and colleagues developed an improved method of computing the $\chi^2$ statistics: the adjusted $\chi^2$ statistic divided by its degree of freedom (*df*). As the $\chi^2$ statistics for individual items allow compensation between local misfits, they are also computed for item pairs and triples. The expected frequency for an item pair in the $(k, k')^{th}$ cell of the two-way contingency table for item *i* and *i'* can be computed as

$$E_{i,i'}(k, k') = N \int P\,(v_i = k \mid \theta = t)\, P\,(v_{i'} = k' \mid \theta = t)\, f(t)dt, \tag{6}$$

and the observed frequencies are counted in each cell. A similar procedure can be performed for item triplets using a multiway contingency table. These expected frequencies are combined with the observed frequencies to produce a $\chi^2$ statistic. To ensure the singles, doubles, and triples $\chi^2$ statistics are comparable across different sample sizes, they are adjusted to what would be expected in a sample size of 3000 and then divided by their degrees of freedom:

$$\text{Adjusted } \chi^2_{3000} = [df + 3000(\chi^2 - df)/N]/df. \tag{7}$$

Drasgow, Levine, Tsien, Williams and Mead (1995) suggested that values of adjusted $\chi^2/df$ smaller than 3.0 indicate good model-data fit.

However, there are a few limitations of the adjusted chi-square statistic as well. Recently, a problem was detected in a simulation study which showed that adjusted $\chi^2/df$ statistics were affected by the sample size used for estimation, and negative values may be obtained after an adjustment to a sample size of 3000 (Guo, Tay, & Drasgow, 2010). This problem is especially prominent in small samples. Therefore, the adjusted $\chi^2/df$ statistics should be applied to large data sets with sample sizes of 3000 or more. Moreover, like many other methods for checking model-data fit for IRT models, these approaches have been based on heuristics and, consequently, lack distribution theory to inform us as to what is a "large" misfit versus inconsequential misfit. For example, a major problem with the adjusted $\chi^2/df$ is that it does not follow a chi-square distribution and this approach does not account for the number of parameters estimated. Although dividing the chi-square statistic by its degree of freedom tries to address the number of parameters issue, it is still a heuristic adjustment.

Another common approach to assessing model-data fit is the likelihood-ratio statistic. The likelihood-ratio statistic can be written as:

$$G^2 = 2\sum_{k=1}^{s} O_i\,(k) \log\frac{O_i\,(k)}{E_i(k)} \tag{8}$$

When the model holds, the likelihood-ratio statistic has an asymptotic chi-square distribution. Otherwise, $\chi^2$ and $G^2$ statistics can have very different values (Agresti, 2002). Moreover, when the expected frequency in each cell is smaller than 5, $\chi^2$ and $G^2$ statistics do not have the expected Type I error rates under their asymptotic distribution. Although both statistics are affected by such sparseness of the contingency table, $\chi^2$ is less affected than $G^2$ (Koehler & Larntz, 1980).

To overcome the problems caused by the sparseness of a contingency table, three approaches have been proposed: pooling cells, resampling methods, and limited information methods (Maydeu-Olivares & Joe, 2006; Maydeu-Olivares, 2013). Pooling cells is the most intuitive approach such that reducing the number of cells in a contingency table automatically increases the expected frequency in most if not all the cells. However, to obtain a statistic with the appropriate asymptotic reference distribution, pooling must be performed *before* the model is fitted. Secondly, empirical sampling distributions of the goodness-of-fit statistics can be generated with a resampling method (e.g., bootstrapping) to produce supposedly trustworthy *p*-values. However, mixed results have been found on the accuracy of *p*-values for the $\chi^2$ and $G^2$ statistics obtained by resampling methods (e.g., Tollenaar & Mooijaart, 2003; von Davier, 1997). Moreover, the resampling method can be very time-consuming if the fit of multiple models needs to be obtained for comparison purposes. Finally, Maydeu-Olivares and colleagues introduced a variety of new model-fit statistics that are based on limited information methods. This approach is similar to pooling cells a priori by using lower-order margins, such as univariate and bivariate probabilities and proportions. Their *p*-values were found to be accurate even for very large models with very small sample sizes. Compared with the heuristic approaches, these new statistics have the degrees of freedom correctly determined and the correct sampling distribution to examine model-data fit. A detailed review of these new statistics is provided in the next section.

## Full- and Limited- Information Statistics for Overall Fit

### Full-Information Statistics

$\chi^2$ and $G^2$ are considered full-information statistics because they use all the information in the contingency table to test the model. That is, the discrepancy between estimated probabilities and sample proportions is examined for *every* cell. The following notations are used throughout this section: *n* is the number of items, *K* is the number of response categories for each item, and *N* is the sample size. The observed *N* responses to these items can generate a contingency table with $C = K^n$ cells. Let $\mathbf{p}$ and $\boldsymbol{\pi}$ denote the *C* dimensional vectors of *observed* proportions and *expected* probabilities, respectively; and let $\boldsymbol{\pi}(\boldsymbol{\theta})$ indicate that $\boldsymbol{\pi}$ has some parametric form that depends on *q* parameters, $\boldsymbol{\theta}$, estimated from the data. To test the null hypothesis $H_0$: $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$ against the alternative hypothesis $H_1$: $\boldsymbol{\pi} \neq \boldsymbol{\pi}(\boldsymbol{\theta})$, Pearson's chi-square statistic can be evaluated in its matrix form:

$$X^2 = N\,(\mathbf{p} - \hat{\boldsymbol{\pi}})'\hat{\mathbf{D}}^{-1}(\mathbf{p} - \hat{\boldsymbol{\pi}}), \qquad (9)$$

where $\hat{\boldsymbol{\pi}}$ and $\hat{\mathbf{D}}$ denote $\boldsymbol{\pi}(\boldsymbol{\theta})$ and $\mathbf{D} = diag(\boldsymbol{\pi}(\boldsymbol{\theta}))$ evaluated at the parameter estimate, $\hat{\boldsymbol{\theta}}$, respectively. When $\hat{\boldsymbol{\theta}}$ is the maximum likelihood (ML) estimator, $\chi^2$ asymptotically follows a chi-square distribution with $C - q - 1$ degrees of freedom. The chi-square statistic can be used for an overall evaluation of all items in a test, but at least three items are required for dichotomous models to maintain a positive degree of freedom. For example, for a pair of dichotomous items, $C = 4$; but if the 2PLM is used, $q = 4$. As a result, $\chi^2$ for the item pair is meaningless due to the negative degrees of freedom (i.e., 4 – 4 – 1).

When $\hat{\boldsymbol{\theta}}$ is not estimated using the ML estimator or any asymptotically optimal estimator, the chi-square distribution is not the correct reference distribution for $\chi^2$ (Maydeu-Olivares & Liu, 2012). Instead, the $M_n$ statistic introduced by Maydeu-Olivares and Joe (2005) has the same reference distribution as Pearson's $\chi^2$ for non-optimal estimators. $M_n$ can be written as

$$M_n = N\,(\mathbf{p} - \hat{\boldsymbol{\pi}})'\hat{\mathbf{U}}(\mathbf{p} - \hat{\boldsymbol{\pi}}), \quad \mathbf{U} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\Delta(\Delta'\,\mathbf{D}^{-1}\,\Delta)^{-1}\,\Delta'\,\mathbf{D}^{-1} \qquad (10)$$

where $\Delta = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$ is a $C \times q$ matrix of derivatives of the probability of each response pattern with respect to each of the model parameters. $M_n = \chi^2$ for the ML estimator; otherwise $M_n < \chi^2$. Similar to $\chi^2$, testing with the $M_n$ statistic must involve at least item triplets for dichotomous models and item pairs for polytomous models to maintain a positive degree of freedom.

In addition to the overall fit assessment, $M_n$ can also be used to examine piecewise fit for item pairs and triples, as long as its degree of freedom is positive. By contrast, $\chi^2$ should not be applied to such subtables because it has an asymptotic chi-square distribution only when the parameter estimates are optimally estimated from data in that subtable rather than from data in the entire table. Therefore, for assessing the source of misfit by examining data in subtables, $M_n$ is preferred to $\chi^2$ (Maydeu-Olivares & Liu, 2012).

**Limited-Information Statistics**

The dimension *C* of observed proportions and expected probabilities depends on the number of cells, which further depends on the number of items and the number of categories for each item because $C = K^n$. Even for binary data, the number of cells equals $2^n$ and it does not require many items before the $2^n$ contingency table becomes too sparse even with a reasonable sample size (e.g., $N = 500$). To solve the problem caused by sparseness in the contingency table, Maydeu-Olivares and Joe (2005) proposed a family of statistics relying on the limited-information method, which focuses only on the information contained in the lower-ordered margins of the contingency table. Different from the full-information estimation which uses observed proportions and expected probabilities in each cell, limited-information estimation uses only the lower-order moments such as the univariate and bivariate moments. In this way, the frequency in each cell is aggregated, which reduces sparseness. Full-information statistics, on the other hand, use *all* moments up to the order *n*, which is equivalent to using information in *each* cell.

The following example demonstrates the difference in the information summarized from the contingency table with full- versus limited-information methods. Suppose $Y_1$ and $Y_2$ are the responses of two items each with a dichotomous outcome (i.e., 0 or 1). The responses then yield a 2 × 2 contingency table:

| | $Y_2 = 0$ | $Y_2 = 1$ |
|---|---|---|
| $Y_1 = 0$ | $\pi_{00}$ | $\pi_{01}$ |
| $Y_1 = 1$ | $\pi_{10}$ | $\pi_{11}$ |

For the full-information method, the table can be characterized by the cell probabilities $\pi' = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$. By contrast, the limited-information method summarizes the information from the contingency table with the univariate $\dot{\pi}_1{}' = (\pi_{1}^{(1)}, \pi_{2}^{(1)})$ and bivariate $\dot{\pi}_2{}' = \left(\pi_{1\ 2}^{(1)(1)}\right)$ probabilities as follows.

| | $Y_2 = 0$ | $Y_2 = 1$ | |
|---|---|---|---|
| $Y_1 = 0$ | | | |
| $Y_1 = 1$ | | $\pi_{1\ 2}^{(1)(1)}$ | $\pi_{1}^{(1)}$ |
| | | $\pi_{2}^{(1)}$ | |

The elements of $\dot{\pi}_1{}'$ and $\dot{\pi}_2{}'$ are univariate and bivariate moments if the variables are binary, because Pr(Y = 1) = E(Y) and Pr($Y_i$ = 1, $Y_j$ = 1) = E($Y_i$ $Y_j$). They are also moments for polytomous items when the indicator variables are used to denote each category except the zero category (see the next section for more details). When all items have the same number of categories, *K*, there are *n*(*K* – 1) univariate moments $\dot{\pi}_1{}'$ and $\binom{n}{2}(K-1)^2$ bivariate moments $\dot{\pi}_2{}'$.

The relationship between the moments $\dot{\boldsymbol{\pi}}$ and cell probabilities $\boldsymbol{\pi}$ can be written as

$$\dot{\boldsymbol{\pi}} = \mathbf{T}_{nc}\boldsymbol{\pi} \tag{11}$$

where $\mathbf{T}_{nc}$ is a (*C* – 1) × *C* matrix of 1's and 0's. The (*C* – 1)-dimensional vector $\dot{\boldsymbol{\pi}}$ includes its joint moment, $\dot{\boldsymbol{\pi}}' = (\dot{\boldsymbol{\pi}}'_{\mathbf{1}}, \dot{\boldsymbol{\pi}}'_{\mathbf{2}}, \ldots, \dot{\boldsymbol{\pi}}'_{\boldsymbol{n}})$, where $\dot{\boldsymbol{\pi}}'_{\mathbf{1}} = (\dot{\pi}_1, \dot{\pi}_2, \ldots, \dot{\pi}_n)'$, $\dot{\boldsymbol{\pi}}'_{\mathbf{2}}$ is the $\binom{n}{2}$-dimensional vector of the bivariate moment with elements E($Y_i$ $Y_j$) = Pr($Y_i$ = 1, $Y_j$ = 1) = $\dot{\pi}_{ij}$, and so on, up to $\dot{\boldsymbol{\pi}}_{\boldsymbol{n}}$ = Pr($Y_1$ = … = $Y_n$ = 1). For the 2 × 2 contingency table illustrated above, the relationship between the *C* – 1 vector of moments $\boldsymbol{\pi}_{\boldsymbol{n}}$ and cell probabilities $\boldsymbol{\pi}$ can be written as:

$$\begin{pmatrix} \pi^{(1)}_{1} \\ \pi^{(1)}_{2} \\ \pi^{(1)(1)}_{1\;\;2} \end{pmatrix} = \begin{pmatrix} 0\ 1\ 0\ 1 \\ 0\ 0\ 1\ 1 \\ 0\ 0\ 0\ 1 \end{pmatrix} \begin{pmatrix} \pi_{00} \\ \pi_{10} \\ \pi_{01} \\ \pi_{11} \end{pmatrix} \tag{12}$$

**T** can be partitioned based on the partition of $\dot{\boldsymbol{\pi}}$ in (11):

$$\begin{pmatrix} \dot{\pi}_1 \\ \dot{\pi}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{\pi}_n \end{pmatrix} = \begin{pmatrix} \mathbf{T}_{\boldsymbol{n1}} \\ \mathbf{T}_{\boldsymbol{n2}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{T}_{\boldsymbol{nn}} \end{pmatrix} \boldsymbol{\pi} \tag{13}$$

Then, the vector of joint moments up to order of $r \leq n$, denoted by $\boldsymbol{\pi_r}$ = ($\dot{\boldsymbol{\pi}}'_{\mathbf{1}}$, $\dot{\boldsymbol{\pi}}'_{\mathbf{2}}$, …, $\dot{\boldsymbol{\pi}}'_{\boldsymbol{r}}$)', can be written as

$$\boldsymbol{\pi_r} = \mathbf{T}_{\boldsymbol{r}}\,\boldsymbol{\pi} \tag{14}$$

where $\mathbf{T}_{\boldsymbol{r}} = (\mathbf{T}'_{\boldsymbol{n1}}, \ldots, \mathbf{T}'_{\boldsymbol{nr}})$' . Let $\mathbf{p}$ and $\dot{\mathbf{p}}$ denote the vector of observed cell proportions and the vector of sample joint moments, respectively. According to Maydeu-Olivares and Joe (2005), for a random sample of size *N* from the multivariate Bernoulli distribution,

$$\sqrt{N}\,(\dot{\mathbf{p}} - \dot{\boldsymbol{\pi}}) = \mathbf{T}\,\sqrt{N}\,(\mathbf{p} - \boldsymbol{\pi}). \tag{15}$$

The multivariate central limit theorem (Rao, 1973, p. 128) implies

$$\sqrt{N}\,(\mathbf{p} - \boldsymbol{\pi}) \overset{d}{\rightarrow} \mathrm{N}(\mathbf{0}, \boldsymbol{\Gamma}), \tag{16}$$

where $\boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\pi}\boldsymbol{\pi}'$ and $\mathbf{D} = diag(\boldsymbol{\pi})$. According to the delta method (Agresti, 1990, p. 579), it follows from (16) that

$$\sqrt{N}\,(\dot{\mathbf{p}} - \dot{\boldsymbol{\pi}}) \overset{d}{\rightarrow} \mathrm{N}(\mathbf{0}, \boldsymbol{\Xi}), \qquad \boldsymbol{\Xi} = \mathbf{T}\,\boldsymbol{\Gamma}\,\mathbf{T}'. \tag{17}$$

Finally, let $\mathbf{p}_r$ be the vector of sample moments up to order *r*, with dimension $s = s(r) = \sum_{i=1}^{r}\binom{n}{i}$. Then we have

$$\sqrt{N}\,(\mathbf{p}_r - \boldsymbol{\pi}_r) \overset{d}{\rightarrow} \mathrm{N}(\mathbf{0}, \boldsymbol{\Xi}_r), \qquad \boldsymbol{\Xi}_r = \mathbf{T}_r\boldsymbol{\Gamma}\mathbf{T}'_{\boldsymbol{r}}. \tag{18}$$

This leads to the overall $M_r$ statistic for both the dichotomous and polytomous models, $r$ = 1, 2, …, *n*, which can be written as

$$M_r = N\,(\mathbf{p}_r - \hat{\boldsymbol{\pi}}_r)'\,\hat{\mathbf{C}}_r\,(\mathbf{p}_r - \hat{\boldsymbol{\pi}}_r), \quad \mathbf{C}_r = \boldsymbol{\Xi}_r^{-1} - \boldsymbol{\Xi}_r^{-1}\,\boldsymbol{\Delta}_r\,(\boldsymbol{\Delta}'_r\,\boldsymbol{\Xi}_r^{-1}\boldsymbol{\Delta}_r)^{-1}\boldsymbol{\Delta}'_r\boldsymbol{\Xi}_r^{-1} \tag{19}$$

where $\mathbf{p}_r$ are the sample moments up to order *r*, $N\boldsymbol{\Xi}_{\mathrm{r}}$ is their asymptotic covariance matrix as shown in (17), and $\hat{\boldsymbol{\pi}}_r$ are the expected moments. $\boldsymbol{\Delta}_r = (\partial\boldsymbol{\pi_r}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}'$is the matrix of derivatives of the moments with respect to the model parameters. When *r* = 2, $M_2$ is a weighted average of the residuals in all bivariate tables that involve the univariate and bivariate probabilities. $M_2$ is asymptotically distributed as chi-square with a *df* of $n(K–1) + n(n–1)/2(K – 1)^2 – q$.

As can be seen, $\{M_r\}$ is a family of test statistics based on residuals up to the *r*-variate margins with members of $\{M_{1,} M_{2,} \ldots, M_n\}$. $M_1$ is a quadratic form in univariate residuals, whereas $M_2$ is a quadratic form in univariate and bivariate residuals, and so on, up to $M_n$. $M_n$ is a full information statistic that can be written as

$$M_n = N\,(\dot{\mathbf{p}} - \dot{\boldsymbol{\pi}})'\,\hat{\mathbf{C}}_n\,(\dot{\mathbf{p}} - \dot{\boldsymbol{\pi}}), \tag{20}$$

where $\dot{\mathbf{p}}$ is the vector of the sample joint moments and $\dot{\boldsymbol{\pi}} = \boldsymbol{\pi_n}$. Maydeu-Olivares and Joe (2005) show that $M_n$ can be alternatively written as a quadratic form in the cell residuals as in (10).

## Choice of Test Statistics

The $M_2$ statistic is just one of the test statistics that use the quadratic form to test the overall goodness-of-fit with only the bivariate information. Alternatively, a quadratic form statistic can be constructed as:

$$Q = N\,(\mathbf{p}_2 - \hat{\boldsymbol{\pi}}_2)'\,\widehat{\mathbf{W}}\,(\mathbf{p}_2 - \hat{\boldsymbol{\pi}}_2), \tag{21}$$

where $\widehat{\mathbf{W}}$ is a real symmetric weight matrix that depends on the model parameters but converges in probability to some constant matrix: $\widehat{\mathbf{W}} \overset{p}{\rightarrow} \mathbf{W}$. For the ML estimator, the asymptotic distribution of the univariate and bivariate residual moment $\mathbf{p}_2 - \hat{\boldsymbol{\pi}}_2$ is asymptotically normal with mean zero and covariance matrix

$$\sum\nolimits_{\mathbf{2}} = \boldsymbol{\Xi}_{\mathbf{2}} - \boldsymbol{\Delta}_2 \; I^{-1} \, \boldsymbol{\Delta}_2' \tag{22}$$

where $\boldsymbol{\Delta}_2 \; = \; (\partial \boldsymbol{\pi}_{\mathbf{2}}(\boldsymbol{\theta}))/\partial \boldsymbol{\theta}'$ is the matrix of derivatives of the univariate and bivariate moments with respect to the parameter vector $\boldsymbol{\theta}$, and $N\boldsymbol{\Xi}_{\mathbf{2}}$ is the asymptotic covariance matrix of the univariate and bivariate sample moment $\mathbf{p}_2$. $\boldsymbol{I}^{-\mathbf{1}}$ divided by the sample size is the asymptotic covariance matrix of the item parameter estimates $\widehat{\boldsymbol{\theta}}$, and $\boldsymbol{I}$ is the information matrix. In general, the asymptotic distribution of $Q$ is a mixture of independent chi-square variates (Maydeu-Olivares, 2013). When $\widehat{\mathbf{W}}$ is chosen so that

$$\sum\nolimits_{\mathbf{2}}\mathbf{W}\sum\nolimits_{\mathbf{2}}\mathbf{W}\sum\nolimits_{\mathbf{2}} = \; \sum\nolimits_{\mathbf{2}}\mathbf{W}\sum\nolimits_{\mathbf{2}} \tag{23}$$

$Q$ is asymptotically chi-square distributed with degrees of freedom equal the rank of $\mathbf{W}\sum_{\mathbf{2}}$ (Rao, 1973, p.188).

There are two ways to choose $\widehat{\mathbf{W}}$ to satisfy (23). One approach is to construct a weight matrix such that $\sum_{\mathbf{2}}$ is a generalized inverse of $\mathbf{W}$. That is, $\mathbf{W}$ satisfies $\mathbf{W}\sum_{\mathbf{2}}\mathbf{W} = \; \mathbf{W}$. This approach is illustrated in the $M_2$ statistic which can be alternatively written as

$$M_2 = N \, (\mathbf{p}_2 - \hat{\boldsymbol{\pi}}_2)' \, \hat{\mathbf{C}}_2 \, (\mathbf{p}_2 - \hat{\boldsymbol{\pi}}_2), \quad \mathbf{C}_2 = \; \boldsymbol{\Delta}_2^{(c)}(\boldsymbol{\Delta}_2^{(c)\prime} \, \boldsymbol{\Xi}_2 \, \boldsymbol{\Delta}_2^{(c)} \,)^{-1}\boldsymbol{\Delta}_2^{(c)\prime} \tag{24}$$

where $\boldsymbol{\Delta}_2^{(c)\prime}\boldsymbol{\Delta}_2 \; = \mathbf{0}$.

Another way to satisfy (23) is to construct a weight matrix such that $\mathbf{W}$ is a generalized inverse of $\sum_{\mathbf{2}}$. That is, $\sum_{\mathbf{2}}\mathbf{W}\sum_{\mathbf{2}} = \sum_{\mathbf{2}}$. This approach leads to the choice $\widehat{\mathbf{W}} = \; \widehat{\sum}_2^{+}$ and the statistic

$$R_2 = N\,(\mathbf{p}_2 - \widehat{\boldsymbol{\pi}}_2)'\,\widehat{\sum}_2^{+}\,(\mathbf{p}_2 - \widehat{\boldsymbol{\pi}}_2). \tag{25}$$

which was proposed by Reiser (1996, 2008) for binary data. The degrees of freedom of $R_2$ equal rank ($\sum_2^{+}\sum_2$) = rank ($\sum_2$).

Between these two statistics, $M_2$ is preferred over $R_2$ due to its computational advantages. $M_2$ does not require the computation of the asymptotic covariance matrix of the parameter estimates. Instead, only the diagonal elements of the information matrix are needed to obtain the standard errors of the parameter estimates. By contrast, $R_2$ is more computationally intensive to obtain its degree of freedom, which depends on the rank of $\sum_2$. In practice, the rank of $\widehat{\sum}_2$ can be estimated by eigendecomposition. As a result, the *p*-value of $R_2$ will depend on how many eigenvalues are numerically judged to be zero. However, in IRT applications where numerical integration is involved, it may be difficult to determine exactly how many eigenvalues are zero when some of them are very close to zero (Maydeu-Olivares & Joe, 2008).

Another way to obtain an overall goodness-of-fit statistic using a weight matrix in the quadratic form $Q$ is to adjust the test statistic by its asymptotic mean and variance so that the asymptotic distribution of the adjusted test statistic can be approximated by a chi-square distribution. This approach was introduced by Bartholomew and Leung (2002) and further developed in Cai, Maydeu-Olivares, Coffman, and Thissen (2006). To be specific, the distribution of $Q$ can be approximated by a $bX_a^2$ distribution. The first two asymptotic moments of $Q$ are

$$\mu_1 = tr(\mathbf{W}\boldsymbol{\sum}_\mathbf{2}),\ \mu_2 =\ 2tr(\mathbf{W}\boldsymbol{\sum}_\mathbf{2})^2. \tag{26}$$

When $a$ and $b$ are solved and $\mu_1$ and $\mu_2$ are evaluated at the parameter estimates, the mean and variance corrected $Q$ statistic can be written as:

$$\bar{Q} = \frac{Q}{b} = \frac{2\hat{\mu}_1}{\hat{\mu}_2} Q, \tag{27}$$

which has an approximate reference chi-square distribution with degrees of freedom

$$a = \frac{2\,\hat{\mu}_1^2}{\hat{\mu}_2}\,. \tag{28}$$

Alternatively, Asparouhov and Muthen (2010) suggested that it is possible to define another mean and variance corrected $\bar{\bar{Q}}$ that has the same degrees of freedom as $M_2$ (i.e., $df_2 = n(K–1) + n(n–1)/2(K–1)^2 – q$) rather than estimate the degrees of freedom as in (28). This statistic can be written as $\bar{\bar{Q}} = a^* + b^*Q$, where $a^*$ and $b^*$ are chosen so that the mean and variance of $\bar{\bar{Q}}$ are $df_2$ and $2df_2$, respectively. The $\bar{\bar{Q}}$ statistics can be written as follows after $a^*$ and $b^*$ are solved:

$$\bar{\bar{Q}} = Q\sqrt{\frac{2df_2}{\hat{\mu}_2}} + df_2 - \sqrt{\frac{2df_2\hat{\mu}_1^2}{\hat{\mu}_2}}. \tag{29}$$

Results from Asparouhov and Muthen (2010) showed that there was a negligible difference in the $p$-values obtained from $\bar{Q}$ and $\bar{\bar{Q}}$. Note that mean and variance corrected statistics also require the computation of an estimate of $\boldsymbol{\Sigma_2}$ (for $\mu_1$ and $\mu_2$), which is the asymptotic covariance matrix of the bivariate residual moments. Therefore, from a computational perspective, $M_2$ is still preferable to these mean and variance corrected statistics.

**Testing Models for Large and Sparse Ordinal Data**

As $n$ and especially $K$ increase, the number of summary statistics in $M_2$ becomes too large for computation. Therefore, the information summarized from the multinomial table needs to be further reduced. A natural choice of statistics in this case is the means and cross-products of the multinomial variables ignoring the multivariate nature of the multinomial variables. This summary statistic can be expressed by the means and cross-products of indicator or dummy variables of the multinomial table instead of the residual $\mathbf{p}_2 - \hat{\boldsymbol{\pi}}_2$ (Joe & Maydeu-Olivares, 2010). For example, suppose $Y_i$ and $Y_j$ are two multinomial variables with three categories each (i.e., $k$ = 0, 1, 2). $Y_i$ and $Y_j$ can each be denoted by the indicator variables $I_{i,1}$, $I_{i,2}$, and $I_{j,1}$, $I_{j,2}$, respectively, as follows:

| $Y_i$ | $I_{i,1}$ | $I_{i,2}$ | $Y_j$ | $I_{j,1}$ | $I_{j,2}$ |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 1 | 1 | 0 |
| 2 | 0 | 1 | 2 | 0 | 1 |

The summary statistics in $M_2$ are the sample means of these indicator variables and the sample cross-products of indicator variables from different variables. The means and cross-products can be summarized as follows:

$$\mathrm{E}[I_{i,1}] = \Pr(Y_i = 1),\ \mathrm{E}[I_{i,1}, I_{j,1}] = \Pr(Y_i = 1, Y_j = 1)$$

$$\mathrm{E}[I_{i,2}] = \Pr(Y_i = 2),\ \mathrm{E}[I_{i,1}, I_{j,2}] = \Pr(Y_i = 1, Y_j = 2)$$

$$\mathrm{E}[I_{j,1}] = \Pr(Y_j = 1),\ \mathrm{E}[I_{i,2}, I_{j,1}] = \Pr(Y_i = 2, Y_j = 1)$$

$$\mathrm{E}[I_{j,2}] = \Pr(Y_j = 2),\ \mathrm{E}[I_{i,2}, I_{j,2}] = \Pr(Y_i = 2, Y_j = 2) \qquad (30)$$

As can be seen, the summary statistics in $M_2$ are the univariate and bivariate proportions that do not include the category zero. In general, the sample means of these indicator variables and the sample cross-products of indicator variables from different variables can be expressed as:

$$\kappa_i = \mathrm{E}[Y_i] = 0 \times \Pr(Y_i = 0) + \ldots + (K_i - 1) \times \Pr(Y_i = K_i - 1), \qquad (31)$$

$$\kappa_{ij} = \mathrm{E}[Y_i\, Y_j] = 0 \times 0 \times \Pr(Y_i = 0, Y_j = 0) + \ldots + (K_i - 1) \times (K_j - 1) \times \Pr(Y_i = K_i - 1, Y_j = K_j - 1) \; (32).$$

For the previous example,

$$\kappa_i = \mathrm{E}[Y_i] = 1 \times \Pr(Y_i = 1) + 2 \times \Pr(Y_i = 2),$$

$$\kappa_j = \mathrm{E}[Y_j] = 1 \times \Pr(Y_j = 1) + 2 \times \Pr(Y_j = 2),$$

$$\kappa_{ij} = \mathrm{E}[Y_i\, Y_j] = 1 \times 1 \times \Pr(Y_i = 1, Y_j = 1) + 1 \times 2 \times \Pr(Y_i = 1, Y_j = 2) + 2 \times 1 \times \Pr(Y_i = 2, Y_j = 1) + 2 \times 2 \times \Pr(Y_i = 2, Y_j = 2). \qquad (33)$$

Note that the quantities in (33) are simply a linear function of those in (30). Therefore, the sample means and cross-products of variables coded as $\{0, 1, \ldots, K_i\}$ in (33) are a further reduction of the sample univariate and bivariate proportions in (30). Furthermore, in the binary case (33) reduces to (30).

Let $\hat{\kappa} = \kappa(\hat{\theta})$ be the statistic that depends on the model parameters and is evaluated at their estimates and let **k** be the sample counterpart of (33). A quadratic-form statistic similar to $M_2$, $M_{ord}$, can be formed:

$$M_{ord} = N\, (\mathbf{k} - \hat{\boldsymbol{\kappa}})'\, \hat{\mathbf{C}}_{ord}\, (\mathbf{k} - \hat{\boldsymbol{\kappa}}), \quad \mathbf{C}_{ord} = \boldsymbol{\Xi}_{ord}^{-1} - \boldsymbol{\Xi}_{ord}^{-1}\, \boldsymbol{\Delta}_{ord}\, (\boldsymbol{\Delta}'_{ord}\, \boldsymbol{\Xi}_{ord}^{-1} \boldsymbol{\Delta}_{ord})^{-1} \boldsymbol{\Delta}'_{ord} \boldsymbol{\Xi}_{ord}^{-1} \; (34)$$

where $N\boldsymbol{\Xi}_{ord}$ is the asymptotic covariance matrix of the sample means and cross-products **k**, $\Delta_{\boldsymbol{ord}}$ is the matrix of derivatives of the population means and cross-products κ with respect to the model parameters $\boldsymbol{\theta}$, and $\mathbf{C}_{ord}$ is evaluated at the parameter estimates. The sample statistics used in $M_{ord}$ are $\mathbf{k}$ = $(\bar{\mathbf{y}}', \mathbf{c}')'$, the *n* sample means $\bar{y}$, and the $n(n - 1)/2$ cross-products $\mathbf{c}$ = *vecr* (**Y'Y**/*N*). Here **Y** denotes the *N* × *n* data matrix and *vecr*() denotes an operator that takes the lower diagonal of a matrix (excluding the diagonal) and stacks it in a column vector. When all variables are binary, $M_{ord}$ reduces to $M_2$. $M_{ord}$ follows an asymptotic chi-square distribution with $df_{ord} = n(n+1)/2 - q$ for any consistent estimator. This means that $M_{ord}$ cannot be used when the number of categories is large and the number of item is small due to the lack of degrees of freedom. Thus, for ordinal data, if the model involves a large number of items and categories per item, $M_{ord}$ must be used because $M_2$ cannot be calculated. On the other hand, when the number of categories is large and the number of items is small, $M_{ord}$ cannot be used due to a lack of degrees of freedom.

## Piecewise Assessment of Fit

After the overall fit of a model is examined, it is important to perform a piecewise goodness-of-fit assessment. If the overall model fit is poor, a piecewise fit assessment might be able to suggest where the problem is. Even if the overall model fit is good, a piecewise fit assessment can still help identify the parts that fit less well.

Many overall model fit statistics can be used for piecewise fit assessment when they are applied to bivariate tables of item pairs. For example, the bivariate Pearson's $\chi^2$ statistic can be computed for each bivariate subtable:

$$X_{ij}^2 = N\left(\mathbf{p}_{\boldsymbol{ij}} - \hat{\boldsymbol{\pi}}_{ij}\right)' \hat{\mathbf{D}}_{ij}^{-1}\left(\mathbf{p}_{\boldsymbol{ij}} - \hat{\boldsymbol{\pi}}_{ij}\right). \tag{35}$$

For a subtable of items *i* and *j* each with *K* categories, $\mathbf{p}_{\boldsymbol{ij}}$ is the $K^2$ vector of observed bivariate proportions; $\hat{\boldsymbol{\pi}}_{ij}$ = $\pi_{ij}$ ($\hat{\boldsymbol{\theta}}_{ij}$) is the vector of the expected probabilities that depend on the $q_{ij}$ parameter estimates, $\hat{\boldsymbol{\theta}}_{ij}$, in the bivariate table and $\mathbf{D}ij$ = *diag*($\hat{\boldsymbol{\pi}}_{ij}$). Although it seems natural to refer $X^2_{ij}$ to a chi-square distribution with $df_{ij} = K^2 - q_{ij} - 1$, Maydeu-Olivares and Joe (2006) showed that the asymptotic distribution of the subtable $X^2_{ij}$ is stochastically larger than this reference distribution. This means that referring $X^2_{ij}$ to a chi-square distribution with $df_{ij}$ may lead to rejecting well-fitting items. Instead, the $M_{ij}$ statistic was shown to be asymptotically distributed as a chi-square with $df_{ij}$ degrees of freedom:

$$M_{ij} = X^2_{ij} - N\left(\mathbf{p}_{\boldsymbol{ij}} - \hat{\boldsymbol{\pi}}_{\boldsymbol{ij}}\right)'\hat{\mathbf{D}}^{-\mathbf{1}}_{\boldsymbol{ij}}\,\hat{\Delta}_{\boldsymbol{ij}}\,(\hat{\Delta}'_{\boldsymbol{ij}}\hat{\mathbf{D}}^{-\mathbf{1}}_{\boldsymbol{ij}}\hat{\Delta}_{\boldsymbol{ij}})^{-\mathbf{1}}\hat{\Delta}'_{\boldsymbol{ij}}\hat{\mathbf{D}}^{-\mathbf{1}}_{\boldsymbol{ij}}\left(\mathbf{p}_{\boldsymbol{ij}} - \hat{\boldsymbol{\pi}}_{\boldsymbol{ij}}\right), \tag{36}$$

where $\boldsymbol{\Delta}_{ij}$ denotes the matrix of derivatives of the bivariate probabilities $\pi_{ij}$ with respect to the parameters involved in the bivariate table, $\hat{\boldsymbol{\theta}}_{ij}$.

As an alternative way to correct $X^2_{ij}$, the distribution of $X^2_{ij}$ can be approximated by a $bX^2_a$ distribution (Maydeu-Olivares, 2013; Liu & Maydeu-Olivares, 2014). The first two asymptotic moments of $X^2_{ij}$ are

$$\mu_1 = tr\left(\hat{\mathbf{D}}^{-\mathbf{1}}_{\boldsymbol{ij}}\hat{\Sigma}_{ij}\right),\ \mu_2 = 2tr\left(\hat{\mathbf{D}}^{-\mathbf{1}}_{\boldsymbol{ij}}\hat{\Sigma}_{ij}\right)^2. \tag{37}$$

Similar to the mean and variance corrected chi-square statistic for the overall fit assessment, the mean and variance corrected $\bar{X}^2_{ij}$ statistic for item pairs in a subtable can be written as:

$$\bar{X}^2_{ij} = \frac{X^2_{ij}}{b} = \frac{2\hat{\mu}_1}{\hat{\mu}_2}X^2_{ij} \tag{38}$$

which has an approximate reference chi-square distribution with degrees of freedom

$$a = \frac{2\,\hat{\mu}_1^2}{\hat{\mu}_2} \tag{39}$$

Again, it is possible to define an alternative mean and variance corrected $X_{ij}^2$ which has $df_{ij} = K^2 - q_{ij} - 1$ degrees of freedom (Asparouhov & Muthen, 2010). This statistic can be written as $\bar{\bar{X}}_{ij}^2 = a^* + b^* X_{ij}^2$ , where $a^*$ and $b^*$ are chosen so that the mean and variance of $\bar{\bar{X}}_{ij}^2$ are $df_{ij}$ and $2df_{ij}$, respectively:

$$\bar{\bar{X}}_{ij}^2 = X_{ij}^2 \sqrt{\frac{2df_{ij}}{\hat{\mu}_2}} + df_{ij} - \sqrt{\frac{2df_{ij}\hat{\mu}_1^2}{\hat{\mu}_2}} \;, \tag{40}$$

When the model parameters have been estimated by maximum likelihood using the full table,

$$\sum_{ij} = diag(\pi_{ij}) - \pi_{ij}\,\pi_{ij}' - \Delta_{ij}(I^{-1})_{ij}\,\Delta_{ij}' \tag{41}$$

multiplied by sample size is the asymptotic covariance matrix of the cell residuals for the pair of items *i* and *j*. $(I^{-1})_{ij}$ denotes the rows and columns of the information matrix corresponding to the item parameters involved in the subtable for variables *i* and *j*.

Similarly, a bivariate subtable counterpart of the overall statistic proposed by Reiser (1996, 2008) can be written as:

$$R_{ij} = N\left(\mathbf{p}_{\boldsymbol{ij}} - \hat{\boldsymbol{\pi}}_{ij}\right)'\hat{\sum}_{ij}^{+}\left(\mathbf{p}_{\boldsymbol{ij}} - \hat{\boldsymbol{\pi}}_{ij}\right). \tag{42}$$

The degrees of freedom of $R_{ij}$ are given by the rank of $\sum_{ij}$, which can be estimated from the data as the number of eigenvalues of $\hat{\sum}_{ij}$ which are nonzero. For example, $10^{-5}$ was suggested as a cutoff when the rank of $\hat{\sum}_{ij}$ and of $\hat{\sum}_2$ were estimated (Liu & Maydeu-Olivares, 2014).

A drawback of $M_{ij}$ is that it cannot be used with dichotomous item pairs due to lack of degrees of freedom (Maydeu-Olivares & Liu, 2012). The $\bar{X}^2_{ij}$ statistic can be used with binary data because its degrees of freedom are estimated, unless the estimate is exactly zero. Moreover, Maydeu-Olivares and Liu (2012) suggested that *z*-statistics for univariate and binary residuals provide a useful approach for dichotomous models. The univariate and bivariate residuals are the sums of the cell residuals. A *z*-statistic is obtained by dividing these univariate and bivariate residuals by their standard errors to identify the source of misfit. The standardized bivariate residuals for binary item pairs can be written as:

$$z_{ij} = \frac{p_{ij} - \hat{\pi}_{ij}}{SE(p_{ij} - \hat{\pi}_{ij})} = \frac{p_{ij} - \hat{\pi}_{ij}}{\sqrt{\hat{\sigma}^2_{ij}/N}} = \frac{p_{ij} - \hat{\pi}_{ij}}{\sqrt{vecdiag(\hat{\Sigma}_{ij})/N}} \tag{43}$$

where $\sum_{ij} = diag(\pi_{ij}) - \pi_{ij}\,\pi'_{ij} - \Delta_{ij}(I^{-1})_{ij}\,\Delta'_{ij}$, $\pi_{ij}$ = Pr($Y_i$ = 1, $Y_j$ = 1), $p_{ij}$ is its corresponding proportion, and $(I^{-1})_{ij}$ can be approximated in different ways. For multinomial models estimated with the maximum likelihood method, one way to approximate this matrix is to use the expected information matrix

$$I_E = \Delta' \mathrm{D}\, \Delta\,, \tag{44}$$

where $\mathbf{D}$ = *diag*($\boldsymbol{\pi}$) is a diagonal matrix of all pattern probabilities, and $\Delta = \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$ is a $C \times q$ matrix of derivatives of all possible response pattern probabilities with respect to the item parameters. The expected information matrix can only be computed for small models, because $\boldsymbol{\pi}$ is of dimension $C$ and $\boldsymbol{\Delta}$ is of dimension $C \times q$, which becomes too large for computation as the size of the model increases.

Another way to approximate the matrix is to use the cross-product (XPD) information matrix

$$I_{XPD} = \Delta_O' diag(\boldsymbol{p}_O \,/\, \boldsymbol{\pi}_O^2\,)\Delta_O \,, \tag{45}$$

where $\boldsymbol{p}_O$ and $\boldsymbol{\pi}_O$ denote the proportions and probabilities of the $C_O$ observed patterns and $\Delta_O$ is the $C_O \times q$ matrix of derivatives of the patterns with respect to the model parameters. In models involving a large number of possible patterns, the matrices involved in the XPD information matrix are smaller than those involved in the expected information matrix. Moreover, the dimension of the vectors in (45) does not increase as a function of test length. Therefore, the covariance matrix of the item parameters of larger models can be approximated if the XPD information matrix is used instead of the expected information matrix. A third way is to use an observed information matrix, which requires a second-order derivatives of the pattern probabilities with respect to the item parameters in the model:

$$I_O = N\sum_{c=1}^{C_O}\frac{p_c}{(\pi_c\,(\theta))^2}\left[\frac{\partial\pi_c\,(\theta)}{\partial\theta}\,\frac{\partial\pi_c\,(\theta)}{\partial\theta'} - \,\pi_c\,(\theta)\,\frac{\partial^2\pi_c\,(\theta)}{\partial\theta\partial\theta'}\right] = \,I_{XPD} - \,N\sum_{c=1}^{C_O}\frac{p_c}{\pi_c\,(\theta)}\,\frac{\partial^2\pi_c\,(\theta)}{\partial\theta\partial\theta'}\,. \tag{46}$$

The *z*-statistic can also be extended to polytomous ordinal data as

$$z_{ord} = \frac{k_{ij} - \hat{\kappa}_{ij}}{SE(k_{ij} - \hat{\kappa}_{ij})} = \frac{k_{ij} - \hat{\kappa}_{ij}}{\sqrt{\hat{\sigma}_{ij}^2 / N}} \tag{47}$$

where $\hat{\sigma}_{ij}^2$ = $\mathbf{v}_{ij}'\,\hat{\Sigma}_{ij}\,\mathbf{v}_{ij}$ , with $\mathbf{v}_{ij}'$ = (0 × 0, 0 × 1, … 0 × (*K* - 1), …, (*K* - 1) × 0, (*K* - 1) × 1, …, (*K* - 1) × (*K* - 1)), and $k_{ij} - \,\hat{\kappa}_{ij}$ = $\mathbf{v}_{ij}'\,(\mathbf{p}_{ij} - \,\hat{\boldsymbol{\pi}}_{ij})$ is the residual mean cross product. The asymptotic distribution of $z_{ord}$ is standard normal, and of $z_{ord}^2$ follows a chi-square distribution with 1 *df*.

Maydeu-Olivares and Liu (2012) demonstrated how $M_n$, $M_2$ and *z*-statistics can be applied to simulation and real data sets. Results from their simulation studies showed that the $M_n$ statistic outperformed $\chi^2$ because the latter over-rejected the model for small sample sizes. For the piecewise fit analysis, the simulation results in Liu and Maydeu-Olivares (2014) showed that the z-statistic has better Type I error rates and more power than the mean and variance adjusted $\chi^2$ statistics in (38) and (40), and the $R_{ij}$ statistics in (42) for both binary and ordinal data when the observed information in (46) is used for computation. But if cross-product information matrix in (45) is used, then the mean and variance adjusted $\chi^2$ is recommended.

Due to the relatively small sample sizes of Maydeu-Olivares and Liu's (2012) real data sets ($N < 900$), the empirical performance of these statistics remains unknown in the context of large-scale assessments. Therefore, the first purpose of this study is to compare the performance of these new overall and piecewise statistics and the traditional heuristic model fit approaches, and see whether these new statistics are superior to previous approaches. Also, the applications of the limited-information fit statistics have mostly focused on data from educational settings. In the current study, their performance on real data from organizational settings is also evaluated.

# METHOD

## Simulation Study

In the simulation study, I examined the performance of both the traditional heuristic and limited-information fit statistics when they detected multidimensionality in data. To be specific, I simulated item responses in the mixed-format exams where items in different test sections loaded on correlated but different factors. For example, most AP® exams are mixed-format tests consisting of a multiple choice (MC) section and a constructed response (CR) section. While examinees can choose one best answer from a list of options provided by the MC items, they need to “construct” their responses to the open-ended essay questions in the CR section. Accordingly, the MC items are usually scored dichotomously as right or wrong (i.e., 1 or 0), whereas as the CR items are scored polytomously with multiple categories from 0 to the maximum score possible for each item. Although the mixed application of MC and CR items brings many psychometric and practical advantages, it raises important questions such as whether the two test sections measure the same latent ability and whether it is appropriate to use a unidimensional IRT model to simultaneously analyze data from the two test formats. To answer these questions, I examined whether the existence of a CR section affected the overall fit of the entire exam and the piecewise fit of the MC items when the MC and CR items either shared one common factor or loaded on their own format factor respectively.

Therefore, the simulation study was conducted with four factors manipulated. First, I used four sample sizes, 200, 500, 1000 and 3000, for examinees. Secondly, I varied the test length by including 10, 20, and 40 MC items. Thirdly, whether the test had a CR section or not was manipulated. Finally, I manipulated whether the IRT model fit the data or not by generating one-dimensional and two-dimensional data. The combination of these four factors led to 48 simulation conditions in total. In each condition, 100 replications were conducted. The average detection of model misfit was compared across different conditions by descriptive statistics such as means and standard deviations of the model fit statistics across replications. The detection rates were also reported as type I error rates for the unidimensional models and as power for the multidimensional models.

The 2PL and SGR item parameter estimates from real exams were used for the simulation of MC and CR item responses. For the simulation of MC items, I used item parameter estimates from the AP® 2011 Calculus AB exam. The 2PL model showed excellent fit for the 45 MC items according to the means of the adjusted $\chi^2/df$ statistics for item pairs (*Mean* = 1.70) and triples (*Mean* = 1.86) (Windsor, Jeon, Cao, & Drasgow, 2013). I removed five items with low *a*-parameters and large *b*-parameters (i.e., items that very few students answered correctly) and kept the item parameters of the other 40 items in the pool for simulation. For the simulation of the CR item responses, I used item parameter estimates from the 2010 US History AP® exam (Wang, Drasgow, & Liu, 2013). In the original exam, there were five CR questions and the highest score possible for each question was 9. Prior to the item parameter estimation, the ten categories (0 to 9) were collapsed into five (0 to 4) to facilitate the model data fit analysis. The means of the adjusted $\chi^2/df$ statistics showed marginal fits for the SGR models for item pairs (*Mean* = 3.47) but good fit for triples (*Mean* = 2.13). For the purposes of this simulation study, the item parameter estimates appeared adequate.

The two-dimensional data were generated to evaluate the power of the fit statistics to detect misspecified models. The item parameters for simulation were the same sets of item parameters as in the unidimensional data. For those conditions with only MC items, the first half of the MC items were simulated to measure the first dimension while the second half of the MC items were simulated to measure the second dimension. The latent trait distribution was bivariate normal with a zero mean and unit variance. The latent trait correlation was set to 0.7, reflecting a moderate level of multidimensionality. For conditions with both MC and CR sections, items from each section loaded on one dimension. The latent trait correlation between the two sections was also set to 0.7.

## Applications to Real Data

**AP® Exams.** I analyzed data from three 2012 AP® exams, each with a random sample of about 20000 examinees. Previous research has shown that different AP® exams have different latent structures of the MC and CR sections (Wang et al., 2013). For some tests, the MC and the CR questions essentially shared one common factor and a unidimensional model provided an excellent fit (e.g., Calculus Exam). Other tests, however, were less unidimensional due to the different material tested by MC and CR items, and thus fit a bifactor model better (e.g., English Language Exam). As a result, if item parameters were estimated simultaneously for the two sections under the assumption of unidimensionality, I expected greater model misfit for exams that fit a bifactor model than those that fit a unidimensional model. To examine this effect, I evaluated model fit for data with different levels of unidimensionality based on the disattenuated correlation between the MC and the CR responses. To be more specific, I analyzed the Physics B Exam which was found to be the most unidimensional ($r$ = 0.96), the English Literature Exam which was found to be the least unidimensional ($r$ = 0.77), and the World History Exam whose disattenuated correlation was intermediate ($r$ = 0.89-0.91; Wang et al., 2013).

**International Personality Item Pool (IPIP).** Model fit statistics for the SGR model were calculated for the International Personality Item Pool (IPIP; Goldberg, 1999). There were 10 items in each of the Big Five personality dimensions. Respondents were asked to rate each item on a 4-point Likert scale, where 1 = "*Strongly Disagree*", 2 = "*Disagree*", 3 = "*Agree*", and 4 = "*Strongly Agree*". Two conditions were examined in the original study: faking and honest (Cao, Tay, Luo, & Drasgow, 2014). In the faking condition, respondents were asked to imagine as if they were "*applying for a job that you want very much*", so they should "*select the response that will make you look like the best job applicant*". In the honest condition, respondents were told that their "*answers will be used for research purposes only*", so they should "*answer the questions as honestly as possible*". Participants were recruited from a large crowdsourcing Internet marketplace. A total of 947 subjects were included in the analyses, with 458 subjects in the faking condition and 489 subjects in the honest condition. The mean adjusted $\chi^2/df$ statistics for item doubles and triplets suggested that the SGR model did not fit well for Emotional Stability and Openness, but fit adequately for Extraversion, Agreeableness, and Conscientiousness. The mean adjusted $\chi^2/df$ statistics also showed that the item response data from the honest condition fit the SGR model better than the item responses from the faking condition for Extraversion and Emotional Stability, but not for Conscientiousness, Agreeableness, and Openness (Cao et al., 2014).

**Counterproductive Work Behavior (CWB) Scale.** Model fit statistics for the SGR model were also calculated for a performance rating scale that is commonly seen in organizational settings. Employees ($N$ = 449) from classes at a large southwestern university provided self-ratings of their CWB using the 19-item Bennett and Robinson (2000) measure, which reflects both interpersonal deviance (CWB-I; $\alpha$ = .88) and organizational deviance (CWB-

O; $\alpha$ = .88) dimensions (Carpenter, 2013). The CWB-I scale included 7 items and the CWB-O scale included 12 items. Respondents were asked to indicate on a 7-point Likert scale the extent to which they had engaged in each of the behaviors described in the scales in the last year. The scale anchors were as follows: 1 *(never),* 2 *(once a year),* 3 *(twice a year),* 4 *(several times a year),* 5 *(monthly),* 6 *(weekly),* and 7 *(daily).* Options 6 and 7 were collapsed into one option in the IRT analysis, because very few respondents said that they engaged in CWB behaviors "daily".

Results from confirmatory factor analysis showed that a two-factor model reflecting CWB-I and CWB-O fit better than a one-factor model. The fit indices were: CFI (Comparative Fit Index) = 0.92, TLI (Tucker Lewis Index) = 0.91, SRMR (Standardized Root Mean Square Residual) = 0.075, and RMSEA (Root Mean Square Error of Approximation) = 0.120 for the two-factor model; and CFI = 0.86, TLI = 0.84, SRMR = 0.092, and RMSEA = 0.162 for the one-factor model (Carpenter & Cao, 2013). Therefore, IRT analyses should be performed on these two subscales separately. Both the CWB-I and CWB-O scales showed good fit for the SGR model based on the mean adjusted $\chi^2/df$ statistics (CWB-I: *Mean* = 1.32 for item pairs, and *Mean* = 2.34 for item triples; CWB-O: *Mean* = -0.66 for item pairs, and *Mean* = 0.76 for item triples) based on the suggested cutoff value of 3.0 (Carpenter & Cao, 2013).

## Analysis

Script files of MATLAB and R (R Core Team, 2014) were used to automate the simulation process. For the simulated data, the dimensionality was known so the dimensionality check was skipped. For each replication, response data were simulated using MATLAB, and 2PL and SGR item parameters were estimated simultaneously by MULTILOG (Thissen, Chen, & Bock, 2003). Because the overall model fit statistics usually are computationally intensive and provide less information than piecewise model fit assessment, especially for large models, only the most powerful and least computationally intensive $M_2$ statistic was calculated for each model with the FlexMIRT software (Cai, 2012). Piecewise model fit statistics were calculated for dichotomous items using the R code from Liu and Maydeu-Olivares (2014). The piecewise fit statistics were not examined for polytomous items when they were included in the model, because the purpose of the simulation study was to evaluate whether the existence of the polytomous items affected the overall fit and the pairwise fit for the dichotomous items. Finally, the heuristic adjusted $\chi^2/df$ statistics were calculated for item pairs and triples with the FORSCORE program (Williams & Levine, 1993).

For the real data sets, I first checked the dimensionality of the data to justify the use of IRT models by conducting a principal component analysis with varimax rotation in SPSS. If the scree plot showed a strong dominant factor, then it was concluded that the data were sufficiently unidimensional for IRT analysis (Drasgow & Parsons, 1983). FlexMIRT (Cai, 2012) was used to estimate item parameters and to obtain overall $M_2$ statistics for the entire model. Piecewise fit statistics, $\bar{X}_{ij}^2$, $R_{ij}$, and $z_{ij}$ were calculated for the dichotomous item pairs, and $M_{ij}$, $\bar{X}_{ij}^2$, $R_{ij}$, and $z_{ord}$ were calculated for polytomous item pairs with Liu and Maydeu-Olivares' (2014) R code. The adjusted $\chi^2/df$ statistics for item pairs and triplets were obtained from the FORSCORE program (Williams & Levine, 1993).

# RESULTS

## Simulation

**Overall fit statistics $M_2$.** Table 1 shows the means, standard deviations and detection rates of the overall $M_2$ statistics across 100 replications in the 48 conditions, as well as the critical values of chi-square statistics at $p$ = .05 with their corresponding *df* for the model in the first column. When all MC items were simulated to load on one factor (i.e., “One factor, MC items only” in Table 1), the $M_2$ statistics showed no misfit for the unidimensional model as expected. None of the means of the $M_2$ statistics were larger than the chi-square critical values at the $p$ = .05 level. The values of the mean $M_2$ statistics increased as the model became larger with more MC items included when the sample size was fixed. But when the number of items in the model was fixed, both the means and standard deviations of the $M_2$ statistics remained almost the same across different sample sizes. The detection rates (i.e., type I error rates here) were all around 0.05, ranging from 0.01 to 0.09. The results were similar for the conditions in which both the MC items and the five CR items were simulated to load on one factor (i.e., “One factor, MC and 5 CR items” in Table 1). Note that in the conditions where both MC and CR items were included in the model, $M_2$ statistics were reported for the entire model (i.e., MC and CR items) rather than just for the MC items.

When MC items were simulated to load on two different factors (i.e., “Two factors, MC items only” in Table 1), the $M_2$ statistics generally detected misfit for unidimensional models when the sample size was at least 500 at the $p$ = .05 level. When the sample size was 200, the $M_2$ statistics often failed to reach statistical significance at $p$ = .05 for small models with only 10 MC items, with a mean of 37.78 with 35 *df*. This result was confirmed by the low power rate of only

0.11 when both the sample size and the model size were small. Different from the one-factor model conditions, the means and standard deviations of $M_2$ statistics increased as the sample sizes became larger, especially when the sample sizes increased from 1000 to 3000. This corresponded to an increase in power when either the sample size or the model size became larger. When the sample size was 3000, the detection rate was 100%. The results were mostly similar for conditions in which the MC and CR items were simulated to reflect two separate factors (i.e., “Two factors, MC and 5 CR items” in Table 1). Again, $M_2$ statistics often failed to detect misfit when the sample size was 200 and the number of item was 10. When the model became larger, misfit was detected frequently even when the sample size was as small as 200. The detection rates were larger than 80% when the sample size was at least 500 and were 100% when the sample size was at least 1000, regardless of the model size. In summary, $M_2$ statistics were able to reliably detect violations of the unidimensionality assumption for small to medium-sized models when the sample size was at least 500. If the sample size was smaller than 500, a larger model was required to detect misfit. Like most statistics, the power of $M_2$ statistics increases with larger samples sizes.

**Piecewise adjusted $\chi^2/df$ statistics for item pairs.** Table 2 shows the means and standard deviations of the adjusted $\chi^2/df$ statistics for 100 replications across all MC item pairs. Again, the pairwise fit statistics were calculated for MC items only, because the CR items were included only for the purpose of examining whether their existence, either on the same or different dimensions of the MC items, would influence the model fit of MC items. As a result, only in the “Two factors, MC items only” conditions in Table 2 was a unidimensional model fit to multidimensional MC item responses; because in the “Two factors, MC and CR items” conditions, model fit was evaluated only for the unidimensional MC items. In the conditions where both MC and CR items were included, the item parameters from both sections were estimated simultaneously and then the parameter estimates of the MC items were used to evaluate the model fit of the MC items. Therefore, I expected the statistics to indicate worse fit in the "Two factors, MC items only" conditions than in other conditions. As the adjusted $\chi^2/df$ statistics were adjusted to what would be expected in a sample size of 3000, negative values were obtained for smaller sample sizes.

As Table 2 shows, the mean adjusted $\chi^2/df$ statistics increased with larger sample sizes in all conditions, whereas the standard deviations decreased. When the sample size was only 200, the mean and standard deviations of the statistics were almost the same across different models in all conditions. When the sample size was 500 or larger, the "Two factors, MC items only" conditions had the largest mean and standard deviations, as expected. The difference became even more obvious when the sample size increased to 1000 and 3000.

When data were simulated with the unidimensional model (i.e., either MC items only or MC items with CR items), the means and standard deviations of the statistics remained almost the same for the same sample size when different numbers of MC items were included in the model. When the data were simulated to reflect two latent factors (i.e., either MC items only or MC items with CR items), there were more fluctuations in the means and standard deviations across different models with the same sample size, especially when the model was small and sample size was as large as 1000 or 3000.

Finally, the fit for the "Two factors, MC and CR items" conditions was better than the multidimensional conditions but worse than the unidimensional conditions. The difference became more obvious as the sample size increased, but especially so in small models with only

10 items. This probably was because when the parameter estimates were calibrated simultaneously with both MC and CR items in the model, the accuracy of the parameter estimates of the MC items was affected by the multidimensional structure of the data. However, when the model was large with 40 MC items, the impact from the five CR items on the other dimension was not as large as when the model was small with only 10 MC items.

In summary, adjusted $\chi^2/df$ statistics for item pairs were, to some extent, sensitive to the multidimensionality in the data, especially when the sample size was large. None of the pairwise statistics exceeded the suggested cutoff value of 3.0, even in the "Two factors MC items only" conditions. This is because this rule of thumb is generally not applied to relatively small samples as studied here, and was developed in the context of cross-validation samples.

**Piecewise $\bar{X}_{ij}^2$ statistics for item pairs.** Table 3 shows the descriptive statistics and detection rates of the mean $\bar{X}_{ij}^2$ statistics for 100 replications across MC item pairs. The *df* for the bivariate $\bar{X}_{ij}^2$ estimated in each replication was approximately 1.0 on average, which corresponds to a critical value of 3.84 under a chi-square distribution. For chi-square distributed statistics, the mean of the statistics (across the 100 replications here) should be their *df* and the variance (across the 100 replications here) should equal two times the *df*. Therefore, when the model holds (as in the one-factor models), it is expected that the mean across all item pairs is close to 1.0; whereas the mean should diverge from 1.0 when the model does not hold (as in the two-factor models). Similar to the adjusted $\chi^2/df$ statistics, it is expected that the fit is better for the one-factor model conditions than for the two-factor model conditions.

The descriptive statistics were almost the same across all of the one-factor model conditions, whether a CR section was included or not: the mean $\bar{X}_{ij}^2$ statistics across all item pairs were all very close to 1.0 and the standard deviations across item pairs were no larger than 0.17 across all sample sizes and models. The type I error rates were all around 0.050. As expected, $\bar{X}_{ij}^2$ was larger in the two-factor conditions than in the one-factor conditions. When only the multidimensional MC items were examined, the mean $\bar{X}_{ij}^2$ fit statistics were all larger than 1.0. While the mean $\bar{X}_{ij}^2$ fit statistics remained relatively stable for different models with the same sample size, they increased as the sample size became larger. The detection rates increased from 0.07 for a sample size of 200 to 0.30 for a sample size of 3000, but the model size did not seem to affect detection rates when the sample size was fixed. Similar results were found for the two-factor models with both MC and CR items included when the model was small with only 10 or 20 MC items. When the model had 40 items, the $\bar{X}_{ij}^2$ fit statistics remained around 1.0 across all sample sizes as in the unidimensional models. Again, this probably is because when the number of MC items was small, the parameter estimates were more likely to be influenced by the other latent factor represented by the CR items if the items from both sections were calibrated simultaneously. Overall, in very few two-factor conditions did the mean $\bar{X}_{ij}^2$ statistics exceed the critical value of 3.84, even for a sample size of 3000, which was confirmed by the low power across all conditions. Therefore, an even larger sample size might be required for $\bar{X}_{ij}^2$ to detect moderate multidimensionality in small to medium models.

**Piecewise $R_{ij}$ statistics for item pairs.** The results for the $R_{ij}$ statistics were similar to those for the $\bar{X}_{ij}^2$ statistics, except that there were more fluctuations in the descriptive statistics (see Table 4). As the *df* was estimated to be approximately 2.0 on average, the mean $R_{ij}$ statistics should be close to 2.0 across all item pairs in the one-factor models, but diverge from 2.0 in the two-factor models. As Table 4 shows, the mean $R_{ij}$ statistics across all item pairs in the model remained around 2.0 across different models and sample sizes for one-factor model conditions. The type I error rates also showed more fluctuations compared with those for the $\bar{X}_{ij}^2$ statistics, but they were still mostly around 0.050. When only MC items were simulated to be multidimensional, the mean $R_{ij}$ statistics were larger than those in the one-factor model conditions, and increasingly diverged from 2.0 as the sample size became larger. The power rates were slightly lower than those for the $\bar{X}_{ij}^2$ statistics, ranging from approximately 0.05 for a sample size of 200 to 0.28 for a sample size of 3000. This was also the case for the two-factor models with both MC and CR items when the model was small with only 10 MC items. When the two-factor model had at least 20 items, the impact caused by the multidimensional structure could hardly be detected by the mean $R_{ij}$ statistics. Only in one condition where the model was small and the sample size was large did the mean $R_{ij}$ statistics exceed the critical value of 5.99 under a chi-square distribution with a *df* of 2. The power rates were mostly quite low in all conditions.

**Piecewise adjusted $z_{ij}$ statistics for item pairs.**The descriptive statistics for $z_{ij}$ statistics are shown in Table 5. When MC items were fit to a unidimensional model, the means of the $z_{ij}$ statistics were approximately zero and the standard deviations were approximately 0.1 across different numbers of MC items and sample sizes. The ranges of the $z_{ij}$ statistics were all within 0.6. The results were very similar for the conditions in which both MC and CR items loaded on one factor. The type I error rates were all around 0.050 for the unidimensional conditions. In conditions where MC items reflected two latent factors, the means were still around zero, but the standard deviations were much larger than those in the one-factor conditions, ranging from 0.67 to 1.75. The range of the $z_{ij}$ statistics was as large as 6.0 for a multidimensional model with 40

MC items and a sample size of 3000. Therefore, the $z_{ij}$ statistics for some item pairs had large absolute values when the data did not fit the unidimensional model. Larger means and standard deviations were also found in the conditions where MC and CR items loaded on two factors, but only in the smaller models with 10 or 20 MC items. Therefore, like $\bar{X}_{ij}^2$ and $R_{ij}$ statistics, the $z_{ij}$ statistics could also detect the misfit in the MC items caused by the simultaneous parameter estimation of both MC and CR items that loaded on two different dimensions, at least in small models. While other statistics detected misfit caused by dimensionality by having larger means, $z_{ij}$ statistics did so by having more items with extremely values indicated by the larger standard deviations. Similar to the $\bar{X}_{ij}^2$ and $R_{ij}$ statistics, the power for the $z_{ij}$ statistics remained low between 0.05 to 0.30 in most of the conditions.

To further explore the relationships between these pairwise fit statistics, I examined the correlations among these mean statistics when they detected misfit in multidimensional MC item pairs (i.e., "Two factors, MC items only" conditions) across different models and sample sizes (see Table 6). The absolute values of the $z_{ij}$ statistics were obtained to correlate with other statistics, so a larger correlation coefficient reflected a closer correspondence between two statistics. The correlation coefficients were all significant at either the $p = .01$ or $p = .05$ level (two-tailed), except the ones between adjusted $\chi^2/df$ and $R_{ij}$ statistics when both the models (i.e., 10 or 20 MC items) and the sample sizes ($N = 200$) were small. The correlation coefficients were all moderate to large when the sample size was at least 500. When the sample size reached 3000, all fit statistics correlated at approximately .90 or larger in all models. Across all models and sample sizes, the adjusted $\chi^2/df$ statistics tended to correlate the highest with $\bar{X}_{ij}^2$, but the other two statistics, especially $R_{ij}$, had lower correlations with these two $\chi^2$-related statistics. But overall, there seemed to be a close correspondence among these pairwise fit statistics when misfit was detected for multidimensionality, especially when the sample size was large.

**Piecewise fit statistics for item triples.** Table 7 shows the means and standard deviations of the adjusted $\chi^2/df$ statistics across MC item triples for 100 replications. Similar to the adjusted $\chi^2/df$ statistics for item pairs, the fit statistics were for MC item triples only. The descriptive statistics were all very similar across models and sample sizes in the one-factor conditions, whether CR items were included or not. When the sample size was 200, the mean adjusted $\chi^2/df$ statistics remained almost the same across item triples in all conditions. When the sample size was at least 500, the means and standard deviations of the adjusted $\chi^2/df$ statistics were larger in the "Two factors, MC items only" conditions than other conditions. The differences became even larger as the sample sizes increased. When the MC and CR items loaded on their respective factor, the descriptive statistics were similar to those in the unidimensional models across different sample sizes except when the model included only 10 MC items. In summary, the adjusted $\chi^2/df$ statistics for item triples were also sensitive to misfit for multidimensionality, although again the rule of thumb did not seem to apply to small sample sizes.

Due to the computational complexity, $M_n$ statistics were calculated for only item triples in small models with 10 MC items (see Table 8). The means were almost comparable across different models and sample sizes in all conditions. The standard deviations increased as the sample sizes became larger, and were comparable in all conditions except for the "One factor, MC only" conditions when the sample size was 3000. The detection rates in all conditions were around 0.050, whether the model was unidimensional or multidimensional. Therefore, the $M_n$

statistics for item triples did not seem to be very effective in detecting the misfit caused by multidimensionality in small models with only 10 MC items.

To compare the performance of the adjusted $\chi^2/df$ statistics for item triples and $M_n$ statistics, I examined the correlation coefficients of the two statistics across four sample sizes in the “Two factors, MC items only” conditions. The correlation coefficients were .417 ($p < .001$) for 200 examinees, .322 ($p < .01$) for 500 examinees, -.053 ($p = .564$) for 1000 examinees, and .057 ($p = .537$) for 3000 examinees. Therefore, there seemed to be a moderate correspondence between the two statistics when the sample size was small ($N = 200$ or 500), but the relationship was not significant when the sample size was large ($N = 1000$ or 3000).

## AP® Exams

**Physics B.** Among the 20000 examinees whose item responses were obtained from the College Board, 17244 answered all questions and thus their responses were included in the analyses. The Physics B Exam had 70 MC items and 7 CR items. Item 26 of the MC section was not scored, so test analyses were performed on the remaining 69 MC items and the 7 CR items. For the 7 CR items, item 1 and item 5 had 16 categories (i.e., 0 - 15) and the other five items had 11 categories (i.e., 0 - 10). To facilitate the IRT and model fit analyses, the 11 or 16 categories were collapsed into 5 categories (i.e., 0 - 4) based on the aggregated frequency of each category.

Although the scree plot in Figure 1 showed potential second and third factors from results of the principal component analysis, the first factor was dominant so the item responses were deemed sufficiently unidimensional for IRT analysis (Drasgow & Parsons, 1983). The parameters of the 69 MC and 7 CR items were then estimated simultaneously. As Table 9 shows, most parameter estimates were within the normal ranges, except that MC items 15 and 23 had extremely low a-parameters and large b-parameters. This indicates that very few examinees answered these two questions correctly, because the items were either extremely difficult or miskeyed.

The overall fit of the 76 items was evaluated by the $M_2$ statistic, which was 36300.07 with a *df* of 4538 ($p < .0001$). The overall assessment showed a severe misfit, probably because it is almost impossible to find exact fit for a large model with so many variables even for well-fitting items. Next, various piecewise model-fit statistics were examined for the MC items. The heuristic adjusted $\chi^2/df$ statistics were calculated for a random sample of 69 MC item pairs and 23 triples. The adjusted $\chi^2/df$ statistics was not calculated for all item pairs and triples due to the large number of all possible combinations. The 2PL model showed excellent fit for the MC items: the mean adjusted $\chi^2/df$ statistics were 1.318 (*SD* = 0.525) for item pairs and 1.620 (*SD* = 0.501) for item triples. Only one item pair (i.e., items 49 & 60) and one item triple (i.e., items 45, 49 & 60) had an adjusted $\chi^2/df$ ratio slightly larger than 3.0. Contrary to the result from the overall fit statistic, the adjusted $\chi^2/df$ statistics indicated that the 2PL model fit the MC items well on average.

As there were too many possible item triples for 76 items (i.e., 70300), $M_n$ statistic was not calculated for such a large model. Instead, piecewise fit was assessed by the 2346 bivariate fit statistics for the 69 MC items (see the descriptive statistics in Table 12). The $\bar{X}_{ij}^2$ statistics had a mean of 11.326 and a standard deviation of 39.887. As the median was only 4.752, there apparently were extreme values for some item pairs. For example, the $\bar{X}_{ij}^2$ statistics for item 6 and 7 was as large as 1550.142. The second and third largest were 624.790 for item 2 and 3, and 571.454 for item 3 and 8. The skewness was also demonstrated in the histogram in Figure 2. Out of the 2346 MC item pairs, 11 had $\bar{X}_{ij}^2$ statistics larger than 100. Slightly less than half (i.e., 1067 pairs) showed acceptable fit at the *p* = .05 level. The skewness pattern was similar for the $R_{ij}$ statistics as shown in Figure 3, but the average statistics were larger (i.e., *Mean* = 146.877) and the values were even more dispersed (i.e., *SD* = 461.009). The majority of the item pairs (i.e., 83.97%) showed misfit to some extent according to the $R_{ij}$ statistics. The largest value was 9657.265 for item 38 and 44 (*df* = 2, *p* < .001). However, the $\bar{X}_{ij}^2$ statistics for this pair was 1.216 (*df* = 0.988, *p* = .267). Also, all the item pairs with the largest $R_{ij}$ statistics were not the ones with the largest $\bar{X}_{ij}^2$ statistics. Therefore, there did not seem to be an exact correspondence between the $R_{ij}$ and $\bar{X}_{ij}^2$ statistics. Finally, across the 2346 $z_{ij}$ statistics, the mean was -0.477 and the standard deviation was 3.455. The largest value of the $z_{ij}$ statistics was 37.468 for items 6 and 7, and the smallest value was -13.831 for items 15 and 16. About 55.54% of the item pairs (1303 out of 2346) had $z_{ij}$ statistics that were significant at the *p* = .05 level. As Figure 4 shows, most $z_{ij}$ statistics were around 0.0 and were within the range of -10.0 to 10.0.

In summary, the heuristic adjusted $\chi^2/df$ statistics suggested that the 2PL model fit the MC items well, but the other piecewise fit statistics showed much worse fit when calculated for all possible combination of item pairs. The $\bar{X}_{ij}^2$ and $z_{ij}$ statistics both suggested that item 6 and 7 was the most problematic pair, whereas the results given by $R_{ij}$ statistics were less consistent with the results from the other two. The $M_2$ statistics showed severe misfit for the overall fit assessment, but it was not clear whether this result was due to a real misfit of the model or simply the fact that the model was large.

**World History.** The World History Exam had 70 MC items and 3 CR items. Originally the CR items had 10 categories (i.e., 0 - 9), but they were collapsed into 5 categories (i.e., 0 - 4) to facilitate the IRT and model fit analyses. The final sample size was 15914. The scree plot in Figure 5 showed a dominant first factor that justified the IRT analysis. As Table 10 shows, no extreme b-parameters were found but the a-parameter of item 60 was only 0.041. Most b-parameters were smaller than zero, which means the items were somewhat easier for the test-takers who were able to answer all the questions.

The $M_2$ statistic for the 73 items was 12437.36 with a *df* of 3230 ($p < .0001$). Again it was expected that the overall fit might not be good for large models even with well-fitting items. The adjusted $\chi^2/df$ statistics for a random sample of 72 pairs and 26 triples showed excellent fit of the 2PL model for the MC items: the means were 1.067 (*SD* = 0.465) for item pairs and 1.233 (*SD* = 0.335) for item triples. Only two item pairs (i.e., items 31 & 36; items 58 & 60) had adjusted $\chi^2/df$ statistics slightly larger than 3.0. Therefore, the adjusted $\chi^2/df$ statistics indicated that the 2PL model fit the MC items well.

Piecewise fit was assessed by the 2415 bivariate fit statistics for the MC items (see Table 12). For the $\bar{X}_{ij}^2$ statistics, the mean was 5.271 and the standard deviation was 34.022. The largest value was 1633.313 for item 58 and 59. The second and the third largest were 132.596 for item 66 and 67, and 129.608 for item 57 and 60. All other values were smaller than 100. The median of 2.243 and the histogram in Figure 6 suggested the existence of extreme values. Out of 2415 MC item pairs, 1616 (66.92%) showed acceptable fit at the $p = .05$ level. The $R_{ij}$ statistics had a relatively large mean of 93.573, standard deviation of 215.550, and extreme values such as 3457.431 for items 19 and 29. The majority of $R_{ij}$ statistics were significant at the $p = .05$ level. Again the correspondence of the extremely large values between the $R_{ij}$ and $\bar{X}_{ij}^2$ statistics was not close, although the distributions of the two statistics across 2415 item pairs were both positively skewed (see Figure 7). Finally, the mean of the $z_{ij}$ statistics was -1.233 and the standard deviation was 2.369. The largest value of $z_{ij}$ statistics was 39.045 for items 58 and 59, and the smallest value was -8.322 for items 18 and 58. About 54.20% of the item pairs (1309 out of 2415) had $z_{ij}$ statistics that were not significant at the *p* = .05 level. The histogram in Figure 8 shows that most $z_{ij}$ statistics clustered around zero and fell in the range of -10.0 and 10.0.

Similar to the results of the Physics B Exam, the heuristics adjusted $\chi^2/df$ statistics suggested a much better fit than the other pairwise fit statistics or the overall fit statistic. The $\bar{X}_{ij}^2$ and $z_{ij}$ statistics detected the same item pair (i.e., item 58 and 59) to have extremely large values, whereas the $R_{ij}$ statistics agreed to a lesser extent with these two statistics. Finally the $M_2$ statistics seemed to always show severe misfit when the model was large.

**English Literature.** The English Literature Exam had 55 MC items and 3 CR items. The original 10 categories (i.e., 0 - 9) of the CR items were collapsed into 5 categories (i.e., 0 - 4). The final sample size was 17243. The scree plot in Figure 9 confirmed that item responses were sufficiently unidimensional for IRT analysis. No extreme item parameter estimates were found in Table 11.

The $M_2$ statistic for the 58 items was 19443.39 with a *df* of 2135 (*p* < .0001). The 2PL model again showed excellent fit for the MC items: the mean adjusted $\chi^2/df$ statistics were 1.265 (*SD* = 0.970) for 57 randomly selected item pairs and 1.492 (*SD* = 0.696) for 21 randomly selected item triples. Only two item pairs (i.e., items 13 & 20; items 48 & 51) had an adjusted $\chi^2/df$ statistic between 5 and 7. Piecewise fit was assessed by the 1485 bivariate fit statistics for the MC items. The mean was 10.504 and the standard deviation was 77.581 for the $\bar{X}_{ij}^2$ statistics. The largest values were 2572.416 for item 38 and 39 and 1118.211 for item 13 and 14. Out of

1485 MC item pairs, 924 (62.22%) showed acceptable fit at the $p$ = .05 level. The $R_{ij}$ statistics had a mean of 59.543 and a standard deviation of 158.214. The largest value was 2580.048 for item 38 and 39. Although this pair had the largest $\bar{X}_{ij}^2$ statistics value as well, the correspondence between the $R_{ij}$ and $\bar{X}_{ij}^2$ statistics for the other pairs with extreme values was not very close. Again, the distributions of the $\bar{X}_{ij}^2$ and $R_{ij}$ statistics were both positively skewed as shown in Figure 10 and 11. Across the 1485 MC item pairs, the mean was -0.429 and the standard deviation was 3.275 for the $z_{ij}$ statistics. The largest value was 50.812 for items 38 and 39, and the smallest value was -11.816 for items 53 and 55. More than half of the item pairs (873 out of 1485 or 58.79%) had $z_{ij}$ statistics that were not significant at the $p$ = .05 level and most $z_{ij}$ statistics were within the range of -10.0 and 10.0 (see Figure 12). All three statistics suggested that item pair 38 and 39 had the most serious misfit among all item pairs.

In summary, the heuristic adjusted $\chi^2/df$ statistics had similar means and standard deviations for the randomly selected item pairs and triples across the three AP® Exams. According to the suggested cutoff value, the data fit the models well for all three exams. The fit was slightly better for the World History exam than the English Literature Exam, followed by the Physics B Exam. This is not consistent with our expectation that the Physics B Exam should fit a unidimensional model better than the other two exams. All other pairwise statistics were more sensitive to misfit than the adjusted $\chi^2/df$ statistics with more item pairs flagged as misfit. The $R_{ij}$ statistics had the most extreme values and rejected most of the item pairs. Based on the percentage of item pairs flagged as misfit, the Physics B Exam again showed worse fit than the other two exams. The $M_2$ statistics showed misfit for all three exams, presumably due to the large model and sample sizes. However, as $M_2$ statistics all had different degrees of freedom, it is difficult to compare overall fit directly across different exams.

To examine why the model misfit was not found to be different across the three exams as expected, I ran a series of post-hoc confirmatory factor analyses to check the structure of the data. I fit unidimensional, two-factor, and bifactor models to the data and examined the fit indices such as the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI) and the Root Mean Square Error of Approximation (RMSEA). As Table 13 shows, all three exams had good fit for all three models based on the recommended cutoff in Hu and Bentler (1999): CFI ≥ .95, TLI ≥ .95, and RMSEA ≤ .06. For all three exams, the bifactor model showed slightly better fit than the unidimensional and two-factor models, especially for Physics B and English Literature exams. This probably is because the bifactor model is more flexible and has more free parameters estimated to accommodate the specificities in the data. Therefore, although previous research (Wang et al., 2013) suggested that AP® Exams had different level of unidimensionality, the difference was not large based on the results from the confirmatory factor analysis. Perhaps this is the reason why the fit indices for the unidimensional IRT models were also found to be similar across the three exams.

**International Personality Item Pool (IPIP).**

Figures 13 to 17 showed the scree plots of the 10 IPIP items in both honest (upper panel) and faking (lower panel) conditions across all five dimensions. In general, one dominant factor emerged in all conditions. A potential second factor was suggested in the faking condition for Agreeableness and in the honest condition for Conscientiousness. But in both cases the second factor was relatively weak such that the first eigenvalue accounted for about 2.5 times the variance of the second one (i.e., 31% for the first eigenvalue and 12-13% for the second one). Therefore, the data were sufficiently unidimensional for IRT analyses in all conditions.

Table 14 shows the means, standard deviations, and parameter estimates from the SGR model of IPIP items in both the honest and faking conditions. Across all five dimensions, the item means were higher in the faking conditions than in the honest conditions, except for item 6 of Agreeableness, and item 8 and 9 of Openness. The instructions to fake seemed to work because respondents did tend to endorse higher categories for each item. In both conditions, the a-parameter estimates were moderate to large and the threshold parameters were mostly within the normal range. The first threshold parameters of a few items, such as item 2 of Emotional Stability and items 1, 8, and 9 of Openness, had values smaller than -4.0 in either the honest or faking condition, because very few respondents endorsed the lowest categories of these items. The threshold parameters also tended to be smaller in the faking conditions than the corresponding parameters in the honest conditions. Even the largest threshold parameters were smaller than zero for more than half of the items in the faking condition of all dimensions except Extraversion. Again this confirms that in the faking conditions respondents endorsed higher categories as instructed, which makes the threshold parameters lower than they should have been.

For the model fit analyses, I first evaluated the overall fit by examining the $M_2$ statistics. Then I compared the $M_2$ statistics between the data from the honest and faking conditions to see which data fit the SGR model better. I also examined the means of the piecewise statistics across item pairs to compare model fit between the two conditions. For adjusted $\chi^2/df$, $M_{ij}$, $\bar{X}^2_{ij}$, and $R_{ij}$ statistics, smaller values indicated a better fit. For $z_{ord}$ statistics, I calculated their means from the absolute values of the original $z_{ord}$ statistics to reflect how the values deviate from the mean zero. The smaller the means of the absolute values for $z_{ord}$ statistics, the better the fit on average for the entire scale.

***Fit Statistics for Agreeableness.*** The overall $M_2$ statistic for Agreeableness was 2227.34 (*df* = 395, *p* < .0001) in the honest condition. Because no respondent endorsed the first category of item 2, the $M_2$ statistics could not be calculated for the full scale in the faking condition due to the different numbers of categories for each item. Instead, I excluded item 2 from the scale in both conditions to compare the overall fit. When the second item was removed from scale for the model fit analysis, $M_2$ statistics were 2056.25 (*df* = 315, *p* < .0001) for the honest condition and 1592.55 (*df* = 315, *p* < .0001) for the faking condition. Therefore, when item 2 was excluded from the fit analyses, the overall fit was better for the faking condition than for the honest condition, although in both conditions the overall fit was less than acceptable.

Tables 15 and 16 show the pairwise fit statistics in the honest and faking conditions, respectively. Again, item 2 was excluded from the analysis in the faking condition, so no piecewise fit statistics were calculated for item pairs that involved item 2 (shown as NA in Table 16). In both conditions, almost all item pairs showed misfit at the *p* = .05 level. When the fit statistics between the two conditions were compared, the fit was better for the honest condition than the faking condition based on the means of the following $\chi^2$-related fit statistics across item pairs: adjusted $\chi^2/df$ = 5.97, $M_{ij}$ = 44.47 (*df* = 7), and $\bar{X}_{ij}^2$ = 53.04 (mean *df* = 8.90) for the honest condition; and adjusted $\chi^2/df$ = 12.13, $M_{ij}$ = 51.76 (*df* = 7), $\bar{X}_{ij}^2$ = 65.08 (mean *df* = 8.84) for the faking condition. For $R_{ij}$ and $z_{ord}$ statistics, the pattern was reversed: the fit was better for the faking condition ($R_{ij}$ = 105.06, mean *df* = 12.78; $z_{ord}$ = 2.21) than the honest condition ($R_{ij}$ = 159.81, mean *df* = 12.51; $z_{ord}$ = 3.95). Therefore, while adjusted $\chi^2/df$, $M_{ij}$ and $\bar{X}_{ij}^2$ fit statistics suggested a better fit for the honest condition, other pairwise fit statistics and the overall fit statistics showed the fit was better for the faking condition.

To examine whether the absence of item 2 in the faking condition but not in the honest condition might affect the fit statistics, I also evaluated the pairwise fit statistics in the honest condition without item 2. The statistics all increased, but the results did not reverse the previous conclusion: adjusted $\chi^2/df$ = 8.53, $M_{ij}$ = 49.07 ($df$ = 7), $\bar{X}_{ij}^2$ = 57.09 (mean $df$ = 8.90), $R_{ij}$ = 128.33 (mean $df$ = 12.42), and $z_{ord}$ = 3.98.

***Fit Statistics for Conscientiousness.*** The overall $M_2$ statistics for Conscientiousness were 1582.02 ($df$ = 395, $p$ < .0001) in the honest condition and 1471.42 ($df$ = 395, $p$ < .0001) in the faking condition. Similar to Agreeableness, the overall fit was not good but better in the faking condition than in the honest condition. Table 17 and 18 show the piecewise fit statistics for the 45 item pairs in the honest and faking conditions respectively. The majority of the item pairs showed misfit at the $p$ = .05 level in both conditions. Consistent with the results of the overall fit, the pairwise fit was also better for the faking condition (adjusted $\chi^2/df$ = 0.12; $M_{ij}$ = 18.84, $df$ = 7; $\bar{X}_{ij}^2$ = 24.85, mean $df$ = 9.32; $R_{ij}$ = 45.70, mean $df$ = 14.00; and $z_{ord}$ = 0.92) than the honest condition (adjusted $\chi^2/df$ = 4.11; $M_{ij}$ = 37.14, $df$ = 7; $\bar{X}_{ij}^2$ = 44.00, mean $df$ = 8.91; $R_{ij}$ = 71.29, mean $df$ = 12.80; and $z_{ord}$ = 1.71). Therefore, both the overall and pairwise fit statistics suggested that data from the faking condition had a better fit than data from the honest condition for the Conscientiousness scale.

***Fit Statistics for Extraversion.*** Unlike Agreeableness and Conscientiousness, the overall fit for Extraversion was better in the honest condition ($M_2$ = 1628.41, $df$ = 395, $p$ < .0001) than in the faking condition ($M_2$ = 2654.00, $df$ = 395, $p$ < .0001). The mean piecewise fit statistics also fully supported this result: adjusted $\chi^2/df$ = 5.30, $M_{ij}$ = 36.22 ($df$ = 7), $\bar{X}_{ij}^2$ = 43.09 (mean $df$ = 9.18), $R_{ij}$ = 84.48 (mean $df$ = 14.20) and $z_{ord}$ = 1.28 for the honest condition; and adjusted $\chi^2/df$

= 11.92, $M_{ij}$ = 54.33 (*df* = 7), $\bar{X}^2_{ij}$ = 63.61 (mean *df* = 9.07), $R_{ij}$ = 99.66 (mean *df* = 13.24), and $z_{ord}$ = 2.61 for the faking condition. Therefore, both the overall and pairwise statistics showed that the fit was better for the honest condition than for the faking condition.

***Fit Statistics for Emotional Stability.*** The overall $M_2$ statistics for Emotional Stability were 1763.24 (*df* = 395, *p* < .0001) in the honest condition and 1642.84 (*df* = 395, *p* < .0001) in the faking condition. Therefore, the overall fit was worse for the honest condition than for the faking condition. The mean piecewise statistics showed comparable fit for the two conditions: adjusted $\chi^2/df$ = 7.04, $M_{ij}$ = 47.66 (*df* = 7), $\bar{X}^2_{ij}$ = 55.53 (mean *df* = 9.03), $R_{ij}$ = 80.35 (mean *df* = 14.11) and $z_{ord}$ = 1.85 for the honest condition; and adjusted $\chi^2/df$ = 7.11, $M_{ij}$ = 45.54 (*df* = 7), $\bar{X}^2_{ij}$ = 52.90 (mean *df* = 9.71), $R_{ij}$ = 56.01 (mean *df* = 14.51), and $z_{ord}$ = 1.51 for the faking condition. While adjusted $\chi^2/df$ suggested that the fit was slightly better for the honest condition, all others suggested the opposite. But the differences in the statistics of the two conditions were quite small except for $R_{ij}$. Therefore, as the overall and most of the piecewise statistics showed, the fit was only slightly better in the faking condition than in the honest condition for the Emotional Stability scale and the fit statistics were almost comparable in the two conditions.

***Fit Statistics for Openness.*** The overall $M_2$ statistics for Openness were 1927.43 (*df* = 395, *p* < .0001) in the honest condition and 1260.34 (*df* = 395, *p* < .0001) in the faking condition. Again, the fit was better for the faking condition than for the honest condition. The mean piecewise statistics supported that the data from the faking condition showed a better fit than data from the honest condition, except for the $z_{ord}$ statistics: adjusted $\chi^2/df$ = 7.41, $M_{ij}$ = 32.75 (*df* = 7), $\bar{X}^2_{ij}$ = 43.21 (mean *df* = 8.83), $R_{ij}$ = 135.40 (mean *df* = 12.49) and $z_{ord}$ = 1.85 for the honest condition; and adjusted $\chi^2/df$ = 6.64, $M_{ij}$ = 28.99 (*df* = 7), $\bar{X}^2_{ij}$ = 36.32 (mean *df* = 8.81),

$R_{ij}$ = 91.74 (mean $df$ = 13.02), and $z_{ord}$ = 1.98 for the faking condition. Again the differences in these statistics were small, and both the overall fit and the majority of the pairwise fit statistics suggested that the fit was better in the faking condition than in the honest condition.

As the overall and pairwise fit statistics showed misfit for items in both honest and faking conditions, it is important to explore whether the misfit is caused by the violation of the unidimensionality assumption or the by the misuse of the SGR model. Therefore, a series of post-hoc confirmatory factor analyses were conducted to examine the fit for a unidimensional model across all personality dimensions and conditions. Fit indices in Table 25 suggested that the fit for a unidimensional model was marginal to acceptable for item-level categorical data across the personality dimensions and conditions, with the honest conditions of Conscientiousness and Openness showing notable problems. Item responses from the faking condition had a better fit for the unidimensional model than those from the honest condition except for Extraversion and Agreeableness, which is mostly consistent with the results from the fit statistics for the SGR model. Interestingly, although the fit for a unidimensional model was good to excellent for the faking conditions of Conscientiousness (CFI = 0.983, TLI = 0.978, RMSEA = 0.082) and Emotional Stability (CFI = 0.991, TLI = 0.989, RMSEA = 0.062), the fit for the SGR model for these two conditions was still less than acceptable (see Table 18 and Table 22). This suggests that the misfit is might be  caused more by the misspecification for the SGR model than by the misfit for the unidimensional model. Therefore, although the assumption of unidimensionality was generally met for the IRT analysis, it appears that an alternative IRT model is needed to fit the personality data better than the SGR model.

In summary, the overall $M_2$ statistics showed severe misfit for both conditions across all five dimensions. Because the model was not large, this result suggested that the SGR model did not fit the personality data well. Contrary to the expectation, the overall fit was better for the faking condition than for the honest condition across all dimensions except Extraversion. The results from the pairwise statistics also partially supported this conclusion, such that the averaged fit across item pairs was clearly better for the faking condition for Conscientious and was almost comparable across two conditions for Emotional Stability and Openness. Therefore, although an increase in item means was observed for almost all items, faking did not change the underlying psychological process that respondents went through to fill out the personality inventory which was reflected by the fitted model.

### Counterproductive Work Behavior (CWB) Scale.

Figure 18 shows the scree plots for the 7 CWB-I and 12 CWB-O items, respectively. In the upper panel, the scree plot of 7 CWB-I items showed a dominant first factor. The first eigenvalue accounted for 56.15% of the total variance. The scree plot for the CWB-O items, on the other hand, showed a secondary factor, although the first factor was still dominant. The first eigenvalue accounted for 44.51% of the total variance, which is much larger than the 13.28% accounted for by the second eigenvalue. Therefore, it is concluded that the responses to the CWB-I and 12 CWB-O items were sufficiently unidimensional for IRT analyses. The internal consistency was .875 for the CWB-I scale and .876 for the CWB-O scale.

Descriptive statistics and parameter estimates for each of the 7 CWB-I and 12 CWB-O items are shown in Table 26. The means were all below 2.0 for a scale of 0 to 5, which means the data were positively skewed. The values of the a-parameters were all moderate to large, showing good discriminating properties. The b-parameters were mostly positive and had large values, which confirmed that the responses were positively skewed. These results were consistent with those obtained from the original analyses in Carpenter and Cao (2013).

The overall $M_2$ statistics were 682.83 (*df* = 518, *p* < .0001) for CWB-I scale and 2259.74 (*df* = 1638, *p* < .0001) for CWB-O scale, which indicate less than acceptable fit. Piecewise statistics were examined for the items that fit less well. For the 21 item pairs in the CWB-I scale (see Table 27), 7 pairs had adjusted $\chi^2/df$ statistics larger than 3.0 (*Mean* = 1.23, *SD* = 5.04). The $M_{ij}$ statistics (*Mean* = 30.15, *SD* = 9.56) identified 6 pairs with misfit at the *p* = .05 level: item pairs (1, 2), (1, 3), (1, 4), (1, 5), (4, 6), and (4, 7). The $\bar{X}_{ij}^2$ statistics (*Mean* = 37.23, *SD* = 14.02) detected one more pair (i.e., item 1 and 6) than the $M_{ij}$ statistics. The $R_{ij}$ statistics (*Mean* = 54.61, *SD* = 18.71) found the most pairs of misfit items (12 in total) while $z_{ord}$ statistics (*Mean* = -0.41, *SD* = 1.01) detected the least (only one). The results from all piecewise fit statistics in Table 27 were quite consistent. Item pairs involved item 1 were flagged by at least three fit statistics, so were item pairs (4, 6) and (4, 7). Therefore, items 1 and 4 seemed most likely to be the source of misfit. In general, the $R_{ij}$ statistics tended to flag more pairs of misfit items than other statistics while $z_{ord}$ statistics was the opposite.

The pattern was less consistent across the piecewise fit statistics for the 66 items pairs in the 12-item CWB-O scale (see Table 28). Overall, the adjusted $\chi^2/df$ statistics (*Mean* = -0.65, *SD* = 3.37) detected 12 pairs, the $M_{ij}$ statistics (*Mean* = 30.75, *SD* = 8.56) 18 pairs, the $\bar{X}_{ij}^2$ statistics (*Mean* = 41.04, *SD* = 10.96) 38 pairs, and the $z_{ord}$ statistics (*Mean* = 1.69, *SD* = 1.68) 28 pairs. The $R_{ij}$ statistics (*Mean* = 75.78, *SD* = 48.68) found almost all pairs to have misfit (59 out of 66). Item pairs that were flagged by all five statistics were: (1, 9), (2, 4), (2, 8), (2, 11), (4, 5), (4, 8), and (8, 11). Therefore, the problematic items were most likely to be items 2, 4, 8 and 11.

In conclusion, the two CWB scales both showed severe overall misfit, probably due to the positive skewness of the data. Pairwise fit statistics suggested that items 1 and 4 in the CWB-I scale, and items 2, 4, 8, and 11 in the CWB-O scale might be the source of the misfit. Interestingly, these 6 misfit items were the items with the highest item means in their scales, which means the misfit items were actually the items with less positively skewed data. However, perhaps the rest of the more skewed items dominated the model for the entire scale, which made the less skewed items fit less well. A content review of these items showed that the deviant workplace behaviors described in some of these 6 items did seem more frequent than those described in items with more extreme means. For example, two of these items were, "Made fun of someone at work" (CWBI-1) and "Spent too much time fantasizing or daydreaming instead of working" (CWBO-2). These two behaviors seem less severe and more commonly encountered compared with the more serious ones such as "Acted rudely toward someone at work" (CWBI-6) and "Taken property from work without permission" (CWBO-1). Therefore, it is not surprising that some of the items that describe less severe CWB had higher means than others simply because of their higher base rates. Moreover, a previous study (Robinson & Bennett, 1995) mapped various deviant workplace behaviors along two dimensions: interpersonal versus organizational, and minor versus serious. This provides further empirical evidence that the few items describing less severe CWB might function differently from those with more extreme means and thus were detected by the fit statistics.

# DISCUSSION

In this paper, I conducted a simulation study and several analyses on real data to compare the performance of the heuristic adjusted $\chi^2/df$ statistics and the recently developed fit statistics that are based on the limited-information method. Results from the simulation studies showed that the overall $M_2$ statistic was sensitive to misfit caused by multidimensionality across different sample sizes and models except when both the model and the sample size were small. The average adjusted $\chi^2/df$ statistics across item pairs also suggested better model-data fit in the one-factor conditions than in the two-factor conditions when the sample size was at least 1000. But in no conditions did the adjusted $\chi^2/df$ statistics exceed the suggested cutoff value of 3.0 to flag any misfit item pairs. The patterns were similar for pairwise fit statistics that are based on the limited-information method: the $\bar{X}_{ij}^2$, $R_{ij}$, and $z_{ij}$ statistics all showed worse fit in the two-factor conditions than in the one-factor conditions, but the detection rates were quite low and the mean statistics rarely exceeded the critical value of the chi-square (with its corresponding *df*) or standard normal (two-tailed) distributions at the $p = .05$ level even in the two-factor conditions. All four pairwise statistics (i.e., the adjusted $\chi^2/df$, $\bar{X}_{ij}^2$, $R_{ij}$, and $z_{ij}$ statistics) examined here showed consistent results when detecting misfit, indicated by the moderate to large correlation coefficients among the four, especially when the sample size was large. The adjusted $\chi^2/df$ and $\bar{X}_{ij}^2$ statistics had the closest correspondence, whereas the $R_{ij}$ statistics correlated the least with other fit statistics. The $R_{ij}$ statistics also seemed to be the least stable with large extreme values. Similar to the pairwise adjusted $\chi^2/df$ statistics, the mean adjusted $\chi^2/df$ statistics for item triples also indicated that responses from the one-factor conditions fit better than those for the two-factor conditions; but again none of the statistics exceed 3.0. The $M_n$ statistic, on the contrary, was not very effective in detecting misfit at least for small models with 10 dichotomous items. The correspondence was also quite low between the two piecewise statistics for items triples.

In summary, the results from the simulation study suggested that $M_2$ was a powerful statistic in evaluating overall fit. All pairwise fit statistics were sensitive to the misfit caused by the moderate multidimensionality in the response data, although either the sample size or the effect size (i.e., level of multidimensionality) needed to be larger for the limited-information statistics to be significant at the $p$ = .05 level or for the heuristic adjusted $\chi^2/df$ statistics to be larger than the suggested cutoff value. Results from all pairwise fit statistics also showed consistent patterns: the fit was the best in the one-factor conditions and the worst in the “Two-factor, MC only” conditions. While it is conceivable that the “Two-factor, MC only” conditions showed the worst fit because responses in these conditions were multidimensional, it is interesting to notice that the fit was worse than the one-factor conditions in the “Two-factor, MC and CR” conditions where the MC item responses were actually unidimensional. Perhaps, due to the concurrent estimation of the MC item parameters with the CR item parameters that were manipulated to load on a different factor, the parameter estimates of the MC items were “contaminated” by this second factor and thus caused the worse fit. The misfit was even more obvious when the number of MC items was small and the sample size was large. This probably was because when the number of MC items was small compared with the number of CR items (which was held constant in the simulation), the parameter estimation of the MC items was influenced to a larger extent by the different factor of the CR items. When the number of MC items was large, such influence was minimal and all pairwise fit statistics were almost the same as those in the one-factor conditions. Finally, due to the computation limitation the piecewise fit statistics for item triples were examined only for the smallest models with 10 MC items.

Therefore, the inconsistent results should be interpreted with caution. However, piecewise fit statistics for item triples tend to become less useful as the model becomes larger for two reasons. First, there are too many combinations of all possible triplets to calculate fit statistics for all of them. Secondly, when the source of misfit needs to be detected, it takes more time to spot the problematic items by examining triplets than pairs.

In addition to evaluating the performance of the overall and piecewise fit statistics in the simulation study, I also applied the fit statistics to real data to solve practical problems. The first application was to the three AP® Exams, which were found to have different levels of unidimensionality across test formats in the past. By evaluating overall and piecewise fit statistics, I expected to find better fit for the more unidimensional exams and worse fit for the less unidimensional exams. The overall $M_2$ statistics showed severe misfit for all three exams, probably because each exam had approximately 50 to 70 test items in the model and the sample sizes were quite large. Due to the different degree of freedom for the $M_2$ statistic in each exam, overall fit among the three exams could not be compared directly. Instead, I examined the mean fit statistics across all MC item pairs in the three exams. The adjusted $\chi^2/df$ statistics showed good fit for MC items pairs in all three exams. The limited-information pairwise fit statistics flagged more item pairs as misfit than the adjusted $\chi^2/df$ statistics in each exam. Contrary to expectation, the MC items in Physics B Exam actually had a slightly worse fit than the MC items in the other two exams based on the limited-information pairwise fit statistics. The results from a post-hoc confirmatory factor analysis showed that the fit indices for both unidimensional and bi-factor models did not differ much across the three exams; all the fits were excellent. Therefore, the three exams did not necessarily have different levels of unidimensionality as a previous study

(Wang et al., 2013) suggested. This might explain why the fit was almost comparable across all three exams.

The other two applications focused on misfit detection of the overall and pairwise fit statistics for polytomous IRT models. Overall and average pairwise fit were compared between honest and faking conditions for the Big Five personality traits. Because faking good in responding to a personality inventory could potentially increase item and scale means, change the item response distribution, or even change the underlying response process, it was expected that the item responses should fit the IRT model better in the honest condition than in the faking condition if the personality inventory was developed in a way to reflect the latent trait based on the honest rather than distorted responses. The results from the overall and piecewise fit analyses showed that item responses did not fit the most popular polytomous model in either condition. The overall fit indicated that the SGR model could not accurately describe the data, and many item pairs were flagged as having misfit by the piecewise fit statistics. Moreover, item responses did not fit the model better in the honest condition than in the faking condition. In fact, the overall fit was better for the faking condition in four dimensions and the piecewise fit statistics were clearly better in the faking condition for Conscientiousness. Therefore, the expectation that item responses from the honest condition should fit better because they should reflect the correct model was probably overly simplified. While a more accurate model is needed to describe item responses from personality inventory in general, it is also important to understand respondents' underlying process of faking good and to examine how faking affects the item responses and their model fit.

The last application focused on the misfit detection of positively skewed ordinal data. Given the relatively small models in this application, the misfit detected by the overall fit statistic should be at least partially due to the positive skewness of the data in both CWB scales. The pairwise fit assessment revealed a few items that seemed to cause the large values in some item pairs. Interestingly, these "problematic" items were actually items describing less extreme behaviors and thus with less skewed responses in each scale. Perhaps these items did not fit the specified model that was dominated by items with more extreme responses and thus stood out as misfit items.

Based on the results from the simulation and real data applications, I concluded that both the new limited-information statistics and the heuristic adjusted $\chi^2/df$ statistics should be considered when examining model-data fit for IRT analysis. The overall $M_2$ statistic is effective in detecting misfit caused by multidimensionality in small to medium-sized models. However, when the sample sizes and/or the model are large, $M_2$ statistics almost always reject the model even for well-fitting items. This is conceivable because it is almost impossible to find exact fit for all possible response patterns from a large number of items. When the $M_2$ statistic becomes "too powerful", average piecewise fit statistics can be examined for overall assessment instead. For example, the adjusted $\chi^2/df$ statistics are to some extent sensitive to misfit caused by multidimensionality. However, when the sample size is smaller than 3000, the average adjusted $\chi^2/df$ statistics for item pairs and triples almost never exceed the suggested cutoff value for data with moderate multidimensionality. Other pairwise fit statistics are more effective in detecting misfit item pairs. Results from all pairwise fit statistics including the adjusted $\chi^2/df$ statistics are highly consistent, except for $R_{ij}$ when the model and/or sample size is small. This is consistent with the conclusion in Liu and Maydeu-Olivares (2014) that the $R_{ij}$ statistic tends to have large sampling variance due to the computation process of its component. Finally $M_n$ statistics does not seem to detect misfit effectively in small models. The adjusted $\chi^2/df$ statistics for item triples can be an alternative if needed, although the interpretation of misfit for item triples is usually less straightforward than that for item pairs.

## Implications

The current study has several theoretical and practical implications. First, the study provided a thorough review of both the heuristic and the limited-information fit statistics. Based on the literature, the limited-information fit statistics are more statistically rigorous and thus are expected to show superior performance to the heuristic statistics that have unknown sampling distribution. Secondly, the performance of these two types of fit statistics was compared in a simulation study. Although all piecewise fit statistics showed consistent results when detecting multidimensionality, limited-information fit statistics tended to have higher power than the heuristic fit statistics in general. Lastly, these two types of fit statistics were applied to three large-scale assessments and two rating scales in organizational settings. The real data applications suggested that the overall fit statistics based on limited information method should be used with caution because they tended to reject well-fitting items.

A few recommendations can be provided based on the results from the current study. Both the overall $M_2$ fit statistic and the averaged adjusted $\chi^2/df$ statistics across item pairs/triplets can be examined for overall fit assessment. When the model is not large (i.e., with fewer than 40 items), the overall $M_2$ fit statistic can distinguish well-fitting items from those that fit less well. But when the model is large, it is very difficult to observe exact fit for a large number of possible response patterns and thus the $M_2$ fit statistic seems to always reject the fitted model. When the sample size is larger than 3000, the overall fit of the data can also be assessed by evaluating the average adjusted $\chi^2/df$ statistics for item pairs and triples. But when the sample size is smaller than 3000, the average adjusted $\chi^2/df$ statistics are usually not effective in detecting misfit based on the suggested cutoff value. While the pairwise adjusted $\chi^2/df$ statistics can also be used to detect item pairs that fit less well than others, the pairwise fit statistics based on limited-information method are more effective especially when the sample size is small. Consistent with the conclusion in Liu and Maydeu-Olivares (2014), statistics that utilize information matrix (e.g., $\bar{X}_{ij}^2$, $R_{ij}$, and $z_{ij}$ or $z_{ord}$) have more power to detect misfit caused by multidimensionality and thus flagged more item pairs in the real data sets than those that do not require the computation of the information matrix (e.g., $M_{ij}$). Among these more powerful statistics, $R_{ij}$ statistics tend to have large sampling variance and thus are more likely to reject well-fitting items due to random error. Liu and Maydeu-Olivares (2014) recommended $z_{ij}$ or $z_{ord}$ statistics based on their type I error rate and power when the observed information matrix is used (as in Mplus; Muthén & Muthén, 2012). Moreover, it is easier to compare $z_{ij}$ or $z_{ord}$ statistics across item pairs because $\bar{X}_{ij}^2$ statistics tend to have slightly different degrees of freedom estimated for each item pair. Finally, piecewise fit statistics for item triples are not only less useful than the pairwise fit statistics, but are also more computational intensive to calculate for all possible item triplets when the model is large. Therefore, pairwise fit statistics are recommended to detect problematic items that are involved in misfit item pairs.

What should researchers or practitioners do when they find model misfit? Here are a few procedures to follow, depending on the characteristics of the items and the scales. If item pairs with extremely large values of fit statistics are detected in the piecewise fit assessment, a content review of these problematic item pairs should be performed by the subject matter experts (SMEs). If the content of these problematic item(s) is confirmed to be different from others (i.e., misfit caused by multidimensional data structure) or to be repetitious among item pairs (i.e., misfit caused by local dependency), the problematic item(s) should be removed to reassess the model fit if the scale is long (e.g., more than 10 items). If the scale is already very short (e.g., fewer than 10 items), removing items might cause a loss of measurement accuracy of the scale. Thus, the problematic item(s) should be revised and reevaluated with new data. Oftentimes, however, it is not clear even to the SMEs why items or item pairs show misfit. In this case, item removal has to rely entirely on statistics. Again if a scale is long, it is recommended that problematic items or item pairs be removed from the scale and model fit reassessed. If any improvement is detected in fit statistics, it is then confirmed that removing the problematic items was appropriate. When no extreme fit statistics are found across all item pairs, it is hard to tell which item pairs cause the problem. Instead, it is highly likely that the data cannot be accurately described by the fitted model in general. In this situation, it is recommended that the data should be fitted with a new statistical model, such as a multidimensional IRT model or a bifactor model.

## Limitations & Future Directions

The current study has some limitations that need to be addressed in future research. For the simulation study, it would be ideal if conditions with skewed item responses could be included in the analyses. As we can see in the real data applications, both positively and negatively skewed data seemed to show misfit to some extent. However, it is not completely certain that the misfit was caused by the skewness because there could be many other unknown factors such as local dependency that influenced the fit in the real data. With more control of the data structure, results from simulation studies would show us a clearer picture of how skewness affects misfit detection by both types of the fit statistics. In addition, more replications in each condition might improve the accuracy of the descriptive statistic, power, and type I error rate of the limited-information fit statistics. When faster computers with larger memory capacity become available, it is recommended that 500 to 1000 replications be conducted.

For overall fit assessment for large models, discrepancies between the true and fitted models can almost always be detected, especially with large model and sample sizes. Therefore, it might not be necessary or even useful to examine exact fit. Rather, approximate fit, which is concerned with "whether the approximation provided by the fitted model is good enough", should be evaluated to solve the over-rejection problem (Maydeu-Olivares & Joe, 2014). For example, in the applications to the three large-scale assessments, the overall $M_2$ fit statistics showed severe misfit ($p < .0001$) for models with about 70 items and sample sizes of about 20000. However, the approximate fit index, RMSEA, was only 0.02 for Physics B and English Literature, and 0.01 for World History. Based on the suggested cutoff of 0.05 as close fit in Maydeu-Olivares and Joe (2014), the 2PL and SGR models showed a close to excellent fit for the item responses. Therefore, in future applications both exact and approximate fit for overall fit assessment in IRT analyses should be examined to avoid over-rejecting well-fitted models, especially when the model includes many variables and the sample size is large.

Finally, it would be interesting to develop limited-information fit statistics for more advanced IRT models such as the ideal point model. As the ideal point model is becoming more and more widely used in the psychometric analysis of personality inventories (e.g., Drasgow, Chernyshenko, & Stark, 2010), it is important to extend the limited-information statistics to the fit assessment for this more mathematically sophisticated model. For example, in the application to IPIP data of the current study, the SGR model did not fit the data well in either the honest or faking conditions. This means an alternative model is needed for item responses to personality inventory in general. In the original study (Cao et al., 2014), the averaged heuristic fit statistics across item pairs and triples showed a better or at least comparable fit for the ideal point model than for the dominance model in both the honest and faking conditions for all five dimensions. If the limited-information fit statistics are available for the ideal point model, they can be applied together with the heuristic fit statistics to personality inventory or attitude rating scales to produce additional empirical evidence for model fit/misfit.

# CONCLUSION

Model-data fit assessment is an important step in the IRT analysis to ensure that the results are interpretable and trustworthy. An overall fit assessment should be evaluated first and then piecewise fit analysis can be conducted to detect the item pairs that fit less well. In spite of their limitations, both the heuristic fit statistics and the limited-information fit statistics provide important information about misfit detection. Researchers and practitioners should choose the fit statistics that possess the best psychometric properties for their data, and examine the model-data fit before they proceed to interpret their results, revise their scale or items, or search for a new model to fit the data.

# REFERENCES

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Asparouhov, T., & Muthén, B. O. (2010). Simple second order chi-square correction scaled chi-square statistics. Technical Report. Los Angeles, CA: Muthén and Muthén.

Bennett, R. J., & Robinson, S. L. (2000). Development of a measure of workplace deviance. *Journal of Applied Psychology, 85*, 349-360.

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

Cai, L. (2012). flexMIRT$^{TM}$ version 1.88: A numerical engine for multilevel item factor analysis and test scoring. [Computer software]. Seattle, WA: Vector Psychometric Group.

Cao, M., Tay. L., Luo. J., & Drasgow, F. (May, 2014). *Does faking influence the process underlying responses to personality measures?* Poster to be presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Carpenter, N. C. (2013). *The substantive validity of work performance measures: Implications for relationships among work behavior dimensions and construct-related validity.* (Unpublished doctoral dissertation). Texas A&M University, College Station, TX.

Carpenter, N. C., & Cao, M. (April, 2013). *Application of item response theory to counterproductive work behavior (CWB).* Poster presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*(4), 465-476.

Drasgow, F., Levine, M. V., Tsien, S., Williams, B. A., & Mead, A. D. (1995). Fitting polytomous item response models to multiple-choice tests. *Applied Psychological Measurement, 19*, 145-165.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to mutidimensional data. *Applied Psychological Measurement*, *7*(2), 189-199.

Embretson, S. E., & Reise, S. P. (2000). *Item response theories for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4,* 26-42.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Guo, Y., Tay, L. & Drasgow, F. (2010, April). *A comparison of IRT item fit statistics for dichotomous responses.* Poster presented at 2010 Society of Industrial and Organizational Psychology Conference, Atlanta, GA.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 3,* 424-453.

Koehler, K. J., & Larntz, K. (1980). An empirical investigation of goodness-of-fit statistics for sparse multinomials. *Journal of the American Statistical Association*, *75*(370), 336-344.

Liu, Y., & Maydeu-Olivares, A. (2014). Identifying the source of misfit in item response theory models. *Multivariate Behavioral Research, 49*(4), 354-371.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives, 11*(3), 71-101.

Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and goodness-of-fit testing in $2^n$ contingency tables: A unified framework. *Journal of the American Statistical Association, 100,* 1009-1020.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*(4), 713-732.

Maydeu-Olivares, A., & Joe, H. (2008). An overview of limited information goodness-of-fit testing in multidimensional contingency tables. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New trends in psychometrics* (pp. 253-262). Tokyo, Japan: Universal Academy Press.

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research, 49*(4), 305-328.

Maydeu-Olivares, A., & Liu, Y. (2012). Item diagnostics in multivariate discrete data. *Manuscript under review*.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide*. (7th ed.). Los Angeles, CA: Muthén & Muthén.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika, 61*, 509-528.

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal, 38*(2), 555-572.

Samejima, F. (1969). Estimation of latent ability using response pattern of graded scores. *Psychometrika Monograph, 17*, 1-100.

Swaminathan, H., Hambleton, R. K., & Rogers, H. J. (2007). Assessing the fit of item response theory. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of statistics, volume 26: Psychometrics* (pp. 683-718). Amsterdam: Elsevier North-Holland.

Thissen, D., Chen, W-H., & Bock, R. D. (2003). MULTILOG 7 for Windows: Multiple-category item analysis and test scoring using item response theory [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Tollenaar, N., & Mooijaart, A. (2003). Type I errors and power of the parametric bootstrap goodness-of-fit test: Full and limited information. *British Journal of Mathematical and Statistical Psychology, 56*(2), 271-288.

Van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika, 47*, 123-140.

von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of Psychological Research*, *2*(2), 29-48.

Wang, W., Drasgow, F., & Liu, L. (2013). Classification accuracy of mixed format tests: A bi-factor approach. Manuscript submitted for publication.

Williams, B. A., & Levine, M. V. (1993). *FORSCORE: A computer program for nonparametric item response theory*. Unpublished manuscript, University of Illinois at Urbana-Champaign, Champaign, IL.

Windsor, M., Jeon, G., Cao, M. & Drasgow, F. (2013). *Analysis of the AP Calculus BC 2010, 2011, and 2012.* College Board Research Report.

## TABLES

**Table 1**. Means, standard deviations (in parentheses), and detection rates of $M_2$ statistics for 100 replications.

| **One factor, MC items only** | | | | |
|---|---|---|---|---|
| | Sample Size (Number of Examinees) | | | |
| Model Size | 200 | 500 | 1000 | 3000 |
| 10MC, *df* = 35<br>$X^2_{.05} = 49.80$ | 35.21 (7.83)<br>0.05 | 35.61 (8.20)<br>0.04 | 34.20 (7.43)<br>0.03 | 35.78 (8.95)<br>0.09 |
| 20MC, *df* = 170<br>$X^2_{.05} = 201.42$ | 168.51 (19.03)<br>0.06 | 170.23 (18.24)<br>0.05 | 169.73 (17.30)<br>0.02 | 166.89 (17.59)<br>0.01 |
| 40MC, *df* = 740<br>$X^2_{.05} = 804.40$) | 743.59 (34.88)<br>0.04 | 740.31 (33.52)<br>0.05 | 747.39 (38.65)<br>0.07 | 737.71 (36.20)<br>0.02 |
| **One factor, MC and 5 CR items** | | | | |
| | Sample Size (Number of Examinees) | | | |
| Model Size | 200 | 500 | 1000 | 3000 |
| 10 MC, *df* = 390<br>$X^2_{.05} = 437.05$) | 396.78 (27.63)<br>0.10 | 388.83 (26.11)<br>0.03 | 393.36 (30.14)<br>0.08 | 392.36 (26.16)<br>0.04 |
| 20 MC, *df* = 725<br>$X^2_{.05} = 788.75$ | 731.73 (35.79)<br>0.02 | 725.66 (42.15)<br>0.05 | 728.13 (37.11)<br>0.04 | 723.69 (41.88)<br>0.07 |
| 40 MC, *df* = 1695<br>$X^2_{.05} = 1791.89$ | 1709.00 (60.04)<br>0.08 | 1704.50(54.46)<br>0.03 | 1698.50 (66.77)<br>0.10 | 1692.20 (59.88)<br>0.05 |
| **Two factors, MC items only** | | | | |
| | Sample Size (Number of Examinees) | | | |
| Model Size | 200 | 500 | 1000 | 3000 |
| 10MC, *df* = 35<br>$X^2_{.05} = 49.80$ | **37.78** (9.51)<br>0.11 | 56.52 (14.25)<br>0.65 | 56.73 (12.92)<br>0.71 | 113.22 (21.46)<br>1.00 |
| 20MC, *df* = 170<br>$X^2_{.05} = 201.42$ | 201.80 (23.76)<br>0.45 | 244.89 (28.31)<br>0.94 | 366.24 (48.21)<br>1.00 | 685.28 (69.46)<br>1.00 |
| 40MC, *df* = 740<br>$X^2_{.05} = 804.40$) | 867.40 (51.95)<br>0.89 | 1074.30 (86.34)<br>1.00 | 1409.00 (126.09)<br>1.00 | 2745.90 (183.72)<br>1.00 |
| **Two factors, MC and 5 CR items** | | | | |
| | Sample Size (Number of Examinees) | | | |
| Model Size | 200 | 500 | 1000 | 3000 |
| 10 MC, *df* = 390<br>$X^2_{.05} = 437.05$) | **425.15** (35.79)<br>0.33 | 474.78 (36.61)<br>0.86 | 603.49 (54.50)<br>1.00 | 851.25 (72.77)<br>1.00 |
| 20 MC, *df* = 725<br>$X^2_{.05} = 788.75$ | 810.45 (52.21)<br>0.63 | 918.00 (51.91)<br>1.00 | 1139.50 (73.30)<br>1.00 | 2048.50 (125.51)<br>1.00 |
| 40 MC, *df* = 1695<br>$X^2_{.05} = 1791.89$ | 1806.80 (73.45)<br>0.56 | 1959.80 (80.62)<br>1.00 | 2208.70 (99.20)<br>1.00 | 3247.10 (130.12)<br>1.00 |

Note. The detection rate is the type I error rate for the unidimensional models (the top two conditions in Table 1) and the power for the multidimensional models (the bottom two conditions in Table 1)

**Table 2**. Means and standard deviations (in parentheses) of mean adjusted $\chi^2 / df$ statistics for 100 replications across MC item pairs.

| **One factor, MC items only** | | | | |
|---|---|---|---|---|
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -10.866 (0.696) | -3.728 (0.246) | -1.393 (0.111) | 0.213 (0.042) |
| 20 | -10.776 (0.669) | -3.632 (0.248) | -1.328 (0.113) | 0.220 (0.037) |
| 40 | -10.564 (0.784) | -3.566 (0.240) | -1.272 (0.124) | 0.239 (0.039) |
| **Two factors, MC items only** | | | | |
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -10.543 (0.626) | -3.060 (0.470) | -0.962 (0.295) | 0.681 (0.238) |
| 20 | -9.910 (0.844) | -2.977 (0.475) | -0.543 (0.418) | 0.919 (0.407) |
| 40 | -9.787 (0.852) | -2.868 (0.446) | -0.602 (0.451) | 0.899 (0.368) |
| **One factor, MC and CR items** | | | | |
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -10.618 (0.555) | -3.602 (0.233) | -1.253 (0.097) | 0.248 (0.042) |
| 20 | -10.416 (0.622) | -3.576 (0.230) | -1.275 (0.118) | 0.244 (0.038) |
| 40 | -10.484 (0.786) | -3.540 (0.243) | -1.273 (0.118) | 0.244 (0.039) |
| **Two factors, MC and CR items** | | | | |
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -9.817 (0.626) | -3.107 (0.274) | -0.841 (0.264) | 0.469 (0.148) |
| 20 | -10.150 (0.684) | -3.414 (0.247) | -1.199 (0.145) | 0.291 (0.063) |
| 40 | -10.457 (0.825) | -3.517 (0.234) | -1.269 (0.118) | 0.247 (0.039) |

**Table 3**. Descriptive statistics of mean $\bar{X}_{ij}^2$ statistics for 100 replications across MC item pairs.

| | | **One factor, MC only** | | | | **One factor, MC and CR** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| 10 MC Items | Mean | 1.061 | 1.040 | 0.992 | 1.027 | 1.016 | 0.965 | 1.018 | 1.020 |
| | SD | 0.159 | 0.144 | 0.149 | 0.148 | 0.153 | 0.134 | 0.115 | 0.167 |
| | MAX | 1.383 | 1.456 | 1.281 | 1.350 | 1.298 | 1.332 | 1.358 | 1.420 |
| | MIN | 0.785 | 0.792 | 0.652 | 0.732 | 0.636 | 0.686 | 0.817 | 0.717 |
| | Type I Error | 0.055 | 0.056 | 0.049 | 0.052 | 0.054 | 0.046 | 0.053 | 0.051 |
| 20 MC Items | Mean | 1.008 | 1.011 | 1.008 | 0.986 | 1.035 | 1.009 | 1.024 | 1.014 |
| | SD | 0.121 | 0.147 | 0.151 | 0.140 | 0.137 | 0.144 | 0.133 | 0.146 |
| | MAX | 1.384 | 1.561 | 1.609 | 1.369 | 1.431 | 1.441 | 1.359 | 1.426 |
| | MIN | 0.667 | 0.691 | 0.659 | 0.661 | 0.693 | 0.706 | 0.721 | 0.657 |
| | Type I Error | 0.048 | 0.051 | 0.050 | 0.047 | 0.053 | 0.049 | 0.052 | 0.051 |
| 40 MC Items | Mean | 1.024 | 1.015 | 1.021 | 1.009 | 1.022 | 1.017 | 1.013 | 1.012 |
| | SD | 0.137 | 0.149 | 0.151 | 0.138 | 0.146 | 0.142 | 0.145 | 0.143 |
| | MAX | 1.571 | 1.568 | 1.535 | 1.540 | 1.478 | 1.599 | 1.527 | 1.562 |
| | MIN | 0.635 | 0.635 | 0.639 | 0.593 | 0.665 | 0.639 | 0.661 | 0.668 |
| | Type I Error | 0.050 | 0.050 | 0.052 | 0.050 | 0.051 | 0.051 | 0.050 | 0.051 |
| | | **Two factors, MC only** | | | | **Two factors, MC and CR** | | | |
| | N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| 10 MC Items | Mean | **1.175** | **1.715** | **1.670** | **3.343** | **1.450** | **2.213** | **3.573** | **7.383** |
| | SD | 0.233 | 0.732 | 0.489 | 1.449 | 0.318 | 0.728 | 2.409 | 3.681 |
| | MAX | 1.904 | 4.553 | 2.943 | 8.622 | 2.090 | 4.393 | 11.973 | 17.109 |
| | MIN | 0.859 | 0.954 | 1.023 | 1.821 | 0.903 | 1.106 | 1.374 | 2.719 |
| | Power | 0.065 | 0.124 | 0.130 | 0.317 | 0.105 | 0.187 | 0.327 | 0.641 |
| 20 MC Items | Mean | **1.213** | **1.464** | **2.176** | **4.075** | **1.131** | **1.205** | **1.346** | **1.749** |
| | SD | 0.232 | 0.334 | 0.803 | 2.051 | 0.166 | 0.197 | 0.303 | 0.535 |
| | MAX | 2.139 | 2.812 | 5.561 | 12.393 | 1.782 | 1.913 | 2.833 | 3.466 |
| | MIN | 0.767 | 0.926 | 1.090 | 1.273 | 0.735 | 0.748 | 0.792 | 0.726 |
| | Power | 0.075 | 0.108 | 0.187 | 0.391 | 0.065 | 0.075 | 0.090 | 0.138 |
| 40 MC Items | Mean | **1.187** | **1.464** | **1.915** | **3.747** | 1.033 | 1.045 | 1.027 | 1.062 |
| | SD | 0.204 | 0.382 | 0.665 | 1.804 | 0.143 | 0.140 | 0.143 | 0.159 |
| | MAX | 1.852 | 3.964 | 6.722 | 11.700 | 1.465 | 1.509 | 1.506 | 1.591 |
| | MIN | 0.713 | 0.802 | 0.847 | 1.208 | 0.658 | 0.681 | 0.692 | 0.658 |
| | Power | 0.070 | 0.103 | 0.158 | 0.363 | 0.052 | 0.054 | 0.052 | 0.056 |

Note. The mean of *df* is approximately 1 in all conditions. $X_{(1).05}^2 = 3.84$. The type I error rates and power are based on the averaged detection rates across all item pairs in the model.

**Table 4**. Descriptive statistics of mean $R_{ij}$ statistics for 100 replications across MC item pairs.

| | | **One factor, MC only** | | | | **One factor, MC and CR** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| 10 MC Items | Mean | 1.548 | 1.518 | 1.717 | 1.947 | 2.610 | 1.983 | 2.216 | 2.474 |
| | SD | 0.355 | 0.366 | 0.426 | 0.442 | 1.290 | 0.504 | 0.665 | 0.661 |
| | MAX | 2.704 | 2.658 | 2.625 | 2.953 | 7.245 | 3.171 | 4.184 | 4.334 |
| | MIN | 0.959 | 0.912 | 1.005 | 1.063 | 1.095 | 0.960 | 0.975 | 1.123 |
| | Type I error | 0.042 | 0.038 | 0.045 | 0.044 | 0.060 | 0.054 | 0.062 | 0.068 |
| 20 MC Items | Mean | 2.101 | 2.182 | 2.411 | 2.129 | 1.881 | 2.383 | 2.523 | 2.426 |
| | SD | 0.765 | 0.698 | 0.781 | 0.550 | 0.504 | 0.784 | 0.700 | 0.595 |
| | MAX | 6.080 | 6.600 | 4.665 | 4.807 | 4.270 | 6.135 | 6.726 | 5.383 |
| | MIN | 1.157 | 1.077 | 0.939 | 1.300 | 1.056 | 1.091 | 1.158 | 1.241 |
| | Type I error | 0.052 | 0.054 | 0.062 | 0.044 | 0.041 | 0.063 | 0.066 | 0.060 |
| 40 MC Items | Mean | 2.933 | 2.718 | 2.328 | 2.530 | 2.283 | 2.167 | 2.990 | 2.719 |
| | SD | 0.985 | 1.036 | 0.686 | 0.532 | 0.882 | 0.682 | 1.039 | 0.637 |
| | MAX | 13.872 | 9.139 | 6.904 | 6.963 | 19.309 | 8.298 | 9.859 | 6.261 |
| | MIN | 1.054 | 0.939 | 1.149 | 1.087 | 1.088 | 0.901 | 1.171 | 1.023 |
| | Type I error | 0.087 | 0.079 | 0.055 | 0.065 | 0.055 | 0.049 | 0.083 | 0.072 |
| | | **Two factors, MC only** | | | | **Two factors, MC and CR** | | | |
| | N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| 10 MC Items | Mean | **1.575** | **2.352** | **2.100** | **4.167** | **1.873** | **2.952** | **4.133** | **8.757** |
| | SD | 0.457 | 0.817 | 0.506 | 1.503 | 0.441 | 1.056 | 2.531 | 3.913 |
| | MAX | 2.917 | 5.480 | 3.234 | 9.180 | 2.834 | 5.647 | 12.476 | 18.265 |
| | MIN | 0.984 | 1.229 | 1.187 | 2.152 | 1.085 | 1.229 | 1.552 | 3.066 |
| | Power | 0.043 | 0.086 | 0.076 | 0.234 | 0.072 | 0.152 | 0.223 | 0.558 |
| 20 MC Items | Mean | **1.941** | **2.718** | **3.140** | **5.120** | 2.030 | 2.222 | 2.143 | 3.036 |
| | SD | 0.745 | 0.793 | 1.109 | 2.131 | 0.663 | 0.748 | 0.485 | 0.750 |
| | MAX | 6.460 | 6.302 | 7.943 | 13.072 | 5.790 | 6.220 | 4.831 | 5.688 |
| | MIN | 1.020 | 1.426 | 1.316 | 1.571 | 0.955 | 0.959 | 1.231 | 1.025 |
| | Power | 0.053 | 0.083 | 0.128 | 0.280 | 0.053 | 0.058 | 0.054 | 0.089 |
| 40 MC Items | Mean | **2.516** | **3.187** | **3.346** | **5.394** | 3.116 | 2.787 | 2.432 | 2.187 |
| | SD | 0.667 | 1.072 | 0.985 | 2.034 | 1.399 | 0.896 | 0.638 | 0.417 |
| | MAX | 7.041 | 8.535 | 8.096 | 13.338 | 25.493 | 8.227 | 6.015 | 4.076 |
| | MIN | 1.255 | 1.054 | 1.285 | 1.541 | 1.299 | 1.011 | 1.041 | 0.976 |
| | Power | 0.076 | 0.113 | 0.124 | 0.280 | 0.085 | 0.077 | 0.058 | 0.040 |

Note. The mean of *df* is approximately 2 in all conditions. $X^2_{(2).05} = 5.99$. The type I error rates and power are based on the averaged detection rates across all item pairs in the model.

**Table 5**. Descriptive statistics of mean $z_{ij}$ statistics for 100 replications across MC item pairs.

| | | **One factor, MC only** | | | | **One factor, MC and CR** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| 10 MC Items | Mean | 0.023 | 0.000 | 0.001 | -0.007 | 0.018 | -0.004 | -0.018 | -0.001 |
| | SD | 0.106 | 0.109 | 0.118 | 0.107 | 0.091 | 0.084 | 0.097 | 0.105 |
| | MAX | 0.161 | 0.298 | 0.316 | 0.207 | 0.175 | 0.175 | 0.243 | 0.237 |
| | MIN | -0.285 | -0.292 | -0.246 | -0.268 | -0.207 | -0.160 | -0.199 | -0.218 |
| | Type I error | 0.055 | 0.056 | 0.049 | 0.052 | 0.054 | 0.046 | 0.052 | 0.052 |
| 20 MC Items | Mean | 0.048 | 0.032 | 0.016 | 0.010 | 0.007 | 0.003 | -0.011 | 0.002 |
| | SD | 0.100 | 0.085 | 0.102 | 0.099 | 0.092 | 0.097 | 0.103 | 0.104 |
| | MAX | 0.300 | 0.268 | 0.309 | 0.268 | 0.279 | 0.267 | 0.268 | 0.242 |
| | MIN | -0.204 | -0.245 | -0.229 | -0.211 | -0.293 | -0.269 | -0.243 | -0.356 |
| | Type I error | 0.046 | 0.051 | 0.051 | 0.047 | 0.052 | 0.050 | 0.052 | 0.051 |
| 40 MC Items | Mean | 0.071 | -0.019 | 0.002 | -0.024 | 0.065 | -0.038 | 0.015 | 0.009 |
| | SD | 0.102 | 0.105 | 0.103 | 0.097 | 0.099 | 0.097 | 0.102 | 0.101 |
| | MAX | 0.398 | 0.283 | 0.287 | 0.244 | 0.369 | 0.237 | 0.311 | 0.351 |
| | MIN | -0.227 | -0.311 | -0.295 | -0.318 | -0.291 | -0.362 | -0.339 | -0.270 |
| | Type I error | 0.051 | 0.051 | 0.051 | 0.050 | 0.050 | 0.051 | 0.050 | 0.051 |
| | | **Two factors, MC only** | | | | **Two factors, MC and CR** | | | |
| | N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| 10 MC Items | Mean | 0.007 | 0.048 | 0.025 | 0.024 | **0.627** | **1.014** | **1.459** | **2.422** |
| | SD | **0.313** | **0.839** | **0.775** | **1.543** | **0.209** | **0.291** | **0.633** | **0.689** |
| | MAX | 0.718 | 1.603 | 1.313 | 2.749 | 1.044 | 1.720 | 3.266 | 4.004 |
| | MIN | -0.425 | -0.965 | -0.901 | -1.859 | 0.158 | 0.453 | 0.601 | 1.294 |
| | Power | 0.066 | 0.125 | 0.130 | 0.316 | 0.104 | 0.180 | 0.326 | 0.639 |
| 20 MC Items | Mean | 0.027 | 0.024 | -0.020 | 0.010 | **0.263** | **0.398** | **0.530** | **0.780** |
| | SD | **0.438** | **0.684** | **1.072** | **1.751** | **0.137** | **0.153** | **0.210** | **0.259** |
| | MAX | 1.031 | 1.343 | 2.041 | 3.298 | 0.648 | 0.821 | 1.320 | 1.513 |
| | MIN | -0.692 | -0.930 | -1.760 | -2.673 | -0.077 | 0.025 | 0.082 | 0.102 |
| | Power | 0.075 | 0.108 | 0.186 | 0.390 | 0.066 | 0.075 | 0.088 | 0.130 |
| 40 MC Items | Mean | 0.056 | -0.051 | -0.011 | -0.033 | 0.150 | 0.045 | 0.125 | 0.199 |
| | SD | **0.421** | **0.670** | **0.949** | **1.645** | 0.111 | 0.105 | 0.103 | 0.111 |
| | MAX | 0.926 | 1.610 | 2.362 | 3.077 | 0.464 | 0.446 | 0.404 | 0.577 |
| | MIN | -0.722 | -1.285 | -1.687 | -2.931 | -0.169 | -0.310 | -0.208 | -0.125 |
| | Power | 0.070 | 0.104 | 0.157 | 0.360 | 0.053 | 0.054 | 0.052 | 0.055 |

Note. The type I error rates and power are based on the averaged detection rates across all item pairs in the model.

**Table 6**. Correlations among mean adjusted $\chi^2/df$, $\bar{X}^2_{ij}$, $R_{ij}$ and $z_{ij}$ statistics for 100 replications across multidimensional MC item pairs.

| *N* = 200 above the main diagonal<br>*N* = 500 below the main diagonal | | | | | *N* = 1000 above the main diagonal<br>*N* = 3000 below the main diagonal | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 10 MC | Adj. $\chi^2$ / *df* | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ | 10 MC | Adj. $\chi^2$ / *df* | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ |
| Adj. $\chi^2$ / *df* | 1 | .368* | **-.115** | .443** | Adj. $\chi^2$ / *df* | 1 | .816** | .565** | .752** |
| $\bar{X}^2_{ij}$ | .906** | 1 | .627** | .437** | $\bar{X}^2_{ij}$ | .964** | 1 | .873** | .918** |
| $R_{ij}$ | .718** | .877** | 1 | .329* | $R_{ij}$ | .894** | .959** | 1 | .854** |
| $z_{ij}$ | .869** | .935** | .834** | 1 | $z_{ij}$ | .946** | .979** | .960** | 1 |
| | | | | | | | | | |
| *N* = 200 above the main diagonal<br>*N* = 500 below the main diagonal | | | | | *N* = 1000 above the main diagonal<br>*N* = 3000 below the main diagonal | | | | |
| 20 MC | Adj. $\chi^2$ / *df* | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ | 20 MC | Adj. $\chi^2$ / *df* | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ |
| Adj. $\chi^2$ / *df* | 1 | .742** | **.075** | .524** | Adj. $\chi^2$ / *df* | 1 | .973** | .785** | .915** |
| $\bar{X}^2_{ij}$ | .940** | 1 | .406** | .721** | $\bar{X}^2_{ij}$ | .978** | 1 | .840** | .960** |
| $R_{ij}$ | .404** | .462** | 1 | .450** | $R_{ij}$ | .937** | .974** | 1 | .852** |
| $z_{ij}$ | .793** | .871** | .516** | 1 | $z_{ij}$ | .955** | .975** | .968** | 1 |
| | | | | | | | | | |
| *N* = 200 above the main diagonal<br>*N* = 500 below the main diagonal | | | | | *N* = 1000 above the main diagonal<br>*N* = 3000 below the main diagonal | | | | |
| 40 MC | Adj. $\chi^2$ / *df* | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ | 40 MC | Adj. $\chi^2$ / *df* | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ |
| Adj. $\chi^2$ / *df* | 1 | .699** | .253** | .397** | Adj. $\chi^2$ / *df* | 1 | .963** | .672** | .900** |
| $\bar{X}^2_{ij}$ | .949** | 1 | .463** | .634** | $\bar{X}^2_{ij}$ | .990** | 1 | .764** | .947** |
| $R_{ij}$ | .402** | .517** | 1 | .275** | $R_{ij}$ | .909** | .926** | 1 | .771** |
| $z_{ij}$ | .762** | .873** | .597** | 1 | $z_{ij}$ | .968** | .976** | .925** | 1 |

Note. There are 45 MC pairs for a 10-MC model, 190 MC pairs for a 20-MC model, and 780 MC pairs for a 40-MC model.

** indicated the correlation coefficient was significant at the *p* = .01 level (two-tailed).

* indicated the correlation coefficient was significant at the *p* = .05 level (two-tailed).

**Table 7**. Means and standard deviations (in parentheses) of mean adjusted $\chi^2 / df$ statistics across MC item triples for 100 replications.

| **One factor, MC items only** | | | | |
|---|---|---|---|---|
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -7.761 (0.639) | -2.459 (0.252) | -0.768 (0.113) | 0.425 (0.041) |
| 20 | -7.992 (0.874) | -2.330 (0.239) | -0.676 (0.108) | 0.437 (0.037) |
| 40 | -7.628 (4.932) | -2.284 (1.979) | -0.610 (1.003) | 0.461 (0.334) |
| **Two factors, MC items only** | | | | |
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -7.359 (0.582) | **-1.541 (0.356)** | **-0.206 (0.243)** | **1.053 (0.202)** |
| 20 | -6.791 (0.967) | **-1.460 (0.444)** | **0.369 (0.435)** | **1.371 (0.417)** |
| 40 | -6.625 (5.550) | **-1.359 (2.561)** | **0.276 (1.554)** | **1.338 (0.815)** |
| **One factor, MC and CR items** | | | | |
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -7.658 (0.719) | -2.317 (0.221) | -0.606 (0.098) | 0.468 (0.042) |
| 20 | -7.396 (0.773) | -2.345 (0.278) | -0.622 (0.110) | 0.465 (0.036) |
| 40 | -7.524 (4.973) | -2.252 (2.002) | -0.613 (1.004) | 0.466 (0.337) |
| **Two factors, MC and CR items** | | | | |
| Number of MC items | Number of Examinees | | | |
| | 200 | 500 | 1000 | 3000 |
| 10 | -6.632 (0.541) | **-1.845 (0.230)** | **-0.315 (0.140)** | **0.627 (0.082)** |
| 20 | -7.225 (0.807) | -2.149 (0.224) | -0.557 (0.122) | 0.491 (0.043) |
| 40 | -7.492 (5.003) | -2.225 (2.019) | -0.611 (1.000) | 0.468 (0.337) |

**Table 8**. Descriptive statistics of mean $M_n$ statistics for 100 replications across 10 MC item triples.

| | **One factor, MC only** | | | | **One factor, MC and CR** | | | |
|---|---|---|---|---|---|---|---|---|
| N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| Mean | 1.023 | 1.023 | 1.066 | 1.119 | 0.862 | 0.906 | 0.979 | 1.299 |
| SD | 0.135 | 0.161 | 0.344 | **0.490** | 0.190 | 0.170 | 0.259 | 1.055 |
| MAX | 1.324 | 1.616 | 3.986 | 5.449 | 1.341 | 1.490 | 2.050 | 10.621 |
| MIN | 0.712 | 0.672 | 0.678 | 0.782 | 0.432 | 0.522 | 0.565 | 0.565 |
| Type I Error | 0.054 | 0.051 | 0.053 | 0.055 | 0.034 | 0.037 | 0.039 | 0.062 |
| | **Two factors, MC only** | | | | **Two factors, MC and CR** | | | |
| N | 200 | 500 | 1000 | 3000 | 200 | 500 | 1000 | 3000 |
| Mean | 1.007 | 1.015 | 1.032 | 1.223 | 0.899 | 0.921 | 0.984 | 1.383 |
| SD | 0.135 | 0.192 | 0.263 | 0.832 | 0.141 | 0.153 | 0.231 | 0.814 |
| MAX | 1.404 | 2.118 | 2.748 | 6.966 | 1.411 | 1.465 | 1.835 | 8.567 |
| MIN | 0.627 | 0.602 | 0.602 | 0.654 | 0.592 | 0.493 | 0.574 | 0.710 |
| Power | 0.051 | 0.050 | 0.047 | 0.049 | 0.039 | 0.041 | 0.047 | 0.078 |

Note. The type I error rates and power are based on the averaged detection rates across all item triplets in the model.

**Table 9**. Parameter estimates for the 69 MC and 7 CR items of the Physics B Exam.

| MC Item | a | b | MC Item | a | b | MC Item | a | b |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.763 | -1.186 | 24 | 0.552 | 0.010 | 48 | 0.803 | 0.778 |
| 2 | 0.787 | -0.707 | 25 | 0.319 | 2.381 | 49 | 0.534 | 0.356 |
| 3 | 0.777 | -0.461 | 27 | 0.773 | 0.107 | 50 | 0.305 | 1.471 |
| 4 | 0.741 | -0.659 | 28 | 0.344 | 1.444 | 51 | 0.642 | 0.272 |
| 5 | 0.458 | -2.683 | 29 | 0.650 | -1.250 | 52 | 0.632 | 0.323 |
| 6 | 0.807 | -1.420 | 30 | 0.565 | -0.606 | 53 | 0.578 | -0.410 |
| 7 | 0.513 | -0.540 | 31 | 0.484 | 0.226 | 54 | 0.327 | -1.181 |
| 8 | 0.626 | -0.051 | 32 | 0.662 | 0.175 | 55 | 0.470 | -0.357 |
| 9 | 0.355 | 2.322 | 33 | 0.432 | 1.343 | 56 | 0.289 | 2.602 |
| 10 | 0.451 | -0.518 | 34 | 0.356 | 0.348 | 57 | 0.424 | 0.961 |
| 11 | 0.470 | -0.184 | 35 | 0.467 | -0.845 | 58 | 0.353 | 0.810 |
| 12 | 0.492 | 0.133 | 36 | 0.682 | -0.942 | 59 | 0.604 | 1.522 |
| 13 | 0.779 | -0.368 | 37 | 0.815 | -1.037 | 60 | 0.511 | 0.361 |
| 14 | 0.328 | 1.976 | 38 | 0.652 | 1.221 | 61 | 0.242 | -0.416 |
| 15 | **0.131** | **4.230** | 39 | 0.647 | 0.311 | 62 | 0.598 | 2.506 |
| 16 | 0.381 | 0.148 | 40 | 0.365 | 1.793 | 63 | 0.370 | 1.373 |
| 17 | 0.331 | 0.280 | 41 | 0.484 | 0.163 | 64 | 0.389 | 0.235 |
| 18 | 0.538 | 0.083 | 42 | 0.443 | 0.970 | 65 | 0.528 | 1.761 |
| 19 | 0.502 | 0.568 | 43 | 0.353 | 1.126 | 66 | 0.525 | 0.066 |
| 20 | 0.748 | -0.923 | 44 | 0.387 | 1.738 | 67 | 0.425 | 1.856 |
| 21 | 0.496 | -0.973 | 45 | 0.510 | 0.101 | 68 | 0.504 | -0.263 |
| 22 | 0.574 | -0.856 | 46 | 0.527 | -0.101 | 69 | 0.392 | 0.879 |
| 23 | **0.160** | **2.998** | 47 | 0.384 | -0.429 | 70 | 0.388 | 1.113 |

| CR Item | a | b1 | b2 | b3 | b4 |
|---|---|---|---|---|---|
| 1 | 1.222 | -1.243 | -0.454 | 0.491 | 1.180 |
| 2 | 1.466 | -1.041 | -0.260 | 0.562 | 1.212 |
| 3 | 1.018 | -0.850 | -0.001 | 0.819 | 1.762 |
| 4 | 0.801 | -1.552 | -0.581 | 0.231 | 1.034 |
| 5 | 1.303 | -1.228 | -0.346 | 0.336 | 1.144 |
| 6 | 0.835 | -0.910 | -0.073 | 1.128 | 2.538 |
| 7 | 1.105 | -1.384 | -0.203 | 0.785 | 1.752 |

Note. 1.702 is not included in the a-parameters of MC and CR items.

**Table 10**. Parameter estimates for the 70 MC and 3 CR items of the World History Exam.

| MC Item | a | b | MC Item | a | b | MC Item | a | b |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.865 | -0.981 | 25 | 0.642 | -1.356 | 49 | 0.668 | -1.140 |
| 2 | 0.714 | -1.503 | 26 | 0.648 | -1.067 | 50 | 0.358 | 0.276 |
| 3 | 0.535 | -0.803 | 27 | 0.487 | -1.244 | 51 | 0.482 | -1.804 |
| 4 | 0.320 | -1.067 | 28 | 0.258 | 2.382 | 52 | 0.764 | -0.558 |
| 5 | 0.453 | -1.703 | 29 | 0.335 | 0.426 | 53 | 0.702 | -2.029 |
| 6 | 0.617 | -1.027 | 30 | 0.442 | -0.863 | 54 | 0.350 | 0.164 |
| 7 | 0.594 | -0.302 | 31 | 0.434 | -0.062 | 55 | 0.830 | -1.378 |
| 8 | 0.568 | -0.323 | 32 | 0.554 | -0.535 | 56 | 0.539 | -0.421 |
| 9 | 0.705 | -2.097 | 33 | 0.693 | -1.447 | 57 | 0.115 | 1.856 |
| 10 | 0.441 | -2.852 | 34 | 0.916 | -1.450 | 58 | 0.808 | -0.964 |
| 11 | 0.597 | -1.298 | 35 | 0.565 | -1.409 | 59 | 0.611 | -0.347 |
| 12 | 0.482 | -0.956 | 36 | 0.702 | -1.522 | 60 | **0.041** | **1.554** |
| 13 | 0.446 | 0.892 | 37 | 0.274 | -1.668 | 61 | 0.235 | -0.260 |
| 14 | 0.404 | 0.744 | 38 | 0.600 | -1.276 | 62 | 0.518 | -0.738 |
| 15 | 0.421 | 0.314 | 39 | 0.952 | -0.970 | 63 | 0.289 | -1.551 |
| 16 | 0.492 | -0.334 | 40 | 0.814 | -1.717 | 64 | 0.615 | -0.117 |
| 17 | 0.418 | 0.907 | 41 | 0.558 | -2.154 | 65 | 0.609 | -0.907 |
| 18 | 0.599 | -0.462 | 42 | 0.652 | -1.143 | 66 | 0.768 | -1.557 |
| 19 | 0.587 | -0.488 | 43 | 0.192 | -2.407 | 67 | 0.353 | 0.186 |
| 20 | 0.414 | -2.268 | 44 | 0.855 | -1.281 | 68 | 0.498 | -0.840 |
| 21 | 0.468 | -0.205 | 45 | 0.739 | -0.677 | 69 | 0.388 | -0.579 |
| 22 | 0.616 | -1.200 | 46 | 0.646 | -0.740 | 70 | 0.552 | -0.347 |
| 23 | 0.551 | -0.296 | 47 | 0.935 | -0.856 | | | |
| 24 | 0.765 | -0.545 | 48 | 1.032 | -0.864 | | | |

| CR Item | a | b1 | b2 | b3 | b4 |
|---|---|---|---|---|---|
| 1 | 0.719 | -1.830 | -0.577 | 0.869 | 2.471 |
| 2 | 0.761 | -1.150 | -0.193 | 0.516 | 1.837 |
| 3 | 0.934 | -0.487 | 0.197 | 0.777 | 1.790 |

Note. 1.702 is not included in the a-parameters of MC and CR items.

**Table 11**. Parameter estimates for the 55 MC and 3 CR items of the English Literature Exam.

| MC Item | a | b | MC Item | a | b | MC Item | a | b |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.756 | -2.427 | 20 | 0.479 | 0.116 | 39 | 0.533 | -0.038 |
| 2 | 0.651 | -0.352 | 21 | 0.790 | -0.750 | 40 | 0.875 | -0.788 |
| 3 | 0.610 | 0.939 | 22 | 0.895 | -1.459 | 41 | 0.393 | -1.133 |
| 4 | 0.430 | 0.411 | 23 | 0.482 | -0.393 | 42 | 0.461 | -0.776 |
| 5 | 0.429 | -0.626 | 24 | 0.574 | -1.703 | 43 | 0.222 | 0.551 |
| 6 | 0.615 | 0.667 | 25 | 0.443 | -0.631 | 44 | 0.314 | 0.208 |
| 7 | 0.332 | 0.950 | 26 | 0.316 | 0.923 | 45 | 0.709 | 0.185 |
| 8 | 0.313 | -0.629 | 27 | 0.653 | 0.068 | 46 | 0.336 | -0.130 |
| 9 | 0.370 | -1.723 | 28 | 0.416 | -1.363 | 47 | 0.983 | -1.681 |
| 10 | 0.727 | 0.074 | 29 | 0.402 | -1.520 | 48 | 0.321 | -2.179 |
| 11 | 0.363 | -1.184 | 30 | 0.358 | 0.251 | 49 | 0.277 | 0.153 |
| 12 | 0.541 | 2.073 | 31 | 0.661 | 0.973 | 50 | 0.528 | -1.542 |
| 13 | 0.480 | 0.229 | 32 | 0.894 | -0.305 | 51 | 0.442 | -0.435 |
| 14 | 0.364 | 0.305 | 33 | 0.471 | -0.827 | 52 | 0.301 | 0.538 |
| 15 | 0.249 | 0.825 | 34 | 0.443 | 0.293 | 53 | 0.488 | -0.411 |
| 16 | 0.576 | -0.635 | 35 | 0.486 | 0.488 | 54 | 0.277 | -0.513 |
| 17 | 0.313 | 0.774 | 36 | 0.820 | -1.145 | 55 | 0.504 | -1.443 |
| 18 | 0.341 | 1.588 | 37 | 0.780 | -0.618 | | | |
| 19 | 0.630 | -1.198 | 38 | 0.576 | 0.403 | | | |

| CR Item | a | b1 | b2 | b3 | b4 |
|---|---|---|---|---|---|
| 1 | 0.801 | -1.720 | -0.716 | 0.465 | 1.532 |
| 2 | 0.626 | -1.454 | 0.001 | 1.321 | 2.561 |
| 3 | 0.682 | -1.623 | -0.399 | 0.670 | 1.707 |

Note. 1.702 is not included in the a-parameters of MC and CR items.

**Table 12**. Fit statistics for MC item pairs in the three AP® Exams.

| | **Physics B (2346 pairs)** | | |
|---|---|---|---|
| Statistics | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ |
| Mean | 11.326 | 146.877 | -0.477 |
| SD | 39.887 | 461.009 | 3.455 |
| Min | 0.000 | 0.000 | -13.831 |
| Max | 1550.142 | 9657.265 | 37.468 |
| Median | 4.752 | 37.152 | -0.624 |
| Number (%) of misfit ($p$ < .05) | 1279 (54.52%) | 1970 (83.97%) | 1303 (55.54%) |
| | **World History (2415 pairs)** | | |
| Statistics | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ |
| Mean | 5.271 | 93.573 | -1.233 |
| SD | 34.022 | 215.550 | 2.369 |
| Min | 0.000 | 0.000 | -8.322 |
| Max | 1633.313 | 3457.431 | 39.045 |
| Median | 2.243 | 41.327 | -1.338 |
| Number (%) of misfit ($p$ < .05) | 799 (33.08%) | 2097 (86.83%) | 1106 (45.80%) |
| | **English Literature (1485 pairs)** | | |
| Statistics | $\bar{X}^2_{ij}$ | $R_{ij}$ | $z_{ij}$ |
| Mean | 10.504 | 59.543 | -0.429 |
| SD | 77.581 | 158.214 | 3.275 |
| Min | 0.007 | 0.003 | -11.816 |
| Max | 2572.416 | 2580.048 | 50.812 |
| Median | 2.109 | 24.713 | -0.657 |
| Number (%) of misfit ($p$ < .05) | 561 (37.78%) | 1188 (80.00%) | 612 (41.21%) |

Note. The $df$ was estimated to be approximately 1 for $\bar{X}^2_{ij}$; and 1, 2, or 3 for $R_{ij}$ statistics.

**Table 13**. Fit indices of confirmatory factor analyses for unidimensional, two-factor, and bi-factor models of AP® Exams.

| Exam | Model | CFI | TLI | RMSEA (90% CI) |
|---|---|---|---|---|
| Physics B | Unidimensional | 0.958 | 0.956 | 0.023 (0.023, 0.024) |
| | Two-factor | 0.959 | 0.958 | 0.023 (0.023, 0.023) |
| | Bi-factor | 0.973 | 0.972 | 0.019 (0.019, 0.019) |
| World History | Unidimensional | 0.981 | 0.980 | 0.015 (0.015, 0.015) |
| | Two-factor | 0.981 | 0.980 | 0.015 (0.015, 0.015) |
| | Bi-factor | 0.989 | 0.988 | 0.011 (0.011, 0.012) |
| English Literature | Unidimensional | 0.942 | 0.940 | 0.026 (0.025, 0.026) |
| | Two-factor | 0.950 | 0.948 | 0.024 (0.023, 0.024) |
| | Bi-factor | 0.968 | 0.965 | 0.019 (0.019, 0.020) |

Note: CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval.

**Table 14**. Descriptive statistics and parameter estimates for International Personality Item Pool (IPIP) items the honest and faking conditions.

| Agreeableness | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Honest | Mean | SD | a | b1 | b2 | b3 | Faking | Mean | SD | a | b1 | b2 | b3 |
| 1 | 2.127 | 0.785 | 0.949 | -2.511 | -1.430 | 0.547 | 1 | 2.467 | 0.794 | 0.878 | -2.672 | -2.012 | -0.527 |
| 2 | 2.119 | 0.671 | 0.973 | -3.081 | -1.545 | 0.843 | 2 | 2.567 | 0.554 | 1.612 | **-4.848** | -2.074 | -0.355 |
| 3 | 2.105 | 0.863 | 0.528 | -3.851 | -1.469 | 0.567 | 3 | 2.790 | 0.512 | 1.385 | -2.969 | -2.159 | -1.207 |
| 4 | 2.160 | 0.700 | 1.898 | -2.055 | -1.319 | 0.500 | 4 | 2.353 | 0.704 | 1.798 | -2.128 | -1.543 | 0.054 |
| 5 | 1.967 | 0.775 | 1.248 | -2.210 | -0.942 | 0.860 | 5 | 2.443 | 0.676 | 1.040 | -2.941 | -2.060 | -0.167 |
| 6 | **2.097** | 0.678 | 1.161 | -2.652 | -1.397 | 0.807 | 6 | **1.969** | 0.752 | 0.629 | -3.457 | -1.471 | 1.327 |
| 7 | 1.963 | 0.757 | 1.108 | -2.363 | -1.014 | 0.962 | 7 | 2.539 | 0.638 | 1.204 | -2.856 | -2.094 | -0.411 |
| 8 | 2.000 | 0.615 | 1.228 | -2.724 | -1.304 | 1.231 | 8 | 2.363 | 0.684 | 1.531 | -2.335 | -1.683 | 0.057 |
| 9 | 2.072 | 0.703 | 1.461 | -2.244 | -1.218 | 0.768 | 9 | 2.156 | 0.689 | 1.379 | -2.304 | -1.485 | 0.641 |
| 10 | 1.825 | 0.661 | 0.687 | -3.580 | -1.093 | 2.100 | 10 | 2.487 | 0.614 | 1.454 | -2.893 | -1.922 | -0.185 |
| Conscientiousness | | | | | | | | | | | | | |
| Honest | Mean | SD | a | b1 | b2 | b3 | Faking | Mean | SD | a | b1 | b2 | b3 |
| 1 | 1.706 | 0.699 | 0.727 | -3.289 | -0.531 | 2.134 | 1 | 2.679 | 0.545 | 1.627 | -2.994 | -2.169 | -0.689 |
| 2 | 1.582 | 0.828 | 1.284 | -1.711 | -0.120 | 1.414 | 2 | 2.511 | 0.669 | 1.123 | -3.157 | -1.976 | -0.353 |
| 3 | 2.213 | 0.627 | 0.664 | **-4.381** | -2.422 | 0.871 | 3 | 2.784 | 0.452 | 1.587 | -3.299 | -2.570 | -1.020 |
| 4 | 1.881 | 0.782 | 0.910 | -2.554 | -0.863 | 1.195 | 4 | 2.705 | 0.597 | 1.736 | -2.512 | -1.950 | -0.883 |
| 5 | 1.496 | 0.798 | 1.011 | -1.850 | 0.015 | 1.855 | 5 | 2.593 | 0.600 | 1.969 | -2.543 | -1.949 | -0.432 |
| 6 | 1.626 | 0.839 | 1.130 | -1.779 | -0.275 | 1.416 | 6 | 2.575 | 0.681 | 1.361 | -2.579 | -1.908 | -0.581 |
| 7 | 2.006 | 0.694 | 0.789 | -3.174 | -1.438 | 1.272 | 7 | 2.483 | 0.653 | 1.021 | -3.056 | -2.330 | -0.205 |
| 8 | 2.049 | 0.735 | 0.696 | -3.543 | -1.476 | 1.066 | 8 | 2.693 | 0.556 | 1.200 | -3.511 | -2.333 | -0.849 |
| 9 | 1.763 | 0.711 | 0.820 | -2.775 | -0.787 | 1.867 | 9 | 2.563 | 0.589 | 1.365 | -3.257 | -2.175 | -0.359 |
| 10 | 1.979 | 0.638 | 0.735 | -3.720 | -1.523 | 1.597 | 10 | 2.547 | 0.598 | 1.062 | -3.391 | -2.511 | -0.343 |
| Extraversion | | | | | | | | | | | | | |
| Honest | Mean | SD | a | b1 | b2 | b3 | Faking | Mean | SD | a | b1 | b2 | b3 |
| 1 | 1.169 | 0.830 | 1.066 | -1.067 | 0.580 | 2.238 | 1 | 1.545 | 0.886 | 0.570 | -2.449 | -0.150 | 2.054 |
| 2 | 1.551 | 0.903 | 1.444 | -1.402 | -0.015 | 1.176 | 2 | 1.985 | 0.810 | 0.913 | -2.585 | -1.006 | 0.830 |
| 3 | 1.778 | 0.774 | 0.967 | -2.312 | -0.683 | 1.456 | 3 | 2.619 | 0.577 | 1.382 | -3.231 | -2.110 | -0.545 |

**Table 14 (cont.)**. Descriptive statistics and parameter estimates for International Personality Item Pool (IPIP) items the honest and faking conditions.

| Extraversion | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Honest | Mean | SD | a | b1 | b2 | b3 | Faking | Mean | SD | a | b1 | b2 | b3 |
| 4 | 1.380 | 0.806 | 1.319 | -1.458 | 0.220 | 1.723 | 4 | 2.225 | 0.715 | 1.271 | -2.629 | -1.445 | 0.372 |
| 5 | 1.684 | 0.796 | 1.347 | -1.795 | -0.411 | 1.379 | 5 | 2.350 | 0.667 | 1.595 | -2.649 | -1.587 | 0.099 |
| 6 | 1.900 | 0.747 | 0.944 | -2.705 | -0.915 | 1.239 | 6 | 2.390 | 0.646 | 1.281 | -2.840 | -1.924 | 0.068 |
| 7 | 1.450 | 0.894 | 1.509 | -1.161 | 0.019 | 1.444 | 7 | 2.250 | 0.746 | 1.157 | -2.520 | -1.518 | 0.272 |
| 8 | 1.157 | 0.790 | 0.818 | -1.363 | 0.787 | 2.618 | 8 | 1.663 | 0.862 | 0.693 | -2.516 | -0.330 | 1.533 |
| 9 | 1.454 | 0.818 | 1.159 | -1.488 | -0.020 | 1.872 | 9 | 1.976 | 0.770 | 0.866 | -2.517 | -1.305 | 1.074 |
| 10 | 1.141 | 0.852 | 1.204 | -0.935 | 0.615 | 1.980 | 10 | 2.068 | 0.766 | 1.250 | -2.355 | -1.111 | 0.627 |
| Emotional Stability | | | | | | | | | | | | | |
| Honest | Mean | SD | a | b1 | b2 | b3 | Faking | Mean | SD | a | b1 | b2 | b3 |
| 1 | 1.446 | 0.868 | 1.469 | -1.289 | 0.074 | 1.493 | 1 | 2.582 | 0.637 | 2.289 | -2.454 | -1.637 | -0.473 |
| 2 | 1.838 | 0.717 | 0.658 | -3.232 | -1.103 | 1.939 | 2 | 2.283 | 0.639 | 0.643 | **-4.156** | -2.831 | 0.602 |
| 3 | 1.072 | 0.753 | 0.934 | -1.164 | 1.009 | 2.646 | 3 | 1.925 | 0.853 | 0.685 | -3.389 | -0.795 | 0.892 |
| 4 | 1.385 | 0.825 | 0.741 | -1.758 | 0.177 | 2.414 | 4 | 2.176 | 0.859 | 0.750 | -2.600 | -1.719 | 0.339 |
| 5 | 1.777 | 0.764 | 0.933 | -2.445 | -0.615 | 1.519 | 5 | 2.615 | 0.617 | 1.795 | -2.486 | -1.917 | -0.557 |
| 6 | 1.756 | 0.835 | 1.796 | -1.623 | -0.431 | 1.047 | 6 | 2.667 | 0.580 | 1.884 | -2.703 | -1.896 | -0.692 |
| 7 | 1.562 | 0.819 | 1.227 | -1.609 | -0.188 | 1.627 | 7 | 2.436 | 0.682 | 1.504 | -2.608 | -1.746 | -0.124 |
| 8 | 1.791 | 0.841 | 1.557 | -1.749 | -0.460 | 1.007 | 8 | 2.677 | 0.614 | 1.895 | -2.529 | -1.809 | -0.795 |
| 9 | 1.703 | 0.815 | 1.589 | -1.742 | -0.341 | 1.181 | 9 | 2.667 | 0.591 | 1.922 | -2.477 | -1.920 | -0.709 |
| 10 | 1.719 | 0.850 | 1.342 | -1.754 | -0.379 | 1.147 | 10 | 2.621 | 0.627 | 1.779 | -2.445 | -1.907 | -0.594 |
| Openness | | | | | | | | | | | | | |
| Honest | Mean | SD | a | b1 | b2 | b3 | Faking | Mean | SD | a | b1 | b2 | b3 |
| 1 | 2.099 | 0.717 | 0.664 | **-4.069** | -1.657 | 0.937 | 1 | 2.537 | 0.606 | 1.362 | -3.032 | -2.036 | -0.321 |
| 2 | 2.039 | 0.784 | 0.881 | -2.655 | -1.260 | 0.808 | 2 | 2.560 | 0.625 | 1.210 | -2.877 | -2.190 | -0.443 |
| 3 | 2.203 | 0.765 | 0.978 | -2.916 | -1.400 | 0.341 | 3 | 2.281 | 0.695 | 0.913 | -3.113 | -1.945 | 0.314 |
| 4 | 2.041 | 0.785 | 0.931 | -2.512 | -1.240 | 0.794 | 4 | 2.414 | 0.699 | 1.067 | -2.856 | -1.926 | -0.107 |
| 5 | 2.172 | 0.626 | 1.227 | -2.993 | -1.599 | 0.704 | 5 | 2.690 | 0.520 | 1.564 | -3.301 | -2.277 | -0.722 |
| 6 | 2.207 | 0.764 | 0.958 | -2.816 | -1.494 | 0.352 | 6 | 2.575 | 0.630 | 1.209 | -2.873 | -2.142 | -0.518 |

**Table 14 (cont.)**. Descriptive statistics and parameter estimates for International Personality Item Pool (IPIP) items the honest and faking conditions.

| Openness | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Honest | Mean | SD | a | b1 | b2 | b3 | Faking | Mean | SD | a | b1 | b2 | b3 |
| 7 | 2.193 | 0.606 | 0.894 | -3.936 | -1.960 | 0.790 | 7 | 2.719 | 0.559 | 1.720 | -2.634 | -2.016 | -0.881 |
| 8 | **1.777** | 0.808 | 0.747 | -2.800 | -0.610 | 1.447 | 8 | **1.541** | 0.880 | 0.307 | **-4.198** | -0.146 | 3.442 |
| 9 | **2.196** | 0.660 | 0.624 | **-4.586** | -2.285 | 0.829 | 9 | **2.152** | 0.699 | 0.717 | -3.299 | -2.070 | 0.842 |
| 10 | 2.171 | 0.666 | 1.572 | -2.573 | -1.332 | 0.547 | 10 | 2.669 | 0.572 | 1.850 | -2.506 | -2.064 | -0.674 |

Note. The a-parameters do not include the scaling constant of 1.702.

**Table 15.** Piecewise fit statistics for IPIP Agreeableness item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}^2_{ij}$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 5.55 | 36.80 | 7 | 0.00 | 62.40 | 9.02 | 0.00 | 188.73 | 12 | 0.00 | -6.27 | 0.00 |
| 1 | 3 | 10.45 | 26.69 | 7 | 0.00 | 36.99 | 9.22 | 0.00 | 371.94 | 14 | 0.00 | -1.92 | 0.05 |
| 1 | 4 | 16.00 | 80.27 | 7 | 0.00 | 87.54 | 8.91 | 0.00 | 225.85 | 12 | 0.00 | -6.62 | 0.00 |
| 1 | 5 | 10.54 | 67.50 | 7 | 0.00 | 78.45 | 9.05 | 0.00 | 153.36 | 12 | 0.00 | -4.94 | 0.00 |
| 1 | 6 | 8.71 | 50.60 | 7 | 0.00 | 58.61 | 8.86 | 0.00 | 166.25 | 12 | 0.00 | -4.32 | 0.00 |
| 1 | 7 | 14.38 | 46.56 | 7 | 0.00 | 52.59 | 9.00 | 0.00 | 118.48 | 12 | 0.00 | -4.14 | 0.00 |
| 1 | 8 | 5.15 | 98.32 | 7 | 0.00 | 111.98 | 8.96 | 0.00 | 248.75 | 12 | 0.00 | -6.31 | 0.00 |
| 1 | 9 | 8.73 | 68.18 | 7 | 0.00 | 76.86 | 8.97 | 0.00 | 201.38 | 12 | 0.00 | -6.66 | 0.00 |
| 1 | 10 | 5.19 | 70.70 | 7 | 0.00 | 88.03 | 9.25 | 0.00 | 194.80 | 14 | 0.00 | -5.97 | 0.00 |
| 2 | 3 | 1.72 | 18.04 | 7 | 0.01 | 26.10 | 9.05 | 0.00 | 372.47 | 14 | 0.00 | -3.36 | 0.00 |
| 2 | 4 | 3.31 | 20.62 | 7 | 0.00 | 24.88 | 8.79 | 0.00 | 448.64 | 13 | 0.00 | -5.00 | 0.00 |
| 2 | 5 | 5.12 | 44.16 | 7 | 0.00 | 49.98 | 8.87 | 0.00 | 125.86 | 12 | 0.00 | -3.65 | 0.00 |
| 2 | 6 | 1.37 | 15.21 | 7 | 0.03 | 23.24 | 8.72 | 0.00 | 218.61 | 12 | 0.00 | -4.65 | 0.00 |
| 2 | 7 | 20.77 | 42.89 | 7 | 0.00 | 64.24 | 8.79 | 0.00 | 110.70 | 12 | 0.00 | -0.25 | 0.81 |
| 2 | 8 | -3.33 | 15.42 | 7 | 0.03 | 18.11 | 8.75 | 0.03 | 86.78 | 12 | 0.00 | -1.20 | 0.23 |
| 2 | 9 | 1.51 | 12.73 | 7 | 0.08 | 24.33 | 8.77 | 0.00 | 108.53 | 12 | 0.00 | -2.20 | 0.03 |
| 2 | 10 | 3.40 | 28.92 | 7 | 0.00 | 35.14 | 9.03 | 0.00 | 83.96 | 13 | 0.00 | -0.73 | 0.46 |
| 3 | 4 | 1.79 | 20.35 | 7 | 0.00 | 28.45 | 8.93 | 0.00 | 95.65 | 13 | 0.00 | -3.82 | 0.00 |
| 3 | 5 | 9.82 | 33.59 | 7 | 0.00 | 39.29 | 9.07 | 0.00 | 75.92 | 13 | 0.00 | -1.95 | 0.05 |
| 3 | 6 | 1.23 | 13.96 | 7 | 0.05 | 21.68 | 8.91 | 0.01 | 175.05 | 13 | 0.00 | -2.69 | 0.01 |
| 3 | 7 | 7.95 | 31.46 | 7 | 0.00 | 36.15 | 9.05 | 0.00 | 199.68 | 13 | 0.00 | -3.32 | 0.00 |
| 3 | 8 | 1.63 | 24.45 | 7 | 0.00 | 29.11 | 8.96 | 0.00 | 161.45 | 13 | 0.00 | -1.98 | 0.05 |
| 3 | 9 | 3.67 | 21.82 | 7 | 0.00 | 30.19 | 8.97 | 0.00 | 120.66 | 13 | 0.00 | -4.18 | 0.00 |
| 3 | 10 | 1.59 | 28.62 | 7 | 0.00 | 32.03 | 9.30 | 0.00 | 175.59 | 15 | 0.00 | -1.96 | 0.05 |
| 4 | 5 | 6.47 | 85.03 | 7 | 0.00 | 93.37 | 8.88 | 0.00 | 287.68 | 12 | 0.00 | -7.17 | 0.00 |
| 4 | 6 | 8.64 | 28.54 | 7 | 0.00 | 38.34 | 8.55 | 0.00 | 85.75 | 12 | 0.00 | -3.40 | 0.00 |
| 4 | 7 | 12.60 | 89.23 | 7 | 0.00 | 100.28 | 8.82 | 0.00 | 213.95 | 12 | 0.00 | -6.82 | 0.00 |
| 4 | 8 | 8.45 | 46.77 | 7 | 0.00 | 51.69 | 8.70 | 0.00 | 136.32 | 12 | 0.00 | -4.20 | 0.00 |
| 4 | 9 | 12.88 | 126.79 | 7 | 0.00 | 135.43 | 8.67 | 0.00 | 150.64 | 12 | 0.00 | -5.16 | 0.00 |

**Table 15 (cont.)**. Piecewise fit statistics for IPIP Agreeableness item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}^2_{ij}$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | 1.51 | 41.09 | 7 | 0.00 | 47.20 | 9.02 | 0.00 | 257.05 | 14 | 0.00 | -5.02 | 0.00 |
| 5 | 6 | 3.20 | 39.76 | 7 | 0.00 | 49.84 | 8.79 | 0.00 | 121.80 | 12 | 0.00 | -5.69 | 0.00 |
| 5 | 7 | 16.43 | 35.48 | 7 | 0.00 | 56.65 | 8.82 | 0.00 | 112.72 | 12 | 0.00 | -1.21 | 0.23 |
| 5 | 8 | -1.81 | 50.60 | 7 | 0.00 | 59.15 | 8.86 | 0.00 | 128.25 | 12 | 0.00 | -5.75 | 0.00 |
| 5 | 9 | 2.20 | 76.25 | 7 | 0.00 | 83.96 | 8.89 | 0.00 | 155.51 | 12 | 0.00 | -6.04 | 0.00 |
| 5 | 10 | 2.61 | 65.06 | 7 | 0.00 | 75.75 | 9.12 | 0.00 | 148.52 | 13 | 0.00 | -5.37 | 0.00 |
| 6 | 7 | 6.84 | 26.28 | 7 | 0.00 | 38.35 | 8.73 | 0.00 | 152.28 | 12 | 0.00 | -5.34 | 0.00 |
| 6 | 8 | 6.41 | 47.36 | 7 | 0.00 | 51.28 | 8.62 | 0.00 | 68.84 | 11 | 0.00 | -2.67 | 0.01 |
| 6 | 9 | 7.50 | 41.21 | 7 | 0.00 | 45.64 | 8.64 | 0.00 | 84.19 | 12 | 0.00 | -3.52 | 0.00 |
| 6 | 10 | -0.07 | 16.25 | 7 | 0.02 | 18.81 | 8.93 | 0.03 | 75.59 | 13 | 0.00 | -3.23 | 0.00 |
| 7 | 8 | 0.85 | 31.43 | 7 | 0.00 | 42.18 | 8.80 | 0.00 | 83.49 | 12 | 0.00 | -4.91 | 0.00 |
| 7 | 9 | 11.44 | 66.13 | 7 | 0.00 | 71.24 | 8.81 | 0.00 | 101.39 | 12 | 0.00 | -4.16 | 0.00 |
| 7 | 10 | 3.09 | 48.05 | 7 | 0.00 | 52.47 | 9.06 | 0.00 | 107.99 | 13 | 0.00 | -3.49 | 0.00 |
| 8 | 9 | 3.92 | 49.92 | 7 | 0.00 | 55.64 | 8.70 | 0.00 | 96.85 | 12 | 0.00 | -2.72 | 0.01 |
| 8 | 10 | 6.88 | 32.84 | 7 | 0.00 | 40.57 | 8.96 | 0.00 | 93.99 | 13 | 0.00 | -1.01 | 0.31 |
| 9 | 10 | 2.25 | 39.02 | 7 | 0.00 | 42.60 | 9.00 | 0.00 | 99.71 | 13 | 0.00 | -2.94 | 0.00 |
| Mean |  | 5.97 | 44.47 | 7.00 | 0.01 | 53.04 | 8.90 | 0.00 | 159.81 | 12.51 | 0.00 | **3.95*** | 0.05 |
| SD |  | 5.08 | 24.87 | 0.00 | 0.02 | 26.07 | 0.16 | 0.01 | 83.22 | 0.78 | 0.00 | **1.79*** | 0.15 |
| Max |  | 20.77 | 126.79 | 7.00 | 0.08 | 135.43 | 9.30 | 0.03 | 448.64 | 15.00 | 0.00 | -0.25 | 0.81 |
| Min |  | -3.33 | 12.73 | 7.00 | 0.00 | 18.11 | 8.55 | 0.00 | 68.84 | 11.00 | 0.00 | -7.17 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 16**. Piecewise fit statistics for IPIP Agreeableness item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | NA | | | | | | | | | | | |
| 1 | 3 | 24.72 | 49.92 | 7 | 0.00 | 65.73 | 8.93 | 0.00 | 79.55 | 13 | 0.00 | -1.70 | 0.09 |
| 1 | 4 | 23.25 | 105.92 | 7 | 0.00 | 141.15 | 9.44 | 0.00 | 297.81 | 14 | 0.00 | -4.34 | 0.00 |
| 1 | 5 | 24.31 | 42.91 | 7 | 0.00 | 67.62 | 9.44 | 0.00 | 77.93 | 14 | 0.00 | -1.43 | 0.15 |
| 1 | 6 | 13.70 | 43.94 | 7 | 0.00 | 63.33 | 9.66 | 0.00 | 93.40 | 15 | 0.00 | -5.14 | 0.00 |
| 1 | 7 | 44.72 | 93.23 | 7 | 0.00 | 117.39 | 9.28 | 0.00 | 126.54 | 14 | 0.00 | -2.33 | 0.02 |
| 1 | 8 | 16.61 | 87.80 | 7 | 0.00 | 97.45 | 9.44 | 0.00 | 114.63 | 13 | 0.00 | -5.35 | 0.00 |
| 1 | 9 | 11.63 | 103.68 | 7 | 0.00 | 120.82 | 9.51 | 0.00 | 122.61 | 13 | 0.00 | -4.62 | 0.00 |
| 1 | 10 | 16.21 | 51.29 | 7 | 0.00 | 66.01 | 8.88 | 0.00 | 95.57 | 14 | 0.00 | -4.55 | 0.00 |
| 2 | 3 | NA | | | | | | | | | | | |
| 2 | 4 | NA | | | | | | | | | | | |
| 2 | 5 | NA | | | | | | | | | | | |
| 2 | 6 | NA | | | | | | | | | | | |
| 2 | 7 | NA | | | | | | | | | | | |
| 2 | 8 | NA | | | | | | | | | | | |
| 2 | 9 | NA | | | | | | | | | | | |
| 2 | 10 | NA | | | | | | | | | | | |
| 3 | 4 | 0.09 | 35.32 | 7 | 0.00 | 41.60 | 8.30 | 0.00 | 104.19 | 13 | 0.00 | -0.65 | 0.52 |
| 3 | 5 | 8.36 | 25.47 | 7 | 0.00 | 29.06 | 8.48 | 0.00 | 30.76 | 12 | 0.00 | -0.24 | 0.81 |
| 3 | 6 | 9.92 | 14.49 | 7 | 0.04 | 31.73 | 8.77 | 0.00 | 67.75 | 13 | 0.00 | -2.46 | 0.01 |
| 3 | 7 | 14.71 | 16.84 | 7 | 0.02 | 27.53 | 8.25 | 0.00 | 47.17 | 12 | 0.00 | 0.45 | 0.65 |
| 3 | 8 | -0.41 | 31.27 | 7 | 0.00 | 47.24 | 8.39 | 0.00 | 91.19 | 12 | 0.00 | -3.57 | 0.00 |
| 3 | 9 | 6.44 | 22.99 | 7 | 0.00 | 44.49 | 8.54 | 0.00 | 193.95 | 12 | 0.00 | -1.55 | 0.12 |
| 3 | 10 | 0.28 | 13.75 | 7 | 0.06 | 18.31 | 7.81 | 0.02 | 49.20 | 12 | 0.00 | -0.36 | 0.72 |
| 4 | 5 | 4.71 | 60.97 | 7 | 0.00 | 65.59 | 8.99 | 0.00 | 145.83 | 13 | 0.00 | -1.54 | 0.12 |
| 4 | 6 | 11.25 | 31.16 | 7 | 0.00 | 37.32 | 9.10 | 0.00 | 61.52 | 13 | 0.00 | 0.19 | 0.85 |
| 4 | 7 | 12.37 | 108.34 | 7 | 0.00 | 119.14 | 8.82 | 0.00 | 117.74 | 12 | 0.00 | -2.82 | 0.00 |
| 4 | 8 | 2.99 | 89.56 | 7 | 0.00 | 114.68 | 8.93 | 0.00 | 406.59 | 13 | 0.00 | -1.85 | 0.06 |
| 4 | 9 | 10.59 | 40.79 | 7 | 0.00 | 48.61 | 8.85 | 0.00 | 64.66 | 12 | 0.00 | 1.03 | 0.31 |

**Table 16 (cont.)**. Piecewise fit statistics for IPIP Agreeableness item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | 4.27 | 28.00 | 7 | 0.00 | 35.68 | 8.38 | 0.00 | 62.34 | 12 | 0.00 | -0.89 | 0.37 |
| 5 | 6 | 21.04 | 63.34 | 7 | 0.00 | 71.36 | 9.19 | 0.00 | 84.54 | 14 | 0.00 | -2.77 | 0.01 |
| 5 | 7 | 34.42 | 53.64 | 7 | 0.00 | 70.60 | 8.79 | 0.00 | 89.43 | 13 | 0.00 | 0.11 | 0.91 |
| 5 | 8 | 8.38 | 61.70 | 7 | 0.00 | 71.07 | 8.98 | 0.00 | 87.83 | 12 | 0.00 | -2.86 | 0.00 |
| 5 | 9 | 2.95 | 67.25 | 7 | 0.00 | 80.78 | 9.05 | 0.00 | 86.27 | 12 | 0.00 | -2.54 | 0.01 |
| 5 | 10 | 3.43 | 41.86 | 7 | 0.00 | 47.08 | 8.43 | 0.00 | 73.57 | 13 | 0.00 | -2.61 | 0.01 |
| 6 | 7 | 15.67 | 33.92 | 7 | 0.00 | 45.22 | 9.07 | 0.00 | 73.47 | 14 | 0.00 | -3.47 | 0.00 |
| 6 | 8 | 15.77 | 42.82 | 7 | 0.00 | 51.56 | 9.09 | 0.00 | 130.81 | 13 | 0.00 | -0.34 | 0.74 |
| 6 | 9 | 18.88 | 40.45 | 7 | 0.00 | 56.19 | 9.07 | 0.00 | 78.39 | 12 | 0.00 | 2.05 | 0.04 |
| 6 | 10 | 13.65 | 39.20 | 7 | 0.00 | 47.58 | 8.62 | 0.00 | 89.02 | 13 | 0.00 | -2.56 | 0.01 |
| 7 | 8 | 8.42 | 99.91 | 7 | 0.00 | 116.10 | 8.83 | 0.00 | 154.87 | 12 | 0.00 | -4.84 | 0.00 |
| 7 | 9 | 5.19 | 77.28 | 7 | 0.00 | 91.28 | 8.90 | 0.00 | 101.13 | 12 | 0.00 | -2.76 | 0.01 |
| 7 | 10 | -0.92 | 26.63 | 7 | 0.00 | 30.12 | 8.25 | 0.00 | 66.09 | 13 | 0.00 | -2.62 | 0.01 |
| 8 | 9 | 12.99 | 69.13 | 7 | 0.00 | 77.16 | 8.88 | 0.00 | 83.62 | 12 | 0.00 | -0.70 | 0.48 |
| 8 | 10 | 12.09 | 34.45 | 7 | 0.00 | 62.83 | 8.31 | 0.00 | 76.84 | 12 | 0.00 | -1.88 | 0.06 |
| 9 | 10 | 4.02 | 14.06 | 7 | 0.05 | 23.35 | 8.42 | 0.00 | 55.28 | 12 | 0.00 | -0.41 | 0.68 |
| Mean | | 12.13 | 51.76 | 7.00 | 0.00 | 65.08 | 8.84 | 0.00 | 105.06 | 12.78 | 0.00 | **2.21*** | 0.22 |
| SD | | 9.66 | 27.77 | 0.00 | 0.01 | 31.32 | 0.42 | 0.00 | 69.10 | 0.82 | 0.00 | **1.51*** | 0.30 |
| Max | | 44.72 | 108.34 | 7.00 | 0.06 | 141.15 | 9.66 | 0.02 | 406.59 | 15.00 | 0.00 | 2.05 | 0.91 |
| Min | | -0.92 | 13.75 | 7.00 | 0.00 | 18.31 | 7.81 | 0.00 | 30.76 | 12.00 | 0.00 | -5.35 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 17**. Piecewise fit statistics for IPIP Conscientiousness item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}^2_{ij}$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1.66 | 31.97 | 7 | 0.00 | 34.43 | 8.87 | 0.00 | 57.90 | 13 | 0.00 | 0.04 | 0.97 |
| 1 | 3 | 7.78 | 18.82 | 7 | 0.01 | 36.15 | 8.90 | 0.00 | 53.17 | 12 | 0.00 | 3.69 | 0.00 |
| 1 | 4 | 5.81 | 30.53 | 7 | 0.00 | 33.89 | 8.97 | 0.00 | 60.98 | 13 | 0.00 | 1.31 | 0.19 |
| 1 | 5 | -1.55 | 21.85 | 7 | 0.00 | 24.32 | 8.96 | 0.00 | 35.90 | 13 | 0.00 | 1.43 | 0.15 |
| 1 | 6 | 17.33 | 133.96 | 7 | 0.00 | 148.88 | 8.98 | 0.00 | 189.09 | 13 | 0.00 | -3.12 | 0.00 |
| 1 | 7 | -2.79 | 16.28 | 7 | 0.02 | 17.65 | 8.93 | 0.04 | 27.16 | 13 | 0.01 | 2.02 | 0.04 |
| 1 | 8 | -1.51 | 59.69 | 7 | 0.00 | 64.68 | 9.01 | 0.00 | 85.53 | 13 | 0.00 | -0.35 | 0.73 |
| 1 | 9 | 4.42 | 15.60 | 7 | 0.03 | 31.09 | 8.95 | 0.00 | 52.47 | 13 | 0.00 | 3.72 | 0.00 |
| 1 | 10 | 0.94 | 29.09 | 7 | 0.00 | 33.27 | 8.93 | 0.00 | 40.33 | 13 | 0.00 | 1.99 | 0.05 |
| 2 | 3 | 6.71 | 34.76 | 7 | 0.00 | 40.96 | 8.82 | 0.00 | 82.91 | 12 | 0.00 | -1.50 | 0.13 |
| 2 | 4 | 4.78 | 17.00 | 7 | 0.02 | 23.10 | 8.70 | 0.01 | 59.00 | 13 | 0.00 | 2.93 | 0.00 |
| 2 | 5 | 3.62 | 43.43 | 7 | 0.00 | 49.00 | 8.85 | 0.00 | 81.17 | 13 | 0.00 | -0.70 | 0.48 |
| 2 | 6 | 12.34 | 40.54 | 7 | 0.00 | 57.43 | 8.47 | 0.00 | 194.96 | 13 | 0.00 | 3.71 | 0.00 |
| 2 | 7 | -1.89 | 10.72 | 7 | 0.15 | 12.30 | 8.80 | 0.18 | 33.83 | 13 | 0.00 | 1.39 | 0.16 |
| 2 | 8 | 6.53 | 26.90 | 7 | 0.00 | 35.51 | 8.84 | 0.00 | 68.64 | 13 | 0.00 | 0.25 | 0.80 |
| 2 | 9 | 5.06 | 45.21 | 7 | 0.00 | 53.76 | 8.94 | 0.00 | 92.34 | 13 | 0.00 | -3.24 | 0.00 |
| 2 | 10 | -1.28 | 11.78 | 7 | 0.11 | 17.32 | 8.86 | 0.04 | 48.77 | 13 | 0.00 | -1.31 | 0.19 |
| 3 | 4 | 4.02 | 31.63 | 7 | 0.00 | 37.30 | 8.92 | 0.00 | 71.28 | 12 | 0.00 | 1.14 | 0.26 |
| 3 | 5 | 1.24 | 16.33 | 7 | 0.02 | 18.89 | 8.90 | 0.02 | 34.28 | 12 | 0.00 | -0.40 | 0.69 |
| 3 | 6 | 8.98 | 73.07 | 7 | 0.00 | 82.13 | 8.91 | 0.00 | 114.95 | 12 | 0.00 | -2.90 | 0.00 |
| 3 | 7 | -0.16 | 26.63 | 7 | 0.00 | 30.94 | 8.87 | 0.00 | 38.79 | 12 | 0.00 | 2.16 | 0.03 |
| 3 | 8 | 4.50 | 38.66 | 7 | 0.00 | 40.98 | 8.95 | 0.00 | 63.64 | 13 | 0.00 | 0.59 | 0.56 |
| 3 | 9 | 4.90 | 25.91 | 7 | 0.00 | 28.78 | 8.93 | 0.00 | 48.56 | 12 | 0.00 | -0.04 | 0.97 |
| 3 | 10 | 7.14 | 17.28 | 7 | 0.02 | 24.95 | 8.87 | 0.00 | 34.23 | 12 | 0.00 | 2.30 | 0.02 |
| 4 | 5 | 6.26 | 66.62 | 7 | 0.00 | 73.43 | 8.99 | 0.00 | 114.22 | 13 | 0.00 | -1.19 | 0.23 |
| 4 | 6 | 10.54 | 36.49 | 7 | 0.00 | 43.54 | 8.82 | 0.00 | 63.62 | 13 | 0.00 | 1.22 | 0.22 |
| 4 | 7 | -0.07 | 28.34 | 7 | 0.00 | 35.33 | 8.96 | 0.00 | 86.92 | 13 | 0.00 | -0.69 | 0.49 |
| 4 | 8 | 11.56 | 29.72 | 7 | 0.00 | 36.74 | 8.95 | 0.00 | 72.85 | 13 | 0.00 | 3.14 | 0.00 |
| 4 | 9 | 7.89 | 47.37 | 7 | 0.00 | 55.30 | 9.02 | 0.00 | 84.69 | 12 | 0.00 | -2.34 | 0.02 |

**Table 17 (cont.)**. Piecewise fit statistics for IPIP Conscientiousness item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | -0.96 | 29.29 | 7 | 0.00 | 34.34 | 8.95 | 0.00 | 63.77 | 13 | 0.00 | 0.23 | 0.82 |
| 5 | 6 | 9.66 | 119.31 | 7 | 0.00 | 126.30 | 8.92 | 0.00 | 153.23 | 13 | 0.00 | -1.10 | 0.27 |
| 5 | 7 | 0.04 | 18.89 | 7 | 0.01 | 21.92 | 8.95 | 0.01 | 36.43 | 13 | 0.00 | -1.21 | 0.23 |
| 5 | 8 | 0.10 | 58.90 | 7 | 0.00 | 67.11 | 9.01 | 0.00 | 104.55 | 13 | 0.00 | -1.77 | 0.08 |
| 5 | 9 | 4.82 | 22.05 | 7 | 0.00 | 29.61 | 8.96 | 0.00 | 39.01 | 13 | 0.00 | 0.86 | 0.39 |
| 5 | 10 | 5.92 | 29.09 | 7 | 0.00 | 31.79 | 8.92 | 0.00 | 44.26 | 13 | 0.00 | 0.48 | 0.63 |
| 6 | 7 | 7.28 | 78.89 | 7 | 0.00 | 87.82 | 8.92 | 0.00 | 114.24 | 13 | 0.00 | -1.82 | 0.07 |
| 6 | 8 | 4.88 | 26.06 | 7 | 0.00 | 28.98 | 8.94 | 0.00 | 54.17 | 13 | 0.00 | 0.29 | 0.77 |
| 6 | 9 | 4.06 | 76.94 | 7 | 0.00 | 92.21 | 9.02 | 0.00 | 146.40 | 13 | 0.00 | -4.83 | 0.00 |
| 6 | 10 | 4.11 | 76.00 | 7 | 0.00 | 88.00 | 8.95 | 0.00 | 124.81 | 13 | 0.00 | -2.87 | 0.00 |
| 7 | 8 | 2.25 | 17.63 | 7 | 0.01 | 22.90 | 8.98 | 0.01 | 42.61 | 13 | 0.00 | -0.68 | 0.50 |
| 7 | 9 | 0.86 | 14.23 | 7 | 0.05 | 21.95 | 8.93 | 0.01 | 34.00 | 13 | 0.00 | 1.95 | 0.05 |
| 7 | 10 | -0.42 | 8.33 | 7 | 0.30 | 13.66 | 8.89 | 0.13 | 18.27 | 13 | 0.15 | 2.75 | 0.01 |
| 8 | 9 | 6.21 | 47.80 | 7 | 0.00 | 52.09 | 9.03 | 0.00 | 69.70 | 13 | 0.00 | -0.46 | 0.64 |
| 8 | 10 | 0.18 | 9.81 | 7 | 0.20 | 15.74 | 8.96 | 0.07 | 27.50 | 13 | 0.01 | 2.59 | 0.01 |
| 9 | 10 | 1.00 | 11.84 | 7 | 0.11 | 23.41 | 8.92 | 0.01 | 43.00 | 13 | 0.00 | 2.32 | 0.02 |
| Mean | | 4.11 | 37.14 | 7.00 | 0.02 | 44.00 | 8.91 | 0.01 | 71.29 | 12.80 | 0.00 | **1.71*** | 0.26 |
| SD | | 4.29 | 27.11 | 0.00 | 0.06 | 28.83 | 0.09 | 0.03 | 40.85 | 0.40 | 0.02 | **1.17*** | 0.30 |
| Max | | 17.33 | 133.96 | 7.00 | 0.30 | 148.88 | 9.03 | 0.18 | 194.96 | 13.00 | 0.15 | 3.72 | 0.97 |
| Min | | -2.79 | 8.33 | 7.00 | 0.00 | 12.30 | 8.47 | 0.00 | 18.27 | 12.00 | 0.00 | -4.83 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 18**. Piecewise fit statistics for IPIP Conscientiousness item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}_{ij}^2$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | -3.57 | 8.09 | 7 | 0.32 | 12.11 | 9.19 | 0.22 | 14.36 | 14 | 0.42 | -0.53 | 0.59 |
| 1 | 3 | -4.61 | 4.71 | 7 | 0.70 | 7.33 | 8.25 | 0.53 | 25.16 | 13 | 0.02 | -0.03 | 0.98 |
| 1 | 4 | -4.73 | 9.71 | 7 | 0.21 | 12.40 | 9.38 | 0.22 | 20.25 | 14 | 0.12 | -0.28 | 0.78 |
| 1 | 5 | 2.24 | 15.97 | 7 | 0.03 | 30.88 | 8.98 | 0.00 | 45.65 | 13 | 0.00 | 0.98 | 0.33 |
| 1 | 6 | 3.85 | 23.74 | 7 | 0.00 | 42.60 | 9.40 | 0.00 | 46.77 | 14 | 0.00 | -1.25 | 0.21 |
| 1 | 7 | -4.84 | 7.48 | 7 | 0.38 | 16.38 | 9.25 | 0.07 | 18.96 | 14 | 0.17 | -0.63 | 0.53 |
| 1 | 8 | -4.99 | 11.79 | 7 | 0.11 | 14.02 | 9.00 | 0.12 | 17.87 | 13 | 0.16 | -0.78 | 0.43 |
| 1 | 9 | -3.92 | 12.25 | 7 | 0.09 | 13.91 | 8.83 | 0.12 | 17.87 | 13 | 0.16 | 0.44 | 0.66 |
| 1 | 10 | -5.48 | 7.68 | 7 | 0.36 | 10.64 | 9.12 | 0.31 | 13.82 | 14 | 0.46 | -0.38 | 0.70 |
| 2 | 3 | -4.32 | 5.43 | 7 | 0.61 | 7.33 | 8.80 | 0.58 | 45.78 | 14 | 0.00 | -1.45 | 0.15 |
| 2 | 4 | 5.41 | 19.39 | 7 | 0.01 | 28.96 | 9.87 | 0.00 | 37.01 | 15 | 0.00 | -0.25 | 0.80 |
| 2 | 5 | -1.20 | 20.72 | 7 | 0.00 | 24.62 | 9.53 | 0.00 | 31.71 | 14 | 0.00 | -0.62 | 0.54 |
| 2 | 6 | 11.26 | 43.19 | 7 | 0.00 | 49.22 | 9.84 | 0.00 | 51.35 | 15 | 0.00 | -1.81 | 0.07 |
| 2 | 7 | -0.98 | 34.53 | 7 | 0.00 | 42.31 | 9.66 | 0.00 | 44.46 | 15 | 0.00 | -2.17 | 0.03 |
| 2 | 8 | -1.24 | 12.23 | 7 | 0.09 | 13.89 | 9.45 | 0.15 | 15.01 | 14 | 0.38 | -1.02 | 0.31 |
| 2 | 9 | 1.85 | 27.14 | 7 | 0.00 | 36.27 | 9.31 | 0.00 | 43.70 | 14 | 0.00 | -1.91 | 0.06 |
| 2 | 10 | -4.39 | 18.89 | 7 | 0.01 | 20.79 | 9.54 | 0.02 | 23.22 | 15 | 0.08 | -1.71 | 0.09 |
| 3 | 4 | -2.13 | 21.48 | 7 | 0.00 | 26.99 | 8.93 | 0.00 | 30.28 | 13 | 0.00 | -0.90 | 0.37 |
| 3 | 5 | 0.60 | 12.37 | 7 | 0.09 | 22.16 | 8.62 | 0.01 | 49.64 | 13 | 0.00 | -0.34 | 0.74 |
| 3 | 6 | 2.68 | 5.54 | 7 | 0.59 | 18.94 | 9.00 | 0.03 | 29.68 | 14 | 0.01 | -2.04 | 0.04 |
| 3 | 7 | -4.85 | 6.52 | 7 | 0.48 | 14.38 | 8.84 | 0.10 | 28.60 | 14 | 0.01 | -0.56 | 0.57 |
| 3 | 8 | -2.71 | 6.53 | 7 | 0.48 | 12.93 | 8.58 | 0.14 | 24.38 | 13 | 0.03 | -0.87 | 0.38 |
| 3 | 9 | -3.62 | 3.84 | 7 | 0.80 | 6.47 | 8.44 | 0.64 | 22.35 | 13 | 0.05 | 0.01 | 0.99 |
| 3 | 10 | -3.27 | 7.24 | 7 | 0.40 | 14.47 | 8.72 | 0.10 | 26.09 | 13 | 0.02 | -0.54 | 0.59 |
| 4 | 5 | 0.35 | 14.76 | 7 | 0.04 | 18.45 | 9.71 | 0.04 | 24.94 | 14 | 0.04 | -0.05 | 0.96 |
| 4 | 6 | 4.12 | 47.98 | 7 | 0.00 | 55.52 | 10.09 | 0.00 | 54.47 | 15 | 0.00 | -2.00 | 0.05 |
| 4 | 7 | -4.73 | 20.00 | 7 | 0.01 | 27.01 | 9.94 | 0.00 | 32.72 | 15 | 0.01 | -1.34 | 0.18 |
| 4 | 8 | 1.02 | 20.52 | 7 | 0.00 | 24.01 | 9.65 | 0.01 | 26.75 | 14 | 0.02 | -0.55 | 0.59 |
| 4 | 9 | -3.38 | 21.76 | 7 | 0.00 | 28.98 | 9.55 | 0.00 | 32.47 | 14 | 0.00 | -0.65 | 0.51 |

**Table 18 (cont.)**. Piecewise fit statistics for IPIP Conscientiousness item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | -3.88 | 7.87 | 7 | 0.34 | 10.55 | 9.83 | 0.38 | 26.50 | 15 | 0.03 | -1.11 | 0.27 |
| 5 | 6 | 19.50 | 60.04 | 7 | 0.00 | 74.31 | 9.76 | 0.00 | 77.51 | 14 | 0.00 | -1.27 | 0.20 |
| 5 | 7 | -0.30 | 15.23 | 7 | 0.03 | 21.08 | 9.57 | 0.02 | 26.78 | 14 | 0.02 | 0.46 | 0.64 |
| 5 | 8 | -0.49 | 27.04 | 7 | 0.00 | 32.16 | 9.34 | 0.00 | 184.08 | 14 | 0.00 | -0.58 | 0.56 |
| 5 | 9 | -1.04 | 16.52 | 7 | 0.02 | 19.96 | 9.16 | 0.02 | 470.75 | 14 | 0.00 | 0.64 | 0.52 |
| 5 | 10 | -1.66 | 23.49 | 7 | 0.00 | 28.23 | 9.46 | 0.00 | 50.40 | 14 | 0.00 | 0.00 | 1.00 |
| 6 | 7 | 0.29 | 17.70 | 7 | 0.01 | 20.15 | 9.87 | 0.03 | 24.23 | 15 | 0.06 | -1.35 | 0.18 |
| 6 | 8 | -0.56 | 23.54 | 7 | 0.00 | 25.97 | 9.66 | 0.00 | 28.79 | 14 | 0.01 | -1.87 | 0.06 |
| 6 | 9 | 8.16 | 42.59 | 7 | 0.00 | 54.88 | 9.51 | 0.00 | 54.83 | 14 | 0.00 | -1.86 | 0.06 |
| 6 | 10 | 5.27 | 33.19 | 7 | 0.00 | 38.04 | 9.78 | 0.00 | 43.18 | 15 | 0.00 | -2.10 | 0.04 |
| 7 | 8 | 1.02 | 26.58 | 7 | 0.00 | 32.79 | 9.48 | 0.00 | 37.45 | 14 | 0.00 | -0.90 | 0.37 |
| 7 | 9 | 16.66 | 19.78 | 7 | 0.01 | 33.01 | 9.25 | 0.00 | 47.78 | 14 | 0.00 | 1.54 | 0.12 |
| 7 | 10 | 3.41 | 23.03 | 7 | 0.00 | 25.16 | 9.55 | 0.00 | 28.29 | 15 | 0.02 | -0.53 | 0.60 |
| 8 | 9 | -4.51 | 9.64 | 7 | 0.21 | 12.00 | 9.11 | 0.22 | 17.97 | 13 | 0.16 | -0.25 | 0.80 |
| 8 | 10 | 1.12 | 16.71 | 7 | 0.02 | 20.60 | 9.37 | 0.02 | 24.36 | 14 | 0.04 | -0.67 | 0.51 |
| 9 | 10 | -2.00 | 13.40 | 7 | 0.06 | 15.60 | 9.19 | 0.08 | 18.42 | 14 | 0.19 | -0.14 | 0.89 |
| Mean | | 0.12 | 18.84 | 7.00 | 0.15 | 24.85 | 9.32 | 0.09 | 45.70 | 14.00 | 0.06 | **0.92*** | 0.45 |
| SD | | 5.37 | 12.11 | 0.00 | 0.22 | 14.12 | 0.42 | 0.16 | 69.16 | 0.67 | 0.11 | **0.64*** | 0.30 |
| Max | | 19.50 | 60.04 | 7.00 | 0.80 | 74.31 | 10.09 | 0.64 | 470.75 | 15.00 | 0.46 | 1.54 | 1.00 |
| Min | | -5.48 | 3.84 | 7.00 | 0.00 | 6.47 | 8.25 | 0.00 | 13.82 | 13.00 | 0.00 | -2.17 | 0.03 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 19**. Piecewise fit statistics for IPIP Extraversion item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}_{ij}^2$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1.36 | 20.62 | 7 | 0.00 | 24.33 | 9.19 | 0.00 | 45.97 | 14 | 0.00 | -1.53 | 0.13 |
| 1 | 3 | 0.40 | 11.35 | 7 | 0.12 | 14.00 | 9.22 | 0.13 | 28.77 | 14 | 0.01 | 0.76 | 0.45 |
| 1 | 4 | -2.04 | 68.46 | 7 | 0.00 | 81.49 | 9.13 | 0.00 | 132.57 | 14 | 0.00 | 0.35 | 0.73 |
| 1 | 5 | -2.12 | 4.59 | 7 | 0.71 | 9.31 | 9.14 | 0.42 | 56.35 | 14 | 0.00 | -0.94 | 0.35 |
| 1 | 6 | 4.14 | 45.27 | 7 | 0.00 | 54.09 | 9.25 | 0.00 | 99.38 | 14 | 0.00 | -0.97 | 0.33 |
| 1 | 7 | -2.12 | 21.38 | 7 | 0.00 | 35.22 | 9.05 | 0.00 | 69.39 | 14 | 0.00 | 3.76 | 0.00 |
| 1 | 8 | 3.44 | 6.10 | 7 | 0.53 | 7.93 | 9.19 | 0.56 | 34.00 | 14 | 0.00 | 0.44 | 0.66 |
| 1 | 9 | 1.65 | 3.95 | 7 | 0.79 | 11.57 | 9.09 | 0.25 | 39.62 | 14 | 0.00 | 3.27 | 0.00 |
| 1 | 10 | 1.93 | 14.41 | 7 | 0.04 | 18.51 | 9.21 | 0.03 | 36.96 | 14 | 0.00 | -2.16 | 0.03 |
| 2 | 3 | 5.00 | 26.78 | 7 | 0.00 | 32.16 | 9.23 | 0.00 | 74.22 | 15 | 0.00 | -1.39 | 0.16 |
| 2 | 4 | 17.16 | 62.02 | 7 | 0.00 | 68.17 | 9.15 | 0.00 | 90.60 | 13 | 0.00 | 0.66 | 0.51 |
| 2 | 5 | 0.41 | 12.73 | 7 | 0.08 | 14.81 | 9.12 | 0.10 | 48.53 | 14 | 0.00 | 0.27 | 0.78 |
| 2 | 6 | -1.26 | 41.60 | 7 | 0.00 | 65.43 | 9.17 | 0.00 | 124.84 | 15 | 0.00 | 3.82 | 0.00 |
| 2 | 7 | 0.23 | 46.48 | 7 | 0.00 | 57.83 | 9.22 | 0.00 | 90.28 | 14 | 0.00 | -2.44 | 0.01 |
| 2 | 8 | -4.08 | 25.13 | 7 | 0.00 | 28.45 | 9.19 | 0.00 | 48.58 | 14 | 0.00 | -0.03 | 0.97 |
| 2 | 9 | 5.72 | 8.97 | 7 | 0.25 | 11.00 | 9.15 | 0.29 | 33.30 | 13 | 0.00 | -0.16 | 0.87 |
| 2 | 10 | 1.74 | 19.20 | 7 | 0.01 | 23.88 | 9.20 | 0.01 | 483.39 | 15 | 0.00 | 0.23 | 0.81 |
| 3 | 4 | 7.72 | 88.63 | 7 | 0.00 | 95.45 | 9.22 | 0.00 | 127.13 | 14 | 0.00 | -0.89 | 0.37 |
| 3 | 5 | 2.57 | 8.25 | 7 | 0.31 | 24.85 | 9.13 | 0.00 | 66.95 | 15 | 0.00 | 3.12 | 0.00 |
| 3 | 6 | 18.38 | 60.23 | 7 | 0.00 | 65.92 | 9.29 | 0.00 | 94.35 | 15 | 0.00 | -0.01 | 1.00 |
| 3 | 7 | 6.48 | 26.55 | 7 | 0.00 | 30.13 | 9.17 | 0.00 | 35.30 | 15 | 0.00 | 1.11 | 0.27 |
| 3 | 8 | 2.55 | 14.46 | 7 | 0.04 | 22.89 | 9.29 | 0.01 | 40.57 | 14 | 0.00 | -2.30 | 0.02 |
| 3 | 9 | 4.06 | 17.63 | 7 | 0.01 | 20.92 | 9.20 | 0.01 | 36.87 | 14 | 0.00 | -0.51 | 0.61 |
| 3 | 10 | 20.34 | 32.96 | 7 | 0.00 | 37.78 | 9.29 | 0.00 | 50.11 | 15 | 0.00 | -1.63 | 0.10 |
| 4 | 5 | 1.17 | 51.29 | 7 | 0.00 | 55.56 | 9.13 | 0.00 | 89.39 | 14 | 0.00 | -0.17 | 0.86 |
| 4 | 6 | 4.96 | 74.67 | 7 | 0.00 | 82.47 | 9.20 | 0.00 | 125.31 | 14 | 0.00 | 2.00 | 0.05 |
| 4 | 7 | 6.01 | 112.26 | 7 | 0.00 | 121.31 | 9.16 | 0.00 | 178.56 | 14 | 0.00 | -0.71 | 0.48 |
| 4 | 8 | 6.17 | 47.97 | 7 | 0.00 | 53.38 | 9.16 | 0.00 | 93.27 | 14 | 0.00 | 1.11 | 0.27 |
| 4 | 9 | 4.74 | 51.26 | 7 | 0.00 | 55.16 | 9.12 | 0.00 | 83.96 | 14 | 0.00 | 0.17 | 0.86 |

**Table 19 (cont.)**. Piecewise fit statistics for IPIP Extraversion item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}^2_{ij}$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | 10.00 | 31.99 | 7 | 0.00 | 37.42 | 9.19 | 0.00 | 73.96 | 14 | 0.00 | 0.74 | 0.46 |
| 5 | 6 | 2.08 | 63.56 | 7 | 0.00 | 68.91 | 9.18 | 0.00 | 105.90 | 15 | 0.00 | 0.79 | 0.43 |
| 5 | 7 | 9.25 | 16.85 | 7 | 0.02 | 23.85 | 9.05 | 0.00 | 63.95 | 14 | 0.00 | 2.44 | 0.01 |
| 5 | 8 | 9.12 | 39.18 | 7 | 0.00 | 52.93 | 9.18 | 0.00 | 98.07 | 14 | 0.00 | -2.17 | 0.03 |
| 5 | 9 | 13.19 | 28.66 | 7 | 0.00 | 32.62 | 9.11 | 0.00 | 66.07 | 14 | 0.00 | -0.33 | 0.74 |
| 5 | 10 | -0.28 | 37.51 | 7 | 0.00 | 41.41 | 9.18 | 0.00 | 107.61 | 15 | 0.00 | 0.20 | 0.84 |
| 6 | 7 | 1.78 | 52.31 | 7 | 0.00 | 59.80 | 9.23 | 0.00 | 99.06 | 15 | 0.00 | 0.17 | 0.86 |
| 6 | 8 | 0.58 | 17.14 | 7 | 0.02 | 20.74 | 9.27 | 0.02 | 45.13 | 14 | 0.00 | 1.62 | 0.11 |
| 6 | 9 | 4.52 | 23.02 | 7 | 0.00 | 29.81 | 9.19 | 0.00 | 55.01 | 14 | 0.00 | 2.19 | 0.03 |
| 6 | 10 | -2.80 | 62.67 | 7 | 0.00 | 70.11 | 9.29 | 0.00 | 122.40 | 15 | 0.00 | -0.39 | 0.70 |
| 7 | 8 | 12.87 | 56.65 | 7 | 0.00 | 68.85 | 9.18 | 0.00 | 81.00 | 14 | 0.00 | -1.32 | 0.19 |
| 7 | 9 | 6.17 | 22.97 | 7 | 0.00 | 24.58 | 9.10 | 0.00 | 73.93 | 14 | 0.00 | 0.77 | 0.44 |
| 7 | 10 | 7.86 | 25.26 | 7 | 0.00 | 28.38 | 9.17 | 0.00 | 35.87 | 14 | 0.00 | 0.29 | 0.77 |
| 8 | 9 | 20.44 | 18.96 | 7 | 0.01 | 34.58 | 9.14 | 0.00 | 61.29 | 14 | 0.00 | 4.07 | 0.00 |
| 8 | 10 | 17.71 | 54.21 | 7 | 0.00 | 61.05 | 9.23 | 0.00 | 79.00 | 14 | 0.00 | -1.16 | 0.25 |
| 9 | 10 | 9.19 | 53.94 | 7 | 0.00 | 59.98 | 9.20 | 0.00 | 74.73 | 14 | 0.00 | -1.89 | 0.06 |
| Mean | | 5.30 | 36.22 | 7.00 | 0.07 | 43.09 | 9.18 | 0.04 | 84.48 | 14.20 | 0.00 | **1.28*** | 0.39 |
| SD | | 6.20 | 23.79 | 0.00 | 0.18 | 25.10 | 0.06 | 0.11 | 68.48 | 0.50 | 0.00 | **1.09*** | 0.33 |
| Max | | 20.44 | 112.26 | 7.00 | 0.79 | 121.31 | 9.29 | 0.56 | 483.39 | 15.00 | 0.01 | 4.07 | 1.00 |
| Min | | -4.08 | 3.95 | 7.00 | 0.00 | 7.93 | 9.05 | 0.00 | 28.77 | 13.00 | 0.00 | -2.44 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 20**. Piecewise fit statistics for IPIP Extraversion item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 33.31 | 85.64 | 7 | 0.00 | 93.48 | 9.36 | 0.00 | 99.50 | 14 | 0.00 | -2.74 | 0.01 |
| 1 | 3 | 7.31 | 15.31 | 7 | 0.03 | 21.64 | 8.89 | 0.01 | 36.42 | 13 | 0.00 | -2.81 | 0.00 |
| 1 | 4 | 17.68 | 48.06 | 7 | 0.00 | 52.66 | 9.31 | 0.00 | 84.24 | 15 | 0.00 | -3.35 | 0.00 |
| 1 | 5 | 9.56 | 28.23 | 7 | 0.00 | 38.10 | 9.04 | 0.00 | 53.55 | 14 | 0.00 | -1.57 | 0.12 |
| 1 | 6 | 44.30 | 89.54 | 7 | 0.00 | 107.78 | 9.27 | 0.00 | 133.96 | 14 | 0.00 | -5.14 | 0.00 |
| 1 | 7 | 8.45 | 30.50 | 7 | 0.00 | 38.20 | 9.32 | 0.00 | 96.67 | 15 | 0.00 | -0.94 | 0.35 |
| 1 | 8 | 28.48 | 72.90 | 7 | 0.00 | 80.66 | 9.39 | 0.00 | 100.53 | 14 | 0.00 | -4.01 | 0.00 |
| 1 | 9 | 13.53 | 27.89 | 7 | 0.00 | 48.69 | 9.43 | 0.00 | 83.85 | 14 | 0.00 | 0.52 | 0.61 |
| 1 | 10 | 27.90 | 63.52 | 7 | 0.00 | 72.40 | 9.35 | 0.00 | 99.73 | 14 | 0.00 | -4.84 | 0.00 |
| 2 | 3 | -0.59 | 6.28 | 7 | 0.51 | 11.96 | 8.78 | 0.20 | 20.05 | 12 | 0.07 | -1.44 | 0.15 |
| 2 | 4 | 12.24 | 80.06 | 7 | 0.00 | 88.57 | 9.13 | 0.00 | 97.81 | 13 | 0.00 | -1.30 | 0.19 |
| 2 | 5 | -0.63 | 124.53 | 7 | 0.00 | 139.18 | 9.01 | 0.00 | 181.43 | 13 | 0.00 | -3.25 | 0.00 |
| 2 | 6 | 10.24 | 61.52 | 7 | 0.00 | 68.01 | 9.07 | 0.00 | 74.78 | 13 | 0.00 | -0.47 | 0.64 |
| 2 | 7 | -0.29 | 27.00 | 7 | 0.00 | 32.83 | 9.26 | 0.00 | 67.35 | 13 | 0.00 | -1.73 | 0.08 |
| 2 | 8 | 18.87 | 47.15 | 7 | 0.00 | 56.57 | 9.19 | 0.00 | 61.85 | 13 | 0.00 | 0.33 | 0.74 |
| 2 | 9 | 27.54 | 65.17 | 7 | 0.00 | 76.15 | 9.35 | 0.00 | 112.36 | 13 | 0.00 | -3.04 | 0.00 |
| 2 | 10 | 9.59 | 40.04 | 7 | 0.00 | 50.12 | 9.14 | 0.00 | 63.96 | 13 | 0.00 | -0.52 | 0.60 |
| 3 | 4 | -1.22 | 23.19 | 7 | 0.00 | 26.84 | 8.68 | 0.00 | 81.44 | 13 | 0.00 | -2.17 | 0.03 |
| 3 | 5 | 14.55 | 36.82 | 7 | 0.00 | 40.54 | 8.20 | 0.00 | 58.61 | 12 | 0.00 | -1.50 | 0.13 |
| 3 | 6 | 0.82 | 20.90 | 7 | 0.00 | 23.07 | 8.54 | 0.00 | 30.25 | 12 | 0.00 | -0.92 | 0.36 |
| 3 | 7 | 6.65 | 27.15 | 7 | 0.00 | 34.54 | 8.65 | 0.00 | 86.39 | 13 | 0.00 | -1.29 | 0.20 |
| 3 | 8 | 14.22 | 33.64 | 7 | 0.00 | 47.79 | 8.80 | 0.00 | 68.53 | 12 | 0.00 | -3.78 | 0.00 |
| 3 | 9 | 4.87 | 26.76 | 7 | 0.00 | 31.39 | 8.85 | 0.00 | 49.85 | 12 | 0.00 | -3.64 | 0.00 |
| 3 | 10 | 0.91 | 36.38 | 7 | 0.00 | 42.43 | 8.75 | 0.00 | 92.94 | 12 | 0.00 | -2.62 | 0.01 |
| 4 | 5 | -4.22 | 86.04 | 7 | 0.00 | 94.26 | 8.94 | 0.00 | 128.60 | 13 | 0.00 | -3.68 | 0.00 |
| 4 | 6 | 8.10 | 41.59 | 7 | 0.00 | 46.18 | 8.97 | 0.00 | 59.13 | 14 | 0.00 | -1.35 | 0.18 |
| 4 | 7 | 1.64 | 67.89 | 7 | 0.00 | 78.30 | 9.24 | 0.00 | 160.65 | 14 | 0.00 | -3.81 | 0.00 |
| 4 | 8 | 9.05 | 33.41 | 7 | 0.00 | 40.75 | 9.11 | 0.00 | 55.12 | 14 | 0.00 | -0.60 | 0.55 |
| 4 | 9 | 16.37 | 70.61 | 7 | 0.00 | 80.13 | 9.32 | 0.00 | 108.56 | 13 | 0.00 | -4.40 | 0.00 |

**Table 20 (cont.)**. Piecewise fit statistics for IPIP Extraversion item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | 8.70 | 47.54 | 7 | 0.00 | 54.85 | 9.06 | 0.00 | 68.28 | 13 | 0.00 | -1.90 | 0.06 |
| 5 | 6 | 1.00 | 76.79 | 7 | 0.00 | 82.83 | 8.80 | 0.00 | 114.99 | 13 | 0.00 | -3.03 | 0.00 |
| 5 | 7 | 21.43 | 54.61 | 7 | 0.00 | 65.08 | 8.74 | 0.00 | 87.57 | 13 | 0.00 | -1.46 | 0.14 |
| 5 | 8 | 20.24 | 90.32 | 7 | 0.00 | 110.64 | 9.02 | 0.00 | 163.60 | 13 | 0.00 | -5.37 | 0.00 |
| 5 | 9 | 13.12 | 34.42 | 7 | 0.00 | 48.07 | 8.99 | 0.00 | 66.72 | 13 | 0.00 | -2.17 | 0.03 |
| 5 | 10 | 2.17 | 137.92 | 7 | 0.00 | 149.08 | 9.02 | 0.00 | 209.73 | 13 | 0.00 | -4.44 | 0.00 |
| 6 | 7 | 0.09 | 53.56 | 7 | 0.00 | 64.21 | 9.16 | 0.00 | 167.47 | 14 | 0.00 | -3.72 | 0.00 |
| 6 | 8 | 19.64 | 46.71 | 7 | 0.00 | 51.33 | 9.08 | 0.00 | 54.89 | 13 | 0.00 | -1.43 | 0.15 |
| 6 | 9 | 10.36 | 57.52 | 7 | 0.00 | 70.22 | 9.27 | 0.00 | 101.39 | 13 | 0.00 | -4.82 | 0.00 |
| 6 | 10 | 18.35 | 44.08 | 7 | 0.00 | 56.41 | 9.03 | 0.00 | 60.30 | 13 | 0.00 | -2.33 | 0.02 |
| 7 | 8 | 7.71 | 33.05 | 7 | 0.00 | 39.35 | 9.26 | 0.00 | 114.25 | 14 | 0.00 | -2.79 | 0.01 |
| 7 | 9 | 10.76 | 32.08 | 7 | 0.00 | 44.04 | 9.28 | 0.00 | 72.91 | 14 | 0.00 | -1.42 | 0.16 |
| 7 | 10 | -1.77 | 80.64 | 7 | 0.00 | 91.65 | 9.29 | 0.00 | 158.87 | 13 | 0.00 | -4.35 | 0.00 |
| 8 | 9 | 29.87 | 93.69 | 7 | 0.00 | 101.63 | 9.35 | 0.00 | 116.32 | 13 | 0.00 | -3.90 | 0.00 |
| 8 | 10 | 18.27 | 47.89 | 7 | 0.00 | 60.25 | 9.13 | 0.00 | 78.88 | 13 | 0.00 | -1.07 | 0.29 |
| 9 | 10 | 17.23 | 96.38 | 7 | 0.00 | 109.37 | 9.37 | 0.00 | 400.52 | 14 | 0.00 | -5.37 | 0.00 |
| Mean | | 11.92 | 54.33 | 7.00 | 0.01 | 63.61 | 9.07 | 0.00 | 99.66 | 13.24 | 0.00 | **2.61*** | 0.13 |
| SD | | 10.56 | 28.41 | 0.00 | 0.07 | 30.22 | 0.26 | 0.03 | 60.85 | 0.74 | 0.01 | **1.47*** | 0.20 |
| Max | | 44.30 | 137.92 | 7.00 | 0.51 | 149.08 | 9.43 | 0.20 | 400.52 | 15.00 | 0.07 | 0.52 | 0.74 |
| Min | | -4.22 | 6.28 | 7.00 | 0.00 | 11.96 | 8.20 | 0.00 | 20.05 | 12.00 | 0.00 | -5.37 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 21**. Piecewise fit statistics for IPIP Emotional Stability item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}^2_{ij}$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 0.61 | 45.37 | 7 | 0.00 | 49.89 | 9.07 | 0.00 | 53.49 | 15 | 0.00 | 0.48 | 0.63 |
| 1 | 3 | 14.23 | 11.29 | 7 | 0.13 | 39.06 | 8.96 | 0.00 | 88.87 | 14 | 0.00 | 4.34 | 0.00 |
| 1 | 4 | 9.79 | 70.86 | 7 | 0.00 | 81.31 | 9.01 | 0.00 | 96.87 | 14 | 0.00 | -3.38 | 0.00 |
| 1 | 5 | 1.05 | 46.47 | 7 | 0.00 | 53.11 | 9.09 | 0.00 | 97.86 | 15 | 0.00 | 0.80 | 0.42 |
| 1 | 6 | 5.77 | 67.89 | 7 | 0.00 | 74.05 | 8.96 | 0.00 | 85.07 | 13 | 0.00 | -0.07 | 0.94 |
| 1 | 7 | 8.07 | 61.96 | 7 | 0.00 | 69.97 | 9.11 | 0.00 | 104.98 | 14 | 0.00 | -3.23 | 0.00 |
| 1 | 8 | -0.54 | 24.58 | 7 | 0.00 | 33.14 | 9.08 | 0.00 | 42.85 | 14 | 0.00 | -2.36 | 0.02 |
| 1 | 9 | 4.59 | 49.03 | 7 | 0.00 | 52.30 | 9.04 | 0.00 | 82.61 | 14 | 0.00 | -1.94 | 0.05 |
| 1 | 10 | 4.79 | 35.55 | 7 | 0.00 | 41.78 | 9.10 | 0.00 | 46.68 | 14 | 0.00 | -3.41 | 0.00 |
| 2 | 3 | 1.46 | 17.02 | 7 | 0.02 | 20.22 | 9.09 | 0.02 | 26.66 | 14 | 0.02 | 0.48 | 0.63 |
| 2 | 4 | 11.52 | 28.58 | 7 | 0.00 | 39.84 | 9.11 | 0.00 | 48.42 | 14 | 0.00 | 2.43 | 0.02 |
| 2 | 5 | 4.17 | 26.76 | 7 | 0.00 | 29.26 | 9.16 | 0.00 | 57.63 | 15 | 0.00 | 0.71 | 0.48 |
| 2 | 6 | 7.17 | 57.49 | 7 | 0.00 | 61.36 | 9.00 | 0.00 | 77.19 | 15 | 0.00 | -0.29 | 0.77 |
| 2 | 7 | 6.93 | 49.50 | 7 | 0.00 | 58.18 | 9.11 | 0.00 | 70.91 | 15 | 0.00 | -2.57 | 0.01 |
| 2 | 8 | 3.23 | 47.92 | 7 | 0.00 | 51.79 | 9.03 | 0.00 | 71.05 | 15 | 0.00 | -0.69 | 0.49 |
| 2 | 9 | 5.18 | 79.97 | 7 | 0.00 | 87.64 | 9.05 | 0.00 | 113.13 | 15 | 0.00 | -2.30 | 0.02 |
| 2 | 10 | 6.96 | 46.25 | 7 | 0.00 | 48.87 | 9.08 | 0.00 | 51.71 | 15 | 0.00 | -0.92 | 0.36 |
| 3 | 4 | 0.69 | 20.57 | 7 | 0.00 | 23.06 | 9.01 | 0.01 | 31.08 | 13 | 0.00 | -0.29 | 0.77 |
| 3 | 5 | 4.02 | 45.66 | 7 | 0.00 | 48.89 | 9.08 | 0.00 | 83.08 | 14 | 0.00 | -0.67 | 0.50 |
| 3 | 6 | -0.47 | 28.63 | 7 | 0.00 | 31.46 | 8.97 | 0.00 | 64.23 | 14 | 0.00 | -2.75 | 0.01 |
| 3 | 7 | 7.36 | 19.58 | 7 | 0.01 | 26.94 | 9.06 | 0.00 | 43.91 | 14 | 0.00 | -3.78 | 0.00 |
| 3 | 8 | 2.58 | 14.35 | 7 | 0.05 | 17.50 | 9.00 | 0.04 | 29.32 | 14 | 0.01 | -2.27 | 0.02 |
| 3 | 9 | 1.48 | 25.41 | 7 | 0.00 | 28.36 | 9.00 | 0.00 | 65.16 | 14 | 0.00 | -3.08 | 0.00 |
| 3 | 10 | -1.46 | 17.38 | 7 | 0.02 | 19.46 | 9.02 | 0.02 | 33.21 | 14 | 0.00 | -1.74 | 0.08 |
| 4 | 5 | 3.90 | 52.20 | 7 | 0.00 | 58.10 | 9.09 | 0.00 | 78.89 | 14 | 0.00 | -0.66 | 0.51 |
| 4 | 6 | 15.46 | 89.07 | 7 | 0.00 | 98.26 | 8.96 | 0.00 | 124.69 | 13 | 0.00 | -4.17 | 0.00 |
| 4 | 7 | 9.32 | 98.14 | 7 | 0.00 | 105.74 | 9.02 | 0.00 | 117.05 | 14 | 0.00 | -1.20 | 0.23 |
| 4 | 8 | 6.06 | 47.20 | 7 | 0.00 | 50.64 | 8.95 | 0.00 | 55.14 | 13 | 0.00 | -1.53 | 0.13 |
| 4 | 9 | 11.01 | 59.88 | 7 | 0.00 | 68.95 | 8.98 | 0.00 | 101.66 | 14 | 0.00 | -3.45 | 0.00 |

**Table 21 (cont.)**. Piecewise fit statistics for IPIP Emotional Stability item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}^2_{ij}$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | 42.29 | 81.96 | 7 | 0.00 | 121.25 | 8.96 | 0.00 | 178.55 | 14 | 0.00 | 4.35 | 0.00 |
| 5 | 6 | 3.13 | 46.96 | 7 | 0.00 | 52.51 | 9.02 | 0.00 | 111.25 | 15 | 0.00 | 0.40 | 0.69 |
| 5 | 7 | 5.15 | 78.98 | 7 | 0.00 | 87.59 | 9.12 | 0.00 | 124.51 | 15 | 0.00 | -1.41 | 0.16 |
| 5 | 8 | 6.26 | 40.17 | 7 | 0.00 | 49.41 | 9.06 | 0.00 | 76.27 | 14 | 0.00 | -0.43 | 0.67 |
| 5 | 9 | 8.29 | 42.23 | 7 | 0.00 | 46.47 | 9.05 | 0.00 | 97.18 | 15 | 0.00 | 1.01 | 0.31 |
| 5 | 10 | 8.29 | 56.90 | 7 | 0.00 | 62.16 | 9.08 | 0.00 | 141.61 | 15 | 0.00 | 0.83 | 0.41 |
| 6 | 7 | -1.06 | 37.38 | 7 | 0.00 | 42.41 | 9.05 | 0.00 | 59.83 | 14 | 0.00 | -2.03 | 0.04 |
| 6 | 8 | -1.40 | 30.51 | 7 | 0.00 | 33.42 | 8.97 | 0.00 | 59.88 | 14 | 0.00 | -0.50 | 0.62 |
| 6 | 9 | 23.35 | 115.00 | 7 | 0.00 | 128.86 | 8.89 | 0.00 | 145.64 | 13 | 0.00 | 0.91 | 0.36 |
| 6 | 10 | 1.65 | 23.07 | 7 | 0.00 | 26.34 | 9.03 | 0.00 | 33.64 | 14 | 0.00 | -2.32 | 0.02 |
| 7 | 8 | 37.01 | 50.09 | 7 | 0.00 | 96.29 | 8.88 | 0.00 | 224.26 | 14 | 0.00 | 6.08 | 0.00 |
| 7 | 9 | 6.95 | 33.09 | 7 | 0.00 | 37.10 | 9.06 | 0.00 | 68.02 | 14 | 0.00 | -1.82 | 0.07 |
| 7 | 10 | 3.06 | 80.98 | 7 | 0.00 | 86.79 | 9.05 | 0.00 | 84.87 | 14 | 0.00 | -1.41 | 0.16 |
| 8 | 9 | 8.25 | 43.85 | 7 | 0.00 | 46.51 | 8.97 | 0.00 | 44.28 | 13 | 0.00 | 0.13 | 0.90 |
| 8 | 10 | 2.45 | 63.15 | 7 | 0.00 | 72.12 | 8.97 | 0.00 | 73.94 | 14 | 0.00 | -0.71 | 0.48 |
| 9 | 10 | 8.19 | 35.77 | 7 | 0.00 | 40.52 | 9.06 | 0.00 | 48.64 | 13 | 0.00 | -3.05 | 0.00 |
| Mean | | 7.04 | 47.66 | 7.00 | 0.00 | 55.53 | 9.03 | 0.00 | 80.35 | 14.11 | 0.00 | **1.85*** | 0.27 |
| SD | | 8.50 | 23.33 | 0.00 | 0.02 | 26.69 | 0.06 | 0.01 | 39.90 | 0.64 | 0.00 | **1.39*** | 0.29 |
| Max | | 42.29 | 115.00 | 7.00 | 0.13 | 128.86 | 9.16 | 0.04 | 224.26 | 15.00 | 0.02 | 6.08 | 0.94 |
| Min | | -1.46 | 11.29 | 7.00 | 0.00 | 17.50 | 8.88 | 0.00 | 26.66 | 13.00 | 0.00 | -4.17 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 22**. Piecewise fit statistics for IPIP Emotional Stability item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}^2_{ij}$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 0.42 | 13.61 | 7 | 0.06 | 17.73 | 9.53 | 0.05 | 31.80 | 15 | 0.01 | 0.73 | 0.46 |
| 1 | 3 | 2.45 | 18.33 | 7 | 0.01 | 21.59 | 9.64 | 0.01 | 31.37 | 15 | 0.01 | -0.34 | 0.73 |
| 1 | 4 | 18.68 | 56.56 | 7 | 0.00 | 64.23 | 9.84 | 0.00 | 64.57 | 15 | 0.00 | -2.72 | 0.01 |
| 1 | 5 | 5.65 | 38.57 | 7 | 0.00 | 43.18 | 9.58 | 0.00 | 49.84 | 14 | 0.00 | -1.16 | 0.24 |
| 1 | 6 | 3.00 | 78.40 | 7 | 0.00 | 86.47 | 9.27 | 0.00 | 91.04 | 14 | 0.00 | -1.12 | 0.26 |
| 1 | 7 | 8.20 | 63.18 | 7 | 0.00 | 69.41 | 9.71 | 0.00 | 67.66 | 14 | 0.00 | -0.51 | 0.61 |
| 1 | 8 | -0.25 | 61.23 | 7 | 0.00 | 78.22 | 9.75 | 0.00 | 87.07 | 14 | 0.00 | -1.93 | 0.05 |
| 1 | 9 | 0.71 | 52.78 | 7 | 0.00 | 58.13 | 9.46 | 0.00 | 61.38 | 14 | 0.00 | -1.17 | 0.24 |
| 1 | 10 | -2.75 | 63.82 | 7 | 0.00 | 73.20 | 9.71 | 0.00 | 71.94 | 14 | 0.00 | -1.26 | 0.21 |
| 2 | 3 | 5.56 | 23.47 | 7 | 0.00 | 25.49 | 9.50 | 0.00 | 25.85 | 15 | 0.04 | -0.53 | 0.60 |
| 2 | 4 | 24.35 | 35.79 | 7 | 0.00 | 58.86 | 9.69 | 0.00 | 59.76 | 15 | 0.00 | -0.68 | 0.50 |
| 2 | 5 | 0.65 | 45.49 | 7 | 0.00 | 55.85 | 9.61 | 0.00 | 60.18 | 15 | 0.00 | -0.84 | 0.40 |
| 2 | 6 | -3.81 | 11.58 | 7 | 0.12 | 14.19 | 9.42 | 0.13 | 22.32 | 14 | 0.07 | -0.78 | 0.44 |
| 2 | 7 | 4.56 | 36.80 | 7 | 0.00 | 57.20 | 9.62 | 0.00 | 58.73 | 15 | 0.00 | -0.97 | 0.33 |
| 2 | 8 | 0.69 | 16.38 | 7 | 0.02 | 22.04 | 9.80 | 0.01 | 26.35 | 15 | 0.03 | -0.63 | 0.53 |
| 2 | 9 | -3.57 | 18.50 | 7 | 0.01 | 21.73 | 9.56 | 0.01 | 31.35 | 14 | 0.00 | 0.09 | 0.93 |
| 2 | 10 | -2.63 | 4.50 | 7 | 0.72 | 6.40 | 9.70 | 0.76 | 10.45 | 15 | 0.79 | 0.23 | 0.82 |
| 3 | 4 | 15.71 | 51.43 | 7 | 0.00 | 62.36 | 9.79 | 0.00 | 60.19 | 15 | 0.00 | -2.22 | 0.03 |
| 3 | 5 | 6.14 | 22.69 | 7 | 0.00 | 26.75 | 9.71 | 0.00 | 37.82 | 15 | 0.00 | -0.56 | 0.58 |
| 3 | 6 | 10.83 | 29.17 | 7 | 0.00 | 34.17 | 9.53 | 0.00 | 41.04 | 14 | 0.00 | -1.84 | 0.07 |
| 3 | 7 | 10.75 | 28.42 | 7 | 0.00 | 33.21 | 9.69 | 0.00 | 35.84 | 15 | 0.00 | 0.07 | 0.94 |
| 3 | 8 | 4.75 | 27.73 | 7 | 0.00 | 35.96 | 9.91 | 0.00 | 40.90 | 15 | 0.00 | -2.35 | 0.02 |
| 3 | 9 | 4.02 | 18.95 | 7 | 0.01 | 22.83 | 9.67 | 0.01 | 27.90 | 14 | 0.01 | -0.97 | 0.33 |
| 3 | 10 | 5.40 | 19.62 | 7 | 0.01 | 21.97 | 9.81 | 0.01 | 23.79 | 15 | 0.07 | -1.16 | 0.24 |
| 4 | 5 | 18.59 | 44.65 | 7 | 0.00 | 53.66 | 9.92 | 0.00 | 55.28 | 15 | 0.00 | -2.79 | 0.01 |
| 4 | 6 | 26.06 | 89.90 | 7 | 0.00 | 98.69 | 9.71 | 0.00 | 98.28 | 14 | 0.00 | -2.62 | 0.01 |
| 4 | 7 | 23.07 | 65.90 | 7 | 0.00 | 73.26 | 9.93 | 0.00 | 70.89 | 15 | 0.00 | -2.88 | 0.00 |
| 4 | 8 | 15.18 | 58.59 | 7 | 0.00 | 66.13 | 10.11 | 0.00 | 63.96 | 15 | 0.00 | -3.66 | 0.00 |
| 4 | 9 | 8.10 | 30.66 | 7 | 0.00 | 37.71 | 9.87 | 0.00 | 41.42 | 14 | 0.00 | -2.78 | 0.01 |

**Table 22 (cont.)**. Piecewise fit statistics for IPIP Emotional Stability item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}^2_{ij}$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | 30.61 | 102.18 | 7 | 0.00 | 114.96 | 10.01 | 0.00 | 106.26 | 15 | 0.00 | -3.00 | 0.00 |
| 5 | 6 | -1.62 | 41.12 | 7 | 0.00 | 49.02 | 9.47 | 0.00 | 53.70 | 14 | 0.00 | -2.28 | 0.02 |
| 5 | 7 | -1.97 | 40.09 | 7 | 0.00 | 45.49 | 9.83 | 0.00 | 59.48 | 15 | 0.00 | -0.93 | 0.35 |
| 5 | 8 | 9.44 | 62.38 | 7 | 0.00 | 70.27 | 9.87 | 0.00 | 65.76 | 14 | 0.00 | -2.19 | 0.03 |
| 5 | 9 | 8.31 | 48.76 | 7 | 0.00 | 53.73 | 9.62 | 0.00 | 53.84 | 14 | 0.00 | -1.67 | 0.10 |
| 5 | 10 | -0.81 | 45.73 | 7 | 0.00 | 51.80 | 9.83 | 0.00 | 71.27 | 15 | 0.00 | -1.75 | 0.08 |
| 6 | 7 | 7.71 | 98.86 | 7 | 0.00 | 108.72 | 9.60 | 0.00 | 108.78 | 14 | 0.00 | -1.57 | 0.12 |
| 6 | 8 | 12.42 | 59.01 | 7 | 0.00 | 67.10 | 9.60 | 0.00 | 70.57 | 14 | 0.00 | -2.56 | 0.01 |
| 6 | 9 | 4.72 | 58.55 | 7 | 0.00 | 64.95 | 9.34 | 0.00 | 67.80 | 14 | 0.00 | -2.16 | 0.03 |
| 6 | 10 | 3.09 | 63.24 | 7 | 0.00 | 69.33 | 9.56 | 0.00 | 71.39 | 14 | 0.00 | -1.55 | 0.12 |
| 7 | 8 | 16.03 | 56.71 | 7 | 0.00 | 64.88 | 10.00 | 0.00 | 61.99 | 15 | 0.00 | -1.15 | 0.25 |
| 7 | 9 | -0.60 | 56.72 | 7 | 0.00 | 66.48 | 9.76 | 0.00 | 67.32 | 14 | 0.00 | -1.44 | 0.15 |
| 7 | 10 | 4.27 | 96.30 | 7 | 0.00 | 107.08 | 9.95 | 0.00 | 102.24 | 15 | 0.00 | -1.50 | 0.13 |
| 8 | 9 | 10.91 | 25.18 | 7 | 0.00 | 29.32 | 9.72 | 0.00 | 32.01 | 14 | 0.00 | -1.55 | 0.12 |
| 8 | 10 | 7.96 | 50.98 | 7 | 0.00 | 56.96 | 9.95 | 0.00 | 56.29 | 15 | 0.00 | -1.95 | 0.05 |
| 9 | 10 | -1.20 | 16.95 | 7 | 0.02 | 19.88 | 9.70 | 0.03 | 22.60 | 14 | 0.07 | -0.95 | 0.34 |
| Mean | | 7.11 | 45.54 | 7.00 | 0.02 | 52.90 | 9.71 | 0.02 | 56.01 | 14.51 | 0.02 | **1.51*** | 0.26 |
| SD | | 8.33 | 23.93 | 0.00 | 0.11 | 26.15 | 0.18 | 0.11 | 23.63 | 0.50 | 0.12 | **0.87*** | 0.26 |
| Max | | 30.61 | 102.18 | 7.00 | 0.72 | 114.96 | 10.11 | 0.76 | 108.78 | 15.00 | 0.79 | 0.73 | 0.94 |
| Min | | -3.81 | 4.50 | 7.00 | 0.00 | 6.40 | 9.27 | 0.00 | 10.45 | 14.00 | 0.00 | -3.66 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 23**. Piecewise fit statistics for IPIP Openness item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}^2_{ij}$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | -2.36 | 19.97 | 7 | 0.01 | 22.78 | 9.03 | 0.01 | 69.71 | 13 | 0.00 | 0.80 | 0.42 |
| 1 | 3 | 6.63 | 19.95 | 7 | 0.01 | 29.27 | 9.05 | 0.00 | 49.67 | 13 | 0.00 | -2.82 | 0.00 |
| 1 | 4 | -1.95 | 13.23 | 7 | 0.07 | 14.55 | 8.99 | 0.10 | 64.67 | 13 | 0.00 | -1.43 | 0.15 |
| 1 | 5 | -0.09 | 17.42 | 7 | 0.01 | 20.03 | 8.70 | 0.02 | 42.11 | 12 | 0.00 | -1.49 | 0.14 |
| 1 | 6 | -2.40 | 30.19 | 7 | 0.00 | 35.01 | 9.05 | 0.00 | 59.18 | 13 | 0.00 | -1.89 | 0.06 |
| 1 | 7 | -1.93 | 9.17 | 7 | 0.24 | 15.18 | 8.56 | 0.07 | 23.25 | 12 | 0.03 | 1.21 | 0.23 |
| 1 | 8 | 64.94 | 48.73 | 7 | 0.00 | 165.87 | 8.99 | 0.00 | 304.69 | 13 | 0.00 | 8.34 | 0.00 |
| 1 | 9 | -0.30 | 13.99 | 7 | 0.05 | 19.47 | 9.10 | 0.02 | 23.61 | 13 | 0.03 | -1.96 | 0.05 |
| 1 | 10 | 6.84 | 21.56 | 7 | 0.00 | 29.00 | 8.62 | 0.00 | 75.86 | 12 | 0.00 | -3.35 | 0.00 |
| 2 | 3 | 7.45 | 28.37 | 7 | 0.00 | 36.23 | 9.15 | 0.00 | 138.73 | 13 | 0.00 | -1.90 | 0.06 |
| 2 | 4 | 38.51 | 49.48 | 7 | 0.00 | 102.52 | 8.96 | 0.00 | 209.17 | 13 | 0.00 | 5.41 | 0.00 |
| 2 | 5 | 10.52 | 86.11 | 7 | 0.00 | 95.61 | 8.85 | 0.00 | 135.71 | 12 | 0.00 | -1.60 | 0.11 |
| 2 | 6 | 0.46 | 20.46 | 7 | 0.00 | 25.39 | 9.13 | 0.00 | 107.48 | 13 | 0.00 | 0.32 | 0.75 |
| 2 | 7 | 17.28 | 44.93 | 7 | 0.00 | 60.49 | 8.62 | 0.00 | 82.65 | 12 | 0.00 | 3.07 | 0.00 |
| 2 | 8 | 6.70 | 18.05 | 7 | 0.01 | 20.55 | 9.14 | 0.02 | 31.09 | 13 | 0.00 | 0.59 | 0.56 |
| 2 | 9 | 2.00 | 21.85 | 7 | 0.00 | 24.64 | 9.18 | 0.00 | 46.96 | 13 | 0.00 | -0.24 | 0.81 |
| 2 | 10 | 19.44 | 67.75 | 7 | 0.00 | 78.24 | 8.73 | 0.00 | 116.88 | 12 | 0.00 | -2.05 | 0.04 |
| 3 | 4 | -0.62 | 36.75 | 7 | 0.00 | 42.02 | 9.07 | 0.00 | 167.05 | 13 | 0.00 | -2.43 | 0.02 |
| 3 | 5 | 2.79 | 19.56 | 7 | 0.01 | 23.81 | 8.68 | 0.00 | 60.07 | 12 | 0.00 | -1.42 | 0.16 |
| 3 | 6 | 39.25 | 52.63 | 7 | 0.00 | 98.39 | 8.89 | 0.00 | 198.43 | 13 | 0.00 | 3.81 | 0.00 |
| 3 | 7 | 9.44 | 37.89 | 7 | 0.00 | 48.06 | 8.68 | 0.00 | 101.01 | 12 | 0.00 | -2.80 | 0.01 |
| 3 | 8 | 7.92 | 28.49 | 7 | 0.00 | 32.96 | 9.16 | 0.00 | 88.16 | 13 | 0.00 | -1.97 | 0.05 |
| 3 | 9 | -2.54 | 7.38 | 7 | 0.39 | 11.84 | 9.09 | 0.23 | 22.04 | 13 | 0.05 | 0.64 | 0.52 |
| 3 | 10 | 11.08 | 26.06 | 7 | 0.00 | 34.92 | 8.37 | 0.00 | 60.79 | 12 | 0.00 | -0.46 | 0.64 |
| 4 | 5 | 5.67 | 49.31 | 7 | 0.00 | 59.31 | 8.79 | 0.00 | 122.50 | 12 | 0.00 | -3.50 | 0.00 |
| 4 | 6 | 1.82 | 38.99 | 7 | 0.00 | 43.42 | 9.07 | 0.00 | 151.40 | 13 | 0.00 | -1.92 | 0.06 |
| 4 | 7 | 7.81 | 37.44 | 7 | 0.00 | 39.07 | 8.60 | 0.00 | 56.76 | 12 | 0.00 | -0.96 | 0.34 |
| 4 | 8 | 9.29 | 28.84 | 7 | 0.00 | 31.39 | 9.10 | 0.00 | 122.13 | 13 | 0.00 | -1.46 | 0.15 |
| 4 | 9 | -0.84 | 23.46 | 7 | 0.00 | 27.84 | 9.12 | 0.00 | 36.79 | 13 | 0.00 | -0.13 | 0.90 |

**Table 23 (cont.)**. Piecewise fit statistics for IPIP Openness item pairs in the honest condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | 0.58 | 39.17 | 7 | 0.00 | 43.11 | 8.64 | 0.00 | 99.29 | 12 | 0.00 | -2.65 | 0.01 |
| 5 | 6 | 0.08 | 28.11 | 7 | 0.00 | 32.94 | 8.75 | 0.00 | 86.92 | 12 | 0.00 | -1.58 | 0.11 |
| 5 | 7 | 6.51 | 50.81 | 7 | 0.00 | 52.46 | 8.32 | 0.00 | 95.78 | 12 | 0.00 | 0.14 | 0.89 |
| 5 | 8 | 5.41 | 29.51 | 7 | 0.00 | 31.97 | 8.81 | 0.00 | 195.01 | 13 | 0.00 | -0.31 | 0.76 |
| 5 | 9 | -0.55 | 23.67 | 7 | 0.00 | 26.42 | 8.81 | 0.00 | 48.99 | 12 | 0.00 | -0.68 | 0.50 |
| 5 | 10 | 21.68 | 30.10 | 7 | 0.00 | 50.07 | 7.97 | 0.00 | 98.92 | 11 | 0.00 | 2.79 | 0.01 |
| 6 | 7 | 8.26 | 37.72 | 7 | 0.00 | 47.47 | 8.70 | 0.00 | 93.96 | 12 | 0.00 | -2.49 | 0.01 |
| 6 | 8 | 5.47 | 42.29 | 7 | 0.00 | 55.03 | 9.18 | 0.00 | 165.63 | 13 | 0.00 | -3.15 | 0.00 |
| 6 | 9 | -1.54 | 16.72 | 7 | 0.02 | 21.49 | 9.13 | 0.01 | 40.76 | 13 | 0.00 | -0.21 | 0.84 |
| 6 | 10 | -0.39 | 26.44 | 7 | 0.00 | 29.89 | 8.49 | 0.00 | 68.58 | 12 | 0.00 | -0.90 | 0.37 |
| 7 | 8 | 9.30 | 33.31 | 7 | 0.00 | 38.19 | 8.66 | 0.00 | 122.16 | 13 | 0.00 | 1.04 | 0.30 |
| 7 | 9 | 0.01 | 27.00 | 7 | 0.00 | 32.68 | 8.72 | 0.00 | 53.28 | 12 | 0.00 | -1.80 | 0.07 |
| 7 | 10 | 7.25 | 70.58 | 7 | 0.00 | 78.55 | 8.24 | 0.00 | 1986.03 | 12 | 0.00 | -2.31 | 0.02 |
| 8 | 9 | -1.60 | 17.70 | 7 | 0.01 | 19.79 | 9.19 | 0.02 | 38.48 | 13 | 0.00 | 0.32 | 0.75 |
| 8 | 10 | 9.74 | 56.46 | 7 | 0.00 | 62.82 | 8.73 | 0.00 | 79.02 | 12 | 0.00 | -2.32 | 0.02 |
| 9 | 10 | 0.21 | 26.05 | 7 | 0.00 | 33.65 | 8.65 | 0.00 | 51.74 | 12 | 0.00 | 0.38 | 0.70 |
| Mean | | 7.41 | 32.75 | 7.00 | 0.02 | 43.21 | 8.83 | 0.01 | 135.40 | 12.49 | 0.00 | **1.85*** | 0.26 |
| SD | | 12.58 | 16.46 | 0.00 | 0.07 | 28.58 | 0.28 | 0.04 | 284.82 | 0.54 | 0.01 | **1.51*** | 0.30 |
| Max | | 64.94 | 86.11 | 7.00 | 0.39 | 165.87 | 9.19 | 0.23 | 1986.03 | 13.00 | 0.05 | 8.34 | 0.90 |
| Min | | -2.54 | 7.38 | 7.00 | 0.00 | 11.84 | 7.97 | 0.00 | 22.04 | 11.00 | 0.00 | -3.50 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 24**. Piecewise fit statistics for IPIP Openness item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}^2_{ij}$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | -4.58 | 35.53 | 7 | 0.00 | 39.95 | 8.52 | 0.00 | 120.70 | 13 | 0.00 | -1.68 | 0.09 |
| 1 | 3 | 0.47 | 11.72 | 7 | 0.11 | 23.34 | 8.65 | 0.00 | 172.34 | 13 | 0.00 | -3.12 | 0.00 |
| 1 | 4 | -2.45 | 18.83 | 7 | 0.01 | 26.51 | 8.67 | 0.00 | 89.61 | 13 | 0.00 | -1.89 | 0.06 |
| 1 | 5 | -2.97 | 14.82 | 7 | 0.04 | 18.80 | 7.66 | 0.01 | 108.48 | 12 | 0.00 | -0.12 | 0.90 |
| 1 | 6 | -1.11 | 27.32 | 7 | 0.00 | 31.54 | 8.53 | 0.00 | 98.78 | 13 | 0.00 | -1.19 | 0.23 |
| 1 | 7 | -1.30 | 12.53 | 7 | 0.08 | 19.76 | 8.49 | 0.01 | 86.42 | 13 | 0.00 | -0.19 | 0.85 |
| 1 | 8 | 19.79 | 37.62 | 7 | 0.00 | 41.08 | 8.75 | 0.00 | 98.24 | 14 | 0.00 | -1.92 | 0.05 |
| 1 | 9 | 7.70 | 49.99 | 7 | 0.00 | 54.08 | 8.72 | 0.00 | 119.01 | 13 | 0.00 | -1.56 | 0.12 |
| 1 | 10 | -2.81 | 13.54 | 7 | 0.06 | 20.77 | 8.46 | 0.01 | 131.33 | 13 | 0.00 | -0.41 | 0.68 |
| 2 | 3 | 3.33 | 26.28 | 7 | 0.00 | 34.14 | 9.05 | 0.00 | 93.12 | 13 | 0.00 | -4.11 | 0.00 |
| 2 | 4 | 23.76 | 31.22 | 7 | 0.00 | 55.63 | 9.01 | 0.00 | 80.23 | 13 | 0.00 | 0.47 | 0.64 |
| 2 | 5 | -2.96 | 11.18 | 7 | 0.13 | 19.31 | 8.11 | 0.01 | 26.37 | 12 | 0.01 | -1.33 | 0.18 |
| 2 | 6 | 3.51 | 23.08 | 7 | 0.00 | 25.74 | 8.93 | 0.00 | 28.13 | 13 | 0.01 | -1.26 | 0.21 |
| 2 | 7 | 1.30 | 25.53 | 7 | 0.00 | 29.98 | 8.88 | 0.00 | 38.20 | 13 | 0.00 | -0.82 | 0.41 |
| 2 | 8 | 21.36 | 34.75 | 7 | 0.00 | 39.67 | 9.17 | 0.00 | 49.13 | 14 | 0.00 | -3.18 | 0.00 |
| 2 | 9 | 6.00 | 43.38 | 7 | 0.00 | 49.16 | 9.12 | 0.00 | 56.70 | 13 | 0.00 | -2.36 | 0.02 |
| 2 | 10 | -0.88 | 29.67 | 7 | 0.00 | 45.35 | 8.96 | 0.00 | 101.61 | 13 | 0.00 | -2.23 | 0.03 |
| 3 | 4 | 2.20 | 53.66 | 7 | 0.00 | 60.26 | 9.18 | 0.00 | 114.43 | 13 | 0.00 | -4.90 | 0.00 |
| 3 | 5 | 20.07 | 29.42 | 7 | 0.00 | 41.74 | 8.27 | 0.00 | 70.05 | 12 | 0.00 | -3.19 | 0.00 |
| 3 | 6 | 13.10 | 44.43 | 7 | 0.00 | 48.57 | 9.05 | 0.00 | 92.90 | 13 | 0.00 | -2.75 | 0.01 |
| 3 | 7 | 7.04 | 25.79 | 7 | 0.00 | 44.10 | 9.10 | 0.00 | 102.56 | 13 | 0.00 | -3.97 | 0.00 |
| 3 | 8 | 14.45 | 35.05 | 7 | 0.00 | 42.94 | 9.23 | 0.00 | 90.19 | 14 | 0.00 | -0.79 | 0.43 |
| 3 | 9 | 10.88 | 24.31 | 7 | 0.00 | 42.53 | 9.15 | 0.00 | 66.81 | 12 | 0.00 | 0.72 | 0.47 |
| 3 | 10 | 12.73 | 36.54 | 7 | 0.00 | 40.85 | 9.08 | 0.00 | 124.28 | 13 | 0.00 | -2.57 | 0.01 |
| 4 | 5 | -1.85 | 5.78 | 7 | 0.57 | 18.07 | 8.30 | 0.02 | 291.91 | 13 | 0.00 | -2.38 | 0.02 |
| 4 | 6 | 0.96 | 29.74 | 7 | 0.00 | 32.43 | 9.10 | 0.00 | 38.47 | 13 | 0.00 | -2.13 | 0.03 |
| 4 | 7 | -4.53 | 10.20 | 7 | 0.18 | 15.13 | 9.09 | 0.09 | 28.00 | 13 | 0.01 | -1.62 | 0.10 |
| 4 | 8 | 15.60 | 39.31 | 7 | 0.00 | 43.86 | 9.29 | 0.00 | 63.13 | 14 | 0.00 | -4.06 | 0.00 |
| 4 | 9 | 13.71 | 76.93 | 7 | 0.00 | 83.88 | 9.24 | 0.00 | 90.75 | 13 | 0.00 | -3.22 | 0.00 |

**Table 24 (cont.)**. Piecewise fit statistics for IPIP Openness item pairs in the faking condition.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 10 | -3.97 | 12.13 | 7 | 0.10 | 24.91 | 9.13 | 0.00 | 71.27 | 13 | 0.00 | -2.51 | 0.01 |
| 5 | 6 | -2.52 | 15.00 | 7 | 0.04 | 24.40 | 8.14 | 0.00 | 30.42 | 12 | 0.00 | -1.55 | 0.12 |
| 5 | 7 | 4.57 | 19.39 | 7 | 0.01 | 21.14 | 7.93 | 0.01 | 23.20 | 12 | 0.03 | 0.48 | 0.63 |
| 5 | 8 | 10.71 | 18.88 | 7 | 0.01 | 23.88 | 8.39 | 0.00 | 587.79 | 14 | 0.00 | -2.82 | 0.00 |
| 5 | 9 | 10.98 | 34.91 | 7 | 0.00 | 45.30 | 8.35 | 0.00 | 69.54 | 12 | 0.00 | -2.05 | 0.04 |
| 5 | 10 | 22.78 | 42.55 | 7 | 0.00 | 49.99 | 7.82 | 0.00 | 92.05 | 12 | 0.00 | 1.09 | 0.28 |
| 6 | 7 | -4.44 | 24.37 | 7 | 0.00 | 28.72 | 8.92 | 0.00 | 42.69 | 13 | 0.00 | -1.29 | 0.20 |
| 6 | 8 | 15.30 | 29.51 | 7 | 0.00 | 35.33 | 9.19 | 0.00 | 51.32 | 14 | 0.00 | -3.53 | 0.00 |
| 6 | 9 | 6.99 | 38.76 | 7 | 0.00 | 43.59 | 9.15 | 0.00 | 53.73 | 13 | 0.00 | -2.28 | 0.02 |
| 6 | 10 | -2.33 | 20.95 | 7 | 0.00 | 27.13 | 8.95 | 0.00 | 68.76 | 13 | 0.00 | -1.53 | 0.13 |
| 7 | 8 | 18.11 | 34.83 | 7 | 0.00 | 38.85 | 9.20 | 0.00 | 52.40 | 14 | 0.00 | -2.79 | 0.01 |
| 7 | 9 | 11.18 | 41.88 | 7 | 0.00 | 50.39 | 9.17 | 0.00 | 68.29 | 13 | 0.00 | -2.24 | 0.02 |
| 7 | 10 | 3.50 | 18.30 | 7 | 0.01 | 21.67 | 8.76 | 0.01 | 59.51 | 13 | 0.00 | -0.30 | 0.77 |
| 8 | 9 | 10.64 | 31.68 | 7 | 0.00 | 36.65 | 9.32 | 0.00 | 48.32 | 14 | 0.00 | -0.26 | 0.80 |
| 8 | 10 | 11.36 | 31.64 | 7 | 0.00 | 35.69 | 9.19 | 0.00 | 70.14 | 14 | 0.00 | -2.42 | 0.02 |
| 9 | 10 | 13.40 | 31.77 | 7 | 0.00 | 37.60 | 9.17 | 0.00 | 66.88 | 13 | 0.00 | -1.74 | 0.08 |
| Mean | | 6.64 | 28.99 | 7.00 | 0.03 | 36.32 | 8.81 | 0.00 | 91.74 | 13.02 | 0.00 | **1.98*** | 0.19 |
| SD | | 8.40 | 13.31 | 0.00 | 0.09 | 13.43 | 0.43 | 0.01 | 87.26 | 0.61 | 0.00 | **1.16*** | 0.27 |
| Max | | 23.76 | 76.93 | 7.00 | 0.57 | 83.88 | 9.32 | 0.09 | 587.79 | 14.00 | 0.03 | 1.09 | 0.90 |
| Min | | -4.58 | 5.78 | 7.00 | 0.00 | 15.13 | 7.66 | 0.00 | 23.20 | 12.00 | 0.00 | -4.90 | 0.00 |

*Note. The mean and standard deviation of the $z_{ord}$ statistics were calculated from the absolute values of the original $z_{ord}$ statistics.

**Table 25**. Fit indices of confirmatory factor analyses for a unidimensional model of International Personality Item Pool (IPIP) items the honest and faking conditions.

| Dimension | Condition | CFI | TLI | RMSEA (90% CI) |
|---|---|---|---|---|
| Agreeableness | **Honest** | 0.943 | 0.926 | 0.121 (0.108, 0.134) |
| | Faking | 0.935 | 0.917 | 0.155 (0.142, 0.168) |
| Conscientiousness | Honest | 0.886 | 0.853 | 0.135 (0.122, 0.148) |
| | **Faking** | 0.983 | 0.978 | 0.082 (0.068, 0.096) |
| Extraversion | **Honest** | 0.949 | 0.935 | 0.127 (0.114, 0.140) |
| | Faking | 0.910 | 0.884 | 0.154 (0.140, 0.167) |
| Emotional Stability | Honest | 0.923 | 0.901 | 0.179 (0.167, 0.192) |
| | **Faking** | 0.991 | 0.989 | 0.062 (0.047, 0.077) |
| Openness | Honest | 0.820 | 0.768 | 0.226 (0.214, 0.239) |
| | **Faking** | 0.955 | 0.943 | 0.112 (0.098, 0.125) |

Note: CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval.

**Table 26**. Parameter estimates for items in the Counter-productive Work Behavior (CWB) scale.

| Item | Mean | SD | a | b1 | b2 | b3 | b4 | b5 |
|---|---|---|---|---|---|---|---|---|
| CWBI-1 | 1.904 | 1.851 | 1.084 | -0.478 | 0.122 | 0.438 | 1.018 | 1.456 |
| CWBI-2 | 0.744 | 1.204 | 2.007 | 0.369 | 0.889 | 1.358 | 1.966 | 2.400 |
| CWBI-3 | 0.770 | 1.471 | 1.162 | 0.744 | 1.158 | 1.449 | 1.701 | 2.190 |
| CWBI-4 | 0.879 | 1.556 | 1.577 | 0.632 | 0.845 | 1.126 | 1.525 | 1.870 |
| CWBI-5 | 0.626 | 1.238 | 1.759 | 0.671 | 1.102 | 1.424 | 1.878 | 2.273 |
| CWBI-6 | 0.835 | 1.317 | 1.610 | 0.351 | 0.868 | 1.278 | 1.699 | 2.548 |
| CWBI-7 | 0.342 | 0.957 | 1.828 | 1.131 | 1.486 | 1.863 | 2.079 | 2.534 |
| CWBO-1 | 0.283 | 0.909 | 1.403 | 1.403 | 1.830 | 2.035 | 2.564 | 3.036 |
| CWBO-2 | 2.051 | 1.754 | 0.643 | -1.230 | -0.061 | 0.521 | 1.246 | 2.106 |
| CWBO-3 | 0.194 | 0.802 | 1.505 | 1.791 | 2.007 | 2.232 | 2.489 | 3.199 |
| CWBO-4 | 1.168 | 1.574 | 0.994 | 0.117 | 0.603 | 1.025 | 1.747 | 2.351 |
| CWBO-5 | 1.279 | 1.494 | 0.790 | -0.318 | 0.648 | 1.309 | 1.917 | 3.056 |
| CWBO-6 | 0.429 | 0.980 | 1.213 | 0.924 | 1.707 | 2.090 | 2.641 | 3.171 |
| CWBO-7 | 0.529 | 0.964 | 1.293 | 0.614 | 1.285 | 2.054 | 2.722 | 3.709 |
| CWBO-8 | 1.022 | 1.430 | 0.827 | 0.153 | 0.934 | 1.568 | 2.163 | 3.010 |
| CWBO-9 | 0.236 | 0.800 | 1.385 | 1.508 | 1.917 | 2.257 | 2.774 | 3.459 |
| CWBO-10 | 0.236 | 0.872 | 1.319 | 1.734 | 2.007 | 2.243 | 2.668 | 3.115 |
| CWBO-11 | 1.205 | 1.390 | 0.978 | -0.269 | 0.624 | 1.224 | 2.040 | 2.871 |
| CWBO-12 | 0.385 | 1.014 | 1.280 | 1.218 | 1.623 | 2.013 | 2.458 | 3.113 |

Note. *N* = 449. The a-parameters do not include the scaling constant of 1.702.

**Table 27**. Piecewise fit statistics for CWB-I item pairs.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}_{ij}^2$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | **10.04** | **44.38** | 23 | 0.00 | **63.77** | 25.63 | 0.00 | **69.31** | 34 | 0.00 | 0.19 | 0.85 |
| 1 | 3 | -1.29 | **36.20** | 23 | 0.04 | **44.42** | 25.44 | 0.01 | **62.94** | 33 | 0.00 | 1.39 | 0.16 |
| 1 | 4 | **13.19** | **47.35** | 23 | 0.00 | **66.88** | 25.52 | 0.00 | **94.39** | 34 | 0.00 | 1.31 | 0.19 |
| 1 | 5 | **6.99** | **40.15** | 23 | 0.01 | **57.95** | 25.50 | 0.00 | **72.96** | 33 | 0.00 | -1.47 | 0.14 |
| 1 | 6 | **5.45** | 30.14 | 23 | 0.15 | **43.76** | 25.55 | 0.01 | **50.85** | 33 | 0.02 | -0.66 | 0.51 |
| 1 | 7 | **5.31** | 28.52 | 23 | 0.20 | 36.43 | 25.48 | 0.07 | **59.94** | 32 | 0.00 | **-2.29** | 0.02 |
| 2 | 3 | -2.36 | 27.41 | 23 | 0.24 | 31.88 | 25.42 | 0.18 | 44.29 | 34 | 0.11 | -0.02 | 0.99 |
| 2 | 4 | -0.49 | 27.72 | 23 | 0.23 | 35.84 | 25.55 | 0.09 | 46.55 | 33 | 0.06 | -1.30 | 0.19 |
| 2 | 5 | -2.97 | 19.76 | 23 | 0.66 | 21.35 | 25.32 | 0.69 | 24.08 | 33 | 0.87 | -0.80 | 0.42 |
| 2 | 6 | -0.32 | 32.06 | 23 | 0.10 | 36.12 | 25.38 | 0.08 | **61.43** | 33 | 0.00 | 0.32 | 0.75 |
| 2 | 7 | -3.15 | 27.63 | 23 | 0.23 | 30.93 | 25.22 | 0.20 | 41.87 | 31 | 0.09 | -1.19 | 0.23 |
| 3 | 4 | -0.94 | 34.31 | 23 | 0.06 | 36.66 | 25.40 | 0.07 | **53.94** | 34 | 0.02 | -1.09 | 0.28 |
| 3 | 5 | -4.95 | 13.52 | 23 | 0.94 | 15.43 | 25.31 | 0.94 | 30.25 | 33 | 0.60 | -1.57 | 0.12 |
| 3 | 6 | -4.68 | 20.59 | 23 | 0.61 | 22.76 | 25.43 | 0.62 | 29.28 | 33 | 0.65 | -1.65 | 0.10 |
| 3 | 7 | -2.92 | 19.54 | 23 | 0.67 | 26.66 | 25.18 | 0.38 | 40.98 | 32 | 0.13 | 0.67 | 0.50 |
| 4 | 5 | 2.81 | 26.11 | 23 | 0.30 | 31.84 | 25.28 | 0.17 | **83.41** | 34 | 0.00 | 0.51 | 0.61 |
| 4 | 6 | **7.50** | **49.23** | 23 | 0.00 | **55.39** | 25.46 | 0.00 | **77.13** | 34 | 0.00 | -0.38 | 0.70 |
| 4 | 7 | **4.45** | **38.85** | 23 | 0.02 | **47.20** | 25.22 | 0.01 | **76.21** | 33 | 0.00 | -0.68 | 0.49 |
| 5 | 6 | -0.66 | 28.41 | 23 | 0.20 | 29.73 | 25.31 | 0.25 | 34.32 | 33 | 0.40 | -1.02 | 0.31 |
| 5 | 7 | -3.78 | 18.35 | 23 | 0.74 | 21.68 | 24.93 | 0.65 | 43.31 | 32 | 0.09 | 0.92 | 0.36 |
| 6 | 7 | -1.37 | 22.88 | 23 | 0.47 | 25.18 | 25.17 | 0.46 | **49.46** | 32 | 0.03 | 0.18 | 0.86 |
| | | | | | | | | | | | | | |
| Mean | | 1.23 | 30.15 | 23.00 | 0.28 | 37.23 | 25.37 | 0.23 | 54.61 | 33.00 | 0.15 | -0.41 | 0.42 |
| SD | | 5.04 | 9.56 | 0.00 | 0.28 | 14.02 | 0.16 | 0.27 | 18.71 | 0.82 | 0.25 | 1.01 | 0.28 |
| Max | | 13.19 | 49.23 | 23.00 | 0.94 | 66.88 | 25.63 | 0.94 | 94.39 | 34.00 | 0.87 | 1.39 | 0.99 |
| Min | | -4.95 | 13.52 | 23.00 | 0.00 | 15.43 | 24.93 | 0.00 | 24.08 | 31.00 | 0.00 | -2.29 | 0.02 |

**Table 28**. Piecewise fit statistics for CWB-O item pairs.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}^2_{ij}$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | -3.66 | 24.26 | 23 | 0.39 | 37.18 | 25.27 | 0.06 | **207.36** | 33 | 0.00 | -0.29 | 0.77 |
| 1 | 3 | -2.65 | 31.29 | 23 | 0.12 | **47.54** | 24.80 | 0.00 | **77.05** | 29 | 0.00 | **4.20** | 0.00 |
| 1 | 4 | -3.21 | 33.29 | 23 | 0.08 | **43.15** | 24.86 | 0.01 | **89.11** | 30 | 0.00 | 0.57 | 0.57 |
| 1 | 5 | 0.32 | 22.00 | 23 | 0.52 | 27.81 | 25.19 | 0.33 | **366.72** | 32 | 0.00 | -0.04 | 0.97 |
| 1 | 6 | -4.04 | 28.79 | 23 | 0.19 | 34.80 | 25.16 | 0.10 | **48.37** | 30 | 0.02 | **2.42** | 0.02 |
| 1 | 7 | -1.62 | 23.06 | 23 | 0.46 | 26.95 | 25.31 | 0.37 | **48.09** | 31 | 0.03 | 1.46 | 0.15 |
| 1 | 8 | -0.73 | 19.72 | 23 | 0.66 | 30.11 | 25.17 | 0.23 | **53.13** | 31 | 0.01 | 0.18 | 0.86 |
| 1 | 9 | **3.40** | **49.70** | 23 | 0.00 | **60.02** | 25.08 | 0.00 | **78.93** | 29 | 0.00 | **4.25** | 0.00 |
| 1 | 10 | -2.67 | 32.33 | 23 | 0.09 | **58.19** | 25.11 | 0.00 | **92.32** | 30 | 0.00 | **4.12** | 0.00 |
| 1 | 11 | -3.14 | 24.34 | 23 | 0.39 | 29.40 | 25.17 | 0.26 | 40.26 | 32 | 0.15 | 0.00 | 1.00 |
| 1 | 12 | -1.34 | 22.64 | 23 | 0.48 | 30.82 | 25.22 | 0.20 | 36.79 | 30 | 0.18 | **2.76** | 0.01 |
| 2 | 3 | -1.28 | 22.94 | 23 | 0.46 | **41.15** | 25.47 | 0.03 | **49.56** | 32 | 0.02 | -0.35 | 0.73 |
| 2 | 4 | **8.15** | **61.11** | 23 | 0.00 | **84.17** | 25.13 | 0.00 | **95.85** | 33 | 0.00 | **2.54** | 0.01 |
| 2 | 5 | **3.70** | 30.99 | 23 | 0.12 | **47.65** | 25.40 | 0.00 | **67.25** | 34 | 0.00 | **3.52** | 0.00 |
| 2 | 6 | -0.54 | 20.47 | 23 | 0.61 | 31.60 | 25.67 | 0.19 | **60.04** | 34 | 0.00 | -1.27 | 0.20 |
| 2 | 7 | -0.91 | 28.07 | 23 | 0.21 | 36.10 | 25.78 | 0.09 | **61.27** | 34 | 0.00 | -0.83 | 0.41 |
| 2 | 8 | **5.27** | **37.68** | 23 | 0.03 | **62.76** | 25.41 | 0.00 | **85.48** | 34 | 0.00 | **5.31** | 0.00 |
| 2 | 9 | **3.38** | 31.90 | 23 | 0.10 | **43.97** | 25.74 | 0.01 | **50.87** | 32 | 0.02 | -0.82 | 0.41 |
| 2 | 10 | -1.31 | **39.80** | 23 | 0.02 | **52.07** | 25.71 | 0.00 | **69.89** | 33 | 0.00 | -0.20 | 0.84 |
| 2 | 11 | **7.58** | **37.05** | 23 | 0.03 | **55.11** | 25.36 | 0.00 | **87.43** | 35 | 0.00 | **5.64** | 0.00 |
| 2 | 12 | 1.36 | 30.41 | 23 | 0.14 | 35.45 | 25.77 | 0.10 | **229.85** | 34 | 0.00 | 0.46 | 0.65 |
| 3 | 4 | -2.49 | 28.43 | 23 | 0.20 | **38.98** | 25.08 | 0.04 | **77.82** | 30 | 0.00 | 0.41 | 0.68 |
| 3 | 5 | -0.43 | 25.54 | 23 | 0.32 | 38.23 | 25.39 | 0.05 | **66.84** | 31 | 0.00 | 0.30 | 0.77 |
| 3 | 6 | -4.52 | 25.45 | 23 | 0.33 | **40.72** | 25.35 | 0.03 | **64.46** | 30 | 0.00 | **3.49** | 0.00 |
| 3 | 7 | -2.19 | **40.34** | 23 | 0.01 | **47.29** | 25.53 | 0.01 | **70.38** | 31 | 0.00 | 2.00 | 0.05 |
| 3 | 8 | -0.38 | 23.85 | 23 | 0.41 | 36.89 | 25.38 | 0.07 | **52.65** | 31 | 0.01 | -0.32 | 0.75 |
| 3 | 9 | -3.71 | 25.56 | 23 | 0.32 | **39.07** | 25.26 | 0.04 | **67.30** | 29 | 0.00 | **3.90** | 0.00 |
| 3 | 10 | 2.57 | **38.95** | 23 | 0.02 | **58.19** | 25.28 | 0.00 | **98.71** | 30 | 0.00 | **3.63** | 0.00 |
| 3 | 11 | -3.67 | 24.13 | 23 | 0.40 | 28.18 | 25.39 | 0.32 | 33.66 | 32 | 0.39 | 0.11 | 0.91 |

**Table 28 (cont.)**. Piecewise fit statistics for CWB-O item pairs.

| Item1 | Item2 | Adjusted $\chi^2$ / *df* | $M_{ij}$ | *df* | *p* | $\bar{X}_{ij}^2$ | *df* | *p* | $R_{ij}$ | *df* | *p* | $z_{ord}$ | *p* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 12 | -5.29 | 22.03 | 23 | 0.52 | 31.06 | 25.40 | 0.20 | 39.60 | 30 | 0.11 | **2.99** | 0.00 |
| 4 | 5 | **3.03** | **40.64** | 23 | 0.01 | **56.93** | 25.03 | 0.00 | **99.72** | 32 | 0.00 | **3.64** | 0.00 |
| 4 | 6 | -2.11 | 28.39 | 23 | 0.20 | 32.83 | 25.24 | 0.14 | **55.12** | 31 | 0.00 | 1.60 | 0.11 |
| 4 | 7 | -1.39 | **35.93** | 23 | 0.04 | **39.32** | 25.37 | 0.04 | **87.41** | 32 | 0.00 | -0.72 | 0.47 |
| 4 | 8 | **3.92** | **47.87** | 23 | 0.00 | **53.41** | 25.08 | 0.00 | **73.30** | 32 | 0.00 | **2.10** | 0.04 |
| 4 | 9 | -2.98 | 30.95 | 23 | 0.12 | **40.88** | 25.34 | 0.03 | **83.25** | 30 | 0.00 | 0.90 | 0.37 |
| 4 | 10 | -3.96 | 23.62 | 23 | 0.42 | 29.89 | 25.33 | 0.24 | **62.10** | 31 | 0.00 | 1.42 | 0.16 |
| 4 | 11 | **4.98** | **43.70** | 23 | 0.01 | **46.68** | 25.07 | 0.01 | **59.47** | 33 | 0.00 | 0.19 | 0.85 |
| 4 | 12 | -3.60 | 7.33 | 23 | 1.00 | 14.01 | 25.35 | 0.97 | 38.83 | 31 | 0.16 | 1.93 | 0.05 |
| 5 | 6 | -1.09 | 25.54 | 23 | 0.32 | **45.56** | 25.57 | 0.01 | **69.26** | 33 | 0.00 | **2.26** | 0.02 |
| 5 | 7 | 1.38 | 29.14 | 23 | 0.18 | **43.23** | 25.69 | 0.02 | **157.60** | 33 | 0.00 | 0.04 | 0.97 |
| 5 | 8 | -0.79 | **36.84** | 23 | 0.03 | **42.40** | 25.37 | 0.02 | **66.39** | 33 | 0.00 | 3.21 | 0.00 |
| 5 | 9 | -1.06 | 28.27 | 23 | 0.21 | **44.71** | 25.66 | 0.01 | **77.01** | 31 | 0.00 | 1.20 | 0.23 |
| 5 | 10 | -1.71 | 19.55 | 23 | 0.67 | 29.72 | 25.63 | 0.26 | **55.46** | 32 | 0.01 | 1.39 | 0.17 |
| 5 | 11 | 2.63 | 27.84 | 23 | 0.22 | 32.15 | 25.34 | 0.16 | **65.43** | 34 | 0.00 | **3.08** | 0.00 |
| 5 | 12 | -1.00 | **36.33** | 23 | 0.04 | **39.64** | 25.69 | 0.04 | **84.39** | 32 | 0.00 | 0.57 | 0.57 |
| 6 | 7 | -1.73 | **45.95** | 23 | 0.00 | **51.15** | 25.75 | 0.00 | **81.52** | 32 | 0.00 | 1.77 | 0.08 |
| 6 | 8 | 0.89 | 28.25 | 23 | 0.21 | 38.78 | 25.59 | 0.05 | **76.89** | 33 | 0.00 | -0.17 | 0.87 |
| 6 | 9 | -4.89 | 29.27 | 23 | 0.17 | 34.82 | 25.64 | 0.11 | **47.03** | 30 | 0.02 | **2.45** | 0.01 |
| 6 | 10 | -2.76 | 27.19 | 23 | 0.25 | **39.85** | 25.64 | 0.04 | **59.77** | 31 | 0.00 | **3.94** | 0.00 |
| 6 | 11 | **4.53** | 33.45 | 23 | 0.07 | 38.42 | 25.60 | 0.05 | **64.04** | 33 | 0.00 | -0.64 | 0.52 |
| 6 | 12 | -2.77 | 22.03 | 23 | 0.52 | 26.89 | 25.73 | 0.40 | 45.36 | 31 | 0.05 | 1.56 | 0.12 |
| 7 | 8 | **7.37** | 27.40 | 23 | 0.24 | **43.10** | 25.66 | 0.02 | **70.50** | 33 | 0.00 | **2.33** | 0.02 |
| 7 | 9 | -1.86 | 34.34 | 23 | 0.06 | **42.12** | 25.78 | 0.02 | **66.36** | 31 | 0.00 | **2.72** | 0.01 |
| 7 | 10 | -5.36 | 31.48 | 23 | 0.11 | 38.06 | 25.81 | 0.06 | **63.55** | 32 | 0.00 | **2.81** | 0.00 |
| 7 | 11 | 0.30 | 20.86 | 23 | 0.59 | 26.11 | 25.65 | 0.44 | **55.04** | 34 | 0.01 | 1.06 | 0.29 |
| 7 | 12 | -2.03 | 31.13 | 23 | 0.12 | **41.22** | 25.89 | 0.03 | **69.88** | 32 | 0.00 | 0.29 | 0.77 |
| 8 | 9 | 1.30 | 24.67 | 23 | 0.37 | **40.05** | 25.65 | 0.03 | **55.08** | 31 | 0.00 | 0.32 | 0.75 |
| 8 | 10 | -1.35 | **36.30** | 23 | 0.04 | **46.46** | 25.62 | 0.01 | **68.36** | 32 | 0.00 | 0.82 | 0.41 |

**Table 28 (cont.)**. Piecewise fit statistics for CWB-O item pairs.

| Item1 | Item2 | Adjusted $\chi^2$ / $df$ | $M_{ij}$ | $df$ | $p$ | $\bar{X}_{ij}^2$ | $df$ | $p$ | $R_{ij}$ | $df$ | $p$ | $z_{ord}$ | $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 11 | **7.66** | **44.92** | 23 | 0.00 | **56.71** | 25.35 | 0.00 | **82.06** | 34 | 0.00 | **4.64** | 0.00 |
| 8 | 12 | -3.12 | 23.30 | 23 | 0.44 | 30.40 | 25.67 | 0.24 | **56.59** | 32 | 0.00 | **2.37** | 0.02 |
| 9 | 10 | -2.13 | **40.44** | 23 | 0.01 | **55.86** | 25.58 | 0.00 | **91.51** | 30 | 0.00 | **3.30** | 0.00 |
| 9 | 11 | -2.90 | 26.10 | 23 | 0.30 | 31.66 | 25.65 | 0.19 | 35.12 | 32 | 0.32 | 0.95 | 0.34 |
| 9 | 12 | -5.47 | 32.49 | 23 | 0.09 | **42.44** | 25.68 | 0.02 | **49.11** | 30 | 0.02 | **3.80** | 0.00 |
| 10 | 11 | -4.08 | 30.65 | 23 | 0.13 | 33.85 | 25.64 | 0.13 | 43.92 | 33 | 0.10 | 1.44 | 0.15 |
| 10 | 12 | -4.40 | 34.42 | 23 | 0.06 | **43.94** | 25.70 | 0.01 | **55.74** | 31 | 0.00 | **2.97** | 0.00 |
| 11 | 12 | -2.20 | **36.98** | 23 | 0.03 | **40.58** | 25.69 | 0.03 | **62.20** | 33 | 0.00 | 0.17 | 0.87 |
| | | | | | | | | | | | | | |
| Mean | | -0.65 | 30.75 | 23.00 | 0.23 | 41.04 | 25.45 | 0.10 | 75.78 | 31.76 | 0.03 | 1.69 | 0.30 |
| SD | | 3.37 | 8.56 | 0.00 | 0.21 | 10.96 | 0.25 | 0.15 | 48.68 | 1.45 | 0.07 | 1.68 | 0.35 |
| Max | | 8.15 | 61.11 | 23.00 | 1.00 | 84.17 | 25.89 | 0.97 | 366.72 | 35.00 | 0.39 | 5.64 | 1.00 |
| Min | | -5.47 | 7.33 | 23.00 | 0.00 | 14.01 | 24.80 | 0.00 | 33.66 | 29.00 | 0.00 | -1.27 | 0.00 |

# FIGURES

**Figure 1**. Scree plot for the 69 MC and 7 CR items of the Physics B Exam.

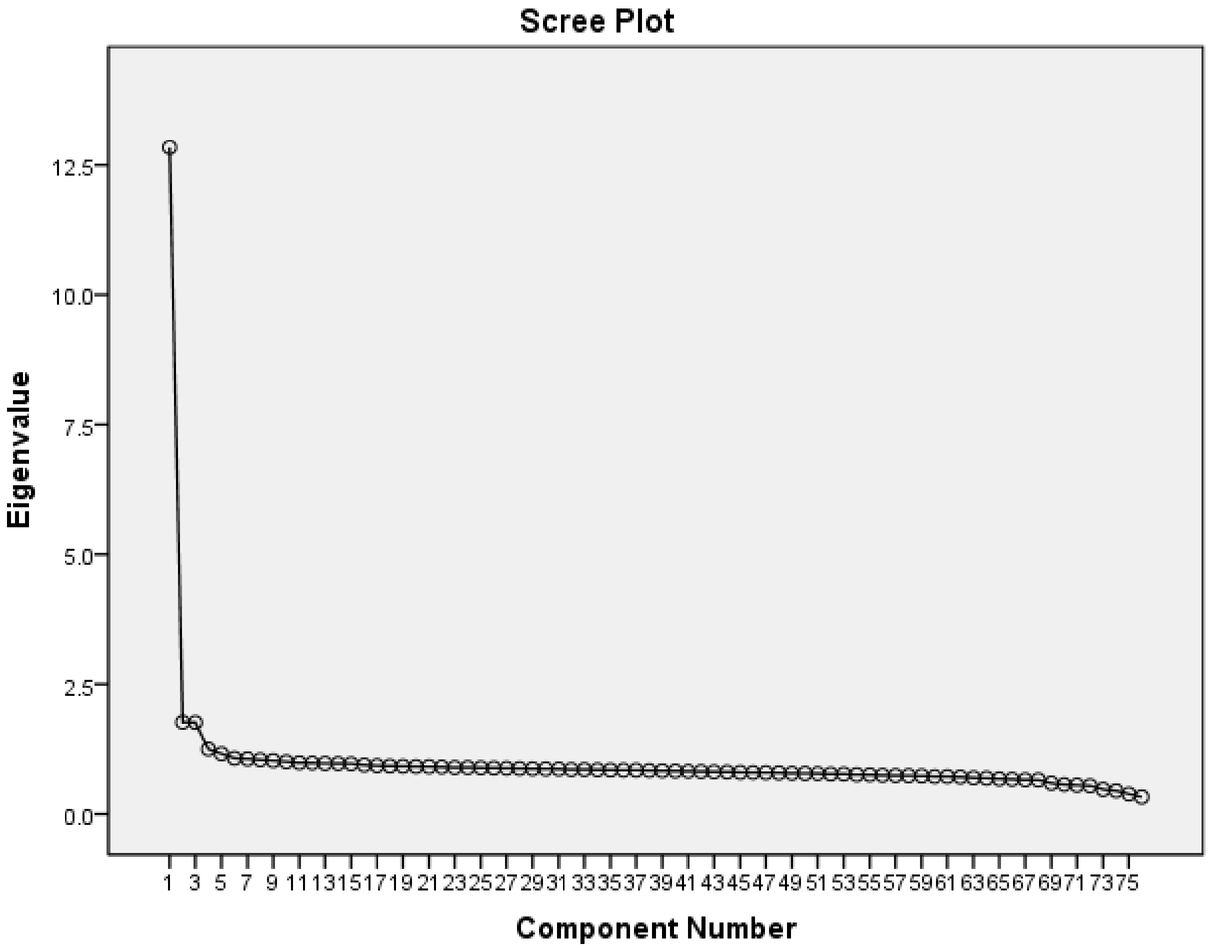


**Figure 2**. Histogram of the $\bar{X}^2_{ij}$ statistics for the 69 MC items of the Physics B Exam.

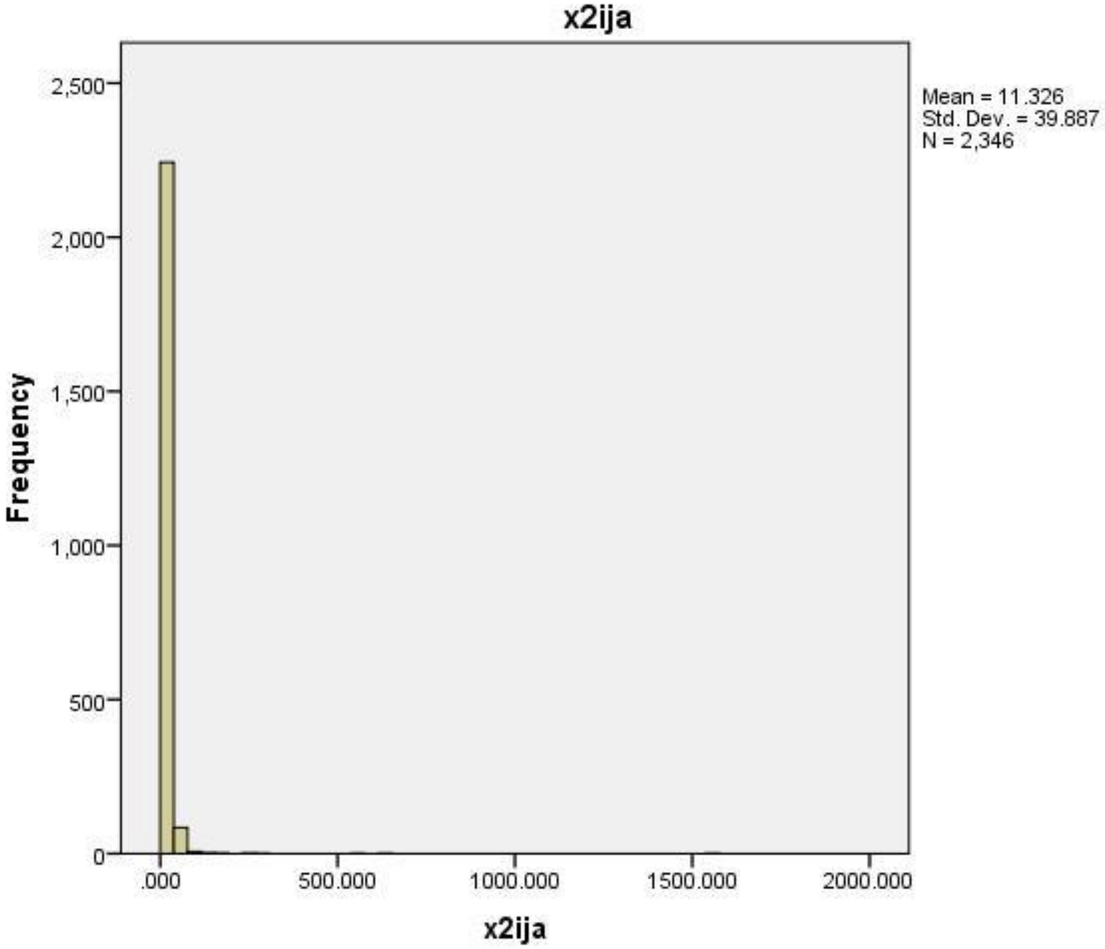


**Figure 3**. Histogram of the $R_{ij}$ statistics for the 69 MC items of the Physics B Exam.

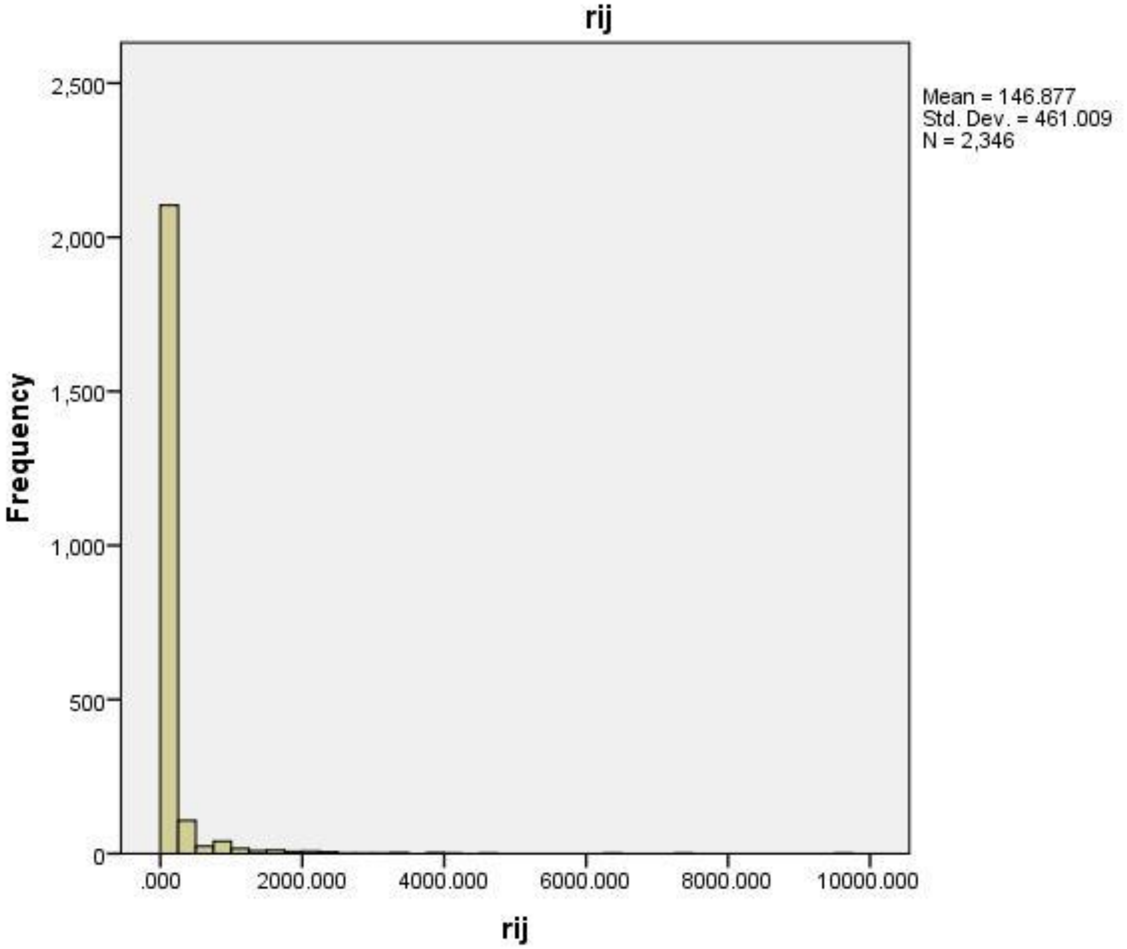


**Figure 4**. Histogram of the $z_{ij}$ statistics for the 69 MC items of the Physics B Exam.

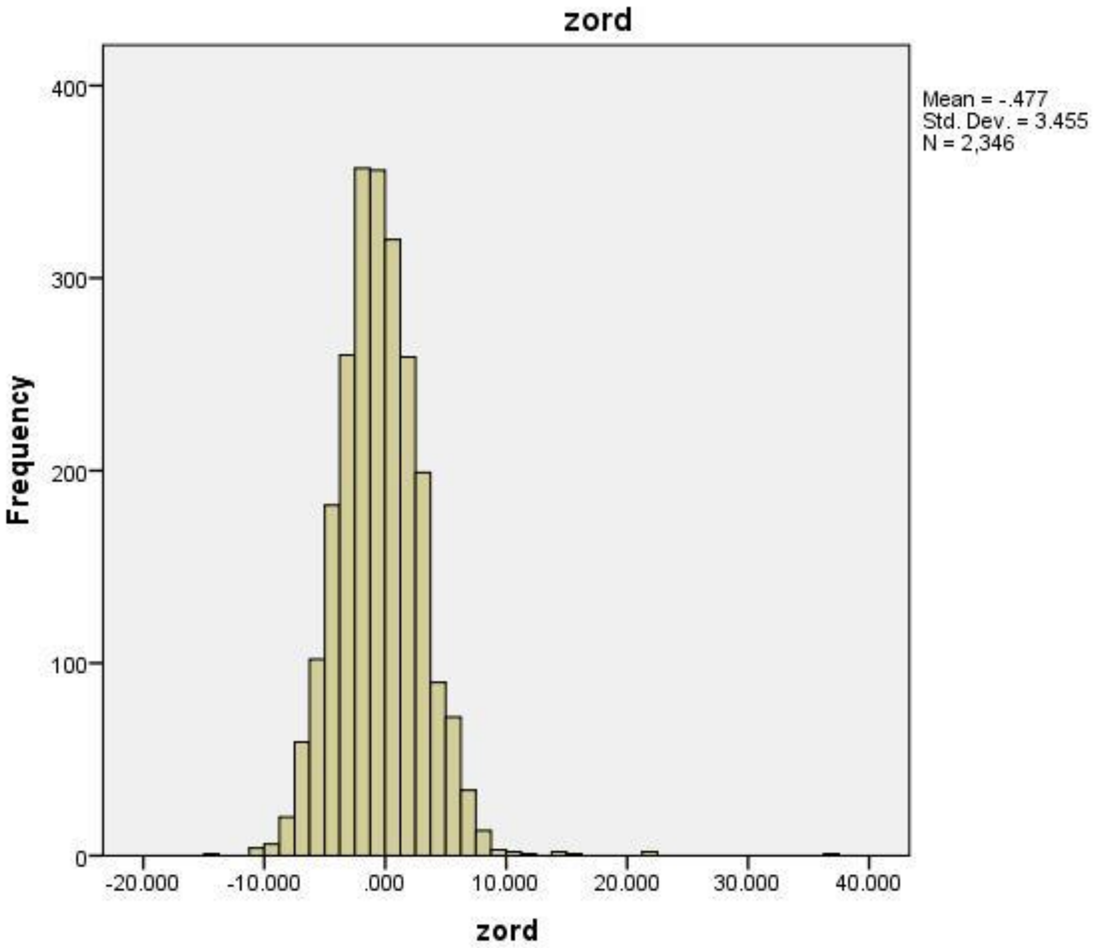


**Figure 5**. Scree plot for the 70 MC and 3 CR items of the World History Exam.

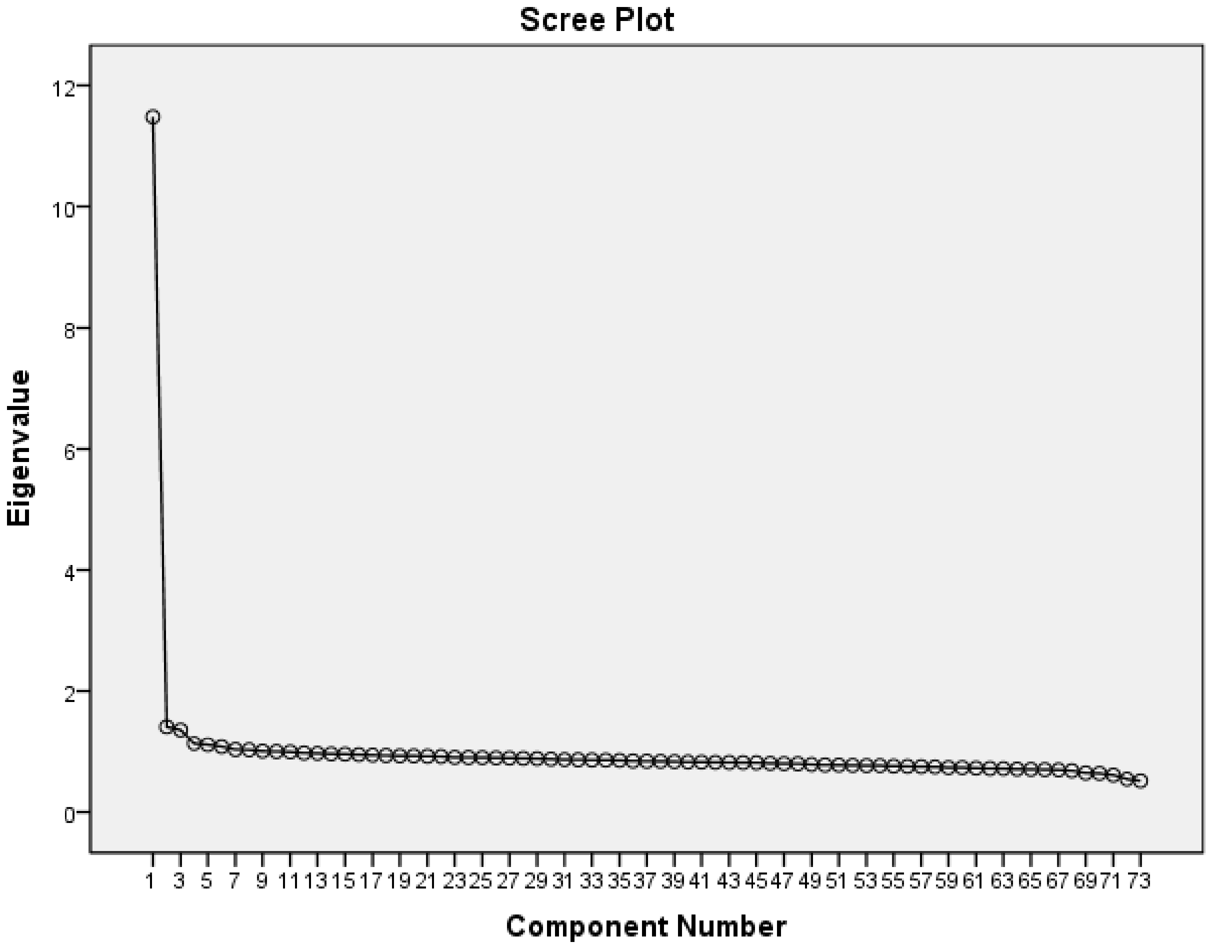


**Figure 6**. Histogram of the $\bar{X}_{ij}^2$ statistics for the 70 MC items of the World History Exam.

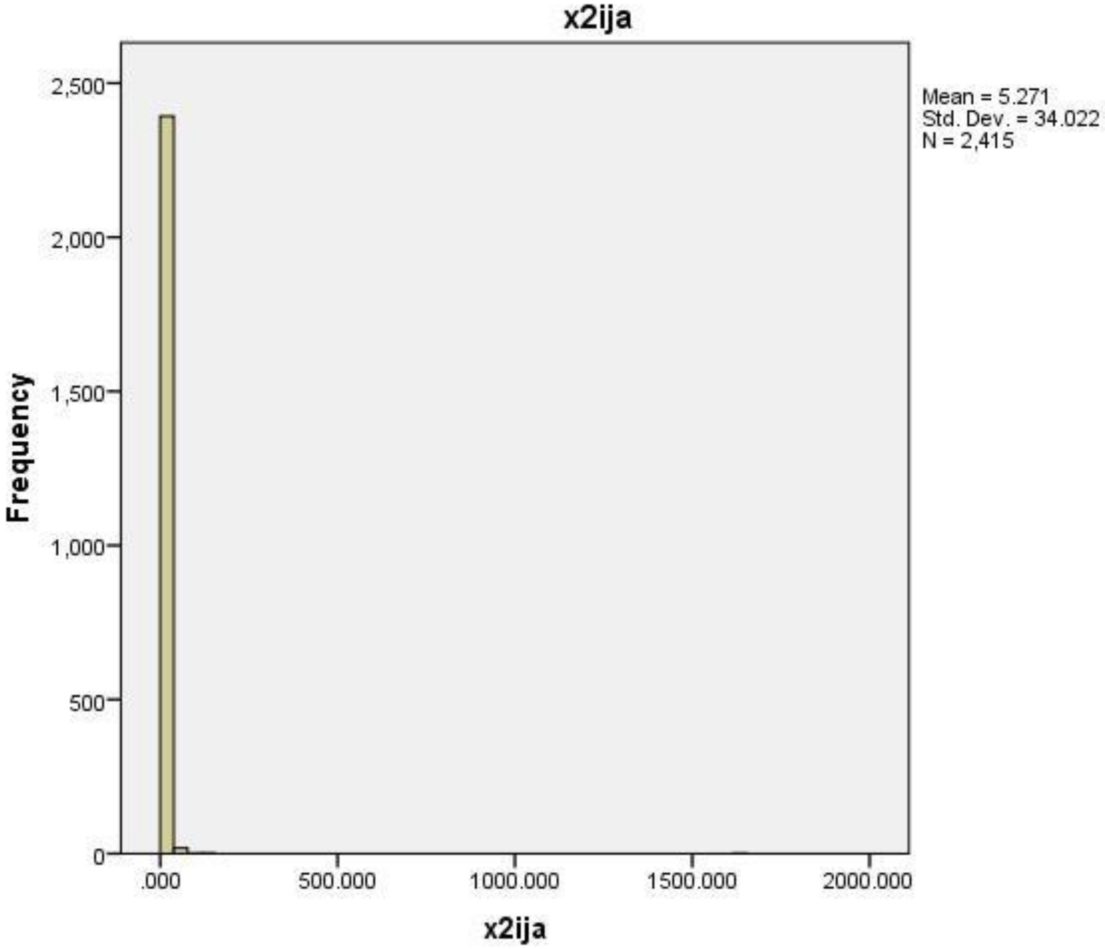


**Figure 7**. Histogram of the $R_{ij}$ statistics for the 70 MC items of the World History Exam.

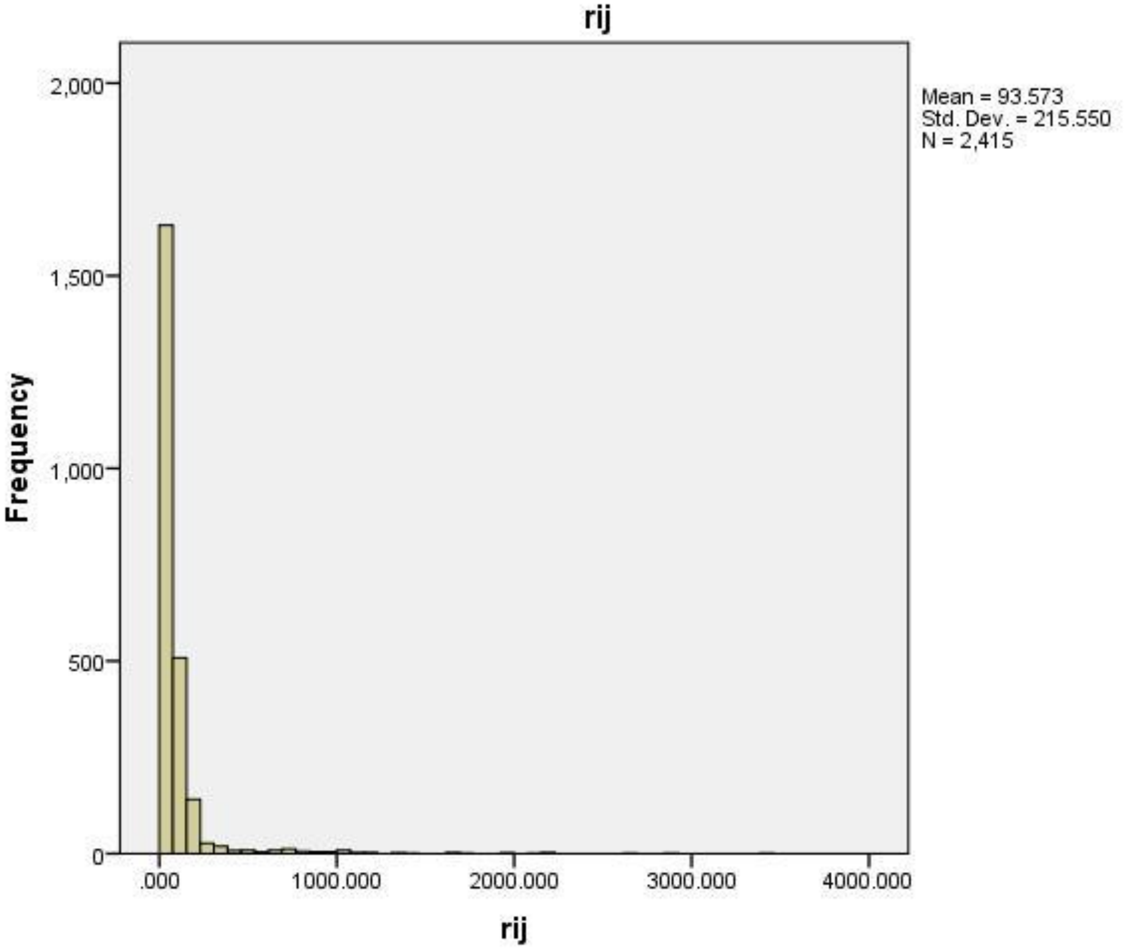


**Figure 8**. Histogram of the $z_{ij}$ statistics for the 70 MC items of the World History Exam.

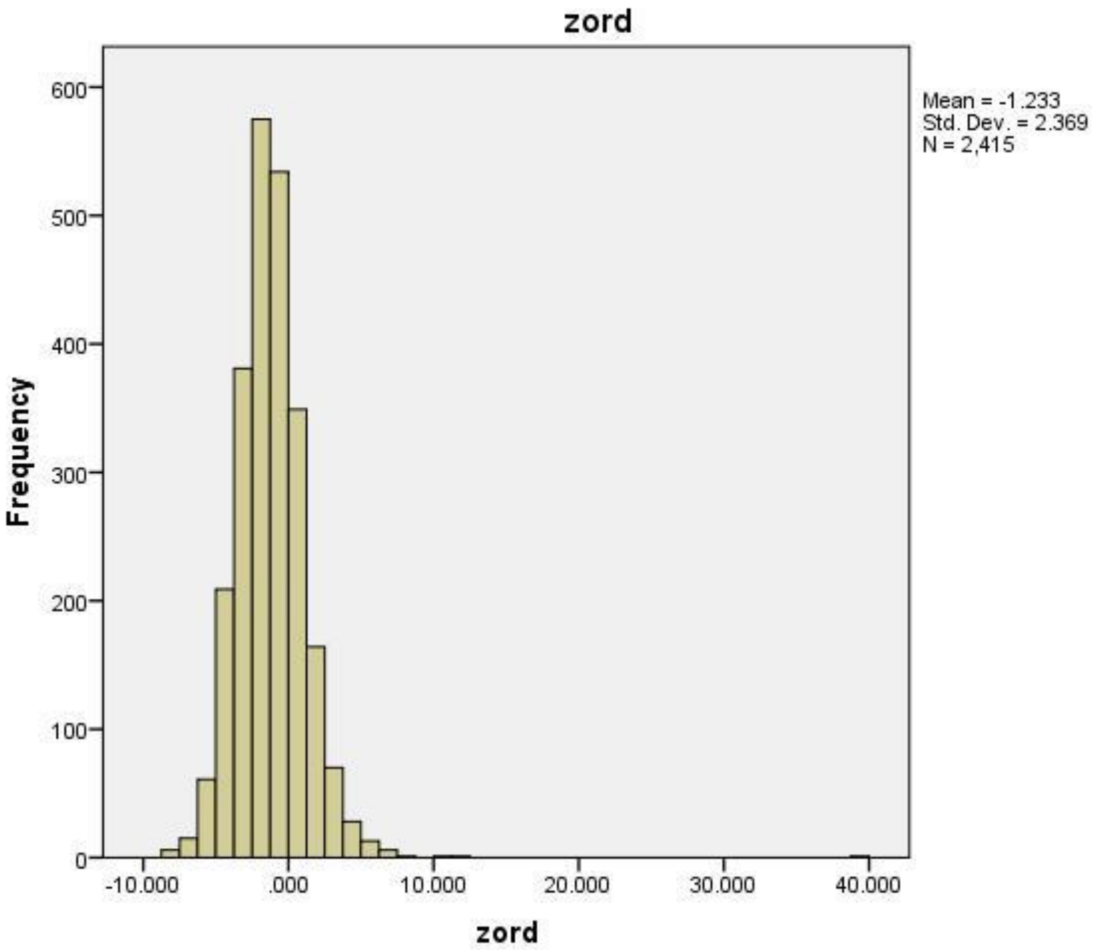


**Figure 9**. Scree plot for the 55 MC and 3 CR items of the English Literature Exam.

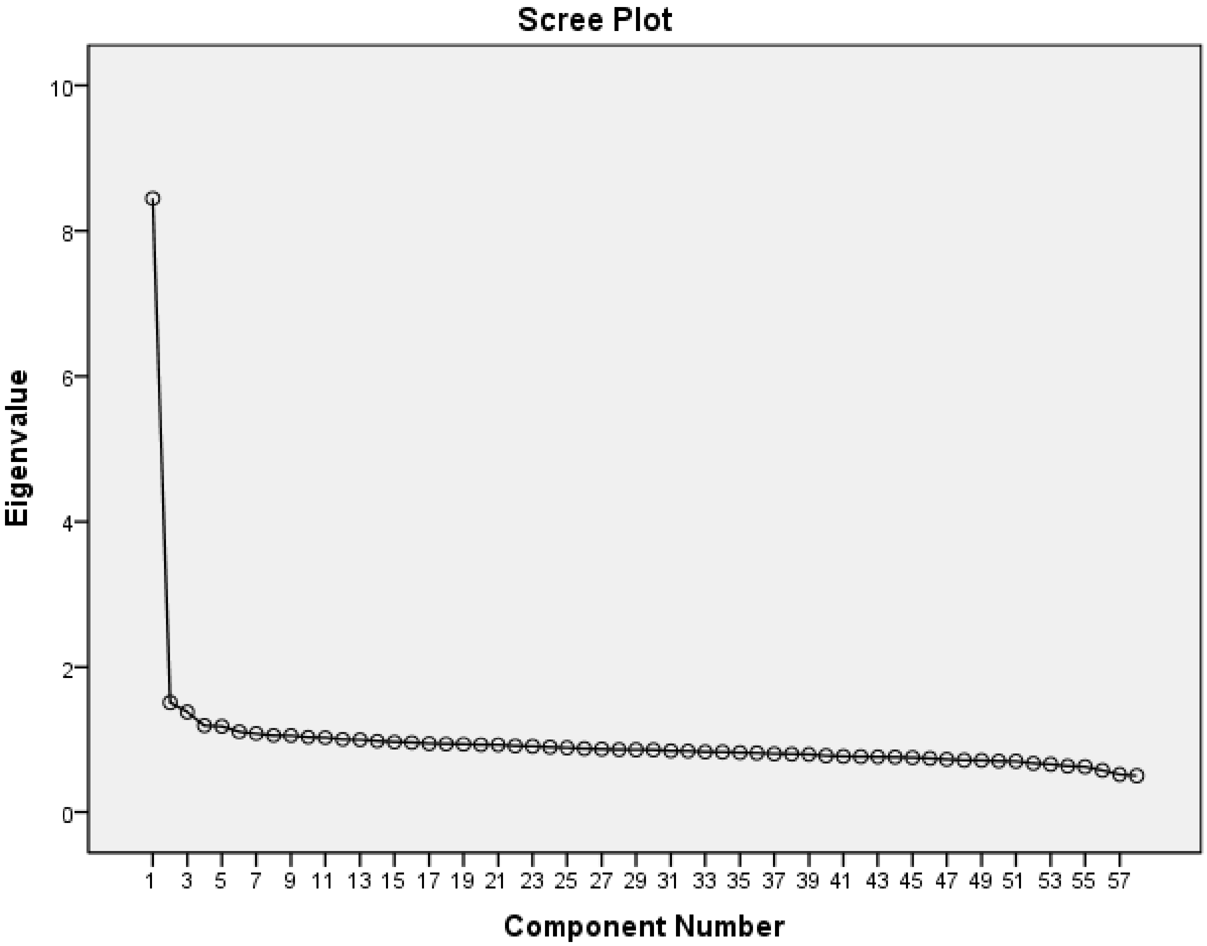


**Figure 10**. Histogram of the $\bar{X}_{ij}^2$ statistics for the 55 MC items of the English Literature Exam.

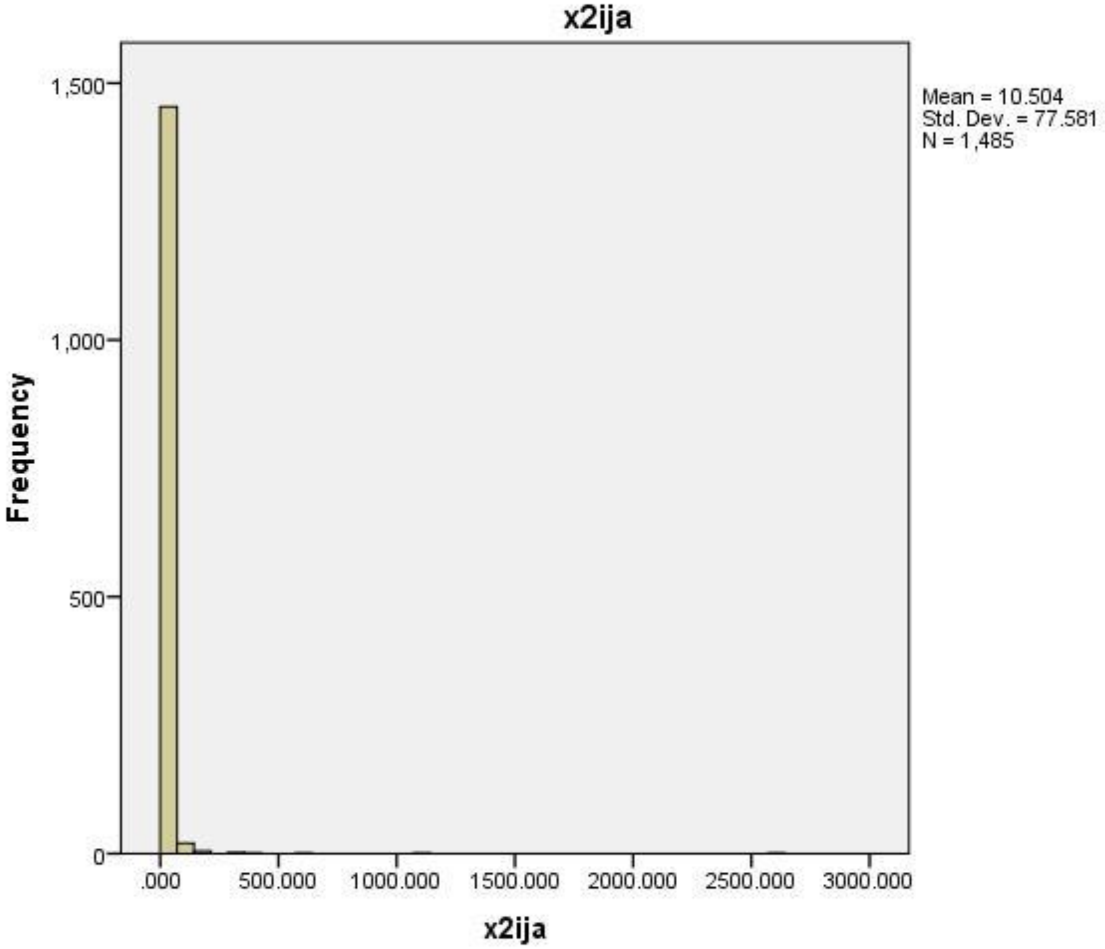


**Figure 11**. Histogram of the $R_{ij}$ statistics for the 55 MC items of the English Literature Exam.

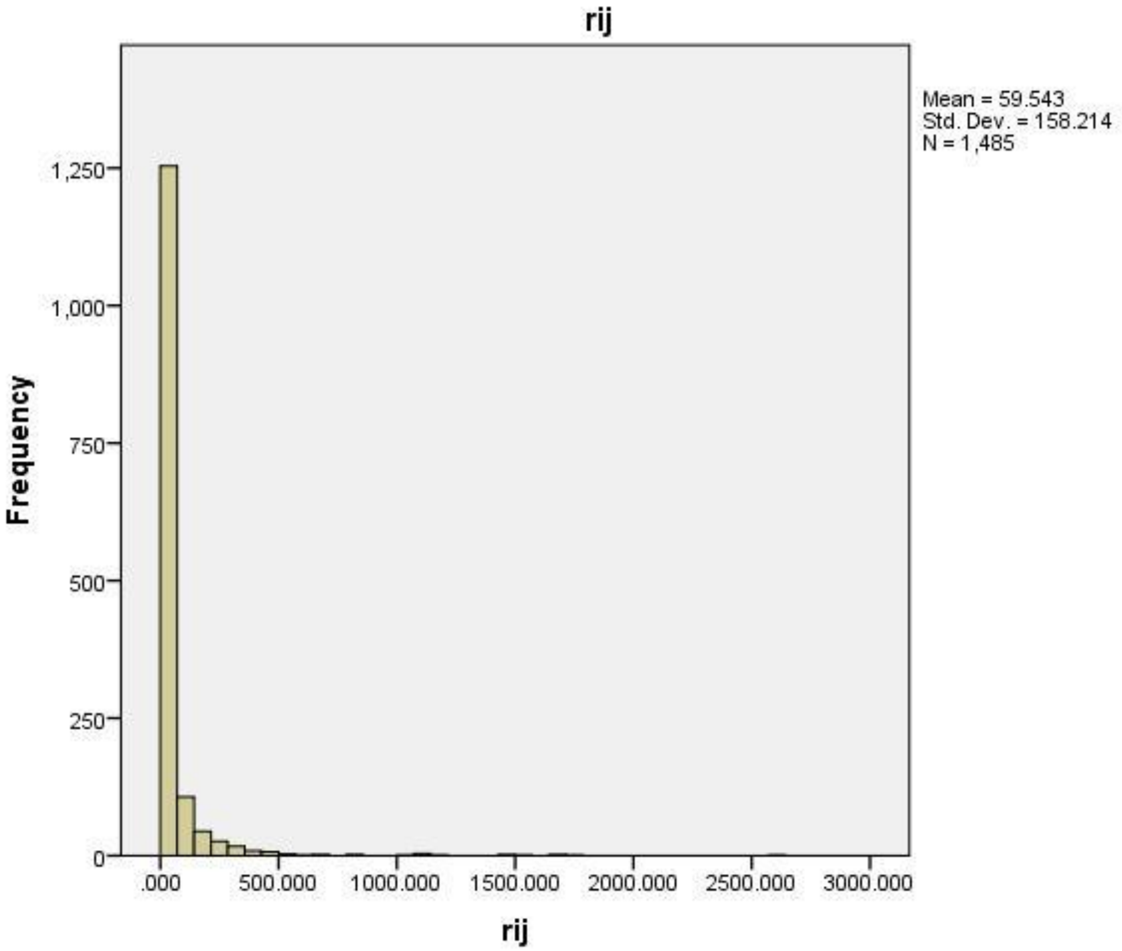


**Figure 12**. Histogram of the $z_{ij}$ statistics for the 55 MC items of the English Literature Exam.

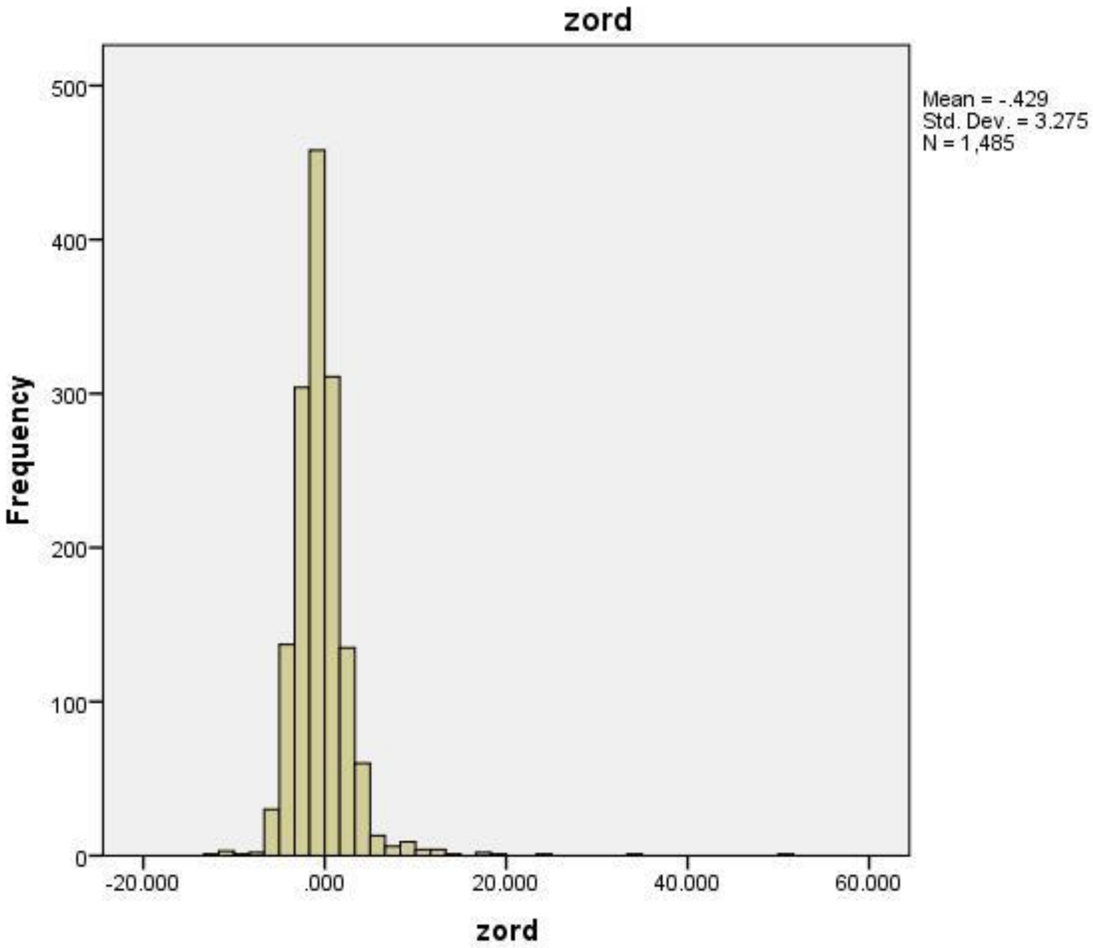
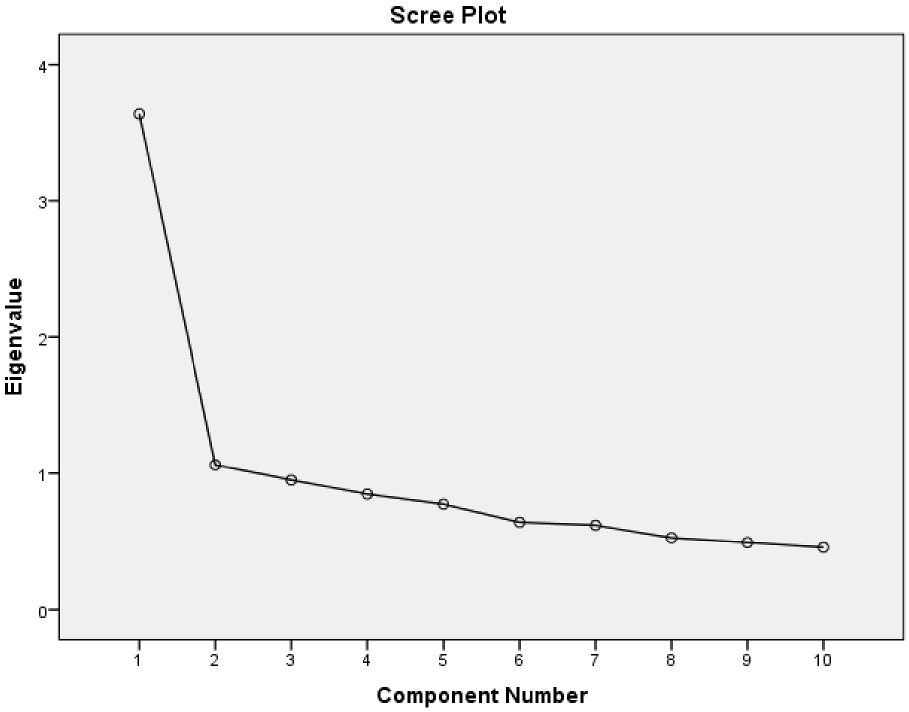


**Figure 13**. Scree plots for the 10 IPIP Agreeableness items in both honest (upper panel) and faking (lower panel) conditions.



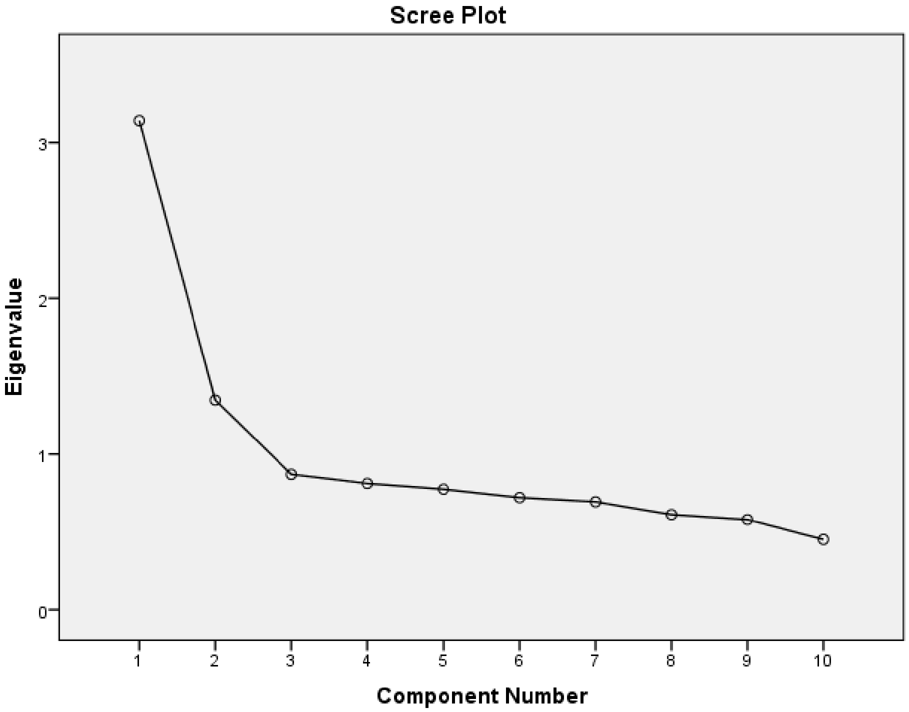


**Figure 14**. Scree plots for the 10 IPIP Conscientiousness items in both honest (upper panel) and faking (lower panel) conditions.

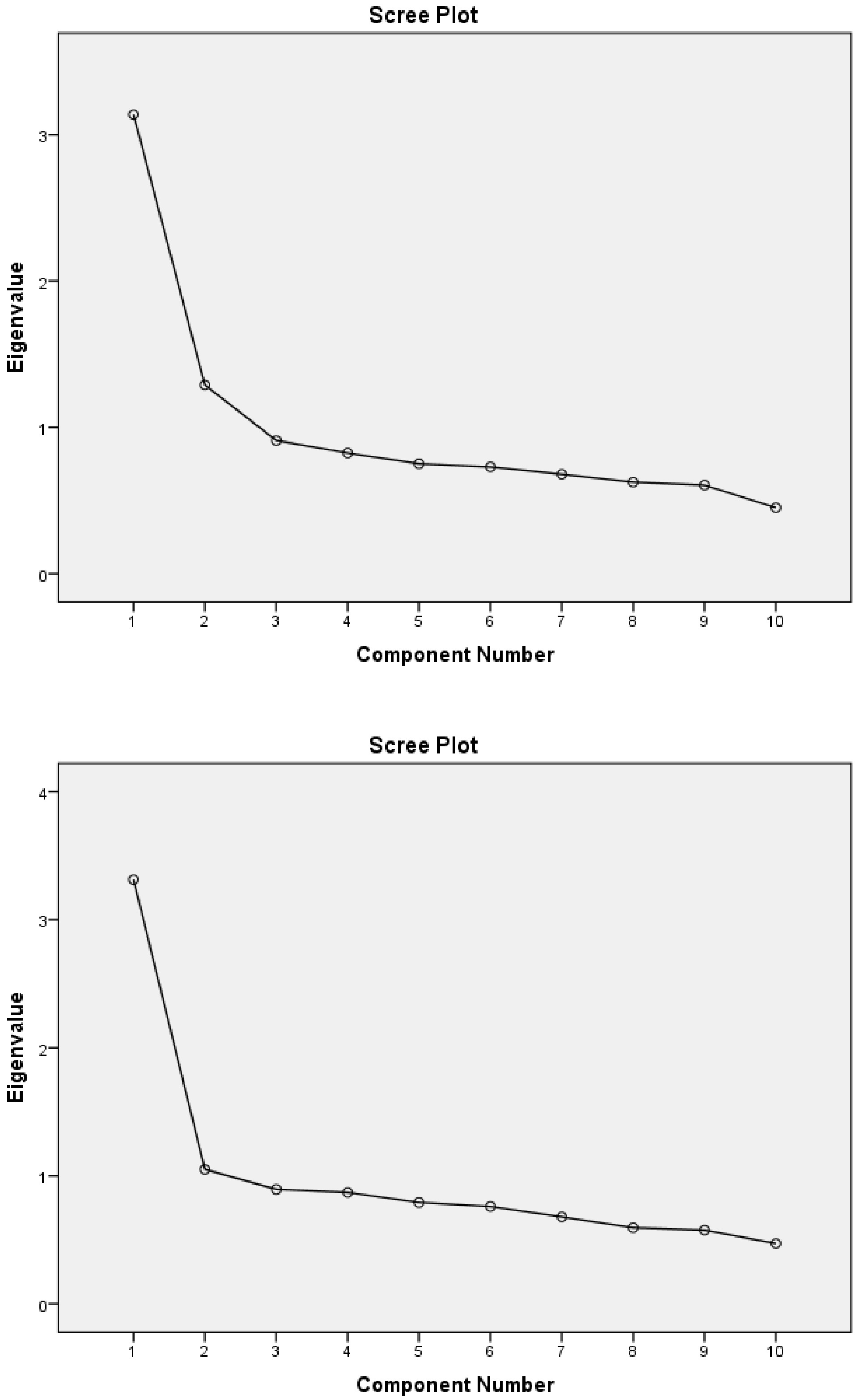
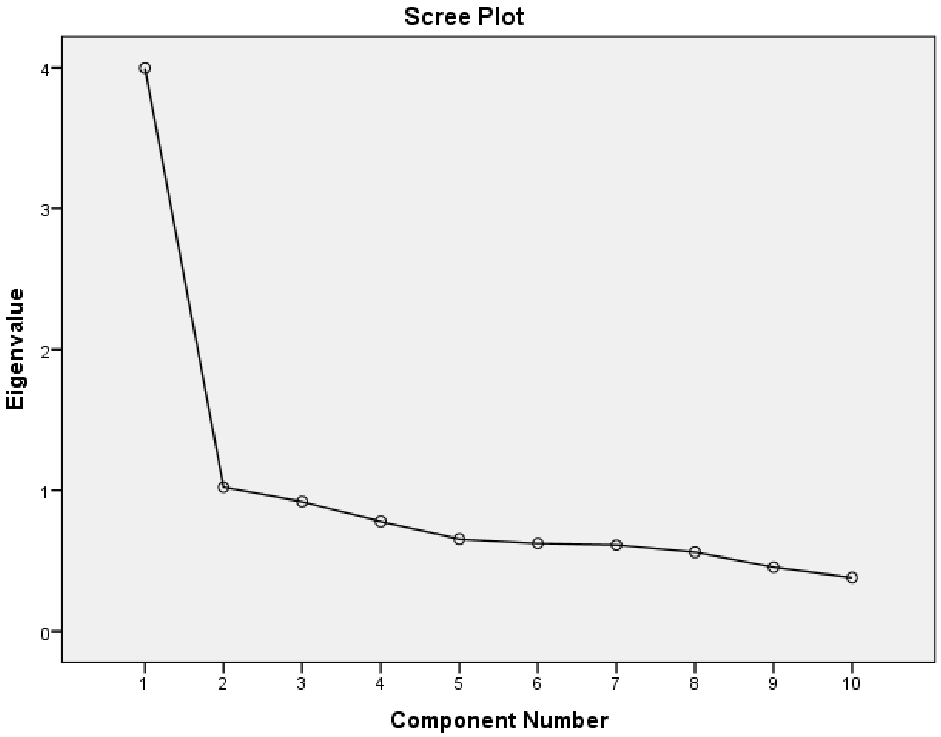


**Figure 15**. Scree plots for the 10 IPIP Extraversion items in both honest (upper panel) and faking (lower panel) conditions.



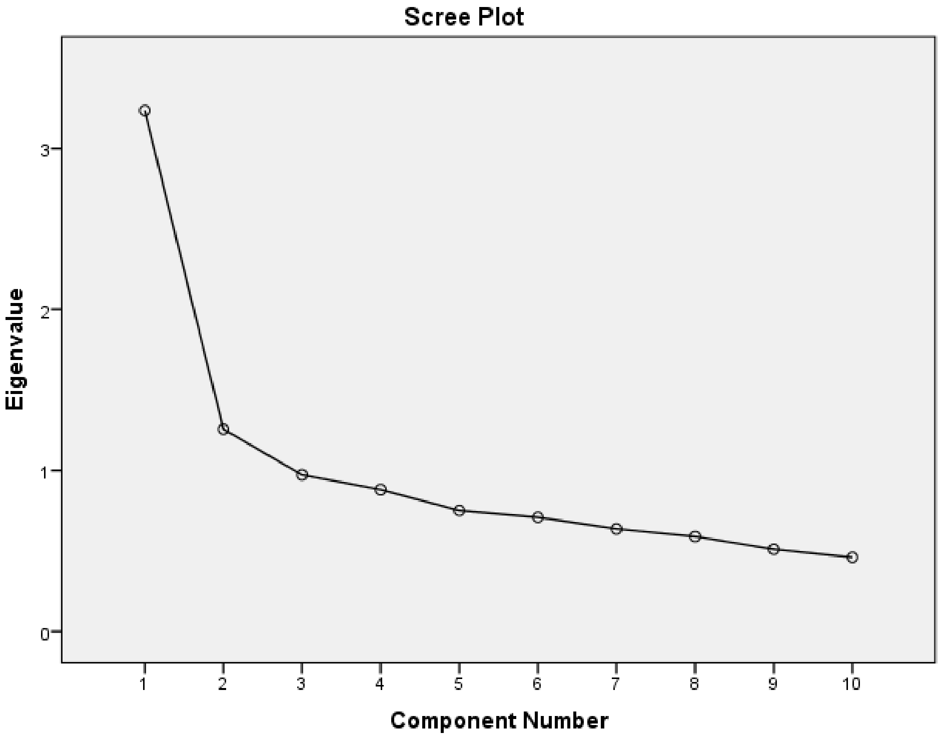


**Figure 16**. Scree plots for the 10 IPIP Emotional Stability items in both honest (upper panel) and faking (lower panel) conditions.

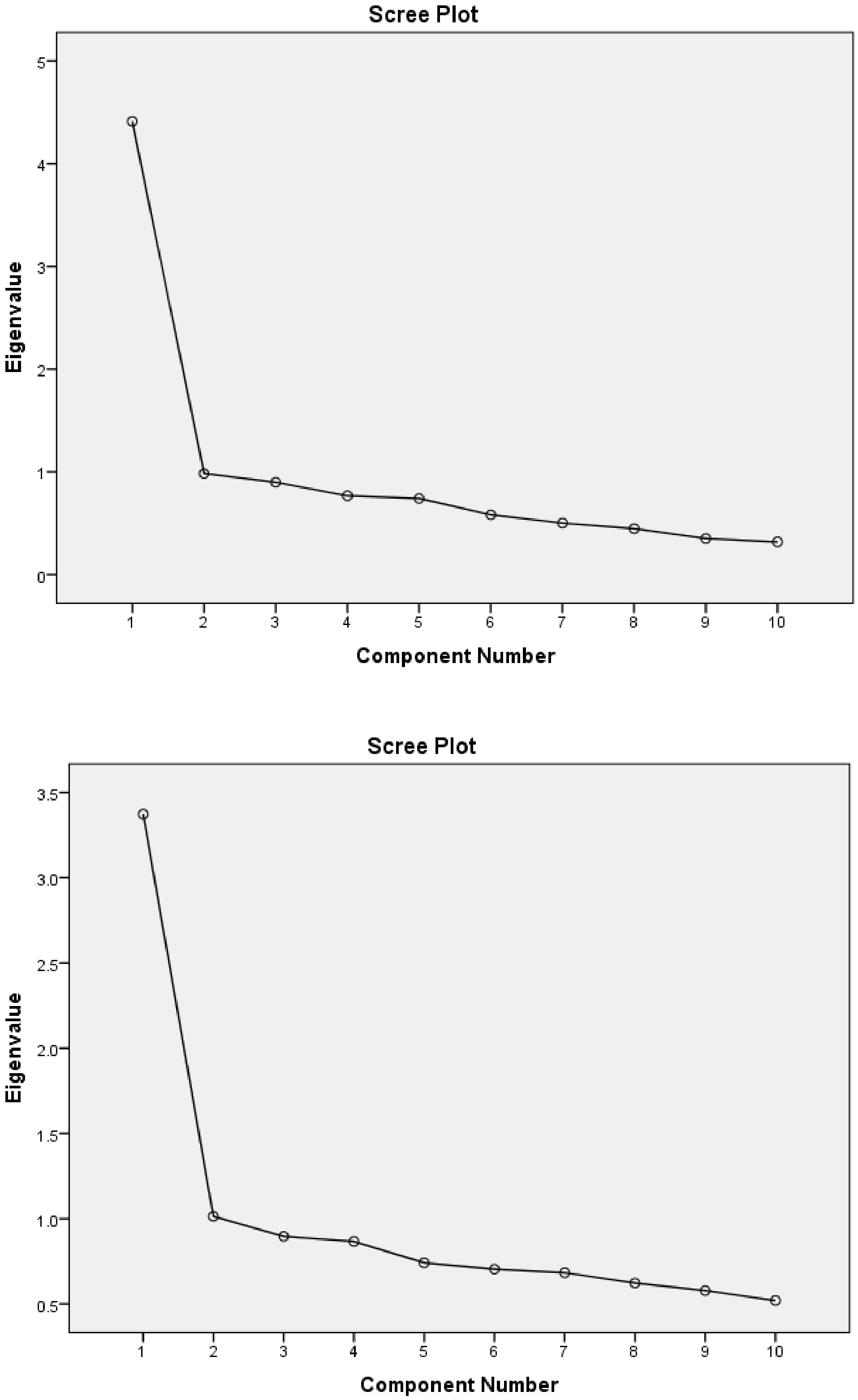


**Figure 17**. Scree plots for the 10 IPIP Openness items in both honest (upper panel) and faking (lower panel) conditions.

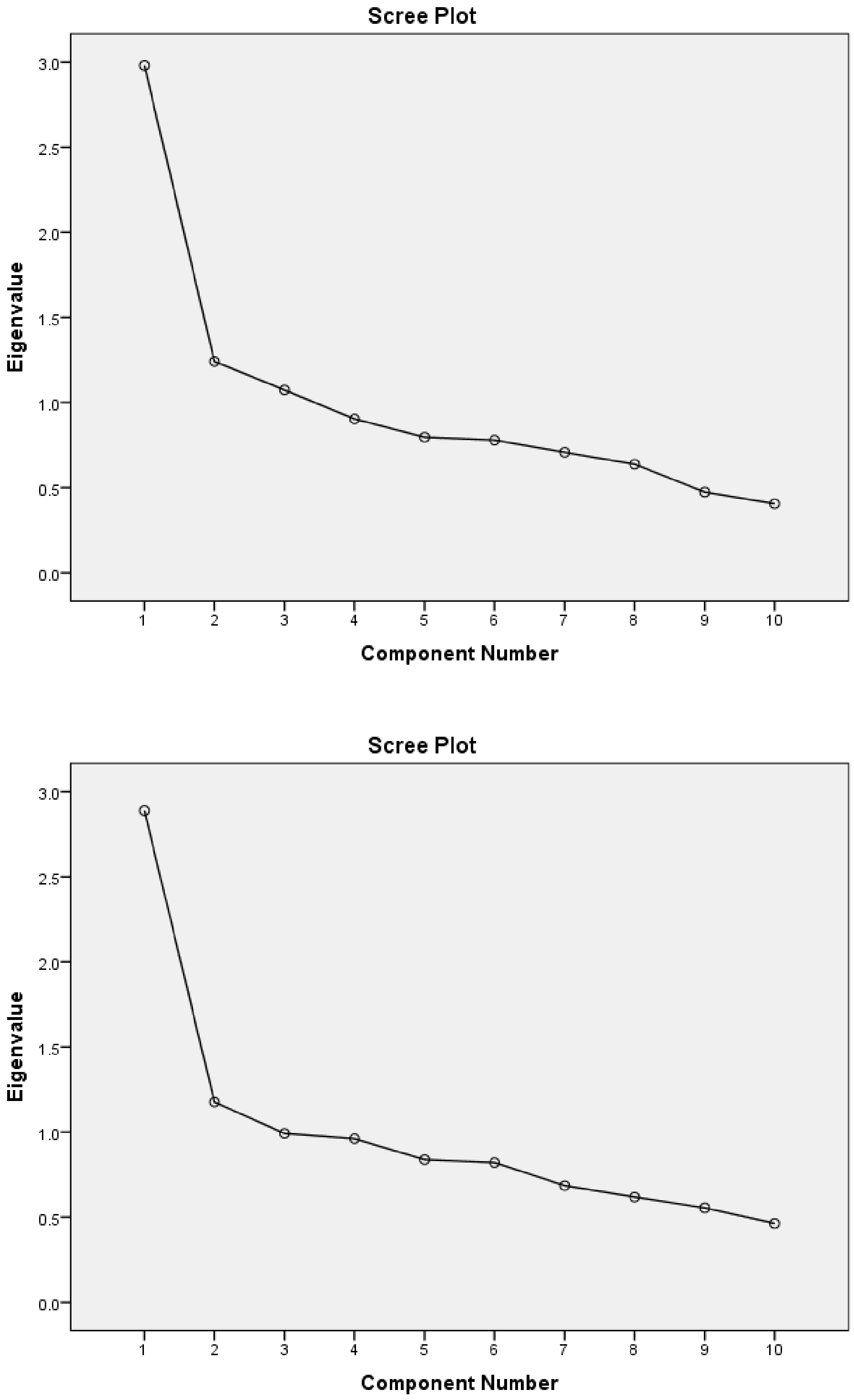
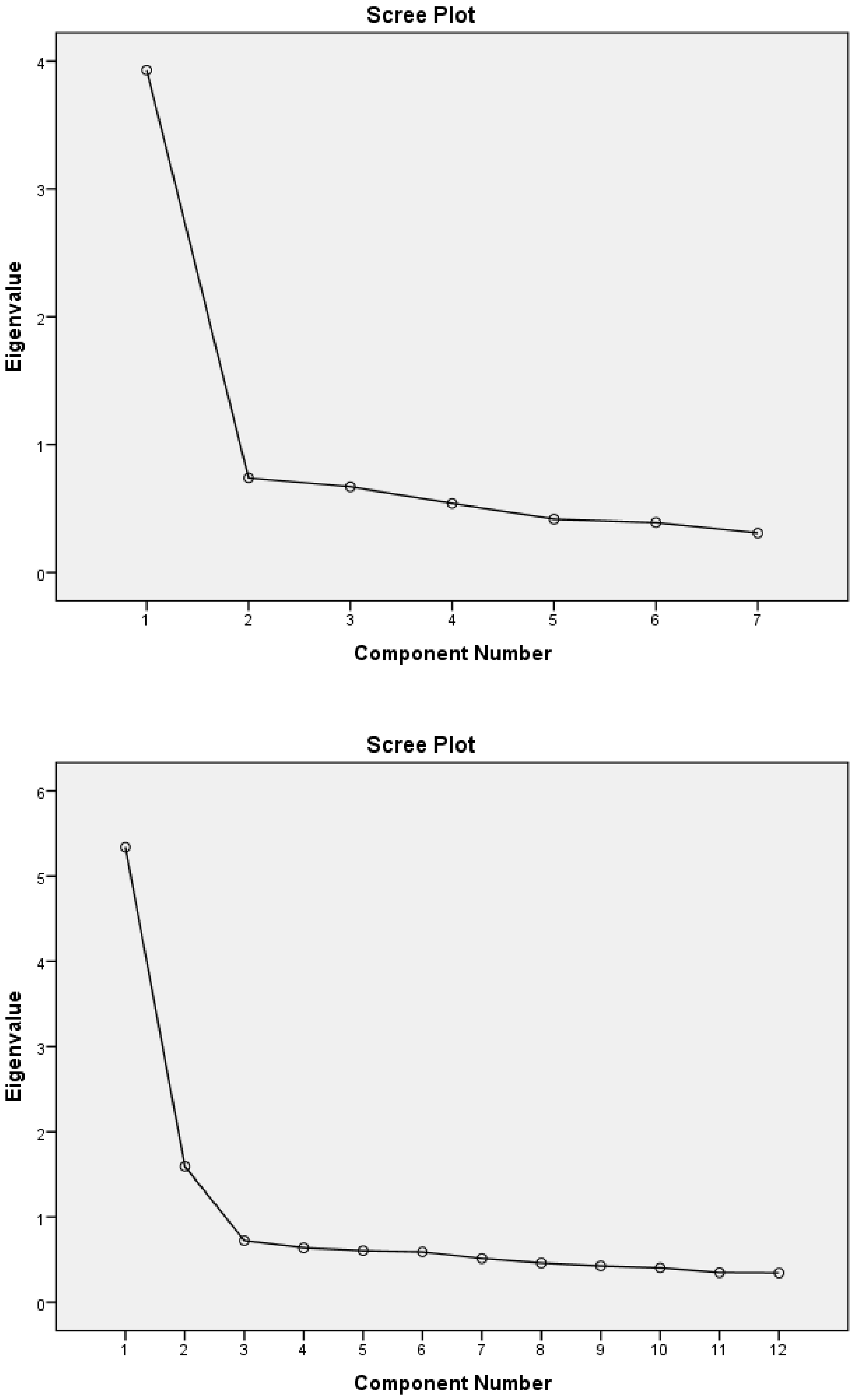


**Figure 18**. Scree plots for the 7 CWB-I (upper panel) and 12 CWB-O (lower panel) items.



Scree Plot

Scree Plot

## APPENDIX A: INTERNATIONAL PERSONALITY ITEM POOL (IPIP) (Goldberg, 1992).

Agreeableness:

1. I feel little concern for others. (R)
2. I am interested in people.
3. I insult people. (R)
4. I sympathize with others' feelings.
5. I am not interested in other people's problems. (R)
6. I have a soft heart.
7. I am not really interested in others. (R)
8. I take time out for others.
9. I feel others' emotions.
10. I make people feel at ease.

Conscientiousness:

1. I am always prepared.
2. I leave my belongings around. (R)
3. I pay attention to details.
4. I make a mess of things. (R)
5. I get chores done right away.
6. I often forget to put things back in their proper place. (R)
7. I like order.
8. I shirk my duties. (R)
9. I follow a schedule.
10. I am exacting in my work.

Extraversion:

1. I am the life of the party.
2. I don't talk a lot. (R)
3. I feel comfortable around people.
4. I keep in the background. (R)
5. I start conversations.
6. I have little to say. (R)
7. I talk to a lot of different people at parties.
8. I don't like to draw attention to myself. (R)
9. I don't mind being the center of attention.
10. I am quiet around strangers. (R)

Emotional Stability:

1. I get stressed out easily. (R)
2. I am relaxed most of the time.
3. I worry about things. (R)
4. I seldom feel blue.
5. I am easily disturbed. (R)
6. I get upset easily. (R)

**APPENDIX A (cont.):**

7. I change my mood a lot. (R)
8. I have frequent mood swings. (R)
9. I get irritated easily. (R)
10. I often feel blue. (R)

Openness:

1. I have a rich vocabulary.
2. I have difficulty understanding abstract ideas. (R)
3. I have a vivid imagination.
4. I am not interested in abstract ideas. (R)
5. I have excellent ideas.
6. I do not have a good imagination. (R)
7. I am quick to understand things.
8. I use difficult words.
9. I spend time reflecting on things.
10. I am full of ideas.

**APPENDIX B: COUNTER-PRODUCTIVE WORK BEHAVIOR (CWB) SCALE** (Bennett & Robinson, 2000).

*CWB-I (Interpersonal Deviance) Scale*

1. Made fun of someone at work

2. Said something hurtful to someone at work

3. Made an ethnic, religious, or racial remark at work

4. Cursed at someone at work

5. Played a mean prank on someone at work

6. Acted rudely toward someone at work

7. Publicly embarrassed someone at work

*CWB-O (Organizational Deviance) Scale*

1. Taken property from work without permission

2. Spent too much time fantasizing or daydreaming instead of working

3. Falsified a receipt to get reimbursed for more money than you spent on business expenses

4. Taken an additional or longer break than is acceptable at your workplace

5. Come in late to work without permission

6. Littered your work environment

7. Neglected to follow your boss's instructions

8. Intentionally worked slower than you could have worked

9. Discussed confidential company information with an unauthorized person

10. Used an illegal drug or consumed alcohol on the job

11. Put little effort into your work

12. Dragged out work in order to get overtime