A Systematic Comparison Between Classical Optimal Scaling and the Two-Parameter IRT Model

Matthijs J. Warrens, Dato N. M. de Gruijter, and Willem J. Heiser, Leiden University, the Netherlands

In this article, the relationship between two alternative methods for the analysis of multivariate categorical data is systematically explored. It is shown that the person score of the first dimension of classical optimal scaling correlates strongly with the latent variable for the two-parameter item response theory (IRT) model. Next, under the assumption that the latent variable has a standard normal distribution, it is derived how the IRT item parameters are related to the optimal category weights. *Index terms: classical optimal scaling, two-parameter IRT model*

1. Introduction

Guttman (1941) presented a method to obtain a representation of the structure of multivariate categorical data. The method gives a multidimensional decomposition of the data with the most informative structural dimension extracted first, then the second most informative dimension, and so on, until the information in the data is exhaustively extracted. A typical use of the method is the construction of geometrical representations of the dependencies in the data in low-dimensional Euclidean space, often two-dimensional, from the extracted dimensions. Given that the data are in a Person $\times$ Item table, each dimension consists of weights for the item categories (known as optimal weights) and scores for the persons.

The discovery or rediscovery of Guttman's (1941) method by many authors has led to the fact that the method is known under many different names: for example, dual scaling (Nishisato, 1980), multiple correspondence analysis (Greenacre, 1984), or homogeneity analysis (Gifi, 1990). In this article, Guttman's method is referred to by the more neutral name, classical optimal scaling (COS), which was coined by McDonald (1983). As modern optimal scaling, one may consider any generalization of COS, as described in Torii (1977), McDonald (1983), or Gifi (1990), for example.

Gifi (1990, pp. 425-440) and Cheung and Mooi (1994) showed that COS is useful for analyzing Likert data. In addition, the latter authors compared the COS findings to an item response theory (IRT) analysis using the rating scale model (Andrich, 1978). They evaluated both the similarities and the differences and concluded that there is great similarity between the two contrasting approaches. A systematic comparison of COS and the IRT approach is lacking, however. The purpose of this article is to explore the relationship between a one-dimensional COS and the logistic

Applied Psychological Measurement, Vol. 31 No. 2, March 2007, 106–120

DOI: 10.1177/0146621605287912

two-parameter model systematically. Using simulated data, it is investigated how the COS person score is related to the latent variable assumed to underlie the IRT model. Furthermore, under an assumed standard normal distribution of the latent variable, it is derived how the IRT item parameters are related to the optimal category weights.

First, COS is further described in the next section. In Section 3, the person score and category weights are related to the IRT parameters for the dichotomous case. An extension of these relations is described in Section 4. Finally, Section 5 contains the discussion.

2. Classical Optimal Scaling

Given the different names for COS, it may not come as a surprise to the reader that there exist different approaches to obtain the category weights and person scores, although they lead mathematically to the same result. Let the multivariate categorical data be collected in a matrix $\mathbf{F}$ where i $(i=1,\dots,I)$ indexes the rows (persons) and j $(j=1,\dots,J)$ indexes the columns (items), and let item j have K_j categories, indexed by k $(k=1,\dots,K_j)$. An example of $\mathbf{F}$ containing the responses of six persons on three items with three, three, and two categories is presented in Figure 1.

The item categories can be nominal or ordinal, in which case the numbers in $\mathbf{F}$ are relatively arbitrary: COS throws the ordinal information out and treats the ordinal data as nominal. In COS, other numbers are assigned to the categories, and these numbers are called "optimal" weights for reasons to be explained later. An essential aspect of the method of COS is that the data matrix $\mathbf{F}$ is not analyzed directly. Instead, $\mathbf{F}$ is converted into a partitioned indicator matrix, denoted by $\mathbf{G}$, which is defined as follows. Let $\mathbf{G}_j$ be an indicator matrix of item j, defined as the order $I \times K_j$ matrix $\mathbf{G}_j = \{g_{ik(j)}\}$, where $g_{ik(j)}$ is a (0,1) variable. Each column of $\mathbf{G}_j$ refers to the K_j possible responses of item j. If person i responded category k on item j, then $g_{ik(j)} = 1$; that is, the cell in the ith row and kth column of $\mathbf{G}_j$ contains a 1, and $g_{ik(j)} = 0$ otherwise. The partitioned indicator matrix $\mathbf{G}$ then consists of all $\mathbf{G}_j$ positioned next to each other. The matrix $\mathbf{G}$ corresponding to matrix $\mathbf{F}$ is also presented in Figure 1.

There are several ways of assigning scores to the rows of $\mathbf{G}$, representing persons, and weights to the columns of $\mathbf{G}$, representing item categories. One way is by means of reciprocal averaging (Horst, 1935; Nishisato, 1980, Section 4.6). The matrix $\mathbf{G}$ is of size $I \times K$ $(m=1,\dots,K)$, where $K = \sum_{j=1}^{J} K_j$. Denote by

$$g_{i.} = \sum_{m=1}^{K} g_{im} \qquad \text{and} \qquad g_{.m} = \sum_{i=1}^{I} g_{im}$$

the row and column totals of $\mathbf{G}$. Let $\mathbf{x} = \{x_1,\dots,x_I\}$ be a set of scores for the rows or persons, and let $\mathbf{y} = \{y_1,\dots,y_K\}$ be a set of weights for the columns of $\mathbf{G}$. Optimal values for $\mathbf{x}$ and $\mathbf{y}$ need to be obtained. These can be obtained iteratively, first beginning with some arbitrary values for the categories y_m. From these category weights, scores for the persons can be determined. The person scores

$$x_i = \sum_{m=1}^{K} \frac{g_{im} y_m}{g_{i.}}$$

are the average values of the category values of the categories chosen by the persons. The derived person scores can now be used to obtain a new set of values for the item weights

$$y'_m = \sum_{i=1}^{I} \frac{g_{im} x_i}{g_{.m}}.$$

Figure 1
Example of a Data Matrix **F** of Three Items With Three, Three, and Two Categories and the Corresponding Partitioned Indicator Matrix **G**

$$\mathbf{F} = \begin{bmatrix} 2 & 1 & 2 \\ 2 & 3 & 1 \\ 1 & 2 & 1 \\ 3 & 1 & 1 \\ 1 & 2 & 2 \\ 3 & 1 & 1 \end{bmatrix} \Rightarrow \mathbf{G} = \left[\begin{array}{ccc|ccc|cc} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{array}\right]$$

This process, called reciprocal averaging, can be iterated with the new scores y'_m in place of the old ones y_m. The column scores are the averages of the row scores; reciprocally, the row scores are the averages of the column scores. With some form of normalization, this process will almost always converge to a unique solution (cf. Gifi, 1990). In this article, a common normalization for the final **x** and **y** is used: **x** is put in standard scores. That is, **x** has zero mean and $\mathbf{x}^T\mathbf{x} = I$, where $\mathbf{x}^T$ denotes the transpose of the column vector **x**; the y_ms are the averages of the corresponding x_is.

The vectors **x** and **y** form the first COS dimension and contain the most informative part of the data in terms of variance. In addition, one may extract higher dimensions to obtain a more complete representation of the categorical data. In many real applications, it turns out that the first two COS dimensions suffice. In this article, the information in **x** and **y** is sufficient because this study attempts to compare COS to one-dimensional IRT models. At this point, several interesting properties of **x** and **y** should be mentioned. First of all, the scores and weights are obtained without recourse to order or distributional assumptions of the data. Furthermore, the **y** are "optimal" given several criteria. The **y** are optimal because they maximize the between-row and between-column discrimination of the matrix $\mathbf{G}^T\mathbf{G}$ simultaneously (Guttman's principle of internal consistency; cf. Nishisato, 1980, pp. 21-27). Another "optimal" criterion was derived by Lord (1958; see Section 3.3). With reciprocal averaging, the values of **x** and **y** are obtained simultaneously, similar to joint maximum likelihood estimation in the one- and two-parameter logistic models (see Section 3.1). The scores and weights can also be obtained separately, for example, by means of eigendecomposition of certain matrices.

3. The Dichotomous Case

3.1 The Two-Parameter Model

The research in the field of IRT has produced a vast amount of models over the past couple of decades. One of the first of these models was the one-dimensional two-parameter model (2-PM; Birnbaum, 1968; Lord, 1952) for dichotomous data. With dichotomous data, $K_j = 2$ for all j. Therefore, let the so-called incorrect response of person i on item j be denoted by $f_{ij} = 1$, and let

$f_{ij} = 2$ denote the correct response (this notation deviates from the (0,1) convention normally used in IRT). With the 2-PM, the item response function has two item parameters—one for location and one for discrimination—for explaining persons' probabilities of answering an item correctly as a function of a latent variable. The normal ogive formulation of the 2-PM comes from Lord (1952). Birnbaum (1968) later proposed the logistic form of the 2-PM. In the latter form, the conditional probability of a correct response to item j $f_j = 2$ is modeled as a logistic function of a latent variable θ, with

$$P_j(\theta) \;=\; \text{Prob}(f_j = 2|\theta, a_j, b_j) \;=\; \frac{\exp[a_j(\theta - b_j)]}{1 + \exp[a_j(\theta - b_j)]},$$

where a_j is used to denote the discrimination parameter and b_j the location parameter of item j. The incorrect response $f_j = 1$ is then modeled by $1 - P_j(\theta)$. A special case of the logistic 2-PM, the Rasch (1960) model, is obtained if $a_j = 1$ for all j.

Note that the 2-PM uses two item parameters, whereas COS produces two category weights. Furthermore, both approaches use one parameter for locating persons. In the next subsection, it is first shown how the IRT person parameter estimate and the COS person score are related.

3.2 Person Parameter and Person Score

Two data sets were generated from both the logistic 2-PM and the Rasch model under the following conditions: the data sets consisted of the responses of 1,000 persons on 50 items; for each data set, the location parameters b_js were sampled from a standard normal distribution; the discrimination parameters for the 2-PM were sampled from a uniform distribution on the range [1, 2], and these were set to unity for the Rasch model; and the latent variable was sampled from a standard normal distribution.

For both data sets, the COS and IRT person estimates were obtained. The IRT analysis was performed using the Multilog software program (Thissen, Chen, & Bock, 2003) to obtain maximum a posteriori (MAP) estimates. The person estimates of both approaches are plotted in Figures 2 and 3 for the Rasch model and the logistic 2-PM, respectively.

The correlations between the two sets of estimates, in both figures, are $> .99$. The root mean squared errors are $< .2$, which concurs with the slight nonlinearity that can be observed upon close inspection. Apart from the nonlinearity, the COS person score seems a reasonable approximation of the latent variable—that is, $\theta_i \approx x_i$—under the 2-PM. In the following subsection, this idea is exploited to functionally relate the COS category weights to the IRT item parameters.

3.3 Discrimination Parameter and Category Weights

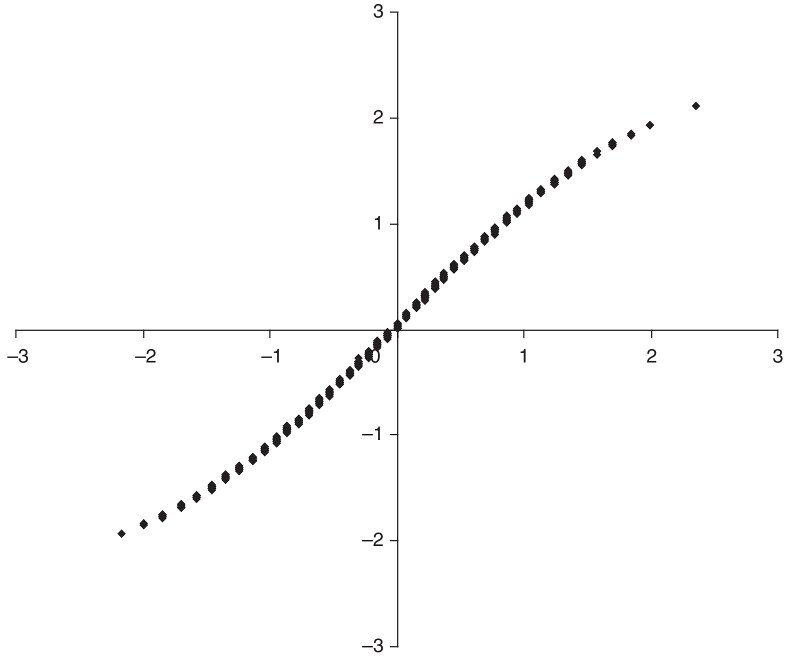
Let y_{j2} and y_{j1} denote the weights for, respectively, the correct and incorrect category of item j. With the normalization in Section 2, **x** is put in standard scores, and the y_ms are the averages of the corresponding x_is. So y_{j2} is the average value of all persons who responded correctly to item j, and y_{j1} is the average of those persons who responded incorrectly. Because $\sum_{i=1}^{I} x_i = 0$, it follows that

$$\pi_j y_{j2} + (1 - \pi_j) y_{j1} \;=\; 0, \tag{1}$$

where π_j is used to denote the proportion of persons who responded correctly to item j. Using equation (1), y_{j1} may be written as

$$y_{j1} \;=\; -\frac{\pi_j \, y_{j2}}{1 - \pi_j}. \tag{2}$$

Figure 2
Plot of Maximum a Posteriori (MAP) Person Estimates (Horizontal) Versus Classical Optimal Scaling (COS) Person Scores (Vertical) for the Rasch Data Set



Lord (1958) showed that the COS category weights **y** maximize coefficient alpha, an important lower bound to reliability. Because, in the dichotomous case, there are only two categories, it is possible to reflect all information on discrimination of item j in a single index. This can be done by translating the weights y_{j1} and y_{j2} into new weights w_{j1} and w_{j2}. With the translations

$$\begin{aligned} w_{j1} &= y_{j1} - y_{j1} \ = \ 0, \\ w_{j2} &= y_{j2} - y_{j1}, \end{aligned} \tag{3}$$

the category weight of the incorrect category is set to zero, and all information on discrimination of item j is reflected in w_{j2}. Because w_{j2} is the weight of item j that maximizes coefficient alpha, it is denoted by $\max(\alpha)_j$ in the following. Inspection of equation (3) reveals that the $\max(\alpha)_j$ item weight becomes greater as the mean values of all persons who responded correctly to item j and those who responded incorrectly become further apart. Hence, $\max(\alpha)_j$ has a clear interpretation as an index of discrimination. With the help of equation (2), $\max(\alpha)_j$ can be written as

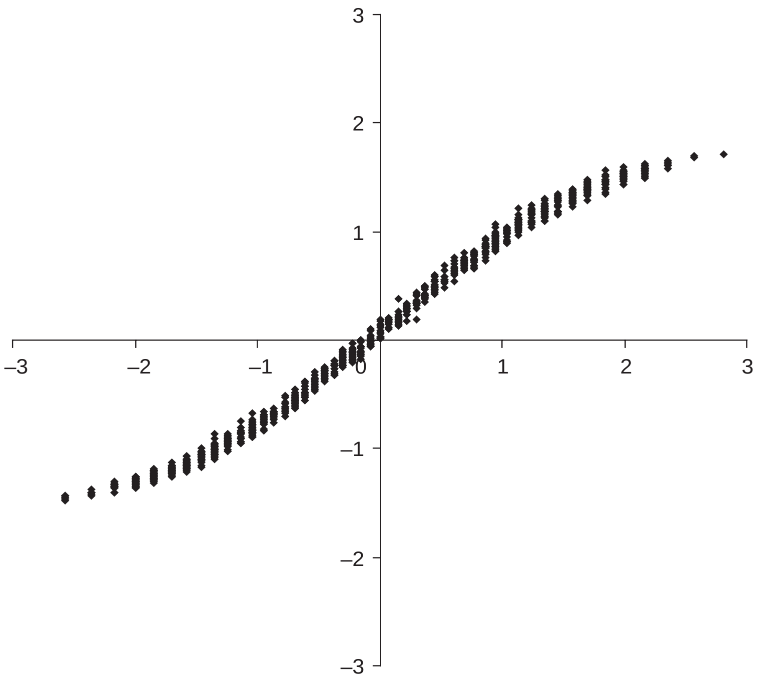
$$\max(\alpha)_j \ = \ \frac{y_{j2}}{1-\pi_j}.$$

In the following, it is assumed that $\theta_i \approx x_i$. In addition, it is assumed that

- the latent variable is normally distributed with zero mean and unit variance and
- the appropriate model is the 2-PM.

Figure 3

Plot of Maximum a Posteriori (MAP) Person Estimates (Horizontal) Versus Classical Optimal Scaling (COS) Person Scores (Vertical) for the Logistic Two-Parameter Model (2-PM) Data Set



Under the above assumptions, the work of Lord and Novick (1968) on the relationship between the IRT item parameters and some indices from classical item analysis becomes available. The relationship between the category weights and the IRT item parameters is elaborated in Appendix A. Let the constant $D = 1.7$, and denote by ρ_j^* the biserial correlation between item j and the latent variable under the logistic 2-PM. Under the above three assumptions, it follows from Appendix A that

$$\max(\alpha)_j \approx D\rho_j^*, \\ \approx \frac{a_j}{\sqrt{1 + D^{-2}a_j^2}}. \tag{4}$$

The functional relationship in equation (4) was derived by De Gruijter (1984) in a different way. Because ρ_j^* has a maximum of unity, the quantity in equation (4) has a maximum value of D. Because the $\max(\alpha)_j$ weight is a function of a_j only, a_j can be expressed as a function of $\max(\alpha)_j$. The resulting function gives an estimate of the discrimination parameter of the logistic 2-PM, given by

$$\hat{a}_j = \frac{D \max(\alpha)_j}{\sqrt{D^2 - [\max(\alpha)_j]^2}}, \quad \text{for } |\max(\alpha)_j| \leq D, \tag{5}$$

which is a function of $\max(\alpha)_j$ only. The relationships between $\max(\alpha)_j$ and two other discrimination indices are presented in Appendix B.

3.4 Location Parameter and Category Weights

Now that the functional relationship between the discrimination indices has been established, the focus turns to the remaining information in the weights y_{j2} and y_{j1}. Because $\max(\alpha)_j$ is given by the difference between y_{j2} and y_{j1}, the remaining information in the weights can be summarized in

$$s_j = y_{j2} + y_{j1},$$

that is, the sum of the two category weights of item j. With the help of equation (2), s_j can be written as

$$s_j = \frac{1 - 2\pi_j}{1 - \pi_j} \, y_{j2}.$$

Under the same three assumptions as used in the previous subsection, it follows from Appendix A that

$$s_j \approx D\rho_j^*(1 - 2\Psi[-b_j D\rho_j^*]). \tag{6}$$

Suppose now that ρ_j^* in equation (6) is constant for all j. For this limited case, it holds that if b_j increases, s_j also increases. Because b_j and s_j are monotonically related under this restriction, s_j can be interpreted as a location parameter for a model of which the discrimination parameters are equal for all j, that is, the Rasch model.

From equation (6), an estimate for the location parameter b_j of the logistic 2-PM can be obtained. This estimate can be simplified. In addition to $\max(\alpha)_j$, only the proportion correct, denoted by π_j, is needed. Let Ψ denote the logistic function. Then, from equation (13) in Appendix A, it follows that

$$\pi_j \approx \Psi[-b_j \max(\alpha)_j]. \tag{7}$$

If one takes the inverse of the logistic function on both sides of equation (7) and rewrites the resulting equation in terms of b_j, one obtains an estimate of location for item j, given by

$$\hat{b}_j = -\frac{\ln\left(\frac{\pi_j}{1-\pi_j}\right)}{\max(\alpha)_j}. \tag{8}$$

Note that the estimate derived in equation (8) is related to the estimate proposed by Cohen (1979) for the Rasch model.

The estimates in equations (5) and (8) were applied to the generated 2-PM data set also used in Figure 3. The IRT analysis was again performed using the Multilog software program, which provides marginal maximum likelihood (MML) estimates for the item parameters of the logistic 2-PM. The two sets of estimates for the discrimination and location parameters are visualized in Figure 4.

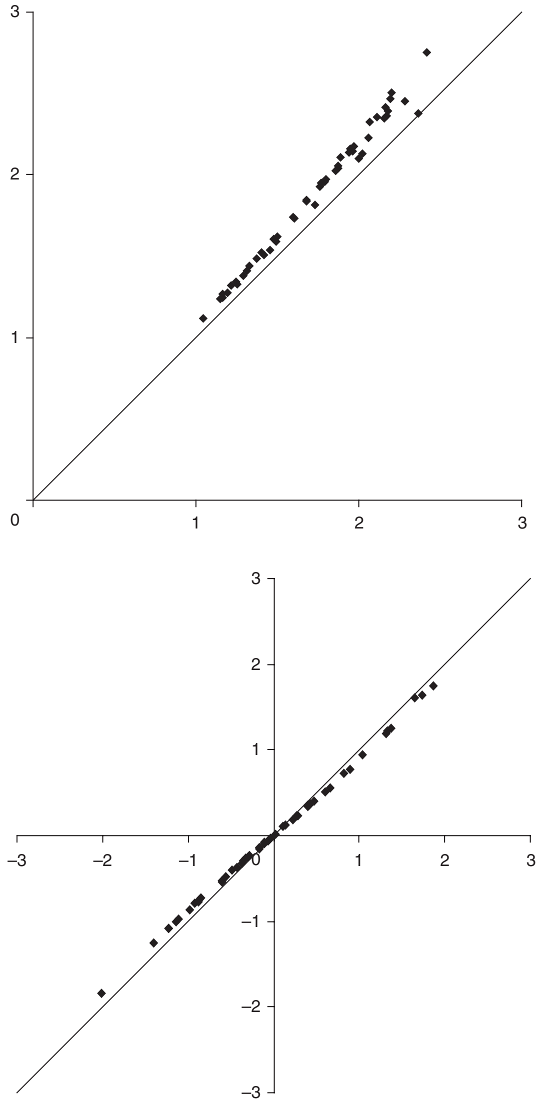
Close inspection of the figure reveals the closeness of the estimates for the simulated data. Furthermore, these findings could be considered empirical support for the functional relationship derived in this section.

4. The Polytomous Case

In this section, a possible extension of the results derived in the previous section to the polytomous case is considered. In the graded response model (GRM; Samejima, 1969), there are $K_j - 1$ ($k = 1, \ldots, K_j - 1$) item steps for item j instead of one item function. The probability of the latent variable taking a value greater than the item step k associated with item j depends on the latent variable θ, the value of the step boundary b_{jk}, and the item's discrimination parameter a_j. With the

Figure 4

Two Plots of Item Parameter Estimates for the Logistic Two-Parameter Model (2-PM) Data Set



Note. Top: Marginal maximum likelihood (MML) discrimination estimates (horizontal) versus classical optimal scaling (COS) estimates based on equation (5). Bottom: MML location estimates (horizontal) versus COS estimates based on equation (8).

logistic formulation of the GRM, the model probability of responding in or above category k to item j is specified as

$$P_{jk}(\theta) = \frac{\exp[a_j(\theta - b_{jk})]}{1 + \exp[a_j(\theta - b_{jk})]},$$

where b_{jk} is the boundary between categories $k-1$ and k associated with item j, with $b_{jk} > b_{jk-1}$. If $P_k(\theta)$ denotes the item step, then $P_k^*(\theta)$ denotes the probability of a category, which can be obtained by

$$\begin{aligned} P_1^*(\theta) &= 1 - P_1(\theta), \\ P_k^*(\theta) &= P_{k-1}(\theta) - P_k(\theta) \quad \text{for } 1 < k < K_j - 1, \\ P_{K_j}^*(\theta) &= P_{K_j-1}(\theta). \end{aligned} \tag{9}$$

Because of equation (9), the GRM is classified by Thissen and Steinberg (1986) as a "difference" model. The essential aspect of the "difference" approach is that the elemental units in theoretical work are the dichotomous item steps $P_k(\theta)$, for $k = 1, \ldots, K_j - 1$. Also, because $P_k(\theta)$ is a logistic function, it is necessary that the discrimination parameter a_j is equal for all the elemental item steps that make up the category response functions that define the probability of a particular categorical response. This is true because logistics or normal ogives with different discrimination values always cross each other somewhere; this would result in negative values for equation (9).

With COS, K_j weights are obtained for the K_j categories of item j. To relate these category weights to the item steps of the GRM, the $K_j - 1$ dichotomies must be distinguished. The mean value of the correct and incorrect groups of the kth dichotomy of item j is denoted by, respectively, z_{jk2} and z_{jk1}. Furthermore, denote by π_{jk} the proportion of persons in category k of item j. The conditional means can then be obtained from the COS category weights and category marginals with

$$z_{jk2} = \frac{\sum_{r=k+1}^{K_j} \pi_{jk} y_{jk}}{\sum_{r=k+1}^{K_j} \pi_{jk}}, \quad \text{and} \quad z_{jk1} = \frac{\sum_{r=1}^{k} \pi_{jk} y_{jk}}{\sum_{r=1}^{k} \pi_{jk}}, \quad \text{for } k = 1, \ldots, K_j - 1.$$

Denote by $\text{dif}_{jk} = z_{jk2} - z_{jk1}$. If $\max(\alpha)_j$ in equation (5) is replaced by dif_{jk}, an estimate for the discrimination parameter of the kth item step of the logistic GRM can be obtained with

$$\hat{a}_{jk} = \frac{D \ \text{dif}_{jk}}{\sqrt{D^2 - \text{dif}_{jk}}}, \quad \text{for } k = 1, \ldots, K_j - 1. \tag{10}$$

Because it is necessary that the discrimination parameter a_j is equal for all the elemental item steps that make up the category response function that defines the probability of a particular categorical response, a difficulty emerges with the estimate in equation (10). With this estimate, $K_j - 1$ discrimination parameters are obtained for item j instead of 1. The mean value can be taken in practical cases to solve the conceptual difficulty. On the other hand, the different COS estimates can also be compared to each other to evaluate the approach described in this section. If all values are close together, this could be taken as empirical evidence for the functional relationship between the two approaches.

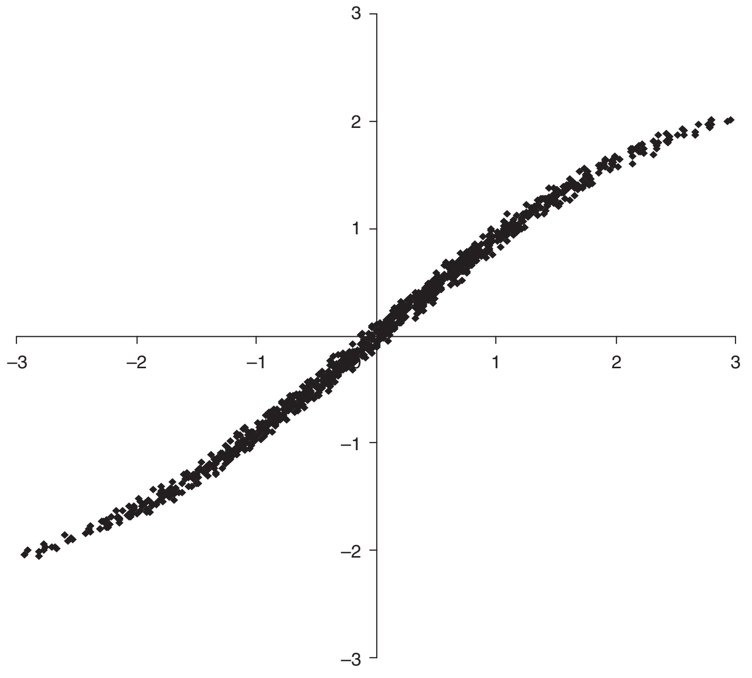
Next, let π_{jk2} denote the proportion of persons in the correct category of the kth item step of item j. Then an estimate for the location parameter of the kth item step of item j can be obtained with

$$\hat{b}_{jk} = -\frac{\ln\left(\frac{\pi_{jk2}}{1-\pi_{jk2}}\right)}{\text{dif}_{jk}}, \quad \text{for } k = 1, \ldots, K_j - 1. \tag{11}$$

To demonstrate the estimates in equations (10) and (11), a data set was generated from the GRM under the following conditions: the data consisted of the responses of 1,000 generated persons on 50 items, the location parameters b_{jk}s were sampled from a standard normal distribution, the discrimination parameters were sampled from a uniform distribution on the range [1, 2], and the number of categories was uniformly sampled to have a range of 3 to 5. The latent trait was taken to be the standard normal.

Figure 5

Plot of Maximum a Posteriori (MAP) Person Estimates (Horizontal) Versus Classical Optimal Scaling (COS) Person Scores (Vertical) for the Graded Response Model (GRM) Data Set



First, the person estimates were obtained for the data set. The COS person scores are plotted against the IRT MAP estimates obtained with Multilog in Figure 5.

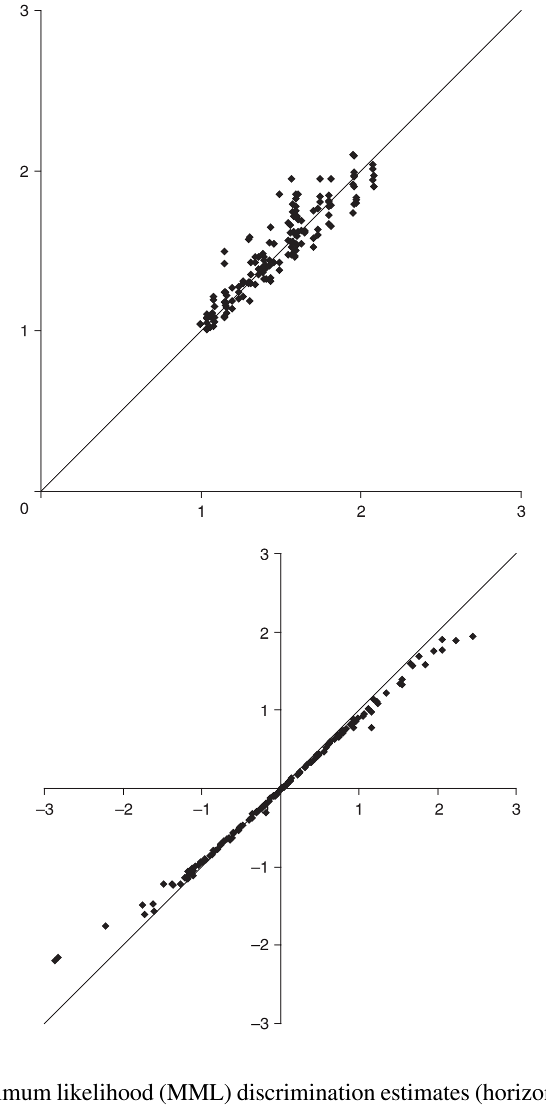
The figure reveals that under the GRM, the COS person score is again a reasonable estimate of the latent variable. Next the item parameters were estimated. The MML estimates for the discrimination and location parameters are plotted in Figure 6 against the estimates obtained with equations (10) and (11). With equation (10), $K_j - 1$different estimates were obtained for the discrimination parameter. All $K_j - 1$ COS discrimination estimates were plotted against the single IRT estimates.

Close inspection of the figure reveals the closeness of the estimates for the data at hand, where the relation between the location estimates appears to be stronger than the relation between the discrimination estimates.

5. Discussion

Several authors (e.g., Cheung & Mooi, 1994; Gifi, 1990) noted strong similarities between an analysis of the multivariate categorical data using either Guttman's (1941) classical optimal scaling or an IRT-based analysis. In this article, the relationship between the two seemingly alternative methods was systematically explored. It was shown that the person score of the first dimension of COS is a close approximation of the latent variable for several IRT models. Next, under an assumed standard normal distribution of the latent variable, it was derived how the item parameters are functionally related to the optimal category weights.

Figure 6
Two Plots of Item Parameter Estimates for the Graded Response Model (GRM) Data Set



Note. Top: Marginal maximum likelihood (MML) discrimination estimates (horizontal) versus classical optimal scaling (COS) estimates based on equation (10). Bottom: MML location estimates (horizontal) versus COS estimates based on equation (11).

At this point, the following question arises: what is the point of knowing the functional relationship between COS and IRT? First of all, it is useful in general to study equivalencies or functional relationships between different methods of data analysis, primarily because this often gives new insight into the methods themselves. More precisely, in Section 3, approximate estimates for the item parameters of the logistic 2-PM were derived, which are based on the conditional means. The estimates were not meant as a possible replacement of the current IRT approach. However, one might be tempted to ask if these estimates may be used to obtain perhaps less biased parameter estimates (maximum likelihood estimation is already most efficient). In nonreported simulation experiments, it turns out that the COS-based estimates do not give less biased estimates or smaller standard errors.

On the other hand, the closeness of the COS person score to the latent variable under a variety of IRT models shows that COS is a useful multipurpose data analysis method. Even without specifying a model, one cannot be far off. A more extensive investigation of the accuracy and usefulness of COS person scores is a possible extension of the present work. The relationship between person and item parameters for the logistic 2-PM and the GRM has also been empirically obtained for other distributions (e.g., uniform or bimodal) of the latent variable than the standard normal. A second possible extension of the present work is the case of multidimensional IRT models.

The findings of this article do give several new insights into the application of COS. A typical use of COS and related optimal scaling methods is the construction of geometrical representations of the dependencies in the data in low-dimensional Euclidean space, often two-dimensional, from the extracted dimensions. The use of two-dimensional (sometimes three-dimensional) plots is embedded so strongly in the optimal scaling community that it is often regarded impossible that all relevant information is in the first dimension only.

For the case of dichotomous data, up to three COS measures of discrimination were distinguished in Appendix B. In Section 3.3, it was argued that the $\max(\alpha)_j$ weight has a clear interpretation as a discrimination parameter, and it was shown that under several assumptions, the $\max(\alpha)_j$ weight is monotonically related to the discrimination parameter of the logistic 2-PM. From Appendix B, it follows that the other two COS measures proposed in the literature do not have a monotonic relationship to the $\max(\alpha)_j$ measure because in both relationships, the variance of item j is involved. Because $\max(\alpha)_j$ has such a straightforward interpretation, it should be the preferred measure of discrimination for the case of dichotomous data.

Appendix A: Approximations

Normal Ogive and Classical Item Analysis

On pages 377 and 378 of their by now-classic book, Lord and Novick (1968) show how the item parameters of the 2-PM normal ogive are related to the indices used in classical item analysis. Two conditions are assumed:

1. The latent variable is normally distributed with zero mean and unit variance.
2. The appropriate model is the 2-PM normal ogive.

Under the above conditions, the mean of θ, conditional on a correct response on item j, equals

$$m_{j2} = \frac{\phi(\gamma_j)\rho'_j}{\pi_j},$$

where $\pi_j = \Phi(-\gamma_j)$ is the item proportion correct, where Φ denotes the cumulative normal distribution function, and $\gamma_j = b_j\rho'_j$; $\phi(\gamma_j)$ is the ordinate of the standard normal distribution; and

$$\rho'_j = \frac{a_j}{\sqrt{1+a_j^2}}$$

is the biserial correlation between item j and the latent variable.

Logistic Approximation

Due to the fact that the logistic formulation of the 2-PM is more tractable then the normal ogive, the former is often preferred in present IRT work. Here it is shown how the above relations, on the basis of the normal ogive, hold under the logistic approximation. The logistic 2-PM and its approximate relation with the normal ogive 2-PM are given by

$$P_j(\theta) \;=\; \Psi[a_j(\theta - b_j)] \;\approx\; \Phi\left[D^{-1}a_j(\theta - b_j)\right],$$

where Ψ denotes the logistic function, and $D = 1.7$ is a constant. Under the logistic approximation, the mean of θ, conditional on a correct response on item j, equals

$$m_{j2} \;\approx\; \frac{\phi(\gamma_j^*)\rho_j^*}{\Psi(-D\gamma_j^*)}, \tag{A1}$$

where

$$\Psi(-D\gamma_j^*) \approx \pi_j, \tag{A2}$$

$\gamma_j^* = b_j\rho_j^*$, and

$$\rho_j^* = \frac{a_j}{D\sqrt{1 + D^{-2}a_j^2}}.$$

Furthermore, under the logistic approximation,

$$\phi(\gamma_j^*) \;\approx\; D\Psi(D\gamma_j^*)[1 - \Psi(D\gamma_j^*)] \;=\; D\Psi(-D\gamma_j^*)[1 - \Psi(-D\gamma_j^*)],$$

and equation (A1) can be rewritten as

$$m_{j2} \;\approx\; (1 - \pi_j)D\rho_j^*.$$

COS Approximation

In Section 3 of this article, it is assumed that the COS person score is a reasonable approximation of the latent variable (i.e., $x_i \approx \theta_i$). The purpose of Figures 2 and 3 is to check if this assumption is a reasonable one. The assumption leads to the approximation $y_{j2} \approx m_{j2}$ on which the results in equations (5) and (7) are based.

Appendix B: Discrimination Measures

A third measure of discrimination for item j, next to a_j and $\max(\alpha)_j$, is described in Gifi (1990, Section 3.8.4). The measure, which will be denoted by η_j^2, is given by

$$\eta_j^2 \;=\; \frac{\mathbf{y}_j^T\mathbf{H}_j\mathbf{y}_j}{I}, \tag{B1}$$

where $\mathbf{H}_j$ is a diagonal matrix containing the diagonal elements of $\mathbf{G}_j^T\mathbf{G}_j$, and I is the total number of persons. It can be shown that η_j is the component loading for item j of ordinary principal component analysis (cf. Gifi, 1990, Section 3.9; Yamada & Nishisato, 1993). With dichotomous data, equation (B1) can be written as

$$\eta_j^2 \;=\; \pi_j y_{j2}^2 \;+\; (1 - \pi_j)y_{j1}^2. \tag{B2}$$

Equation (B2) can be reexpressed in terms of y_{j2} and y_{j1}, with the help of equation (2), which gives

$$y_{j2} = \eta_j \left[\frac{1-\pi_j}{\pi_j}\right]^{1/2},$$
$$-y_{j1} = \eta_j \left[\frac{\pi_j}{1-\pi_j}\right]^{1/2}.$$

Hence, the following is obtained:

$$\max(\alpha)_j = \frac{\eta_j}{[\pi_j(1-\pi_j)]^{1/2}},$$

or

$$\eta_j^2 = \pi_j(1-\pi_j)(\max(\alpha)_j)^2,$$

a result also reported in Yamada and Nishisato (1993, p. 60). In words, η_j^2 is the squared $\max(\alpha)_j$ weight of item j times the variance of item j.

A fourth measure of discrimination is described in McDonald (1983). In a more general context than the one considered in this article, McDonald argued not to interpret the category weights themselves but the regression weights of each category on the person score $\mathbf{x}$, given by

$$r_{jk} = \pi_{jk} y_{jk},$$

where π_{jk} is the proportion of persons in category k of item j. Note that with McDonald's (1983) formulation, there is not one discrimination measure for each item j but multiple measures for the K_j categories. Note that equation (1) can be written as

$$\pi_j y_{j2} = (\pi_j - 1) y_{j1},$$

from which it follows that

$$r_{j2} = -r_{j1}.$$

Because, with dichotomous data, the two regression weights contain the same information, it suffices to look at r_{j2}, assumed to be positive, only. From Subsection 3.3,

$$y_{j2} = (1-\pi_j)\max(\alpha)_j. \tag{B3}$$

Multiplication of both sides of equation (B3) with π_j gives

$$r_{j2} = \pi_j(1-\pi_j)\max(\alpha)_\mathrm{j}.$$

Thus, in words, the regression weight of item j is either plus or minus the variance of item j times the $\max(\alpha)_j$ weight.

References

Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika, 43*, 561-573.

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.

Cheung, K. C., & Mooi, L. C. (1994). A comparison between the rating scale model and dual scaling for Likert scales. *Applied Psychological Measurement, 18*, 1-13.

Cohen, L. (1979). Approximate expressions for parameter estimates in the Rasch model. *British Journal of Mathematical and Statistical Psychology, 32*, 113-120.

de Gruijter, D. N. M. (1984). Homogeneity analysis of test score data: A confrontation with the latent trait approach. *Applied Psychological Measurement, 8*, 385-390.

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester, UK: Wiley.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* New York: Academic Press.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst (Ed.), *The prediction of personal adjustment*. New York: SSRC.

Horst, P. (1935). Measuring complex attitudes. *Journal of Social Psychology, 6*, 369-374.

Lord, F. M. (1952). A theory of mental test scores. *Psychometrika, Monograph No. 7*.

Lord, F. M. (1958). Some relations between Guttman's principal components analysis and other psychometric tests. *Psychometrika, 36*, 109-133.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

McDonald, R. P. (1983). Alternative weights and invariant parameters in optimal scaling. *Psychometrika, 48*, 377-391.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Samejima, F. (1969). Estimation of latent ability using a response pattern for graded scores. *Psychometrika*, *Monograph No. 17*.

Thissen, D., Chen, W. H., & Bock, D. (2003). *Multilog 7: Analysis of multiple-category response data.* Lincolnwood, IL: Scientific Software International.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51, 567-577.

Torii, Y. (1977). *Some generalizations of optimal scaling*. Unpublished doctoral dissertation.

Yamada, F., & Nishisato, S. (1993). Several mathematical properties of dual scaling as applied to dichotomous item-category data. *Japanese Journal of Behaviormetrics, 20*, 56-63.

Author's Address

Address correspondence to Matthijs J. Warrens, Leiden University, Wassenaarseweg 52, P.O. Box 9555, 2300 RB Leiden, Leiden, the Netherlands; e-mail: warrens@fsw.leidenuniv.nl.