

Tilburg University

## Selection of Alzheimer symptom items with manifest monotonicity and manifest invariant item ordering

Ligtvoet, R.; van der Ark, L.A.; Sijtsma, K.

*Published in:*  
New Trends in Psychometrics

*Publication date:*  
2008

[Link to publication in Tilburg University Research Portal](#)

### *Citation for published version (APA):*

Ligtvoet, R., van der Ark, L. A., & Sijtsma, K. (2008). Selection of Alzheimer symptom items with manifest monotonicity and manifest invariant item ordering. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 235-532). Universal Academy Press.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

---

## Selection of Alzheimer Symptom Items with Manifest Monotonicity and Manifest Invariant Item Ordering

---

Rudy Ligtoet, L. Andries van der Ark, and Klaas Sijtsma

Department of Methodology and Statistics, Tilburg University, P.O. Box 90153,  
5000 LE Tilburg, The Netherlands.

---

### Abstract

A procedure is proposed for selecting items from a test for which the assumptions of both manifest monotonicity and manifest invariant item ordering hold. The use of the procedure is illustrated by means of an application to data from an Alzheimer disease assessment.

### 1. Introduction

Nonparametric item response theory (IRT) is based on a minimal set of assumptions necessary to obtain useful measurement properties. A regularly used assumption is local independence (LI) of the item scores given the latent variables underlying these item scores. Often, IRT models assume only one latent variable, an assumption that agrees with the practical requirement that the test measures only one trait or ability. This assumption is known as unidimensionality (UD). Other assumptions concern the relationships between the items and this latent variable. One assumption is latent monotonicity (M; Junker, 1993) and another is latent invariant item ordering (IIO; Sijtsma & Junker, 1996). Latent M means that the higher the score on the latent variable, the higher the expected item score. IRT models that assume LI, UD, and latent M allow ordinal measurement of persons (Sijtsma & Molenaar, 2002). A latent IIO means that the ordering of the items in a test according to their attractiveness is the same for all levels of the latent variable. Sijtsma & Junker (1996) discuss several practical testing situations in which latent IIO is important, such as intelligence testing and person-fit analysis.

This study provides the outline of a bottom-up procedure which assists the researcher in selecting from a larger set of items a subset of items for which both latent M and latent IIO hold. The proposed procedure is illustrated by an application to data from Alzheimer disease assessment at the Southern Illinois University School of Medicine\*.

### 2. Theory

We assume that LI and UD hold for the test under consideration. For  $I$  items indexed  $i$  ( $i = 1, \dots, I$ ; indices  $j$  and  $k$  are also used), let  $X_i$  denote the item score, and let  $X_i$  have realizations  $x_i \in \{0, \dots, m_i\}$ . For dichotomous scoring,  $m_i = 1$ , and for polytomous scoring  $m_i \geq 2$ . Let  $\theta$  denote the latent variable. The conditional expected value  $E(X_i|\theta)$  is known as the item response function (IRF).

*Latent and manifest monotonicity.* The assumption of latent M means that the IRFs are nondecreasing in  $\theta$  (i.e., no strict increasingness is required); that is,

$$E(X_i|\theta) \text{ is nondecreasing in } \theta.$$

For dichotomously scored items for which the assumptions of LI, UD, and latent M hold, IRFs may be estimated from the test data as follows. Let  $Y$  be a conveniently

---

\*The authors would like to thank Larry Hughes for providing the data.

chosen ordinal estimator of latent variable  $\theta$ , and let us consider the IRF of item  $i$ . We define the total score on  $I - 1$  items in the test excluding item  $i$  for which we seek to estimate the IRF, as  $Y = R_{(i)} = \sum_{j \neq i} X_j$ . It has been shown (Junker, 1993) that LI, UD, and latent M together imply that

$$E(X_i | R_{(i)} = r) \text{ is nondecreasing in } r, r = 0, \dots, I - 1.$$

This observable property, known as manifest M, can be estimated from the test data for each item. Manifest M does not imply latent M; this means that manifest M is a necessary condition for latent M. Thus, if manifest M holds in the data, this provides support for latent M but no proof, whereas deviations from manifest M are in conflict with latent M. Because the proof of manifest M does not use the number of items  $I$ , manifest M also holds for a two-item test containing items  $i$  and  $k$ , for which  $Y = R_{(i)} = X_k$ ; that is,  $E(X_i | X_k = 0) \leq E(X_i | X_k = 1)$ . A violation of manifest M occurs when we find that  $E(X_i | X_k = 0) > E(X_i | X_k = 1)$ .

Unfortunately, for polytomously scored items for which LI, UD, and latent M hold it has been shown that latent M does not imply manifest M (B. T. Hemker, in Junker & Sijtsma, 2000). This means that a sequence of expected values  $E(X_i | R_{(i)} = r)$ , that is nondecreasing in  $r$ , need not support latent M, and a sequence that is not monotone need not be in conflict with latent M (this conclusion also holds when  $R_{(i)} = X_k$ ). In practical data analysis, however, it seems reasonable to assume that little harm is done when researchers use such sequences heuristically for assessing latent M (Sijtsma & Meijer, 2007).

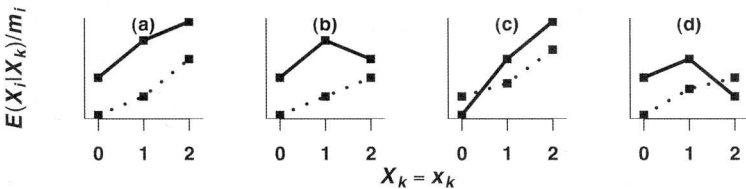
In this study, for polytomous items we use this heuristic strategy for expected item score  $X_i$  conditioned on only one item  $X_k$ , and define manifest M as

$$E(X_i | X_k = x_k) \text{ is nondecreasing in } x_k, x_k = 0, \dots, m_k. \tag{1}$$

A violation of manifest M occurs each time we find for two item scores,  $0 \leq x_{k,a} < x_{k,b} \leq m_k$ , that

$$E(X_i | X_k = x_{k,a}) > E(X_i | X_k = x_{k,b}). \tag{2}$$

Examples of violations are found in Figure 1b (solid curve; one violation) and Figure 1d (solid curve; two violations). Figure 1a shows two monotone curves.



**Fig. 1** Graphs of  $E(X_i | X_k) / m_i$  values with (a) no violations, (b) violation of manifest M, (c) violation of manifest IIO, and (d) violations of both manifest M and IIO.

*Latent and manifest invariant item ordering.* Latent IIO has been defined for polytomously scored items (Sijtsma & Hemker, 1998) with  $m + 1$  answer categories, and dichotomous scoring as a special case. This definition can be generalized to items from the same test having different numbers of answer categories, by considering the conditional expectation of item  $i$  adjusted for the number of answer categories,  $E(X_i | \theta) / m_i$ . Let the attractiveness of item  $i$  be defined as the unconditional expectation,  $E(X_i) / m_i$ , and let the items be numbered such that  $i < j$

means that item  $i$  is less attractive than item  $j$ ,  $E(X_i)/m_i < E(X_j)/m_j$ . A set of  $I$  items has latent IIO if

$$E(X_i|\theta)/m_i \leq E(X_j|\theta)/m_j, \text{ for all } i < j, \text{ and all } \theta. \quad (3)$$

Equation 3 allows possible ties. Van der Ark and Bergsma (2006) showed that for conditioning variable  $Y$ , item score  $X_i$ , and item score  $X_j$ , all three independent of one another conditional on  $\theta$ , latent IIO implies manifest IIO; that is

$$E(X_i|Y = y)/m_i \leq E(X_j|Y = y)/m_j, \text{ for all } i < j, \text{ and all } y, \quad (4)$$

again allowing possible ties (proof in Appendix A). Manifest IIO does not imply latent IIO; thus, manifest IIO is a necessary condition for latent IIO. Analogous to manifest M, one may replace  $Y$  by a single item score, such that  $Y = X_k$  and, as a result, Equation 4 is defined for a triplet of items; that is,

$$E(X_i|X_k = x_k)/m_i \leq E(X_j|X_k = x_k)/m_j, \text{ all } i < j, \text{ and } x_k = 0, \dots, m_k. \quad (5)$$

In this triplet, each of the three items may play the role of conditioning variable  $Y$ . Thus, for a triplet of items manifest IIO in Equation 5 needs to be evaluated three times, conditioned once on each of the three items  $i$ ,  $j$ , and  $k$ . A violation of manifest IIO occurs each time we find for at least one value  $x_k$ , that

$$E(X_i|X_k = x_k)/m_i > E(X_j|X_k = x_k)/m_j. \quad (6)$$

Examples are found in Figure 1c (one violation at  $x_k = 0$ ) and Figure 1d (one violation at  $x_k = 2$ ).

*Combination of monotonicity and invariant item ordering.* For dichotomously scored items, latent M and latent IIO can be combined into one set of inequalities (Proposition 2.1 in Sijtsma & Junker, 1996). First, define the total score on  $I - 2$  items excluding the items  $i$  and  $j$  for which we seek to establish non-intersection of the IRFs, such that  $Y = R_{(i,j)} = \sum_{k \neq i,j} X_k$ . By assuming that LI, UD, and latent M hold, and assuming that also latent IIO holds for items  $i$  and  $j$ , for two arbitrarily chosen values  $\theta_a$  and  $\theta_b$ , we have that if

$$\theta_a < \theta_b \implies E(X_i|\theta_a)/m_i \leq E(X_j|\theta_b)/m_j, \quad (7)$$

it follows that, for  $Y = R_{(i,j)}$ ,

$$r_{(i,j)a} < r_{(i,j)b} \implies E(X_i|R_{(i,j)} = r_{(i,j)a})/m_i \leq E(X_j|R_{(i,j)} = r_{(i,j)b})/m_j. \quad (8)$$

Equation 7 may be called latent M&IIO, and Equation 8 may be called manifest M&IIO. For finite test length  $I$ , manifest M&IIO does not imply latent M&IIO. Manifest M&IIO is a necessary condition for latent M&IIO and, logically, if manifest M&IIO holds for the data this supports but does not prove latent M&IIO, whereas failure of manifest M&IIO disproves latent M&IIO. The proof that Equation 7 implies Equation 8 does not depend on the number of items  $I$ ; thus, it also holds for  $I = 3$  and  $Y = R_{(i,j)} = X_k$ .

Because for polytomous scoring latent M does not imply manifest M, the implication in Equation 7 and Equation 8 does not straightforwardly generalize to polytomous items. Here we propose to use as a heuristic for investigating latent M&IIO (Equation 7) in real data the following *manifest M&IIO* property. Let the items  $i$ ,  $j$ , and  $k$  be polytomously scored and let the number of ordered scores be

variable across the items. Then, for two arbitrarily chosen item scores for which  $0 \leq x_{k,a} < x_{k,b} \leq m_k$ , we propose to use the heuristic

$$E(X_i|X_k = x_{k,a})/m_i \leq E(X_j|X_k = x_{k,b})/m_j. \quad (9)$$

A violation of manifest M&IIO occurs each time we find for at least one pair  $x_{k,a} < x_{k,b}$ , that

$$E(X_i|X_k = x_{k,a})/m_i > E(X_j|X_k = x_{k,b})/m_j. \quad (10)$$

Figure 1b shows a violation of manifest M, which does not lead to a reversal of expectations as in Equation 10, Figure 1c shows a violation of manifest IIO, which again is not picked up, and Figure 1d shows a violation of both manifest M and manifest IIO, which is reflected by a reversal of expected values as in Equation 10. Thus, Equation 9 may not be a powerful tool for investigating violations of manifest M and manifest IIO.

*Estimation of expected values.* The  $E(X_i|X_k)/m_i$  values for assessing the inequalities in the Equations 1, 5, and 9 can be estimated from the data as follows. Let  $N_{x_k}$  denote the number of respondents out of a sample of size  $N$  who have a score equal to  $x_k$  on item  $k$ , and let  $N_{x_i x_k}$  denote the number of respondents who have a score  $x_i$  on item  $i$  and a score  $x_k$  on item  $k$ ; then

$$\hat{E}(X_i|X_k = x_k)/m_i = \frac{1}{m_i} \sum_{x_i} x_i \frac{N_{x_i x_k}}{N_{x_k}}.$$

### 3. A Bottom-Up Item Selection Procedure

The goal of this study is to suggest a procedure for finding subsets of items in a larger set, for which manifest M and manifest IIO are satisfied. The first step of the procedure is finding all triplets consisting of three different items that satisfy manifest M and manifest IIO. The second step of the procedure entails combining the triplets found in the first step into quartets for which manifest M and manifest IIO hold. The third step entails combining the quartets found in the second step into 5-tuples for which manifest M and manifest IIO hold. The procedure ends when the largest  $n$ -tuple is found for which manifest M and manifest IIO holds.

The first step of the procedure can be executed using two different methods for investigating manifest M and manifest IIO. The first method (*Method I*) investigates both manifest M (Equation 1) and manifest IIO (Equation 5). The second method (*Method II*) investigates manifest M&IIO (Equation 9). Method I and Method II are discussed next.

#### 3.1. Technical details of first step

*Method I.* For the triplet of items,  $i$ ,  $j$ , and  $k$ , the procedure is as follows. Arbitrarily, let item  $k$  be the conditioning variable in Equation 1 and Equation 5. All violations of manifest M (Equation 2) for item  $i$  and item  $j$  are tested using a one-sided independent  $t$ -test: The null hypothesis  $E(X_i|X_k = x_{k,a}) \leq E(X_i|X_k = x_{k,b})$  is tested against the alternative that  $E(X_i|X_k = x_{k,a}) > E(X_i|X_k = x_{k,b})$  (Equation 2). Rejection of the null hypothesis means that a violation of manifest M is found.

All violations of manifest IIO (Equation 6) are tested using a one-sided dependent  $t$ -test. For  $E(X_i)/m_i < E(X_j)/m_j$ , the null hypothesis  $E(X_i|X_k = x_k)/m_i \leq E(X_j|X_k = x_k)/m_j$  is tested against the alternative that  $E(X_i|X_k = x_k)/m_i > E(X_j|X_k = x_k)/m_j$  (Equation 6). Rejection of the null hypothesis means that

a violation of manifest IIO is found. It may be noted that testing may be problematic when  $X_i$  en  $X_j$  have not been measured on the same scale, but given the heuristic nature of this research this is ignored here.

Hypothesis testing is done with item  $i$ ,  $j$ , and  $k$  consecutively playing the role of conditioning variable while the expected values of the other two items are evaluated. If no violations are found, this result supports latent M and latent IIO for the item triplet. If at least one violation is found, it is concluded that latent M and latent IIO are not valid for the triplet. Because each triplet of items is investigated three times, and because the number of triplets is large for realistic test length  $I$ , it is reasonable to expect large numbers of sample violations of manifest M and manifest IIO. Method I picks up all violations and allows the researcher to adjust the level of significance so that the violations considered to be important can be assessed for item selection with an eye to optimal decision-making.

Molenaar & Sijtsma (2000) noted that practical data analysis often yields large numbers of violations of manifest M and manifest IIO but argue that many of the relatively small violations are not damaging for the measurement of persons on an ordinal scale. They suggested ignoring violations smaller than a value  $minvi$  for statistical significance testing. The computer program MSP (Molenaar & Sijtsma, 2000) uses default option  $minvi = .03$ . If a large power of the statistical test is considered to be undesirable, the user could choose larger values of  $minvi$ . For the violations that remain after selection by  $minvi$ , the nominal Type I error rate of the statistical test may be adapted. The requirement not to reject items too easily is accomplished using large  $minvi$  and small Type I error rate.

Another possibility arises in applications in which having the best item subset possible has priority over reliable person ordering using large numbers of items. For example, when the importance of individual decision-making is paramount, as in medical diagnosis, items are selected that show no more than minor violations of manifest M and manifest IIO. This is accomplished using small  $minvi$  and large Type I error rate. In the research to be reported shortly both  $minvi$  and the nominal Type I error rate are manipulated.

*Method II.* Method II tests Equation 9, which is true when both manifest M and manifest IIO hold, using a dependent  $t$ -test. Again,  $minvi = .03$  may be used for ignoring small sample violations, and statistical testing may be used for the remaining violations. However, the method may well overlook serious violations of manifest M and manifest IIO in the data and, as a result, it is expected to have less power than Method I.

### 3.2. Technical details of the next steps

The next steps combine item triplets into larger items sets without further statistical testing. In the second step, item quartets are identified of which all  $\binom{4}{3} = 4$  constituent item triplets were found in the first step by means of either Method I or Method II. In the third step, item 5-tuples are identified of which all  $\binom{5}{4} = 5$  constituent item quartets were found in the second step; and so on, until no larger sets can be identified. The end result of this procedure may consist of several item subsets that overlap, and that may contain different numbers of items. It is up to the researcher to interpret this result with respect to his/her research question.

## 4. Simulation Study

A study was done to investigate the effects of different values of  $minvi$  and the significance level  $\alpha$  on the number of item triplets identified using either Method I or Method II, and on the end result of item selection.

#### 4.1. Method

*Data.* The data were sampled from a real-data set consisting of the scores of 200 persons on the eleven items from the Mini-Mental State exam (MMS; Folstein, Folstein, & McHugh, 1975). The number of score categories varied across items (see Appendix B). These data were collected during an Alzheimer disease assessment at the Southern Illinois University School of Medicine between 1994 and 2000 (Hughes, Perkins, Wright, & Westrick, 2003). The items assess several cognitive functions such as orientation, registration, and attention (see Appendix B for item labels). The assumptions UD and LI were checked on the dichotomized item scores using the DETECT index (Zhang & Stout, 1999) and the scalability coefficient  $H$  (Mokken, 1971). The resulting values of DETECT and  $H$  suggested that the assumptions of UD and LI held for these data.

*Independent variables.* Four independent variables were used in this study:

1. Method for investigating manifest M and manifest IIO: Method I and Method II were used.
2. Minimum violation to be considered for statistical testing: Two values were investigated,  $minvi = 0$ , which implies that all observed violations are tested for significance; and  $minvi = .03$ , which is the MSP default value.
3. Nominal Type I error rate:  $\alpha = .05$  is the MSP default, and  $\alpha = .10$ , which leads to a more frequent rejection of the null hypothesis.
4. Sample size: A relatively small ( $N = 200$ ) sample and a relatively large ( $N = 500$ ) sample were drawn with replacement from the real data.

*Dependent variables.* Two dependent variables were used in this study:

1. The proportion of item triplets for which manifest M and manifest IIO could not be rejected. This proportion was computed as the number of item triplets for which manifest M and manifest IIO could not be rejected divided by  $\binom{11}{3} = 165$ , the maximum number of item triplets in this study.
2. The magnitude of the largest item set that resulted from the item selection procedure. The maximum value that can be obtained is 11.

*Design characteristics.* The design had size  $2 \times 2 \times 2 \times 2 = 16$ . In each cell, 10 data sets were generated. Given a particular sample size, a given data matrix was used across all 8 combinations of method,  $minvi$ , and Type I error rate. As a result, sample size is a between-factor, and the other independent variables are within-factors.

#### 4.2. Results

Table 1 shows the proportions and the standard deviations of the number of item triplets in each cell of the design, based on the theoretical maximum of 165 item triplets. Furthermore, Table 1 shows the modal number of items in the largest item set that resulted from the selection procedure. Most item triplets were found using Method II; between 93% and 96% of the theoretical maximum. Here, the smaller standard deviation is due to the proportions being close to 1. This higher number of item triplets leads to larger modal  $n$ -tuples; 8-tuples for  $\alpha = .10$  and 9-tuples for  $\alpha = .05$ . The mean number of item triplets was close to the theoretical maximum; thus, few triplets violated M&IIO. For Method I, nominal Type I error rate and sample size had a relatively small effect. The modal number of items in the largest item set was equal to 6. Nominal Type I error  $\alpha = .05$  yielded higher means than nominal Type I error  $\alpha = .10$ , and  $N = 200$  yielded higher proportions than  $N = 500$ . These effects were smaller for Method II. No effect of factor  $minvi$  was found.

**Table 1** Proportion of Item Triplets (Standard Deviation Between Parentheses) and Modal Number of Items in the Largest Set (bold).

Method	<i>minvi</i>	Alpha	Sample Size					
			200			500		
I	.00	.05	.654	(.080)	<b>6</b>	.623	(.076)	<b>6</b>
		.10	.549	(.073)	<b>6</b>	.530	(.072)	<b>6</b>
	.03	.05	.654	(.080)	<b>6</b>	.624	(.076)	<b>6</b>
		.10	.552	(.075)	<b>6</b>	.532	(.071)	<b>6</b>
II	.00	.05	.957	(.011)	<b>9</b>	.948	(.018)	<b>9</b>
		.10	.939	(.017)	<b>8</b>	.929	(.023)	<b>8</b>
	.03	.05	.957	(.011)	<b>9</b>	.948	(.018)	<b>9</b>
		.10	.939	(.017)	<b>8</b>	.929	(.023)	<b>8</b>

### 4.3. Discussion

Method II identified few violations of latent M and latent IIO, and may have had too little power to be useful here. Lower bound *minvi* did not have any effect. This may be due to small sample size yielding no significance for small violations when *minvi* = 0. For *minvi* = .03, the same small violations were not tested or they were not significant, yielding the same result as found for *minvi* = 0. For larger sample sizes *minvi* is probably more effective for reducing power.

## 5. Real-Data Analysis

### 5.1. Results

Because the MMS is used to identify symptoms of Alzheimer disease, we considered reliably identifying the largest subset of items that is characterized by latent M and latent IIO to be of greatest importance here. Based on the results found in the previous section, we thus used Method I with *minvi* = 0 and  $\alpha = .10$ . These choices produced a conservative item selection, thus avoiding unnecessary risk of selecting items for which latent M and latent IIO did not hold.

The 11 items from the MMS were numbered from least attractive to most attractive (see Appendix B for item numbers). Method I was used to test all 165 item triplets three times for violations of manifest M and manifest IIO, using *minvi* = 0 and  $\alpha = .10$ . This resulted in 107 item triplets. Next, triplets were combined to produce larger item sets fulfilling both manifest M and manifest IIO. The result was one 7-tuple containing the items 1, 2, 5, 6, 8, 9, and 11.

### 5.2. Discussion

The items 1, 2, 5, 6, 8, 9, and 11 can be interpreted as follows with respect to latent M and latent IIO. Given the severity of the symptoms associated with the items, we conclude that the first problems of Alzheimer disease occur with recalling recently learned series (item 11), naming the date (item 9), naming a location (item 8), following a sequence of instructions (item 6), copying a drawing (item 5), registration of words read aloud (item 2), and writing a sentence read aloud (item 1). This ordering seems to agree with the onset of symptoms (memory impairment, disorientation, impaired judgement, and language disturbance) reported by Kang, Jeong, Lee, Baek, Kwon, Chin and Na (2004). Much research with respect to the progression of Alzheimer disease remains to be done (and is beyond our expertise).



## 6. General discussion

A bottom-up item selection procedure was proposed which assists the researcher in selecting items from a larger set in which items satisfy the requirements of latent M and latent IIO. Two methods were used to assess these properties. Method I, which assesses both manifest M (Equation 1) and manifest IIO (Equation 5) is more demanding than Method II, which assesses a weaker version of manifest M and manifest IIO, here denoted *manifest M&IIO* (Equation 9). Any subset of items that is selected using Method I is also selected using Method II but not the other way round. Method II probably will gain power if it is used to assess expected values conditional on, for example,  $Y = R_{(i,j)}$ , as in Equation 8. Restscore  $Y = R_{(i,j)}$  is certainly a more fine-grained ordinal estimator of latent trait  $\theta$  than the coarse estimator provided by single item score  $Y = X_k$  (Equation 9). Thus, restscore  $Y = R_{(i,j)}$  will reveal violations of *manifest M&IIO* more easily than item score  $Y = X_k$ . Use of the restscore requires the number of item scores to be the same across the items; else, scores from different items are incomparable.

For the analysis of the MMS we chose  $minvi = 0$  and  $\alpha = .10$ . This way, we reduced the risk of selecting items for which manifest M and/or manifest IIO do not hold. The resulting item subset has these properties with much certainty; thus, we infer latent M and latent IIO to hold for these items. In other applications, in which a large number of items is needed to accurately order respondents for whom the items have the same ordering, a larger value of  $minvi$  and a smaller value of  $\alpha$  allow more items into the scale but the selection is less stringent and accepts more violations of latent M and latent IIO. Future research should clarify which values of  $minvi$  and  $\alpha$  are acceptable for producing scales that allow for accurate person and item ordering, while rejecting as few items as possible.

This study did not provide a benchmark indicating whether Method I or Method II is more appropriate for identifying violations of latent M and latent IIO. Such a study would require knowledge about the presence of latent M and latent IIO in the population, but this information is unavailable in real-data analysis. A well controlled simulation study to investigate the extent to which the procedure yields correct conclusions concerning latent M and latent IIO, is a topic for future research.

Furthermore, it has been suggested to use scalability coefficient  $H$  to evaluate manifest M (Molenaar, 1991; Mokken, 1971, pp. 148–153) and coefficient  $H^T$  to evaluate manifest IIO (Ligtvoet, Van der Ark, & Sijtsma, 2007; Sijtsma & Meijer, 1992). For the item subset 1, 2, 5, 6, 8, 9, and 11, we found that  $H = .598$ , which indicates a “strong” scale (Mokken, 1971, p. 185). This result supports manifest M. We also found that  $H^T = .747$ , which indicates a strong agreement of individual item-score patterns with the ordering of items in the group. This result supports manifest IIO. These results lend credibility to the results obtained by means of our proposed item selection procedure.

## Appendix

### *Appendix A: Proof that latent IIO implies manifest IIO*

The proof that latent IIO (Equation 3) implies manifest IIO (Equation 4) is based on a proof by Van der Ark and Bergsma (2006). Let  $G(\theta)$  be the distribution function of  $\theta$ .

First,  $E(X_i|Y)/m_i$  (see Equation 4) is rewritten. Standard algebra shows that

$$\begin{aligned} E(X_i|Y)/m_i &= \frac{1}{m_i} \sum_{x_i} x_i P(X_i = x_i|Y) = [m_i \cdot P(Y)]^{-1} \sum_{x_i} x_i P(X_i = x_i, Y) \\ &= [m_i \cdot P(Y)]^{-1} \int_{\theta} \sum_{x_i} x_i P(X_i = x_i, Y, \theta) dG(\theta) \\ &= [m_i \cdot P(Y)]^{-1} \int_{\theta} \sum_{x_i} x_i P(X_i = x_i, Y|\theta) G(\theta) dG(\theta). \end{aligned} \quad (11)$$

Because of LI,  $P(X_i = x_i, Y|\theta)$  in Equation 11 can be further reduced to

$$P(X_i = x_i, Y|\theta) = P(X_i = x_i|\theta)P(Y|\theta). \quad (12)$$

Substituting Equation 12 into Equation 11 gives

$$\begin{aligned} E(X_i|Y)/m_i &= [m_i \cdot P(Y)]^{-1} \int_{\theta} P(Y|\theta) G(\theta) \sum_{x_i} x_i P(X_i = x_i|\theta) dG(\theta) \\ &= \frac{1}{m_i} \int_{\theta} E(X_i|\theta) G(\theta|Y) dG(\theta). \end{aligned} \quad (13)$$

Second, it is shown that a latent IIO implies a manifest IIO. Consider Equation 3. Multiplying both sides of Equation 3 by  $G(\theta|Y)$  and taking the integral over  $G(\theta)$  leaves the inequality unchanged. Hence, Equation 3 implies

$$\frac{1}{m_i} \int_{\theta} E(X_i|\theta) G(\theta|Y) dG(\theta) \leq \frac{1}{m_j} \int_{\theta} E(X_j|\theta) G(\theta|Y) dG(\theta). \quad (14)$$

It follows from Equation 13 that the left-hand side of Equation 14 equals  $E(X_i|Y)/m_i$ , and the right-hand side of Equation 14 equals  $E(X_j|Y)/m_j$ . Hence, Equation 3 implies Equation 4. This completes the proof.

#### Appendix B: The 11 Alzheimer assessment items.

For the 11 Alzheimer assessment items from the Mini-Mental State exam, Appendix B shows the item numbers, their content, and their number of ordered answer categories between parentheses.

- |                     |                        |                |
|---------------------|------------------------|----------------|
| 1. Writing (2)      | 5. Copying (2)         | 9. Date (5)    |
| 2. Registration (2) | 6. 3-Stage Command (3) | 10. Naming (2) |
| 3. Repetition (2)   | 7. Attention (4)       | 11. Recall (4) |
| 4. Reading (2)      | 8. Location (4)        |                |

#### References

- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State" A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189-198.
- Hughes, L. F., Perkins, K., Wright, B. D., & Westrick, H. (2003). Using a Rasch scale to characterize the clinical features of patients with a clinical diagnosis of uncertain, probable, or possible Alzheimer disease at intake. *Journal of Alzheimer's Disease*, 5, 367-373.

- Junker, B. W. (1993). Conditional association, essential independence and monotone unidimensional item response models. *The Annals of Statistics*, *21*, 1359–1378.
- Junker, B. W. & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement*, *24*, 65–81.
- Kang, S. J., Jeong, Y., Lee, B. H., Baek, M. J., Kwon, J. C., Chin, J. & Na, D. L. (2004). How early are initial symptoms recognized in Korean patients with Alzheimer's disease? *International Journal of Geriatric Psychiatry*, *19*, 699–700.
- Ligtvoet, R., Van der Ark, L. A. & Sijtsma, K. (2007). Investigating an invariant item ordering for polytomously scored items. Manuscript submitted.
- Mokken, R. J. (1971). *A Theory and Procedure of Scale Analysis*. The Hague: Mouton.
- Molenaar, I. W. (1991). A weighted loevinger H-coefficient extending Mokken scaling to multicategory items. *Kwantitatieve Methoden*, *12*, 97–117.
- Molenaar, I. W., & Sijtsma, K. (2000). *User's Manual MSP5 for Windows*. Groningen: iecProGAMMA.
- Sijtsma, K., & Hemker, B. T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika*, *63*, 183–200.
- Sijtsma, K., & Junker, B. W. (1996). A survey of theory and methods of invariant item ordering. *British Journal of Mathematical and Statistical Psychology*, *49*, 79–105.
- Sijtsma, K., & Meijer, R. R. (1992). A method for investigating the intersection of item response functions in Mokken's nonparametric IRT model. *Applied Psychological Measurement*, *16*, 149–157.
- Sijtsma, K., & Meijer, R. R. (2007). Nonparametric item response theory and special topics. In: C. R. Rao & S. Sinharay (Eds.). *Handbook of Statistics*, *26*, 719–746. Amsterdam: Elsevier.
- Sijtsma, K. & Molenaar, I. W. (2002). *Introduction to Nonparametric Item Response Theory*, Thousand Oaks, CA: Sage.
- Van der Ark, L. A., & Bergsma, W. P. (2006). Investigating item orderings in polytomous test data. *Unpublished Manuscript*.
- Zhang, J., & Stout, W. F. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structures. *Psychometrika*, *64*, 213–249.