Psychological Methods
2005, Vol. 10, No. 3, 285–304

1082-989X/05/$12.00 DOI: 10.1037/1082-989X.10.3.285

Structural Equation Modeling of Paired-Comparison and Ranking Data

Albert Maydeu-Olivares
University of Barcelona and Instituto de Empresa

Ulf Böckenholt
McGill University

L. L. Thurstone's (1927) model provides a powerful framework for modeling individual differences in choice behavior. An overview of Thurstonian models for comparative data is provided, including the classical Case V and Case III models as well as more general choice models with unrestricted and factor-analytic covariance structures. A flow chart summarizes the model selection process. The authors show how to embed these models within a more familiar structural equation modeling (SEM) framework. The different special cases of Thurstone's model can be estimated with a popular SEM statistical package, including factor analysis models for paired comparisons and rankings. Only minor modifications are needed to accommodate both types of data. As a result, complex models for comparative judgments can be both estimated and tested efficiently.

Keywords: Lisrel, preference data, choice models, random utility models, factor analysis

The methods of ranking and paired comparison play an essential role in the measurement of preferences, attitudes, and values. In a ranking task, respondents are presented with a set of alternatives (which, in the literature, are also referred to as *options, stimuli,* or *items*) and are asked to order them from most to least preferred. In a paired-comparison task, respondents are presented with pairs selected from the set of available alternatives and are instructed to select the more preferred alternative from each pair. The popularity of paired-comparison and ranking methods can be traced back to three reasons. First, by asking for a comparison of choice alternatives, these methods impose minimal constraints on the response behavior of a respondent. Especially when differences between choice alternatives are small, these methods are likely to provide more information about individual preferences than is obtainable by rating methods. Second, internal consistency checks are available that facilitate the identification of respondents who do not have well-defined preferences, values, or attitudes. If respondents can be shown to be consistent in their judgments, one can have much greater confidence in the obtained measurements and the predictive value of the derived scales in further applications. Third, paired-comparison and ranking data provide a rich source of information about the effects of individual differences and perceived similarity relationships among choice alternatives. This article presents two applications that illustrate this important feature.

Because of their versatility, the methods of paired comparison and rankings are used in a wide range of studies. Recent applications include visual paired-comparison studies involving young children (Pascalis, de Haan, & Nelson, 2002; Turati & Simion, 2002; Younger & Furrer, 2003), rankings of risk perceptions (Florig et al., 2001; Morgan et al., 2001), food characteristics (Oakes & Slotterback, 2002), and clinical services (Hazell, Tarren-Sweeney, Vimpani, Keatinge, & Callan, 2002). Most statistical models for the analysis of paired-comparison and ranking data are based on Thurstone's (1927) work, which emphasized that decisions should be viewed as probabilistic to account for apparent inconsistencies in choice outcomes. For example, when confronted several times with the same choice alternative pair, respondents may not consistently choose the same choice alternative each time. Similarly, respondents may not always be consistent in the comparison of several choice alternative pairs. Consider, for example, the choice alterna-

Albert Maydeu-Olivares, Faculty of Psychology, University of Barcelona, Barcelona, Spain, and Marketing Department, Instituto de Empresa, Madrid, Spain; Ulf Böckenholt, Faculty of Management, McGill University, Montreal, Quebec, Canada.

Additional materials are on the Web at http://dx.doi.org/10.1037/1082-989X.10.3.285.supp

Albert Maydeu-Olivares was supported by the Department of Universities, Research and Information Society of the Catalan Government and by grants BSO2000-0661 and BSO2003-08507 from the Spanish Ministry of Science and Technology. Ulf Böckenholt was supported by the Social Sciences and Humanities Research Council of Canada.

We are indebted to David Gallardo-Pujol for supplying the figures and to Carlos García-Forero for comments on earlier versions of the article.

Correspondence concerning this article should be addressed to Albert Maydeu-Olivares, Faculty of Psychology, University of Barcelona, Paseo Valle de Hebrón, 171, 08035 Barcelona, Spain. E-mail: amaydeu@ub.edu

tives set {A, B, C}. A respondent may choose B when given the pair {A, B}, A when given the pair {A, C}, and B when given the pair {B, C}. These choices are consistent with a {B, A, C} ordering of the items, and the pattern of paired comparisons is said to be transitive. In contrast, an intransitive pattern results when B is chosen for the pair {A, B}, A for the pair {A, C}, but C for the pair {B, C}.

To account for these apparent inconsistencies, Thurstone (1927) proposed a framework that in today's terms may be called a discrete latent utility[1] model. He argued that in a choice task, (a) each choice alternative elicits a latent continuous utility judgment as a result of a discriminal process, (b) respondents choose the choice alternative with the largest utility value at the moment of comparison, and (c) the utility values are normally distributed in the population of respondents. Thus, his approach may be viewed as a latent variable model in which each latent variable corresponds to the utility judgment for one of the choice alternatives (Takane, 1987, 1994). As it is shown in the next section, the observed preferences or choices are obtained on the basis of a comparison of these latent variables. Thurstone (1927) discussed several special versions of this approach, including the Case V and Case III models. The Case III model assumes that the latent utilities are uncorrelated in the population of judges. The Case V model assumes further that the latent utilities have a common variance. Although he focused initially on paired comparisons, Thurstone (1931) recognized later that many other types of choice data, including rankings, could be modeled in a similar way.

Thurstone's early conceptualizations gave rise to an extensive body of methodological and empirical research. Surveys about developments in choice theories (Böckenholt, 2001b; Luce, 1977; Marley, 2002; Suppes, Krantz, Luce, & Tversky, 1992) and choice modeling (Böckenholt, 1993; Fligner & Verducci, 1993; Train, 2003) indicate that Thurstone's approach continues to play a fundamental role in the structural investigation of choice data. However, applications of Thurstonian choice models in psychological work are rare because, until recently, model estimation proved to be a complex task and required considerable statistical expertise. Fortunately, this situation has changed. We show that if Thurstonian choice models are embedded within a structural equations framework for discrete data, the estimation of Thurstonian models for paired comparisons and rankings becomes straightforward because of their similarity to a confirmatory common factor model with binary indicators. As a result, computer programs for the estimation of structural equation models can be used for both estimating and testing the adequacy of Thurstonian choice models. In addition, by framing Thurstonian models as a special class of structural equation models, researchers gain access to a full array of new modeling possibilities. On the one hand, less restrictive models than Thurstone's Case III and Case V models are readily apparent. For instance, a model in which the latent utilities are all intercorrelated can be fitted. We refer to this model as an *unrestricted Thurstonian model.* Alternatively, a model in which the latent utilities are structured according to a factor model can be considered. This model is similar to a second-order factor model with binary indicators. Finally, when background information about the respondents and/or the choice alternatives is available, those variables can be readily incorporated into the model.

The purpose of this article is thus twofold. We present a comprehensive nontechnical account of Thurstonian choice modeling that reviews, integrates, and expands on recent technical research on Thurstonian choice modeling. In addition, we embed Thurstonian models within a structural equation modeling (SEM) framework and show how these models can be estimated with a popular SEM package, Mplus (L. Muthén & Muthén, 2001). Technical accounts on a subset of Thurstonian ranking and paired-comparisons models within an SEM framework were given by Maydeu-Olivares (1999, 2001, 2003b). We go beyond this work by including factor models with and without restrictions on the estimated utility means of the choice alternatives in our presentation.

The present article also complements and extends Böckenholt's (2001a) work, which showed that the Bradley–Terry–Luce paired-comparison model can be fitted with multilevel software. The Bradley–Terry–Luce model differs from Thurstone's model by assuming that the within-pair variability follows a logistic instead of a normal distribution. We go beyond Böckenholt (2001a) by (a) allowing for heterogeneous within-pair variances in paired-comparisons models, (b) including models for ranking data, and (c) considering factor models for both paired comparisons and rankings. These models for paired comparisons cannot be analyzed with the estimation approach presented in that article.

Factor models play a fundamental role in choice modeling. On the one hand, they help researchers to identify sources of individual differences in utility judgments. On the other hand, we show in this article that they enable researchers to overcome the interpretational problems recently outlined by Tsai (2003) that are inherently present when modeling comparative data. Thurstonian factor analytic models for paired comparisons were considered previously by Tsai and Böckenholt (2001; see also Takane, 1987). For ranking data, they were considered by Chan and

[1] *Utility* is a term that has a long history in philosophy and economics to explain the phenomenon of value. Most frequently, it has been given the connotation of "desiredness," which is also the meaning that we apply here (see also, Kahneman, 2003).

Bentler (1998) and Maydeu-Olivares (1999). However, these authors used a specialized estimation and testing approach that is cumbersome to use in applied work. The SEM framework presented here is less complex and considerably more user friendly. Equally important, it allows us to consider more general factor-analytic models than those considered in these references.

In sum, this article extends the current literature in three important ways: First, we use the same estimation framework for the analysis of paired comparisons and rankings. Only minor modifications are necessary to accommodate these two different types of choice data within a SEM approach. The joint consideration of both data types highlights their inherent similarities. Second, we present a comprehensive treatment of parameter interpretation and model selection. Despite their importance, these topics have been neglected in the literature. Third, with little loss in statistical efficiency, the proposed SEM approach can accommodate models with a much larger number of choice alternatives than currently feasible under alternative estimation methods. Estimation is fast, and even complex models can be estimated within seconds. Also, goodness-of-fit statistics with accurate p values can be obtained even when the data are sparse (Maydeu-Olivares, 2003a), overcoming a problem with previous approaches.

The article is structured into nine sections. In the first section (Binary Coding of Comparative Judgments), we describe how to code the observed paired comparisons and rankings in a suitable form for subsequent SEM analyses. In the second section (Thurstonian Choice Models for Rankings), we review three Thurstonian ranking models (Cases III and V and the unrestricted model) using a matrix formulation and discuss the restrictions that are imposed by these models on the thresholds and tetrachoric correlations that are estimated from the data. In the third section (SEM of Ranking Data), we describe how to estimate Thurstonian ranking models using Mplus (L. Muthén & Muthén, 2001). In the fourth section (A Ranking Application: Modeling Career Preferences Among Spanish Undergraduates), we analyze a ranking data set about career preferences of psychology undergraduate students using this software package. In the fifth section (Thurstonian Paired-Comparison Models), we describe an extension, originally proposed by Takane (1987), to accommodate paired-comparison data within Thurstonian choice models. In this section, we also discuss (following Tsai, 2003) the model interpretation problems that occur in Thurstonian modeling because of the existence of equivalent covariance structures. The sixth section (A Paired-Comparison Application: Modeling Preferences for Compact Cars) presents a paired-comparison application. As in the ranking example, we use Mplus to analyze pairwise preferences for compact cars. The seventh section (Thurstonian Factor Models for Paired Comparisons and Ranking Data) discusses Thurstonian factor models for both paired-comparisons and ranking data, including models with structured means. In this section, we also show how these models can help in overcoming the interpretation problems discussed in the fifth section, and we illustrate the approach by fitting a factor model with structured means to the car data in the eighth section (The Paired-Comparisons Application Revisited: Modeling Preferences for Compact Cars Using a Factor Model). The final section (Conclusion) summarizes the main points of this article, discusses further extensions, and provides a set of guidelines for choosing among the different Thurstonian ranking and paired-comparison models considered in this article.

Binary Coding of Comparative Judgments

This section discusses how to code the observed paired-comparison and ranking data in a form suitable for estimating Thurstonian choice models when standard software packages are used for SEM.

Paired Comparisons

In a paired-comparison task, respondents are presented with pairs selected from an item set and are instructed to select the more preferred item from each pair. With n items, there are $\tilde{n} = [n(n-1)]/2$ pairs of items. For instance, $\tilde{n} =$ 6 pairs can be constructed with $n = 4$ items. If the $n = 4$ items are labeled {A, B, C, D}, the following pairs can be constructed: {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}. A presentation of the pairs in this order may result in strong carry-over effects. To control for this effect, it is important to randomize the presentation order of the pairs as well as the order of items within each pair (Bock & Jones, 1968). The observed paired-comparison responses can be coded as follows:

$$y_l = \begin{cases} 1 & \text{if item } i \text{ is preferred over item } k \\ 0 & \text{if item } k \text{ is preferred over item } i \end{cases}, \tag{1}$$

where l indicates the pair $\{i, k\}$. Thus, we obtain a pattern of $\tilde{n}$ binary responses from each respondent.

Ranking Data

In a ranking task, all choice alternatives are presented at once (in a randomized order), and respondents are asked to either rank or order them. A ranking is obtained when the choice alternatives are presented, and the respondent is asked to assign ranking positions to each choice alternative. For instance, for the $n = 4$ choice alternatives {A, B, C, D}, a ranking task consists of filling the following blanks with numbers from 1 (*most preferred*) to 4 (*least preferred*).

Ranking

A	B	C	D
—	—	—	—

Alternatively, the respondents may be asked to order the choice alternatives. An ordering is obtained when the ranking positions are presented, and the respondent is asked to assign the choice alternatives to them. An ordering task for our previous example consists out of filling the following blanks with the given choice alternatives {A, B, C, D}.

Ordering			
1st	2nd	3rd	4th
—	—	—	—

Any ordering or ranking of n choice alternatives can be coded equivalently using $\tilde{n}$ paired comparisons. Thus, to continue our example, the ordering {A, D, B, C} (or its equivalent ranking) can be coded with the following paired comparisons:

Ranking				Ordering			
A	B	C	D	1st	2nd	3rd	4th
1	3	4	2	A	D	B	C

Pairwise Outcomes					
{A, B}	{A, C}	{A, D}	{B, C}	{B, D}	{C, D}
1	1	1	1	0	0

The converse is not true because not all paired comparison outcomes that can be observed can be transformed into rankings or orderings. Intransitive paired comparisons cannot be converted to an ordering of the choice alternatives. In the following, we analyze rankings and orderings after transforming them to paired comparisons. Although both paired comparisons and rankings can be transformed into binary variables, we show later that the two data types have different covariance structures that need to be taken into account in a data analysis.

Thurstonian Choice Models for Rankings

This section discusses the Thurstonian response model for ranking data. We present a matrix formulation of the ranking model and explain the tetrachoric correlations and threshold parameters that are implied by it. We also describe some basic covariance structures that can be used (Case V, Case III, and the unrestricted model). For these covariance structures, we explain the identification constraints necessary to estimate model parameters.

Response Model for Rankings

Consider a random sample of respondents sampled from the population of interest. According to Thurstone (1927), when a respondent is confronted with a ranking task, each of the n items to be ranked elicits a utility. This utility is unobserved and varies across respondents. A latent random variable may be associated with each of the items to represent individual differences in the utilities in the population of interest. We shall denote by t_i the latent random variable associated with the utilities for item i. Therefore, in Thurstone's model there are exactly n latent variables when modeling n choice alternatives.

A respondent prefers item i over item k, if for that respondent, her or his utility for item i is larger than for item k, and consequently ranks item i before item k. Otherwise, the respondent prefers item k over item i and ranks item k before item i. The former outcome is coded as 1, and the latter is coded as 0. That is,

$$y_l = \begin{cases} 1 \text{ if } t_i \geq t_k \\ 0 \text{ if } t_i < t_k \end{cases}, \tag{2}$$

where the equality sign is arbitrary, as the latent utilities are assumed to be continuous, and thus, by definition, two latent utilities can never take on exactly the same value. The response process (see Equation 2) can be alternatively described by the computation of differences between the latent utilities. Let

$$y_l^* = t_i - t_k \tag{3}$$

be a variable[2] that represents the difference between choice alternatives i and k. Because t_i and t_k are not observed, y_l^* is also unobserved. Then, the relationship between the observed comparative response y_l and the latent comparative response y_l^* is

$$y_l = \begin{cases} 1 \text{ if } y_l^* \geq 0 \\ 0 \text{ if } y_l^* < 0 \end{cases}. \tag{4}$$

It is convenient to write the response process in matrix form. Let $\mathbf{t}$ be the $n \times 1$ vector of latent utilities and $\mathbf{y}^*$ be the $\tilde{n} \times 1$ vector of latent difference responses, where $\tilde{n} = [n(n-1)]/2$. Then we can write the set of $\tilde{n}$ equations (see Equation 3) as

$$\mathbf{y}^* = \mathbf{A}\mathbf{t}, \tag{5}$$

where $\mathbf{A}$ is an $\tilde{n} \times n$ design matrix. Each column of $\mathbf{A}$ corresponds to one of the n choice alternatives, and each row of $\mathbf{A}$ corresponds to one of the $\tilde{n}$ paired comparisons. For example, when $n = 2$, $\mathbf{A} = (1 \quad -1)$, whereas when $n = 3$ and $n = 4$,

$$\underset{n=3}{\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}}, \text{ and } \underset{n=4}{\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}}, \tag{6}$$

[2] Notice that there is no error term in this equation. The latent comparative responses are determined solely by the latent utilities.

respectively. For instance, in the design matrix for $n = 4$ choice alternatives, each column corresponds to one of the four choice alternatives {A, B, C, D}. The corresponding rows give the six possible paired comparisons {A, B}, {A, C}, {A, D}, {B, C}, {B, D}, {C, D}. Row 4 indicates that B is compared with C; and Row 6 indicates that C is compared with D.

Thurstone's model assumes that each of the latent variables in the vector of latent utilities **t** is normally distributed in the population of respondents. Thus, we can write

$$\mathbf{t} \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t), \tag{7}$$

where $\boldsymbol{\mu}_t$ contains the means of the n latent utilities in the population of respondents. The diagonal elements of $\boldsymbol{\Sigma}_t$ contain the variances, and the off-diagonal elements contain the covariances of the n latent utilities in the population of respondents.

The latent difference responses **y*** are a linear combination of the latent utilities **t**, as indicated in Equation 5. Given the assumption of multivariate normality for **t**, the mean and covariance structure of **y*** is

$$\boldsymbol{\mu}_{y^*} = \mathbf{A}\boldsymbol{\mu}_t, \text{ and } \boldsymbol{\Sigma}_{y^*} = \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}'. \tag{8}$$

$\boldsymbol{\mu}_{y^*}$ contains the means of the latent difference responses. The diagonal elements of $\boldsymbol{\Sigma}_{y^*}$ contain the variances, and the off-diagonal elements the covariances of the latent difference responses. Moreover, because the latent utilities **t** are assumed to be multivariate normally distributed, the latent difference responses **y*** also follow a multivariate normal distribution because they can be expressed as a linear function of the latent utilities. The latent difference responses are linked to the observed comparative responses **y** via the threshold relationship expressed in Equation 4. Thus, only the mean and covariance parameters of the latent difference responses, $\boldsymbol{\mu}_{y^*}$ and $\boldsymbol{\Sigma}_{y^*}$, are related directly to the observed choice data. The mean vector $\boldsymbol{\mu}_t$ and the corresponding covariance matrix $\boldsymbol{\Sigma}_t$ of the latent utilities have to be derived from Equation 8.

Figure 1 depicts graphically as an SEM model the covariance structure represented in Equation 8 for a ranking model with four choice alternatives. With $n = 4$ choice alternatives, there are four latent utilities **t** and $\tilde{n} = 6$ comparisons **y***. Consequently, in Figure 1 there are six "observed" variables and four latent variables. The relationship between the "observed" and latent variables is given by Equation 3. Yet, the "observed" variables **y*** are not actually observed, only their dichotomizations **y** are observed (the actual choices), where the relationship between the **y*** and **y** variables is given by Equation 4. Note in this figure that the relationship between the **y*** and **t** variables is deterministic; the residual variances for the **y*** variables are zero. Also, no structure has been imposed on the covariance matrix of the latent variables **t**. Some constraints on the parameters of the model depicted in Figure 1 are needed to identify it. Also, different structures can be imposed on $\boldsymbol{\Sigma}_t$

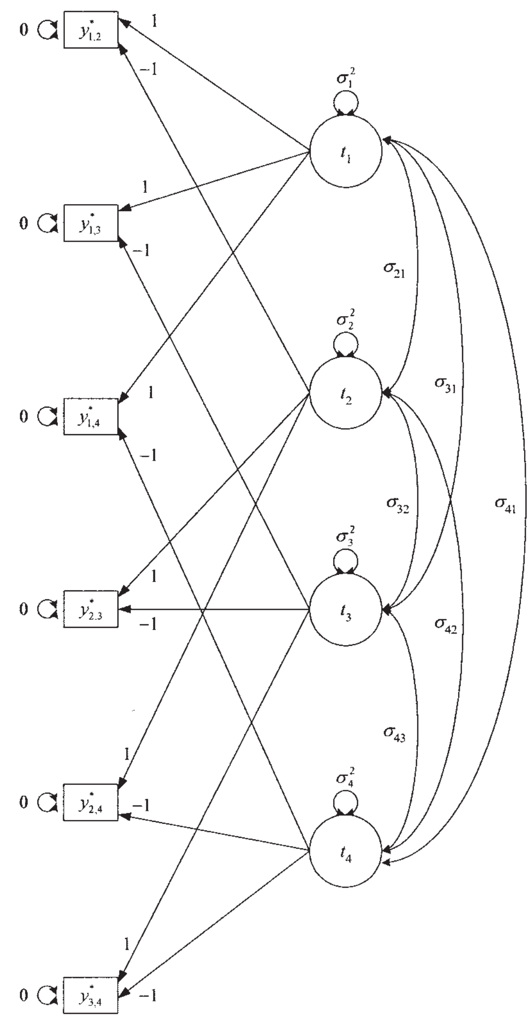

Figure 1. Covariance structure of an unrestricted Thurstonian choice model for ranking data for $n = 4$ choice alternatives. Some identification restrictions are needed to identify the model. These are described in the text.

to test hypotheses about possible sources of individual differences in the evaluation of the utilities **t**. We return to these issues below. Before doing so, we have to consider how to embed in the model the fact that the only observed data in the Thurstonian choice model are categorical.

Thresholds and Tetrachoric Correlations Implied by the Ranking Model

In SEM, structured multivariate normal distributions that have been dichotomized according to a set of thresholds are estimated in several stages (B. Muthén, 1978). First, the thresholds and the tetrachoric correlations[3] among the un-

[3] When two normally-distributed variables are dichotomized at some threshold values, the correlation between the two underlying normal variables (i.e., before the discretization) is called a tetrachoric correlation.

derlying normal variables are estimated. Then, the parameters of interest are estimated from the threshold and tetrachoric correlations. In this section we provide the restrictions imposed by the Thurstonian choice model on the population thresholds and tetrachoric correlations. These are needed to estimate the model, as the only observed data in the Thurstonian choice model are categorical.

To obtain the thresholds and tetrachoric correlations implied by this model it is necessary to standardize the latent difference responses $\mathbf{y}^*$. The standardization is performed by computing $\mathbf{y}^* - \boldsymbol{\mu}_{y^*}$ and then dividing the result by the corresponding standard deviation. In matrix form, the standardization can be written as $\mathbf{z}^* = \mathbf{D}(\mathbf{y}^* - \boldsymbol{\mu}_{y^*})$. $\mathbf{z}^*$ are the standardized latent difference responses, and $\mathbf{D} = [\text{Diag}(\boldsymbol{\Sigma}_{y^*})]^{-1/2}$ is a diagonal matrix with the reciprocals of the standard deviations of $\mathbf{y}^*$ in the diagonal. The standardized latent difference responses are multivariate normal with a $\mathbf{0}$ mean vector and correlation matrix $\mathbf{P}_{z^*}$, where

$$\mathbf{P}_{z^*} = \mathbf{D}(\boldsymbol{\Sigma}_{y^*})\mathbf{D} = \mathbf{D}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}')\mathbf{D}. \tag{9}$$

The standardized latent difference responses $\mathbf{z}^*$ are related to the observed comparative responses $\mathbf{y}$ via the threshold relationship

$$y_l = \begin{cases} 1 \text{ if } z_l^* \geq \tau_l \\ 0 \text{ if } z_l^* < \tau_l \end{cases}. \tag{10}$$

Because there are $\tilde{n}$ observed comparative responses, there are $\tilde{n}$ thresholds τ_l. Collecting all thresholds τ_l into an $\tilde{n} \times 1$ vector, $\boldsymbol{\tau}$, we show in Appendix A that this vector of thresholds has the following structure:

$$\boldsymbol{\tau} = -\mathbf{D}\mathbf{A}\boldsymbol{\mu}_t. \tag{11}$$

Finally, the $\tilde{n} \times \tilde{n}$ matrix $\mathbf{P}_{z^*}$ in Equation 9 is a matrix of so-called tetrachoric correlations because the latent comparative responses $\mathbf{z}^*$, which have been dichotomized according to Equation 10, are multivariate normal.

In summary, the estimation process proceeds as follows. First, the ranking data are transformed into binary paired comparisons. Then, the thresholds and tetrachoric correlations of the paired comparisons are computed. Finally, the parameters of interest, $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$, are estimated from the estimated thresholds and correlations with Equations 9 and 11.

Because rankings give rise to only a subset of all possible paired comparison data (no intransitivities can be observed), an adjustment is needed to the number of degrees of freedom of the model for ranked data. When the analyzed binary data are obtained by transforming ranking patterns, there will be

$$r = \frac{n(n-1)(n-2)}{6} \tag{12}$$

redundancies among the thresholds and tetrachoric correlations estimated from the binary variables (Maydeu-Olivares, 1999). For this reason, the correct number of degrees of freedom, when the parameters of the model are being estimated for ranking data, is the number of thresholds plus the number of tetrachoric correlations minus the number of estimated parameters (say, q) minus r, the number of redundancies; that is, $df = [\tilde{n}(\tilde{n} + 1)]/2 - q - r$.

Covariance Structures

The covariance matrix of the latent utilities $\boldsymbol{\Sigma}_t$ can be restricted in various ways to test hypotheses about possible sources of individual differences in the evaluation of the latent utilities. The classical covariance structures that were proposed by Thurstone (1927) include (a) the unrestricted model, in which the mean vector $\boldsymbol{\mu}_t$ and the covariance matrix $\boldsymbol{\Sigma}_t$ are unrestricted; (b) the Case III model, which assumes that the latent utilities are uncorrelated[4] (i.e., $\boldsymbol{\Sigma}_t$ is specified to be diagonal but otherwise unrestricted); and (c) the Case V model in which, in addition, the latent utilities are assumed to have a common variance (i.e., $\boldsymbol{\Sigma}_t = \sigma^2\mathbf{I}$).

Identification Constraints

Some identification constraints are needed for one to estimate $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$, the parameters of the Thurstonian choice model. Here we provide identification constraints for the three classical Thurstonian covariance structures. The identification constraints provided here were selected because they are convenient and easy to implement. However, they are not unique, and other identification constraints can be specified that yield equivalent model fits. The issue of equivalent models and parameter interpretation in Thurstonian choice models is deferred to the section on model identification for paired-comparison models.

Unrestricted Thurstonian model. The unrestricted Thurstonian model requires three identification constraints: (a) fix one of the item means, say $\mu_n = 0$; (b) fix all the covariances involving the last latent utility to 0; and (c) fix the variance of the first and last latent utilities to 1. In the next section we present an empirical study in which career preferences among four broad psychology areas (academic, clinical, educational, and industrial) were investigated with a ranking task. For this example, $n = 4$, and $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ are to be specified as

$$\boldsymbol{\mu}_t = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ 0^* \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_t = \begin{pmatrix} 1^* & \sigma_{21} & \sigma_{31} & 0^* \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & 0^* \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & 0^* \\ 0^* & 0^* & 0^* & 1^* \end{pmatrix}. \tag{13}$$

[4] Thurstone's choice model assumes that latent utilities are normally distributed. Therefore, uncorrelated utilities imply independent utilities, and we use both terms interchangeably throughout the manuscript.

In Equation 13, the parameters fixed for identification are marked with an asterisk. That is, for identification purposes (a) the mean of the latent utilities for industrial psychology is set to 0, (b) the variances of the latent utilities for academic and industrial psychology are set to 1, and (c) the covariances between the utilities for industrial psychology and the utilities for all remaining areas are set to 0.

Thurstone's Case III model. Thurstone's Case III model is identified by fixing one of the means of the latent utilities, say $\mu_n = 0$, and fixing one of the variances of the latent utilities to 1, say $\sigma_n^2 = 1$. Thus, for the career's example, $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ are to be specified as

$$\boldsymbol{\mu}_t = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ 0^* \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_t = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & 1^* \end{pmatrix}. \quad (14)$$

Thurstone's Case V model. Thurstone's Case V is identified by fixing one of the means, say $\mu_n = 0$, and fixing the common variance of the latent utilities to 1 (i.e., $\sigma^2 = 1$). Again, for the careers example, $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ are to be specified in this model as

$$\boldsymbol{\mu}_t = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ 0^* \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_t = \begin{pmatrix} 1^* & 0 & 0 & 0 \\ 0 & 1^* & 0 & 0 \\ 0 & 0 & 1^* & 0 \\ 0 & 0 & 0 & 1^* \end{pmatrix}. \quad (15)$$

SEM of Ranking Data

In this section, we first describe how to embed Thurstonian ranking models within an SEM framework. Then, we provide details on how estimation is performed with a popular SEM program, Mplus (L. Muthén & Muthén, 2001).

Thurstonian Ranking Models as SEM Models

For modeling the linear relations of $\tilde{n}$ indicators $\mathbf{y}^*$ on p latent variables $\boldsymbol{\eta}$, Lisrel (Jöreskog & Sörbom, 2001) and Mplus use the following model when there are no exogenous variables:

$$\mathbf{y}^* = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad (16)$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}. \quad (17)$$

Here, $\boldsymbol{\nu}$ is an $\tilde{n}$-dimensional vector containing the intercepts for the measurement equation (see Equation 16), and $\boldsymbol{\alpha}$ is a p-dimensional vector of intercepts for the structural equation (see Equation 17). $\boldsymbol{\Lambda}$ is an $\tilde{n} \times p$ matrix of factor loadings, and $\mathbf{B}$ is a $p \times p$ parameter matrix of slopes for regressions of latent variables on other latent variables. $\boldsymbol{\varepsilon}$ is the $\tilde{n}$-dimensional vector of residuals for the measurement equation, and $\boldsymbol{\zeta}$ is a p-dimensional vector of residuals for the structural equation. In this model, it is assumed that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}$ have mean 0 and that they are mutually uncorrelated. We denote the covariance matrix of $\boldsymbol{\varepsilon}$ by $\boldsymbol{\Theta}$ (measurement error variances), and the covariance matrix of $\boldsymbol{\zeta}$ by $\boldsymbol{\Psi}$.

Then, the mean and covariance matrices of $\mathbf{y}^*$ implied by this SEM model are

$$\boldsymbol{\mu}_{y^*} = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha},$$

and (18)

$$\boldsymbol{\Sigma}_{y^*} = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1\prime}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}.$$

The Thurstonian choice model for ranking data is a special case of Equations 16 and 17, where $p = n$, $\boldsymbol{\eta} = \mathbf{t}$, and $\boldsymbol{\varepsilon} = \mathbf{0}$. As a result, the mean and covariance structures implied by Thurstonian models for rankings—given in Equation 8—is a special case of Equation 18, where $\boldsymbol{\nu} = \mathbf{0}$, $\boldsymbol{\Lambda} = \mathbf{A}$, $\mathbf{B} = \mathbf{I}$, $\boldsymbol{\alpha} = \boldsymbol{\mu}_t$, $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_t$, and $\boldsymbol{\Theta} = \mathbf{0}$. Imposing these constraints on the matrices of Equation 18 yields Thurstonian models for ranking data. One may estimate the three basic Thurstonian covariance structure models described previously (unrestricted, Case III, and Case V models) by imposing suitable constraints on $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_t$.

SEM Estimation of Models With Binary Observed Variables

SEMs with categorical indicators are estimated in similar ways in Lisrel, Mplus, and EQS (Bentler, 1995). First, the thresholds and tetrachoric correlations are estimated. In a second stage, the model parameters are estimated from the estimated thresholds and tetrachoric correlations. When a single population is involved, Mplus, but not current versions[5] of Lisrel and EQS, can estimate models with categorical indicators that have mean or threshold structures (such as the ones in Thurstonian choice models). Therefore, we apply Mplus in subsequent sections of this paper.

Mplus performs the second stage of the estimation by minimizing

$$F = [\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta})]'\hat{\mathbf{W}}[\hat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}(\boldsymbol{\theta})], \quad (19)$$

where $\hat{\boldsymbol{\kappa}}$ denotes the set of all estimated thresholds and tetrachoric correlations collected in a vector, and $\boldsymbol{\kappa}(\boldsymbol{\theta})$ denotes the restrictions imposed on the thresholds and tetrachoric correlations by the model parameters $\boldsymbol{\theta}$. For Thurstonian ranking models, $\boldsymbol{\theta}$ is the set of parameters estimated in $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$. Thus, $\boldsymbol{\kappa}(\boldsymbol{\theta})$ contains the set of population thresholds from Equation 11 and the set of tetrachoric correlations below the diagonal of Equation 9 expressed as a function of the elements of $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$.

Let $\hat{\boldsymbol{\Xi}}$ be the estimated asymptotic covariance matrix of

[5] At the time of this writing, the current versions of Mplus, Lisrel, and EQS are 3.12, 8.7, and 6.1, respectively.

the sample thresholds and tetrachoric correlations. Two choices for $\hat{\mathbf{W}}$ in Equation 19 are $\hat{\mathbf{W}} = \hat{\boldsymbol{\Xi}}^{-1}$ (weighted least squares [WLS]; B. Muthén, 1978, 1984), and $\hat{\mathbf{W}} = [\mathrm{Diag}(\hat{\boldsymbol{\Xi}})]^{-1}$ (diagonally weighted least squares [DWLS]; B. Muthén, du Toit, & Spisic, 1997). Consistent and asymptotically normal parameter estimates (as well as standard errors) can be obtained for both estimation methods.

Let N denote the sample size and $T = N\hat{F}$ denote the usual chi-square statistic. Restrictions imposed by a model on the thresholds and tetrachoric correlations can be tested as follows: For WLS estimation, T follows a chi-square distribution in large samples. In the case of DWLS estimation, T is not asymptotically distributed as a chi-square variable. However, when DWLS is employed, a goodness-of-fit test can be obtained, following Satorra and Bentler (1994) and B. Muthén (1993), by scaling T by its asymptotic mean or by its asymptotic mean and variance. We denote these two test statistics by T_M and T_{MV}, respectively. As shown by B. Muthén (1993), larger samples are needed to obtain adequate parameter estimates, standard errors, and goodness of fit tests with WLS than with DWLS. Thus, DWLS seems to be preferable in practical applications and is used in this article.

For the comparison of nested models when DWLS is employed, the Satorra–Bentler mean adjusted statistic, T_M-(dif), can be used. This test is computed as described in Satorra and Bentler (2001). A test for comparing nested models with the T_{MV} statistic is yet to be developed. Therefore, in our applications, which involve comparing nested models, T_M, rather than T_{MV}, will be used.

A Ranking Application: Modeling Career Preferences Among Spanish Psychology Undergraduates

A Spanish university wished to investigate career preferences among its undergraduate psychology students. A pilot study was performed in which a sample of 57 psychology sophomores were asked to express their preferences for four broad psychology career areas (A = Academic, C = Clinical, E = Educational, and I = Industrial) using a ranking task.

Using Mplus, we fit three ranking models to the data: $\boldsymbol{\Sigma}_t$ unrestricted, Case III, and Case V. The goodness-of-fit results from the Satorra–Bentler scaled statistic T_M are presented in Table 1. The degrees of freedom in the tests reported in this table have been adjusted by use of Equation 12. For instance, for the unrestricted model, Mplus reports $T_M = 16.91$, $df = 13$, $p = .53$. However, the correct number of degrees of freedom is $13 - 4 = 9$, which yields a p value of .08.

As can be seen in Table 1, all three models adequately reproduce the estimated thresholds and tetrachoric correlations. However, they differ in the number of estimated parameters. Especially in view of the small sample size, it is desirable to identify the more parsimonious representation.

Table 1
Goodness-of-Fit Tests for Some Basic Models Applied to the Career Ranking Data

Model and comparison	T	T_M	df	p
Overall fit				
Model				
Unrestricted	6.26	15.23	9	.08
Case III	13.76	22.93	11	.02
Case V	15.42	16.91	14	.26
Nested tests				
Comparison				
Unrestricted vs. Case III	7.50	5.17	2	.08
Unrestricted vs. Case V	9.16	5.05	5	.41
Case III vs. Case V	1.69	0.81	3	.85

Note. Diagonally weighted least squares estimation; $T = N\hat{F}$; T_M is the Satorra–Bentler mean adjusted statistic T; $p > .05$ indicates acceptable overall fit to the data. In the unrestricted model, $\boldsymbol{\Sigma}_t$ is unrestricted; in the Case III model, $\boldsymbol{\Sigma}_t$ is specified to be diagonal but otherwise unrestricted; and in the Case V model, $\boldsymbol{\Sigma}_t = \sigma^2\mathbf{I}$.

To choose among these models, we performed nested tests using the Satorra-Bentler mean adjusted statistic, T_M-(dif). The results of these nested tests are included in Table 1. They suggest that the fit of the Case V model cannot be improved upon by the additional parameters estimated under Case III or the unrestricted model specifications. Therefore, a Case V model yields the most parsimonious fit of the data.[6]

The Case V model assumes that the latent utilities are uncorrelated and that their variances are equal. The common variance cannot be identified, and it is therefore set equal to 1. Thus, in this model the only estimated parameters are the means of the latent utilities. The estimated means (with standard errors in parentheses) are $\hat{\mu}_{\mathrm{A}} = -1.01$ (0.23), $\hat{\mu}_{\mathrm{C}} = 0.99$ (0.22), and $\hat{\mu}_{\mathrm{E}} = 0.30$ (0.23), where $\mu_{\mathrm{I}} = 0$ for identification purposes.

In conclusion, these analyses reveal that preferences for the different academic areas appear to be mutually independent, with equal between-judge variances for each of the four career areas. The most preferred career area is clinical, followed by educational and industrial, which do not appear to differ significantly from each other. The least preferred career area is academic. Because preferences appear to be unrelated for the four areas, there is no particular academic area that can serve as a substitute for other areas.

We return to this example in the next section, which discusses equivalent covariance structures, because—as is shown—all that can be said is that the available information

[6] A larger sample size would provide greater power to distinguish the three models.

in the data is consistent with the interpretation of mutually independent preferences but that other interpretations are consistent with the data as well.

Thurstonian Paired-Comparison Models

This section presents the Thurstonian response model for paired comparisons as a straightforward extension of the ranking model. We discuss the thresholds and tetrachoric correlations implied by the paired-comparison model, and we describe the identification constraints needed to identify its parameters. The section concludes with a discussion of equivalent choice models in analyzing paired-comparisons and ranking data.

Response Model for Paired Comparisons

In a paired-comparisons task, respondents need not be consistent in their pairwise choices, yielding intransitive patterns. Inconsistent pairwise responses can be accounted for by adding an error term, e_l, to the difference judgment (see Equation 3),

$$y_l^* = t_i - t_{i'} + e_l. \tag{20}$$

This random error, e_l, is assumed to be normally distributed with 0 mean, variance ω_l^2, uncorrelated across pairs, and uncorrelated with the latent utilities. The error term accounts for intransitive responses by reversing the sign of the difference between the preference responses t_i and $t_{i'}$.

As in the case of ranking data, the relationship between the observed comparative response y_l and the latent difference judgment y_l^* is given by Equation 4. Similarly, the response process can be written in matrix form as

$$\mathbf{y}^* = \mathbf{At} + \mathbf{e}, \tag{21}$$

where $\mathbf{e}$ is the $\tilde{n} \times 1$ vector of random errors with covariance matrix $\mathbf{\Omega}^2$, which is a diagonal matrix with elements $\omega_1^2, \ldots, \omega_{\tilde{n}}^2$.

Because the latent utilities $\mathbf{t}$ and the random errors $\mathbf{e}$ are assumed to be normally distributed, the latent difference responses $\mathbf{y}^*$ are normally distributed. Their mean vector and covariance matrix are

$$\boldsymbol{\mu}_{y^*} = \mathbf{A}\boldsymbol{\mu}_t \text{ and } \mathbf{\Sigma}_{y^*} = \mathbf{A}\mathbf{\Sigma}_t\mathbf{A}' + \mathbf{\Omega}^2. \tag{22}$$

Equation 22 follows the form of Equation 8 for rankings but adds the covariance matrix of the random errors $\mathbf{\Omega}^2$. Clearly, the smaller the elements of the error covariance matrix $\mathbf{\Omega}^2$ are, the more consistent the respondents are in evaluating the choice alternatives. In the extreme case, when all the elements of $\mathbf{\Omega}^2$ are 0, the paired-comparison data are effectively rankings, and no intransitivities would be observed in the data. A more restricted model that is often found to be useful in applications involves setting the error variances to be equal for all pairs (i.e., $\mathbf{\Omega}^2 = \omega^2\mathbf{I}$). This restriction implies that the number of intransitivities is approximately equal for all pairs, provided the mean differences are small.

Thresholds and Tetrachoric Correlations Implied by the Models

In this section we provide the restrictions imposed by the Thurstonian choice model on the population thresholds and tetrachoric correlations. These are needed to estimate the model, as the paired comparisons are dichotomous variables. To obtain the thresholds and tetrachoric correlations for the paired comparison model, it is necessary to standardize the latent comparative responses $\mathbf{y}^*$. As in the case of rankings, the standardization is performed in matrix form as $\mathbf{z}^* = \mathbf{D}(\mathbf{y}^* - \boldsymbol{\mu}_{y^*})$, where $\mathbf{D} = [\mathrm{Diag}(\mathbf{\Sigma}_{y^*})]^{-1/2}$ is a diagonal matrix with the reciprocals of the standard deviations of $\mathbf{y}^*$—given in Equation 22—and $\mathbf{z}^*$ are the standardized latent difference responses.

These standardized latent difference responses are multivariate normal with mean 0 and correlation matrix

$$\mathbf{P}_{z^*} = \mathbf{D}(\mathbf{\Sigma}_{y^*})\mathbf{D} = \mathbf{D}(\mathbf{A}\mathbf{\Sigma}_t\mathbf{A}' + \mathbf{\Omega}^2)\mathbf{D}. \tag{23}$$

Again, this equation parallels Equation 9 for rankings but adds the error covariance matrix. Thus, this formula provides the restrictions assumed by the Thurstonian choice model on the tetrachoric correlations estimated from the observed paired comparisons. The standardized latent difference responses $\mathbf{z}^*$ are related to the observed comparative responses $\mathbf{y}$ via the threshold relationship presented in Equation 10, and the structure imposed by the model on the thresholds $\boldsymbol{\tau}$ in Equation 10 is given by Equation 11.

In summary, paired-comparison responses may deviate from rankings by containing intransitivities. To allow for such inconsistencies, random errors are introduced in the response model. Because the random errors have a mean of 0, they do not change the threshold structure given by Equation 11. However, the structure of the tetrachoric correlations is affected by this additional source of variation in the data, as can be seen by comparing Equation 23 for paired comparisons and Equation 9 for rankings. This comparison also reveals the strong structural similarities between rankings and paired comparisons. Under a Thurstonian framework, paired comparisons provide the same structural representation of the choice alternatives' utilities as rankings. From this perspective, rankings seem preferable to paired comparisons because they are easier to administer and less time consuming in experimental work. However, the estimation of ranking and paired comparison models is the same. For four or more choice alternatives, there are no differences in the identification constraints needed to estimate the unrestricted, Case III and Case V covariance structures for paired-comparisons and ranking data. Note that the mean and covariance structures implied by Thurstonian models for paired comparisons can be specified within the SEM framework of Equations 17 and 18 in the same way that we specified models for rankings. The only difference is that for paired comparisons, $\mathbf{\Theta} = \mathbf{\Omega}^2$, whereas for rankings, $\mathbf{\Theta} = \mathbf{0}$.

Equivalent Covariance Structures and Model Interpretation

For any SEM model there are a number of equivalent models that cannot be distinguished in terms of overall fit. However, each of these models may have a different substantive interpretation (MacCallum, Wegener, Uchino, & Fabrigar, 1993). Over the years, a set of rules has been developed that can be applied to find equivalent SEM models (e.g., Lee & Hershberger, 1990; Stelzl, 1986). For Thurstonian choice models, Tsai (2003) has provided a rule that can be applied to find the full set of models that are equivalent to a given estimated model. Suppose a Thurstonian model for paired comparisons has been estimated. For this specific model, we denote the covariance matrix of the latent utilities as $\mathbf{\Sigma}_1$ and the error covariance matrix as $\mathbf{\Omega}_1^2$. Both $\mathbf{\Sigma}_1$ and $\mathbf{\Omega}_1^2$ must be positive definite. Then, any other model with $\mathbf{\Sigma}_2$ and $\mathbf{\Omega}_2^2$ of the form

$$\mathbf{\Sigma}_2 = c\mathbf{\Sigma}_1 + \mathbf{d1}' + \mathbf{1d}', \text{ and } \mathbf{\Omega}_2^2 = c\mathbf{\Omega}_1^2, \tag{24}$$

will be equivalent to the estimated model (Tsai, 2003, Corollary 1). In Equation 24, c is a positive constant, and $\mathbf{d}$ is an $n \times 1$ vector of constants. These constants are arbitrary, but $\mathbf{\Sigma}_2$ and $\mathbf{\Omega}_2^2$ must be positive definite. Equation 24 also can be used to find equivalent models for ranking data. In this case $\mathbf{\Omega}_1^2 = \mathbf{0}$.

To illustrate our discussion, consider the Case V model that was selected as the most parsimonious model for the career ranking data. One can obtain an equivalent model by applying Equation 24 to the estimated $\mathbf{\Sigma}_t$ matrix (which was an identity matrix). For instance, arbitrarily using $c = 2$ and $\mathbf{d} = (-0.2, 0.5, -0.2, 0.6)'$, we find that a new model with

$$\mathbf{\Sigma}_t^* = c\mathbf{\Sigma}_t + \mathbf{d1}' + \mathbf{1}'\mathbf{d} = 2 \times \mathbf{I} + \begin{pmatrix} -0.2 \\ 0.5 \\ -0.2 \\ 0.6 \end{pmatrix}(1, 1, 1, 1) + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}(-0.2, 0.5, -0.2, 0.6) = \begin{pmatrix} 1.6 & 0.3 & -0.4 & 0.4 \\ 0.3 & 3 & 0.3 & 1.1 \\ -0.4 & 0.3 & 1.6 & 0.4 \\ 0.4 & 1.1 & 0.4 & 3.2 \end{pmatrix} \tag{25}$$

is equivalent to the Case V model that was estimated in the career data application; this new model has the same correlation structure as in Equation 9. Additional equivalent models can be found by use of different constants c and $\mathbf{d}$.

The existence of equivalent models affects, in important ways, the substantive interpretation of Thurstonian choice models. On the one hand, it creates interpretational problems for Case III and Case V models. As can be seen in Equation 25, the latent utilities in $\mathbf{\Sigma}_t^*$ are all correlated. Thus, it cannot be claimed in the reported application that career preferences are uncorrelated. All that can be stated is that this model is the most parsimonious representation for these data and that, according to this model, the data are consistent with the hypothesis of uncorrelated career preferences. For this interpretation to be validated, additional data must be gathered. See Böckenholt (2004) for a detailed discussion of this issue.

On the other hand, the existence of equivalent models also creates interpretational problems for the unrestricted model. The size of the estimated covariances in this model cannot be interpreted, as an equivalent model with the same goodness of fit but different estimates can always be found. The estimated covariances can only be interpreted relative to $\sigma_{1n} = 0$ (the covariance between the first and last alternatives). This covariance sets the scale for the estimated covariances. Positive covariances are to be interpreted as "stronger degree of association than σ_{1n}," whereas negative covariances are to be interpreted as "weaker degree of association than σ_{1n}." To see this, we transform the covariance matrix into a matrix of squared distances $\mathbf{\Delta}$ between the latent utilities. The elements of this $\mathbf{\Delta}$ matrix, δ_{ik}^2 (the squared distance between choice alternatives i and k), can be obtained with the formula (Mardia, Kent, & Bibby, 1979)

$$\delta_{ik}^2 = \sigma_i^2 + \sigma_k^2 - 2\sigma_{ik}, \tag{26}$$

where σ_i^2 and σ_k^2 are the variances, and σ_{ik} is the covariance of the two latent utilities. Let c be the arbitrary positive constant used to obtain an equivalent model with Equation 24. The matrix of relative distances $\mathbf{\Delta}/c$ is unique for the full set of equivalent models defined by Equation 24. Therefore, relative distances can be interpreted uniquely, and covariances should only be interpreted in relative terms. As can be seen in Equation 26, a positive covariance implies a smaller dissimilarity (distance) between the utilities, whereas a negative covariance implies a larger dissimilarity (Böckenholt, 2003).

Clearly, much care needs to be taken both in the parameter interpretation and the comparison of different Thurstonian models. Estimated parameters (means, variances, and covariances) can be interpreted only in relative terms. Also, if a Case III or V model is found to be a parsimonious representation of the data, it is not possible to infer that the latent utilities are independent. However, as shown in the next application, the distinction between the three cases (unrestricted and Cases III and V) is still a useful starting point in an exploration of different parsimonious forms of the utilities' covariance matrices.

A Paired Comparisons Application: Modeling Preferences for Compact Cars

The goal of this application was to model purchasing preferences for compact cars among Spanish college students. The following seven compact cars were considered: Citroën AX, Fiat Punto, Nissan Micra, Opel Corsa, Peugeot 106, Seat Ibiza, and Volkswagen Polo. Binary paired comparison data were collected from a random sample of 294 undergraduates in the fall of 1996. The students were asked to indicate for each of the 21 pairs formed on the basis of the seven cars which one they would prefer to purchase. The presentation order of pairs as well as the order of cars within each pair was randomized.

Four basic models were fitted to the data with Mplus. The four models were obtained by crossing two structures for $\boldsymbol{\Sigma}_t$ (unrestricted and Case III) and two structures for $\boldsymbol{\Omega}^2$ ($\boldsymbol{\Omega}^2$ = diagonal, and $\boldsymbol{\Omega}^2 = \omega^2\mathbf{I}$). The goodness-of-fit results are presented in Table 2. Table 2 shows that the unrestricted model reproduces reasonably well the thresholds and tetrachoric correlations. However, the Case III model fails to adequately fit these data. Therefore, we conclude that preferences for these compact cars are not consistent with the hypothesis of independently evaluated choice alternatives.

To further investigate the error variances of the unrestricted model, we compared the unrestricted model with $\boldsymbol{\Omega}^2$ diagonal and the unrestricted model with equal error variances. The test of the difference in fit of these two nested models yielded T_M(dif) = 13.54, df = 20, p = 0.85, which indicates that a model with equal error variances is satisfactory for these data. Thus, there is no evidence to suggest that some car pairs are compared with more within-pair variability than others.

The parameter estimates and standard errors for the selected model are shown in Table 3. The estimated parameters shown in this table are to be interpreted relative to the fixed parameter that sets the scale for the remaining parameters. Thus, the estimated means (presented in the last column of the table) are to be compared relative to the mean of Volkswagen Polo, which is fixed to 0. All estimated means are statistically different from the mean of this car, except for Peugeot 106. Constructing 95% confidence intervals for the car means we see that Seat Ibiza is the most preferred car, followed by Volkswagen Polo, Peugeot 106, and so forth until the least preferred car, which is the Citroën AX. The estimated common variance of the pairwise errors, $\hat{\omega}^2 = 0.29$, is statistically larger than 0. The variances of the latent utilities for the different cars are also greater than 0. The estimated variances of the latent utilities also can be compared (with 95% confidence intervals) with the two variances, which are fixed to 1, and set the scale for the remaining variances. Finally, the estimated covariances can be compared with the covariance that sets their scale. With the identification constraints we have chosen, this is the covariance between the first and last choice alternatives (in this case Citroën AX and Volkswagen Polo). In Table 3 there is one estimated covariance that is positive and statistically larger than 0. We interpret this covariance as "the association between the utilities for Citroën AX and Peugeot 106 is stronger than the association between Citroën AX and Volkswagen Polo." Also, there are two estimated covariances that are statistically significant but negative in Table 3. Because a negative covariance implies a greater distance between the utilities, we interpret these covariances as "the association between the utilities for Citroën AX and Seat Ibiza is weaker than the association between Citroën AX and Volkswagen Polo" and "the association between the utilities for Nissan Micra and Seat Ibiza is weaker than the association between Citroën AX and Volkswagen Polo." The remaining associations are not stronger or weaker than the association between Citroën AX and Volkswagen Polo given the sample size employed.

In sum, the following substantive conclusions can be drawn from the analysis of these data: (a) a Thurstonian framework is appropriate for modeling these data because the unrestricted model fits; (b) the number of intransitivities is approximately equal across pairs, as the variances of the error terms can be set equal to each other; (c) the most preferred car model is the Seat Ibiza, followed by the Volkswagen Polo, Peugeot 106, and so forth, until the least preferred car which is the Citroën AX; and (d) the latent preferences are not independent, as the Case III model does not fit.

Because the cars were not evaluated independently of each other, respondents may have used one or several attributes in arriving at their preference judgments. These attributes cannot be derived directly by inspecting the estimated covariance matrix. In other words, it is not immediately apparent why there is a stronger association between preferences for Citroën AX and Peugeot 106 than between preferences for Citroën AX and Seat Ibiza. In the next section, we consider fitting a factor-analytic model to the utilities' covariance matrix (Takane, 1994; Tsai & Böcken-

Table 2
Goodness-of-Fit Tests for Some Basic Models Applied to the Cars Paired-Comparisons Data

Model	T	T_M	df	p
Unrestricted, $\boldsymbol{\Omega}^2$ diagonal	98.95	168.28	184	.79
Unrestricted, $\boldsymbol{\Omega}^2 = \omega^2\mathbf{I}$	106.03	182.30	204	.86
Case III, $\boldsymbol{\Omega}^2$ diagonal	346.91	389.74	198	<.01
Case III, $\boldsymbol{\Omega}^2 = \omega^2\mathbf{I}$	381.16	431.30	218	<.01

Note. Diagonally weighted least squares estimation; $T = N\hat{F}$; T_M is the Satorra–Bentler mean adjusted statistic T; $p > .05$ indicates acceptable overall fit to the data. In the unrestricted model, $\boldsymbol{\Sigma}_t$ is unrestricted; in the Case III model, $\boldsymbol{\Sigma}_t$ is specified to be diagonal but otherwise unrestricted; and in the Case V model, $\boldsymbol{\Sigma}_t = \sigma^2\mathbf{I}$.

Table 3
Unrestricted Model for the Cars Paired-Comparisons Data: Parameter Estimates of $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ and Standard Errors

$\hat{\boldsymbol{\Sigma}}_t$	1	2	3	4	5	6	7	$\hat{\boldsymbol{\mu}}_t$
1. Citroën AX	—							**−1.43**
	(fixed)							(0.11)
2. Fiat Punto	−0.04	**0.89**						**−0.58**
	(0.14)	(0.26)						(0.09)
3. Nissan Micra	0.29	0.05	**1.60**					**−0.78**
	(0.18)	(0.20)	(0.34)					(0.11)
4. Opel Corsa	0.13	−0.14	−0.07	**0.47**				**−0.30**
	(0.12)	(0.15)	(0.16)	(0.20)				(0.08)
5. Peugeot 106	**0.25**	−0.12	0.12	−0.01	**0.96**			−0.15
	(0.13)	(0.16)	(0.19)	(0.15)	(0.26)			(0.09)
6. Seat Ibiza	**−0.35**	0.12	**−0.40**	−0.11	−0.06	**1.29**		**0.22**
	(0.17)	(0.21)	(0.19)	(0.18)	(0.18)	(0.35)		(0.10)
7. Volkswagen Polo	0	0	0	0	0	0	—	0
	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)	(fixed)

Note. Diagonally weighted least squares estimation; standard errors in parenthesis; parameters in bold are significantly different from 0 at $\alpha = .05$. $\hat{\omega}^2 =$ 0.29 (0.04). A number of parameters have been fixed for identification purposes.

holt, 2001). We refer to these models as Thurstonian factor models. By using these factor analytic models, we may be able to uncover latent attributes that systematically influenced the respondents' judgments. These latent attributes may prove useful in explaining why associations for some cars are stronger than for others.

Thurstonian Factor Models for Paired Comparisons and Ranking Data

Thurstonian factor models are well suited to represent dependencies among choice alternatives because of three significant advantages. First, Thurstonian factor models are more parsimonious than Thurstonian models with an unconstrained covariance matrix. As a result, parameters can be estimated with greater precision, especially when the number of factors is small. Second, they can provide a graphical representation of the similarity structure underlying the choice data that is considerably easier to interpret than the estimated elements of the unconstrained covariance matrix. Third, and most important, relative distances between items in this graphical representation may be unique. That is, when a Thurstonian factor model fits the data, we can overcome, to some extent, the interpretational problems caused by the existence of equivalent models.

In the following, we present the threshold and correlation structure implied by Thurstonian factor models, and we provide identification restrictions for these models. Next, we discuss how to meaningfully interpret the model provided it fits the data. We conclude the section by introducing a special case of this model in which the mean preferences are specified to depend on the choice alternatives' position in the factor space.

Factor Models for Choice Data

We begin by considering the unobserved latent utilities $\mathbf{t}$ for the n choice alternatives. In a Thurstonian factor model, the vector $\mathbf{t}$ is represented by the following exploratory factor analysis model:

$$\mathbf{t} = \boldsymbol{\mu}_t + \boldsymbol{\Lambda}_t\boldsymbol{\xi} + \boldsymbol{\varsigma}. \tag{27}$$

When the number of common factors is m, $\boldsymbol{\mu}_t$ contains the n means of the latent utilities, $\boldsymbol{\Lambda}_t$ is an $n \times m$ matrix of factor loadings, $\boldsymbol{\xi}$ is an m-dimensional vector of common factors, and $\boldsymbol{\varsigma}$ is an n-dimensional vector of unique factors. This factor model assumes that the common factors have mean 0, have unit variance, and are uncorrelated. The model also assumes that the unique factors have mean 0 and are uncorrelated, so that their covariance matrix, $\boldsymbol{\Psi}_t^2$, is diagonal. In concordance with the distributional assumptions of Thurstonian choice models, we assume that the common and unique factors are normally distributed.

Consequently, the covariance matrix of the latent utilities can be written as

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t' + \boldsymbol{\Psi}_t^2. \tag{28}$$

The mean vector and covariance matrix of the latent difference responses $\mathbf{y}^*$ are normally distributed. Their mean vector and covariance matrix are obtained by substituting the utilities' mean vector $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$, implied by the factor model represented by Equation 28 into Equation 22.

In important ways, the Thurstonian factor model is similar to a second-order factor-analytic model. Under a Thurstonian factor model, the unobserved utilities $\mathbf{t}$ can be viewed as first-order factors, and the common factors $\boldsymbol{\xi}$ can be viewed as second-order factors. Furthermore, the structure for the latent utility judgments is confirmatory (given by the contrast matrix **A**), but the structure for the common factors is exploratory. To illustrate our discussion, consider Figure 2. In this figure, we have depicted the covariance structure of the latent utilities **y*** for a one-factor Thurstonian model for paired-comparisons data. Notice that the structure for the first-order factors (the unobserved utilities **t**) is confirmatory and that Thurstonian factors are similar to second-order factors.

Again, as the observed data is binary, the model must be estimated with tetrachoric correlations. Equations 11 and 23 contain the restrictions imposed by Thurstonian paired-comparisons models on the thresholds and tetrachoric correlations of the binary outcomes, respectively. To obtain the thresholds and tetrachoric correlations implied by a Thurstonian factor model, we substitute the utilities' mean vector $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$ that are implied by the factor model into Equations 11 and 23. First, we obtain the vector of thresholds,

$$\boldsymbol{\tau} = -\mathbf{DA}\boldsymbol{\mu}_t. \tag{29}$$

Then, we obtain the correlation matrix of the standardized latent difference responses,

$$\begin{aligned}\mathbf{P}_{z^*} &= \mathbf{D}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}' + \boldsymbol{\Omega}^2)\mathbf{D}\\ &= \mathbf{D}[\mathbf{A}(\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t' + \boldsymbol{\Psi}_t^2)\mathbf{A}' + \boldsymbol{\Omega}^2]\mathbf{D},\end{aligned} \tag{30}$$

where $\mathbf{D} = [\mathrm{Diag}(\boldsymbol{\Sigma}_{y*})]^{-1/2} = \{\mathrm{Diag}[\mathbf{A}(\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t' + \boldsymbol{\Psi}_t^2)\mathbf{A}' + \boldsymbol{\Omega}^2]\}^{-1/2}$. With $\boldsymbol{\Omega}^2 = \mathbf{0}$, Equations 29 and 30 represent the common factor model for ranking data.

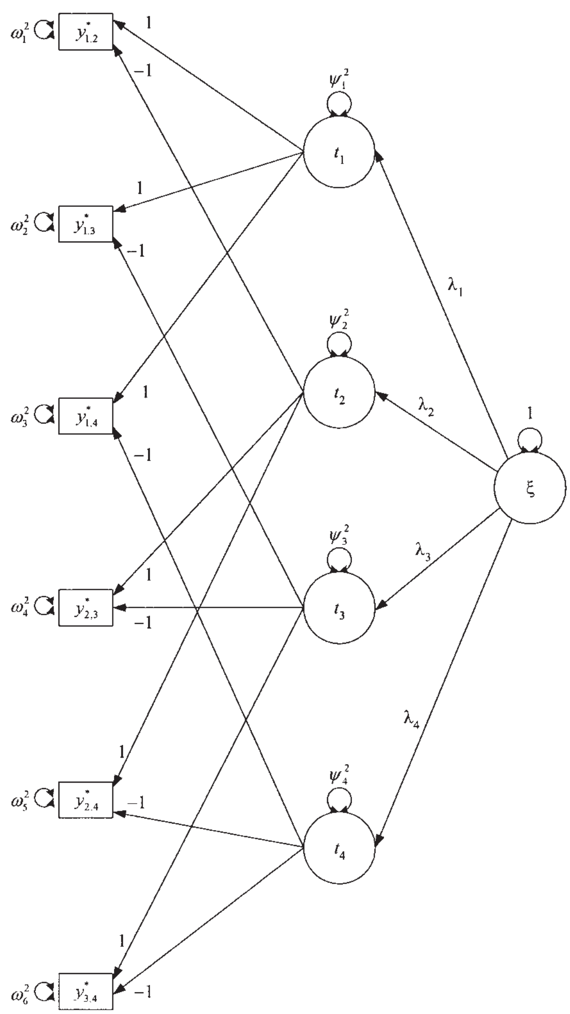

Figure 2. Covariance structure of a Thurstonian one-factor model for paired-comparisons data for $n = 4$ choice alternatives. Some identification restrictions are needed to identify the model. These are described in the text. A one-factor model for ranking data is obtained by setting $\omega_1, \ldots, \omega_6 = 0$.

Identification Constraints for Thurstonian Factor Models

Some identification restrictions are needed to be able to estimate Thurstonian factor models. Here we provide one possible set of restrictions chosen because they are easy to implement.

The identification restrictions needed to identify the factor part of the Thurstonian model are similar to those used to identify a standard exploratory factor-analytic model. McDonald (1999, p. 181) notes that the simplest way to solve the rotational indeterminacy problem in exploratory factor models with SEM software consists of setting the upper triangular part of the factor loading matrix equal to 0. This suggestion amounts to setting $\lambda_{ij} = 0$, $i = 1, \ldots, m = -1$; $j = i + 1, \ldots, m$. For example, with these constraints, the factor loading matrix for a three-factor model has the following form:

$$\boldsymbol{\Lambda}_t = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \\ \vdots & \vdots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \lambda_{n3} \end{pmatrix}. \tag{31}$$

The use of Equation 31, together with the assumption of uncorrelated factors, provides an arbitrary but simple way of obtaining a just-identified exploratory factor model. Additional constraints are needed for paired-comparisons and ranking factor models, because the data represent differences between latent utilities. The identification constraints for the parameters of the Thurstonian factor models $\boldsymbol{\mu}_t$, $\boldsymbol{\Lambda}_t$, $\boldsymbol{\Psi}_t$, and $\boldsymbol{\Omega}$[7] are the same for both data types. In addition to the constraints on the loading matrix given by Equation 31, we suggest using the following constraints to obtain an

[7] Recall that for ranking models, $\boldsymbol{\Omega} = \mathbf{0}$. Also, the Thurstonian factors are assumed to be uncorrelated.

identified Thurstonian factor model: (a) fix the mean of the last item in the item set to 0, $\mu_n = 0$; (b) fix all factor loadings involving the last item to 0, $\lambda_{ni} = 0$, $i = 1, \ldots, m$; and (c) fix the unique variance of the last item to 1, $\psi_n^2 = 1$. These identification constraints for the Thurstonian factor models define the scales of the means of the latent utilities, the factor loadings, and the unique factor variances, respectively. As an illustration, the identification restrictions needed to estimate a Thurstonian two-factor model for paired-comparisons and ranking data are

$$\boldsymbol{\mu}_t = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n-1} \\ 0^* \end{pmatrix}, \boldsymbol{\Lambda}_t = \begin{pmatrix} \lambda_{11} & 0^* \\ \lambda_{21} & \lambda_{22} \\ \vdots & \vdots \\ \lambda_{n-1,1} & \lambda_{n-1,2} \\ 0^* & 0^* \end{pmatrix}, \text{ and}$$

$$\boldsymbol{\Psi}_t^2 = \begin{pmatrix} \psi_1^2 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \psi_{n-1}^2 & 0 \\ 0 & \cdots & 0 & 1^* \end{pmatrix}. \quad (32)$$

Note that at least $n = 5$, 6, 8, and 9 items are required to estimate Thurstonian factor models with $m = 1$, 2, 3, and 4 common factors in both paired-comparisons and ranking data. The minimal number of items results from the constraint that the number of identified factor loadings and variances of the unique factors cannot exceed the number of identified elements in the unconstrained covariance matrix. Thus, in a Thurstonian m-factor model with $\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t' + \boldsymbol{\Psi}_t^2$, there are $n - 1$ parameters in $\boldsymbol{\Psi}_t^2$ and $nm - m - m(m-1)/2 = m(2n - m - 1)/2$ parameters in $\boldsymbol{\Lambda}_t$. Yet, the total number of estimated parameters cannot exceed $\{[n(n-1)]/2\} - 1$, the number of identifiable parameters in $\boldsymbol{\Sigma}_t$ for unrestricted models.

Fortunately, despite the large number of identification constraints, the interpretation of factor models is straightforward and appealing. As shown in the next example, the factor loadings provide us with information about similarity relationships between choice alternatives that can be depicted graphically (Takane, 1994).

Interpretation of Thurstonian Factor-Analytic Models

Equation 24 demonstrates that different covariance matrices are empirically indistinguishable because they can lead to the same paired-comparison or ranking outcomes. When a Thurstonian m-factor model holds, the model interpretation is easier than for an unrestricted model because we can interpret the factor loadings as coordinates in an m-dimensional similarity space (the factor space). As with any exploratory factor model, its axes can be shifted and rotated in any way to simplify interpretation.

Furthermore, when the variances of the unique factors can be set equal to each other, the relative distances between the choice alternatives in the m-dimensional space are unique (see Appendix B). In other words, they remain invariant for any of its equivalent models. This is a very important property, as when this model fits, the interpretational problems caused by the existence of equivalent models are resolved. Also, with this constraint, higher dimensional factor models can be estimated for a fixed number of choice alternatives than when the variances of the unique factors are not set equal to each other. Thus, only $m + 3$ choice alternatives are required to estimate a Thurstonian m-factor model in both paired-comparisons and ranking data when the variances for the unique factors are set equal to each other. These models are identified with the constraints depicted in Equation 32, where the ψ_i^2s are set equal to each other.

To illustrate the approach, we estimate a factor model for the car preference data in the next section. This analysis will demonstrate that it is not necessary to consider the unrestricted covariance matrix for these data sets. Instead, a two-dimensional factor model with equal variances for the unique factors provides a satisfactory description of the data.

Factor Models With Restricted Means

In standard factor-analytic models it is possible to restrict the means of the observed variables so that they depend on the factor loadings and the factor means (see Bentler & Yuan, 2000). However, this is done rarely in factor-analytic applications, as generally there is little interest in modeling means. In contrast, in analysis of choice data the means of the latent utilities are important parameters because they define stochastically the ordering of the choice alternatives from most to least preferred. By relating the factor loadings to the choice alternatives' utility means, we can investigate whether the mean preferences can be explained by individual differences in the evaluation of the choice alternatives. For example, if individuals differ in their perceptions of the leadership qualities of the political candidates, it is possible that this source of individual differences also determines the mean preferences for the candidates. Such hypotheses can be tested by formulating a Thurstonian factor model in which the utilities' means depend on the common factors. This is a constrained version of the Thurstonian factor model represented in Equation 27, given by

$$\mathbf{t} = \mathbf{1}\gamma + \boldsymbol{\Lambda}_t\boldsymbol{\xi} + \boldsymbol{\varsigma}. \quad (33)$$

Here, $\mathbf{1}$ is an $n \times 1$ vector of 1s, and γ is a constant. This model differs from the common-factor model (see Equation 27) in two respects. First, the vector of utilities' means $\boldsymbol{\mu}_t$ in Equation 27 has been replaced by the vector of common intercepts, $\mathbf{1}\gamma$, for regressing the latent utilities on the common factors. Second, the means of the common factors, $\boldsymbol{\mu}_\xi$, are now parameters to be estimated. In standard factor-analytic models, these means are set equal to 0.

The correlation structure implied by this special case of the Thurstonian factor model is equal to the one given in Equation 30. However, the restrictions introduced by this model on the means constrain the threshold structure, leading to the expression

$$\boldsymbol{\tau} = -\mathbf{D}\mathbf{A}\boldsymbol{\mu}_t = -\mathbf{D}\mathbf{A}(\mathbf{1}\gamma + \boldsymbol{\Lambda}_t\boldsymbol{\mu}_\xi). \quad (34)$$

When the latent means $\boldsymbol{\mu}_t$ are regressed on the factor means $\boldsymbol{\mu}_\xi$, the factor loading matrix becomes the matrix of regression slopes. In addition, a common intercept, γ, needs to be estimated. This model is identified with the same constraints used for other Thurstonian factor models. However, to identify the threshold structure in Equation 34, the last element of $\mathbf{1}\gamma$ is replaced by a 0.

The Paired-Comparisons Application Revisited: Modeling Preferences for Compact Cars With a Factor Model

Previously, we found that the best fitting basic model for the cars paired-comparison example was an unrestricted model with $\boldsymbol{\Omega}^2 = \omega^2\mathbf{I}$. In this section, we reanalyze these data to investigate whether a common factor model accounts for the observed associations among the car preferences.

Because the interpretation of Thurstonian factor models is simplified when the variances for the unique factors are set equal to each other, here we consider models with this constraint. The maximum number of dimensions that can be fitted in this application with this constraint is three. Therefore, we estimated models from one to three factors to these data, assuming that the within-pair error variance is homogenous (i.e., $\boldsymbol{\Omega}^2 = \omega^2\mathbf{I}$) and that the unique factors have the same variance with $\boldsymbol{\Psi}_t^2 = \psi^2\mathbf{I}$ (except for the variance of the last choice alternative, which is fixed to 1 for identification purposes). We provide in Table 4 the goodness-of-fit statistics for these three models. Although the one-factor model does not fit, the two-factor model does provide a satisfactory fit to the data. The three factor model does not fit significantly better than the two factor model, $T_M(\text{dif}) = 5.56$, $df = 4$, $p = .23$. Furthermore, a comparison between the two-factor model and an unrestricted model revealed that the additional parameters of the unrestricted model do not improve the fit of the two-factor model: $T_M(\text{dif}) = 11.34$ ($df = 8$, $p = .18$). Thus, the two-factor model gives the most parsimonious representation for these data. We provide in Table 5 the parameter estimates and standard errors for this two-factor model. The ordering of the mean utilities changes little for the different covariance structures. Thus, the estimated means for the two-factor model given in Table 5 are very similar to those estimated under the unrestricted model (see Table 3).

Table 4
Goodness-of-Fit Tests for Some Factor Models Applied to the Cars Paired-Comparisons Data

Model	T	T_M	df	p
One factor	226.91	286.16	217	<.01
Two factors	146.82	211.14	212	.50
Three factors	125.13	197.47	208	.69

Note. Diagonally weighted least squares estimation. $\boldsymbol{\Omega}^2 = \omega^2\mathbf{I}$; $\boldsymbol{\Psi}_t^2 = \psi^2\mathbf{I}$ (except for the last choice alternative, which is fixed to 1 for identification); $T = N\hat{F}$; T_M is the Satorra–Bentler mean adjusted statistic T; $p > .05$ indicates acceptable fit to the data.

Figure 3 provides a plot of the factor loadings reported in Table 5. Because, in this model, the variances of the unique errors are set equal to each other, the relative distances between the compact cars remain invariant when identification restrictions are changed. Factor models with equal variances for the unique errors are unaffected by the interpretation problem caused by Equation 24. Thus, the relative positions of the compact cars in Figure 3 can be interpreted directly.

In Figure 3, we have drawn two orthogonal rotated axes that are more amenable to interpretation. The first rotated dimension—the dashed diagonal line going from the lower left to upper right of the figure—appears to be related to car size. Note that the positive end of this dimension is at the lower left. The largest model (Seat Ibiza) loads positively on this dimension, and the two smallest cars (Nissan Micra and Citroën AX) load negatively. The second dimension—the dashed line going from the upper left to the lower right of the figure—appears to be related to the perceived sturdiness of the cars. The positive end of this dimension is at the lower right. The cars loading positively on this dimension (e.g., Citroën AX) are perceived to be more sturdy, whereas the cars loading negatively (Seat Ibiza, Nissan Micra, and Fiat Punto) are perceived to be less sturdy. Thus, individual differences in the comparisons of these car models are explained by these dimensions. Respondents who prefer smaller car models are less likely to prefer larger car models, and conversely. Similarly, respondents who prefer more sturdy car models are less likely to prefer less sturdy cars, and conversely.

We next investigated whether these dimensions also explain the means of the latent utilities. As a first check of the relationship, we regressed the utilities' means on the estimated factor loadings. The regression model yielded a R^2 of .52. Although the variance accounted for is substantial, this result suggests that a model in which the utilities' means are exclusively represented by the two-dimensional factor structure is unlikely to fit the data. However, we could not obtain a fit statistic for a model in which the mean preferences depend on the two common factors. The estimation algorithm failed to converge after 1,000 iterations. On the basis of the R^2 statistic, we conclude that although car size and car sturdiness explain the associations among the preferences for the different car models, they do not fully account for the mean preferences. Other considerations, in

Table 5
Two-Factor Model With Common Unique Variances for the Cars Paired-Comparisons Data

	Λ_t				μ_t		Ψ_t^2	
Car	Parameter estimate	*SE*	Parameter estimate	*SE*	Parameter estimate	*SE*	Parameter estimate	*SE*
Citroën AX	0.55	0.14	0	fixed	−1.44	0.13	0.69	0.11
Fiat Punto	−0.23	0.16	0.24	0.21	−0.57	0.57	0.69	0.11
Nissan Micra	0.48	0.28	0.79	0.15	−0.77	0.11	0.69	0.11
Opel Corsa	0.15	0.13	−0.09	0.13	−0.29	0.08	0.69	0.11
Peugeot 106	0.31	0.15	−0.14	0.12	−0.16	0.09	0.69	0.11
Seat Ibiza	−0.67	0.16	−0.17	0.32	0.21	0.10	0.69	0.11
Volkswagen Polo	0	fixed	0	fixed	0	fixed	1	fixed

Note. Diagonally weighted least squares estimation; $\hat{\omega}^2 = 0.27$ (0.05). The factor loadings Λ_t correspond to the positions of the car models in the preference map shown in Figure 3.

addition to car size and sturdiness, played a role in the students' car choices.

Conclusion

In this article, we reviewed Thurstonian choice modeling of paired comparisons and ranking data by embedding them within an SEM framework. In addition, we showed how these models can be estimated with a popular SEM package, Mplus.

Guidelines for Model Selection in Thurstonian Choice Modeling

Model selection and interpretation require much care because of the comparative nature of the observed data. Our recommended sequence of analyses in applications is summarized in the flow chart presented in Figure 4.

Step 1: Estimate an unrestricted model. The goals of this analysis are to determine whether the unrestricted model with equal or unequal pair-specific variances provides a satisfactory fit to the data and, in the case of misfit, to identify the systematic sources of misfit. For example, a Thurstonian model may not fit the data if the experimenter ignored an important individual difference variable in the analysis (Böckenholt, 1993). Consider a ranking of U.S. political candidates. If the political orientation of the voters (Democrat vs. Republican) is not taken into account, the Thurstonian model is likely to provide a poor fit of the data, as it may not be possible to capture all of the individual differences with a single covariance matrix. Instead, it may be necessary to estimate a separate covariance matrix for each group of the voters. However, if the unrestricted model provides a satisfactory fit of the data, it is important to consider special cases of the model, to reduce both the number of parameters to be estimated, and to simplify the interpretation of the results.

Step 2: Estimate a Case III model. It is useful to compare the fit of the unrestricted model with the fit of a Case III model. If the Case III does not fit well, the hypothesis of independent latent utilities is rejected. In contrast, when a Case III model fits, we cannot conclude that the independence hypothesis holds because models with nonzero covariances can yield an equivalent fit of the data. If the hypothesis of dependent preferences is to be rejected, additional data need to be collected that provide information on the origin of the scale values (see Böckenholt, 2004, for further details).

Step 3a: Case III model fits well. If the Case III model provides a satisfactory fit to the data, investigate whether the more parsimonious representation provided by the Case V model can fit the data as well.

Step 3b: Case III fits poorly. A poor fit of the Case III model indicates that the hypothesis of independent individual differences can be rejected. In this case, factor models

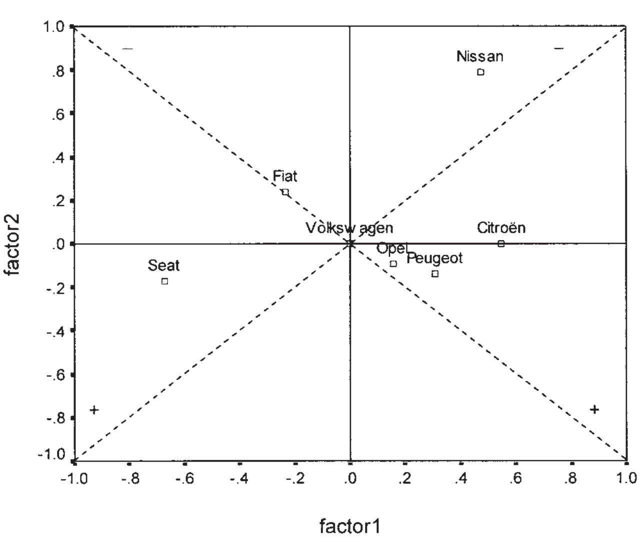

Figure 3. Two-dimensional preference map for the compact cars data. The car models are Citroën AX, Fiat Punto, Nissan Micra, Opel Corsa, Peugeot 106, Seat Ibiza, and Volkswagen Polo. The solid lines represent the unrotated factors. The dashed lines represent the rotated factors (car size and sturdiness). The + and − signs denote the high and low ends of the rotated dimensions.

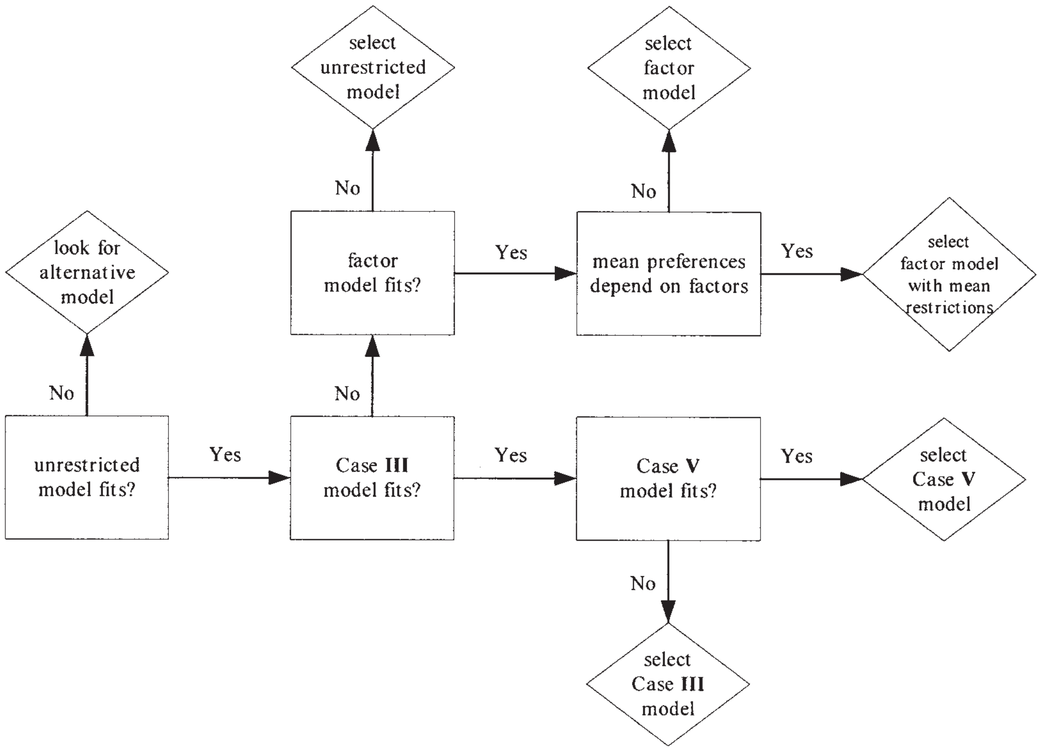

Figure 4. Suggested strategy for model selection. Flow chart summarizing the decision tree for selecting a Thurstonian model for paired-comparisons or ranking data.

with equal unique variances can be fitted to the data. These models facilitate the interpretation of the individual preferences' structure. A rejection of these models requires fitting a factor model with unequal unique variances. If these models are rejected as well, then the unrestricted model would be selected as the final model.

Step 4: Thurstonian factor model fits well. If a satisfactory fit of the Thurstonian factor model is obtained, possibly with different unique variances, then the fit of a factor model, in which preference means depend linearly on the factor loadings, should be investigated. A satisfactory fit of this model provides an attractive interpretation about the source of the mean preferences.

Thurstonian Choice Modeling Within an SEM Framework

We have demonstrated in this article that Thurstonian models can be fitted to ranking and paired-comparison data within a structural equation framework. Using Mplus, we showed that the classical hypotheses of an unrestricted covariance, Case III, and Case V structures are straightforward to specify and estimate. Moreover, with additional modifications, factor-analytic models can be estimated that facilitate testing dimensional theories about individual differences. A description of how to use Mplus to estimate Thurstonian models, the input files used in the examples, and the data files are available as supplementary materials that can be downloaded from http://dx.doi.org/10.1037/1082-989X.10.3.285.supp.

With little loss in statistical efficiency, the proposed SEM approach can accommodate models with a much larger number of choice alternatives than currently feasible under alternative estimation methods. Estimation is fast, and even complex models can be estimated within seconds. Also, goodness-of-fit statistics with accurate p values can be obtained even when the data are sparse.[8] For example, Maydeu-Olivares (2003a) reports that when SEM procedures are

[8] The p values reported in Mplus assess how well Thurstonian choice models reproduce the estimated thresholds and tetrachoric correlations (i.e., the structural restrictions imposed by the model). The tetrachoric correlations assume a normal distribution underlying the observed data. However, recent research (Maydeu-Olivares, in press) has shown that the tests for structural restrictions in tetrachoric correlations are somewhat robust to violations of the underlying normality assumption. It is possible to test the underlying normality assumption with procedures described in B. Muthén and Hofacker (1988) but only for three variables at a time. A more fruitful avenue may be to simultaneously test the structural restrictions and the underlying normality assumption with an overall test described in Maydeu-Olivares (2001). Neither the Muthén–Hofacker nor the Maydeu-Olivares procedure is currently implemented in Mplus.

used, 300 observations suffice to obtain accurate parameter estimates, standard errors, and goodness-of-fit tests for an unrestricted ranking model for seven choice alternatives, and as few as 100 suffice to estimate and test an unrestricted paired-comparisons model with seven choice alternatives. In contrast, p values obtained by use of the usual goodness-of-fit tests for maximum-likelihood estimation (e.g., the likelihood ratio test G^2 and Pearson's χ^2) are notoriously inaccurate when choices among many objects are modeled and the sample size is small. Taken together, these advantages demonstrate that the proposed framework has effectively overcome past estimation and inference problems in the analysis of ranking and paired-comparison data.

Also, our specification of Thurstonian scaling models as structural equation model facilitates a number of further extensions that are valuable in applications of this approach. For example, covariates may be included to explain individual differences in the evaluation of choice alternatives. Because Mplus assumes that the latent responses are multivariate normal conditional on the values of the covariates, binary and nonnormal continuous covariates can be considered. Moreover, multivariate paired comparisons and rankings that involve comparisons of stimuli with respect to multiple attributes (Böckenholt, 1990) can be analyzed with the same approach. Applications in which respondents compare only subsets of all possible pairs (i.e., incomplete paired comparisons) or rank some but not all of the choice alternatives (i.e., partial rankings) can also be handled, requiring only minor modifications to the approaches presented in this article. However, this outlook for future work should not distract from the fact that an important and large class of choice models has become accessible to researchers. We expect and look forward to many more choice modeling applications in the future.

References

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Yuan, K. H. (2000). On adding a mean structure to a covariance structure model. *Educational and Psychological Measurement, 60,* 326–339.

Bock, R. D., & Jones, L. V. (1968). *The measurement and prediction of judgment and choice.* San Francisco: Holden-Day.

Böckenholt, U. (1990). Multivariate Thurstonian models. *Psychometrika, 55,* 391–403.

Böckenholt, U. (1993). Applications of Thurstonian models to ranking data. In M. A. Fligner & J. S. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 157–172). New York: Springer-Verlag.

Böckenholt, U. (2001a). Hierarchical modeling of paired comparison data. *Psychological Methods, 6,* 49–66.

Böckenholt, U. (2001b). Preference models with latent variables. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 17, pp. 11965–11969). Amsterdam, the Netherlands: Elsevier.

Böckenholt, U. (2003). *Analyzing comparative judgments with and without a scale origin.* Unpublished manuscript, McGill University, Montreal, Quebec, Canada.

Böckenholt, U. (2004). Comparative judgments as an alternative to ratings: Identifying the scale origin. *Psychological Methods, 9,* 453–465.

Chan, W., & Bentler, P. M. (1998). Covariance structure analysis of ordinal ipsative data. *Psychometrika, 63,* 369–399.

Fligner, M. A., & Verducci, J. S. (Eds.). (1993). *Probability models and statistical analyses for ranking data.* New York: Springer-Verlag.

Florig, H. K., Morgan, M. G., Morgan, K. M., Jenni, K. E., Fischhoff, B., Fischbeck, P. S., & DeKay, M. L. (2001). A deliberative method for ranking risks (I): Overview and test bed development. *Risk Analysis, 21,* 913–921.

Hazell, P. L., Tarren-Sweeney, M., Vimpani, G. M., Keatinge, D., & Callan, K. (2002). Children with disruptive behaviours: II. Clinical and community service needs. *Journal of Pediatrics and Child Health, 38,* 32–40.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.50* [Computer software]. Chicago, IL: Scientific Software.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review, 93,* 1449–1475.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in structural equation modeling. *Multivariate Behavioral Research, 25,* 313–334.

Luce, R. D. (1977). Thurstone's discriminal processes fifty years later. *Psychometrika, 42,* 461–490.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114,* 185–199.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis.* London: Academic Press.

Marley, A. A. J. (2002). Random utility models and their applications: Recent developments. *Mathematical Social Sciences, 43,* 289–302.

Maydeu-Olivares, A. (1999). Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika, 64,* 325–340.

Maydeu-Olivares, A. (2001). Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika, 66,* 209–228.

Maydeu-Olivares, A. (2003a). On Thurstone's model for paired comparisons and ranking data. In H. Yanai, A. Okada, K. Shigematu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics* (pp. 519–526). Tokyo: Springer.

Maydeu-Olivares, A. (2003b). *Thurstonian covariance and correlation structures for multiple judgment paired comparison data* (Working Paper 04-03). Madrid, Spain: Instituto de Empresa.

Maydeu-Olivares, A. (in press). Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika.*

McDonald, R. P. (1999). *Test theory: A unified approach.* Mahwah, NJ: Lawrence Erlbaum.

Morgan, K. M., DeKay, M. L., Fischbeck, P. S., Morgan, M. G., Fischhoff, B., & Florig, H. K. (2001). A deliberative method for ranking risks (II): Evaluation of validity and agreement among risk managers. *Risk Analysis, 21,* 923–937.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43,* 551–560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115–132.

Muthén, B. (1993). Goodness of fit with categorical and other non normal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.* Unpublished manuscript, College of Education, University of California, Los Angeles.

Muthén, B., & Hofacker, C. (1988). Testing the assumptions underlying tetrachoric correlations. *Psychometrika, 53,* 563–578.

Muthén, L., & Muthén, B. (2001). *Mplus 2* [Computer software]. Los Angeles, CA: Muthén & Muthén.

Oakes, M. E., & Slotterback, C. S. (2002). The good, the bad, and the ugly: Characteristics used by young, middle-aged, and older men and women, dieters and non-dieters to judge healthfulness of foods. *Appetite, 38,* 91–97.

Pascalis, O., de Haan, M., & Nelson, C. A. (2002, May 17). Is face processing species-specific during the first year of life? *Science, 296,* 1321–1323.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications to developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514.

Stelzl, I. (1986). Changing the causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research, 21,* 309–331.

Suppes, P., Krantz, D., Luce, D., & Tversky, A. (1992). *Foundation of measurement: Vol 2. Geometric, threshold, and probabilistic representations.* San Diego: Academic Press.

Takane, Y. (1987). Analysis of covariance structures and probabilistic binary choice data. *Communication and Cognition, 20,* 45–62.

Takane, Y. (1994). A review of applications of AIC in Psychometrics. In H. Bozdogan (Ed.), *Proceedings of the first US/Japan conference on the frontiers of statistical modeling: An informational approach* (pp. 379–403). Dordrecht, the Netherlands: Kluwer.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 79,* 281–299.

Thurstone, L. L. (1931). Rank order as a psychological method. *Journal of Experimental Psychology, 14,* 187–201.

Train, K. E. (2003). *Discrete choice methods with simulation.* Cambridge, England: Cambridge University Press.

Tsai, R. C. (2003). Remarks on the identifiability of Thurstonian paired comparison models under multiple judgment. *Psychometrika, 68,* 361–372.

Tsai, R. C., & Böckenholt, U. (2001). Maximum likelihood estimation of factor and ideal point models for paired comparison data. *Journal of Mathematical Psychology, 45,* 795–811.

Turati, C., & Simion, F. (2002). Newborns' recognition of changing and unchanging aspects of schematic faces. *Journal of Experimental Child Psychology, 83,* 239–261.

Younger, B. A., & Furrer, S. D. (2003). A comparison of visual familiarization and object-examining measures of categorization in 9-month-old infants. *Infancy, 4,* 327–348.

(*Appendixes follow*)

Appendix A

Threshold and Correlation Structure Implied by Thurstonian Models

The mean and covariance structure of the latent comparative responses **y*** for any Thurstonian model for paired comparisons is given by Equation 22. This equation also applies to ranking models with $\boldsymbol{\Omega}^2 = \mathbf{0}$. Because the latent comparative responses **y*** are linked to the observed binary outcomes **y** via the threshold relation (see Equation 4), it follows that the probability of observing an $\tilde{n}$-dimensional vector of binary outcomes is

$$\Pr\left(\bigcap_{l=1}^{\tilde{n}} y_l\right) = \int \cdots_{\mathbf{R}} \int \phi_{\tilde{n}}(\mathbf{y}^* : \boldsymbol{\mu}_{y^*}, \boldsymbol{\Sigma}_{y^*}) d\mathbf{y}^*. \quad \text{(A1)}$$

In Equation A1, $\phi_{\tilde{n}}(\bullet)$ denotes an $\tilde{n}$-variate normal density, and **R** is the multidimensional rectangular region formed by the product of intervals

$$R_l = \begin{cases} (0, \infty) & \text{if } y_l = 1 \\ (-\infty, 0) & \text{if } y_l = 0 \end{cases}. \quad \text{(A2)}$$

The probabilities (see Equation A1) are unchanged when we standardize **y*** using

$$\mathbf{z}^* = \mathbf{D}(\mathbf{y}^* - \boldsymbol{\mu}_{y^*}) \text{ and } \mathbf{D} = [\mathrm{Diag}(\boldsymbol{\Sigma}_{y^*})]^{-1/2}, \quad \text{(A3)}$$

in which case we can write

$$\Pr\left(\bigcap_{l=1}^{\tilde{n}} y_l\right) = \int \cdots_{\check{\mathbf{R}}} \int \phi_{\tilde{n}}(\mathbf{z}^* : \mathbf{0}, \mathbf{P}_{z^*}) d\mathbf{z}^*, \quad \text{(A4)}$$

where $\check{\mathbf{R}}$ is the multidimensional rectangular region formed by the product of intervals

$$\check{R}_l = \begin{cases} (\tau_l, \infty) & \text{if } y_l = 1 \\ (-\infty, \tau_l) & \text{if } y_l = 0 \end{cases}, \quad \text{(A5)}$$

and the constraints on the vector of thresholds, $\boldsymbol{\tau}$, and on the matrix of tetrachoric correlations, $\mathbf{P}_{z^*}$, are given by Equations 11 and 23. This result follows from using Equation A3, with $\boldsymbol{\mu}_{z^*} = \mathbf{0}$,

$$\boldsymbol{\tau} = -\mathbf{D}\boldsymbol{\mu}_{y^*}, \text{ and } \mathbf{P}_{z^*} = \mathbf{D}\boldsymbol{\Sigma}_{y^*}\mathbf{D}. \quad \text{(A6)}$$

The first equation within Equation A6 is obtained as follows: Let μ_l^* be an element of $\boldsymbol{\mu}_{y^*}$, and let σ_l^{2*} be a diagonal element of $\boldsymbol{\Sigma}_{y^*}$. Then, at $y_l^* = 0$, $\tau_l := (-\mu_l^*)/\sqrt{\sigma_l^{2*}} = -d_l\mu_l^*$, where d_l is a diagonal element of **D**.

Appendix B

Invariance of Relative Distances Between Choice Alternatives in the Factor Space When Variances of Unique Factors Are Equal

Letting c be the arbitrary positive constant in Equation 24, we pointed out that the matrix of relative squared distances $\boldsymbol{\Delta}/c$ is unique. That is, it is not affected by the linear transformation given in Equation 24. The elements of this matrix are

$$\delta_{ik}^2 = (\sigma_i^2 + \sigma_k^2 - 2\sigma_{ik})/c. \quad \text{(B1)}$$

Under the factor model (see Equation 28), $\sigma_{ik} = \boldsymbol{\lambda}_i\boldsymbol{\lambda}_k'$ and $\sigma_i^2 = \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' + \psi_i^2$, where $\boldsymbol{\lambda}_i$ contains the m factor loadings of choice alternative i, and ψ_i^2 is the corresponding variance component of the unique factors. By inserting these expressions in Equation B1, we can write this equation as

$$\delta_{ik}^2 = [(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_k)'(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_k) + \psi_i^2 + \psi_k^2]/c. \quad \text{(B2)}$$

Now, when the variances of the unique factors can be set equal to each other, we can write

$$c\delta_{ik}^2 + 2\psi^2 = (\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_k)'(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_k) = \sum_{j=1}^{m} (\boldsymbol{\lambda}_{ji} - \boldsymbol{\lambda}_{jk})^2. \quad \text{(B3)}$$

Equation B3 shows that the Euclidean distance between the loadings of choice alternatives i and k is unique up to a linear transformation.

Received July 2, 2002
Revision received February 28, 2005
Accepted April 27, 2005 ■