Research Report

Measuring Multidimensional Latent Growth

Frank Rijmen

November 2010

ETS RR-10-24

Listening. Learning. Leading.®

Measuring Multidimensional Latent Growth

Frank Rijmen
ETS, Princeton, New Jersey

November 2010

Technical Review Editor: Matthias von Davier

Technical Reviewers: John Donoghue and Kentaro Yamamoto

Abstract

As is the case for any statistical model, a multidimensional latent growth model comes with certain requirements with respect to the data collection design. In order to measure growth, repeated measurements of the same set of individuals are required. Furthermore, the data collection design should be specified such that no individual is given the same item twice, while at the same time allowing for common items over time so that different measurement occasions can be linked. An example of such a data collection design is presented.

A computational challenge arises due to the high dimensional nature of a multidimensional latent growth model. Not only are there multiple dimensions within each measurement occasion, but insofar not all individuals change at the same rate for a given construct, that construct will also give rise to multiple dimensions over time. Fortunately, the computational burden can be overcome insofar as one is willing to incorporate assumptions on the latent structure, such as a bifactor or higher order structure within measurement occasions, and the assumption that the construct at a particular time point is independent of the construct at all previous time points given the construct at the immediately preceding time point (first order Markov assumption).

Key words*:* item response theory, growth, longitudinal data, data collection designs, graphical models, bifactor model

Table of Contents

List of Figures

The last two decades have witnessed a spurt in the development and application of statistical models for repeated measurement data (or more in general longitudinal data) throughout various scientific fields including biostatistics (e.g., Verbeke & Molenberghs, 2000), quantitative social and behavioral sciences (e.g., Skrondal & Rabe-Hesketh, 2004), and educational measurement (e.g., Braun & Wainer, 2007). In a repeated measurement data collection design, the same dependent variable and a collection of background variables are recorded for a sample of cases at several occasions. For example, height may be measured for a sample of kids on a yearly basis, together with a set of covariates such as gender, diet, age, physical activity level, and so on. Then, height and its evolution over time (i.e., growth) can be modeled as a function of age and the other covariates. A natural framework for this modeling effort is the linear mixed model (Verbeke & Molenberghs, 2000), or the closely related multilevel (Goldstein, 1995) and hierarchical linear model frameworks (Raudenbush & Bryk, 2002).

In an educational context, the dependent variables are typically measures of achievement. Two important differences between measures of achievement and measures of physical attributes such as height bear consequences for both the data collection design and the statistical framework.

First, opposed to measuring height, it is not straightforward to ensure one is using the same measure over measurement occasions. Whereas in measuring height one can simply use the same measuring rule over and over, the repeated use of the same test materials in an achievement test may lead to a change in measurement characteristics of the test due to item exposure effects. On the other hand, using a different collection of test materials for each occasion makes it challenging to express the different measurements in time on the same scale.

In the next section, a data collection design is presented that prevents the same test material from being presented to the same persons twice, while maintaining common items over test occasions. Such a design may lend itself to link the measures stemming from different measurement occasions.

Second, whereas height is a unidimensional construct, achievement measures may be multidimensional. By implication, growth or in general, change over time, may operate on a multidimensional construct rather than on a unidimensional measure. This change in turn may result in a complex dependence structure for the joint collection of measures across all measurement occasions. Item response theory (IRT) models that accommodate multidimensionality both within and across measurement occasions are not too difficult to specify. However, parameter estimation for high dimensional models may become computationally intractable. Generally speaking, the computational complexity of a multidimensional model is inversely related to the number of conditional independence relations one is willing to assume. Graphical model theory turns out to be extremely useful in this regard, as it provides algorithms to determine the computational complexity involved in estimating the parameters of a given model.

In the third section of this paper, the underlying principles are explained and illustrated with a relatively simple model for repeated measurements. In the fourth section, several multidimensional model structures are presented, and their computational complexity with respect to parameter estimation is derived.

A Repeated Measurement Linking Design

In this section, a data collection design is presented that lends itself to establishing a link between the measurement occasions of a repeated measurement design.

Insofar the measurements at different occasions are targeted at different achievement levels, one can think of linking those measures as vertical linking, although vertical scales have typically been established on the basis of a cross-sectional data collection design, in which a different group of persons is sampled for each achievement level so that each person is measured only once. Readers should keep in mind that vertically linked measures are not needed for many purposes and that all procedures for vertical linking rely on a strong set of assumptions that may or may not be met in a particular situation. Good overviews of methodological pitfalls and caveats involving vertical linking procedures can be found in Braun and Wainer (2007) and Kolen and Brennan (2004). In this paper, it is assumed that the conditions under which vertical linking procedures are meaningful are met. Only assumptions that pertain to the use of a repeated measurement data collection design are mentioned.

Common linking procedures incorporate a common item, a single group, or an equivalent groups design (Kolen & Brennan, 2004). In the common item design, a common scale is established through the inclusion of common item blocks between test forms. In the two other designs, a common population can be assumed, either because the persons are common (single group) or because students are randomly assigned to one of several test forms (equivalent groups design). Then, differences in performance can be attributed to differences in item characteristics.

None of these designs is applicable to a longitudinal data collection context without modifications. Presenting the same item twice may distort the linking due to memory effects. Specifically, an item is likely to become easier if it has been presented before to the same group of students. In other words, one cannot assume that common items are common in a statistical sense. On the other hand, a person may change (mature, learn) between measurement occasions, so that one cannot assume that the person stays the same, ruling out the single group design. Because people may change, one cannot assume that the population at one measurement occasion is equivalent to the population at another measurement occasion, ruling out the equivalent groups design.

However, repeated measurement linking designs can be constructed by combining the equivalent groups and the common item design. Rijmen (2009c) presented two such designs. The second one is more suited to the context of IRT modeling. Because IRT models for longitudinal data are discussed in the second part of this paper, it is presented in the following. Table 1 presents the design in its basic form.

Table 1

Repeated Measurement Linking Design: Equivalent Groups With Common Items Over Time

Measurement occasion	Equivalent group	
	G1	G2
T1	A1B1	A2B2
T2	B2C1	B1C2
T3	C2D1	C1D2
T4	D2E1	D1E2

Note. The letters A through E indicate increasing levels of difficulty.

Without loss of generality, let us assume there are four measurement occasions. Also, let letters A through E indicate increasing levels of difficulty. At the first measurement occasion, two randomly equivalent groups are formed. Each group is presented with one of two forms that are constructed to be parallel. Each form consists of two parts: one part that is unique to the first measurement occasion and one part that will be used in the subsequent measurement occasion as well. The two parts of the first form are denoted by A1 and B1, and the two parts of the second form by A2 and B2. B1 and B2 are not allowed to have items in common, but A1 and A2 may have some or all items in common.

Both forms of the first measurement occasion can be linked horizontally through the equivalent group design. If A1 and A2 have sufficient items in common, they can also be linked through a common item design.

At the second measurement occasion, the test consists again of items pertaining to two different levels: B and C. Each group receives the B items that were not administered to that group at the previous measurement occasion.

Again, both forms at the second measurement occasion can be linked horizontally through an equivalent groups design. The scale at Measurement Occasion T2 can be aligned vertically with the scale at Measurement Occasion T1 through the common items B1 and B2.

The same two assumptions are made as in the previous design: Groups stay equivalent over time, and common items are truly common.

In order to mitigate the risk that groups become increasingly less equivalent over time, new random groups can be formed at each measurement occasion for the administration of the new level. For example, at Measurement Occasion T2, groups can be redefined with respect to the administration of C1 and C2. Then, there are four rather than two parallel forms at the second test occasion: B1C1, B1C2, B2C1, and B2C2. This procedure can be repeated at all subsequent test occasions.

Under this design, a common scale is established through several links. Each measurement occasion has two sets of items in common with both the previous and the next occasion. In addition, at all but the first measurement occasion, the two parallel forms can be linked through both an equivalent groups design or through linking both forms back to the scale of the previous measurement occasion using the set of common items. This property of the data collection design allows for some of the linking assumptions to be tested. For example, if there are indications that some or all of the items of B1 show item drift, one can refrain from using B1 as a set of items that is common between the first two measurement occasions and rely solely on B2 for linking the first two measurement occasions and on the equivalent group design for linking the two forms at Test Occasion T2.

It would be worthwhile to investigate to which degree these links can be put to work in concert and how this finding relaxes the requirements for each of these links. For example, it may be that a robust linking can be obtained with only a few common items between successive test occasions, as long as the total number of common items is sufficiently large.

In principle, both designs could function under a classical test theory framework as well as an IRT framework. However, when forming new randomized groups at each measurement occasion to mitigate the risk of groups becoming less equivalent over time, and when putting all links to work in concert as discussed in the previous, a classical test theory framework may be less suited. Both situations can be handled in a relatively straightforward way within an IRT framework because it easily allows for incomplete designs and for equality constraints between item parameters.

Graphical Models for Investigating the Computational Complexity of Multidimensional Item Response Theory (IRT) Models: Principles and Leading Example

Statement of the Problem

At the end of the previous section, it was argued that IRT is the statistical modeling framework of choice when implementing a repeated measurement linking design. A repeated measurement IRT model should in principle be equipped to accommodate two sources of individual differences. First, insofar not all individuals change at the same rate for a given construct, that construct will give rise to multiple dimensions over time. Second, unlike physical measures such as height, achievement measures may constitute different sources of individual differences and hence give rise to multiple dimensions within a given measurement occasion. Individuals may change at a different rate on each of these dimensions over time, giving rise to a high dimensional space for the joint collection of measures across all measurement occasions.

Several further complications that are not discussed in this paper may arise. An item may be an indicator of multiple constructs, as opposed to the simple structure assumed in this paper in which every item is an indicator of a single dimension. Furthermore, the number of dimensions and the degree to which they are represented in a given assessment may change over time (i.e., construct shift; Martineau, 2006).

For now, let’s keep to the assumptions of simple structure and the lack of construct shift. Even in this simplified situation, technical challenges arise for high dimensional IRT models. In a nutshell, because item responses are discrete variables, one cannot rely on linear (mixed) models as a statistical framework. Instead, IRT can be modeled as generalized and nonlinear mixed models (Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003).

Maximum likelihood estimation of model parameters in nonlinear mixed models involves numerical integration over the space of all random effects, for which no closed form solution is available (Tuerlinckx, Rijmen, Verbeke, & De Boeck, 2006). Brute force numerical integration over the joint space of all latent variables becomes computationally very demanding as the number of dimensions grows exponentially with measurement occasions. The exponential increase in dimensionality also quickly becomes prohibitive for the naïve application of Monte Carlo integration techniques, and for Markov chain Monte Carlo techniques in a Bayesian framework.

As an alternative, so-called limited information techniques can be used to estimate the parameters of multidimensional latent variable models for categorical data (Jöreskog, 1994; Muthén, 1984). Limited information techniques have been developed in the field of structural equation modeling. Unlike maximum likelihood estimation methods, the limited information techniques do not take into account the complete joint contingency table of all categorical manifest variables, but only marginal tables up to the fourth order (Mislevy, 1985).

Notwithstanding the widespread use of limited information techniques and ongoing efforts for further improvements in these methods, one can safely assume that many researchers would prefer or at least consider full information maximum likelihood estimation methods if they were to converge to a solution in reasonable time.

Often, the researcher will have a set of assumptions about the underlying structure of the multidimensional latent space. That is, rather than assuming that everything is related to everything in a completely unconstrained way, the correlational structure between dimensions may be assumed to stem from an underlying set of basic relations. For example, it is often reasonable to assume that the association between two different ability dimensions at two different measurement occasions (e.g., geometry at Measurement Occasion 1 and algebra at Measurement Occasion 2) can be accounted for by the associations between those dimensions within a given measurement occasion (geometry at Measurement Occasion 1 and algebra at Measurement Occasion 1, geometry at Measurement Occasion 2 and algebra at Measurement Occasion 2) on the one hand, and the association between different measurements of the same ability over time (geometry at Measurement Occasion 1 and geometry at Measurement Occasion

2, algebra at Measurement Occasion 1 and algebra at Measurement Occasion 2). Obviously, incorporating such a set of conditional independence assumptions, if corroborated by the data, is preferable in that it provides a more parsimonious and hence more easily interpreted statistical model.

The crucial tenet of this paper is to show how incorporating conditional independence assumptions not only results in a more parsimonious statistical model, but can also be exploited for the purpose of parameter estimation. In particular, the set of conditional independence relations implied by a model can often be used to partition the joint space of all latent variables into smaller subsets that are conditionally independent. As a consequence, brute force numerical integration over the joint latent space can be replaced by a sequence of integrations over smaller subsets of latent variables. In the context of Monte Carlo techniques, sampling schemes can be constructed in an analogous way, which will be more efficient than their naïve counterparts (Chib, 1996; Scott, 2002). The gain in efficiency may be dramatic in some cases.

In the following sections, it will be explained how graphical models can be used to obtain a general procedure for partitioning the joint space of all latent variables into smaller subsets that are conditionally independent. A thorough account of the general procedure involves a substantial amount of graph theory and is outside the scope of this paper. The main results will be stated without proof. Instead, a more intuitively based account is presented. The interested reader is referred to Cowell, Dawid, Lauritzen, and Spiegelhalter (1999) for a more in-depth account of graphical models. (For the use of graphical models in the context of latent variable models, see Rijmen, Vansteelandt, & De Boeck, 2008; Rijmen, Ip, Rapp, & Shaw, 2008; Rijmen, 2009a, 2009b, 2010; and Jeon & Rijmen, 2010.)

Representing the Model by a Directed Acyclic Graph

A first step is to represent the statistical model in a directed acyclic graph in which the nodes correspond to random variables and the directed edges represent conditional dependence relations. Directed acyclic graphs have been used extensively in the literature to visualize statistical models. They offer a convenient way of representing and communicating the structure of a statistical model.

Let's illustrate the use of directed acyclic graphs with an overly simplistic model for repeated measurements of achievement. In this, we assume a unidimensional IRT model for each measurement occasion. For the current purpose, there is no need to choose a specific IRT model.

It simply specifies a conditional probability distribution $\Pr\left(\mathbf{Y}_{it} = \mathbf{y}_{it} \,\middle|\, Z_{it} = z_{it}\right)$ at each measurement occasion, where $\mathbf{y}_{it} = \left(y_{it1}, \ldots, y_{itj}, \ldots, y_{itJ}\right)'$ denotes the response vector of person i (i = 1,…, n) at occasion t (t = 1,…, T), z_{it} denotes the position of person i on the latent variable for Measurement Occasion t, and capitals denote the corresponding random variables.

Furthermore, it is assumed that the latent variable at Measurement Occasion t depends on the past through the latent variable at Measurement Occasion t-1 only,

$$\Pr\left(Z_{it} = z_{it} \,\middle|\, Z_{i1} = z_{i1}, \ldots, Z_{i\,t-1} = z_{i\,t-1}\right) = \Pr\left(Z_{it} = z_{it} \,\middle|\, Z_{i\,t-1} = z_{i\,t-1}\right).$$

This is the so-called (first-order) Markov assumption.

A graphical representation of the model is given in Figure 1 for T = 3. In the graph, the conditional dependence of $\mathbf{Y}_{it}$ on Z_{it} is represented by the edges from Z_{it} to $\mathbf{Y}_{it}$ for t = 1, …, 3. It is said that Z_{it} is a (the) parent of $\mathbf{Y}_{it}$. Analogously, the edges from $Z_{i\,t-1}$ to Z_{it} for t = 2, 3 represent the conditional dependence of Z_{it} on $Z_{i\,t-1}$. A directed graph represents certain conditional independence relations as well. For example, the first-order Markov assumption is implied by the directed acyclic graph in Figure 1 by the fact that it shows a directed edge from Z_{i1} to Z_{i2}, and a directed edge from Z_{i2} to Z_{i3}, but no directed edge from Z_{i1} to Z_{i3} (and no other paths from Z_{i1} to Z_{i3}).

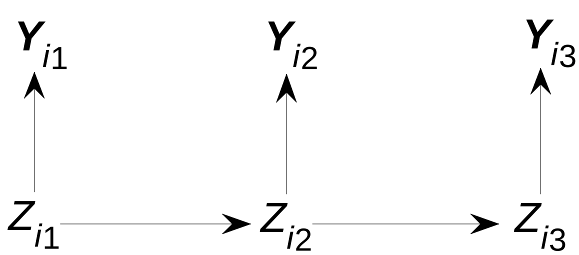

***Figure 1*. Directed acyclic graph representing a model with one latent variable at each of three measurement occasions and incorporating a first-order Markov assumption across measurement occasions.**

The reason that directed acyclic graphs from a convenient way of representing a statistical model is that the joint probability function of all (latent and observed) variables always can be factorized into a set of conditional probability functions according to the directed acyclic graph. Formally, for a set of random variables X_1, …, X_m, …, X_M,

$$\Pr\left(\mathbf{x}\right)=\prod_{m=1}^{M}\Pr\left(x_m\left|pa\left(x_m\right)\right.\right), \tag{1}$$

where $pa(x_m)$ denotes the realization set of variables that are parents of X_m in the directed acyclic graph. For our leading example, factorizing the joint probability $\Pr\left(\mathbf{y}_i,\mathbf{z}_i\right)$ according to the graph results in

$$\Pr\left(\mathbf{y}_i,\mathbf{z}_i\right)=\Pr\left(z_{i1}\right)\Pr\left(\mathbf{y}_{i1}\left|Z_{i1}=z_{i1}\right.\right)\prod_{t=2}^{T}\Pr\left(z_{it}\left|Z_{it-1}=z_{it-1}\right.\right)\Pr\left(\mathbf{y}_{it}\left|Z_{it}=z_{it}\right.\right) \tag{2}$$

where $\mathbf{y}_i=\left(\mathbf{y}_{i1}',\ldots,\mathbf{y}_{it}',\ldots,\mathbf{y}_{iT}'\right)'$ and $\mathbf{z}_i=\left(z_{i1},\ldots,z_{it},\ldots,z_{iT}\right)'$.

Note that the results presented further on require the latent variables to be discrete. However, the latent variables in most IRT models are continuous variables. Therefore, each Z_{it} is to be considered as a discrete approximation of a continuous latent variable θ_{it}. This is not a strong limitation of the approach. As a matter of fact, replacing the vector of continuous latent variables $\boldsymbol{\theta}_i=\left(\theta_{i1},...,\theta_{iT}\right)'$ with a vector of discrete latent variables $\mathbf{Z}_i=\left(Z_{i1},...,Z_{iT}\right)'$ is tantamount to what is done when evaluating the integral over $\boldsymbol{\theta}_i$ using numerical integration. That is, from a computational viewpoint, there is no difference at all between having $\boldsymbol{\theta}_i$ in the model formulation and approximating the integrals over $\boldsymbol{\theta}_i$ through numerical integration over a discrete grid $\mathbf{Z}_i$ on the one hand, and approximating the model through the estimation of its discrete counterpart incorporating $\mathbf{Z}_i$ on the other hand.

Maximum likelihood estimation involves the computation of the marginal probabilities of the response vectors:

$$\Pr\left(\mathbf{y}_i\right)=\sum_{\mathbf{z}_i}\Pr\left(\mathbf{z}_i\right)\prod_{t=1}^{T}\Pr\left(\mathbf{y}_{it}\left|Z_{it}=z_{it}\right.\right), \tag{3}$$

where the summation is over all possible trajectories $\mathbf{z}_i$ in the latent space. Clearly, calculating the marginal probabilities through a direct application of Equation 3 becomes computationally intractable for large *T*, as the number of possible trajectories is exponential in the number of measurement occasions.

Luckily, exploiting the first-order Markov assumption, the marginal probabilities can be calculated far more efficiently by partitioning the joint latent space $\mathbf{Z}_i$ into subsets that are conditionally independent of each other and carrying out a sequence of computations on those subsets. Here, graphical model theory shows its true benefits.

Transforming the Directed Acyclic Graph

The core of the construction of efficient computational schemes relies on the transformation of a directed acyclic graph into a triangulated graph and the subsequent construction of a junction tree.

A first step is transforming the directed acyclic graph into an undirected graph. The undirected graph is called the moral graph. It is obtained by adding an undirected edge between all nodes with a common child that are not yet joined and dropping directions from all edges. Figure 2 displays the moral graph of the directed acyclic graph of Figure 1.

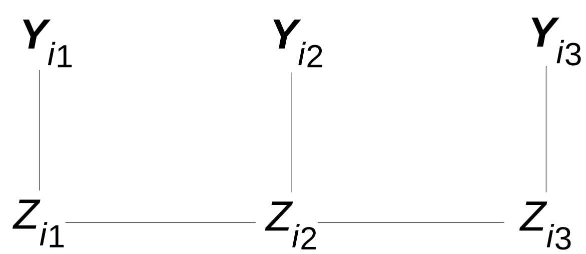

Figure 2. **Moral graph for a model with one latent variable at each of three measurement occasions and incorporating a first-order Markov assumption across measurement occasions.**

The moralization step ensures that a probabilistic model that satisfies the conditional independence relations implied by a directed acyclic graph also satisfies the conditional independence relations implied by the undirected moral graph of the directed acyclic graph. In the process of moralization, conditional independence relations that were implied by the directed acyclic graph might lose their representation in the moral graph by the process of adding edges.

Second, the moral graph is triangulated by adding edges so that chordless cycles contain no more than three nodes. A chordless cycle is a cycle in which there are only edges between consecutive nodes. In general, a triangulated graph can be obtained in many different ways, but one tries to add as few edges as possible to retain the graphical representation of the conditional independence relations that were implied by the directed acyclic graph. Finding an optimal triangulation is nondeterministic polynomial-time (NP) hard (Yannakakis, 1981; for the reader not familiar with computational complexity theory, NP-hard is *very hard*), but well performing heuristic algorithms are available (Kjærulff, 1992). The moral graph in Figure 2 contains only cycles with two nodes and thus is already triangulated.

A graph being triangulated is a necessary and sufficient condition for the existence of an associated junction tree. A tree is a graph whose undirected version (obtained by dropping all the directions from the edges) has a path between all pairs of nodes and has no cycles. In a junction tree, the nodes correspond to cliques. Cliques are complete subsets of nodes. A set of nodes is complete if there is an edge between every pair of nodes. The intersection between two neighboring cliques C_k and C_l is called a separator, $S_{kl} = C_k \cap C_l$.

A junction tree possesses the running intersection property: The intersection $C_k \cap C_l$ of a pair C_k, C_l of cliques is contained in every node on the unique path in the junction tree between C_k and C_l. Figure 3 shows a junction tree of cliques obtained from the triangulated moral graph of Figure 2. Again, more than one junction tree can be constructed in general.

Factorizing the Joint Probability Function According to the Junction Tree

A crucial result is that a junction tree offers an alternative factorization of the joint probability function. In particular, the joint probability can be factorized as the product of all marginal clique probabilities over the product of all marginal separator probabilities:

$$\Pr(\mathbf{x}) = \frac{\prod_C \Pr(\mathbf{x}_C)}{\prod_S \Pr(\mathbf{x}_S)}, \qquad (4)$$

where $\mathbf{x}_c$ and $\mathbf{x}_s$ denote realizations of the random variables that constitute clique *c* and separator *s*, respectively. Applying this result to our leading example, the probability of the complete data vector $\Pr\left(\mathbf{y}_i, \mathbf{z}_i\right)$ can be written as

$$\begin{aligned}\Pr\left(\mathbf{y}_i, \mathbf{z}_i\right) &= \frac{\left(\prod_{t=1}^{T} \Pr\left(\mathbf{y}_{it}, z_{it}\right)\right)\left(\prod_{t=2}^{T} \Pr\left(z_{it-1}, z_{it}\right)\right)}{\left(\prod_{t=1}^{T} \Pr\left(z_{it}\right)\right)\left(\prod_{t=2}^{T-1} \Pr\left(z_{it}\right)\right)} \\ &= \left(\prod_{t=1}^{T} \Pr\left(\mathbf{y}_{it} \middle| z_{it}\right)\right) \Pr\left(z_{i1}\right) \prod_{t=2}^{T} \Pr\left(z_{it} \middle| z_{it-1}\right)\end{aligned} \tag{5}$$

The last line shows that Equation 5 is equivalent to Equation 2 indeed.

***Figure 3.* Junction tree obtained from the triangulated graph in Figure 2.**

The factorization of Equation 5 serves as the basis for a computational scheme using local computations that are carried out on the cliques and separators of the junction tree in a sequential way. This scheme can be incorporated within an EM-algorithm, resulting in an efficient EM-algorithm. The algorithm is efficient in the sense that it circumvents the brute force integration over the joint space of all latent variables that is carried out in the E-step of a traditional EM-algorithm. The complexity of the efficient EM-algorithm scales with the number of latent variables (and the number of categories for each latent variable) within the cliques, as opposed as with the total number of latent variables. For our leading example, the number of computations is of order $2 \times T \times S^2$ (with S the number of categories for each latent variable Z_{it}) for the E-step of the efficient EM-algorithm, as opposed to S^T for the E-step of a traditional EM-algorithm. So, a complexity that is exponential in the number of measurement occasions is reduced to a complexity that is linear in the number of measurement occasions.

To conclude, the dimensionality of the latent space over which has to be integrated in maximum likelihood estimation is not determined by the number of latent variables per se, but by the dimensionality of the latent spaces of the subsets of variables that are conditionally independent. These subsets are a function of the conditional independence assumptions the researcher is willing to make. Because one can rely on algorithms defined on a graphical representation of the statistical model, the sets of conditionally independent variables can be obtained in an automatic way and for a whole family of statistical models.

Instead of using maximum likelihood estimation methods with numerical integration techniques, one may opt for Monte Carlo integration techniques or even for Markov chain Monte Carlo techniques in a fully Bayesian framework. The factorization of the joint probability function according to the cliques in the junction tree may still be worthwhile in constructing the sampling scheme. A Gibbs sampler based on the junction tree has been proposed by Chib (1996) for the hidden Markov model, and Scott (2002) presented empirical and mathematical results showing that such a Gibbs sampler mixes more rapidly than a traditional Gibbs sampler.

In the next section, the approach is used to determine the complexity for several other multidimensional IRT models for repeated measurements. The models are less restrictive than the model that was used throughout this section as a leading example, in which unidimensionality was assumed within each measurement occasion.

Graphical Models for Investigating the Computational Complexity of Multidimensional Item Response Theory (IRT) Models: Applications

Unidimensional Model Within Measurement Occasions—Bifactor Model With a Markov Structure Model Across Occasions

The model that was used throughout the previous section as a leading example incorporated a unidimensional IRT model within each measurement occasion. The associations between the latent variables across measurement occasions were taken into account by a first-order Markov structure. Under a first-order Markov structure, the association between measurement occasions diminishes the further the measurement occasions are apart. However, it may well be the case that abilities are more stable over time than can be accounted for by the Markov process alone. This possibility can be taken into account by incorporating for all measurement occasions a general latent variable Z_{ig} that is independent of all occasion-specific dimensions. The directed acyclic graph for such a model is presented in Figure 4. This model is a generalization of the bifactor model. Z_{ig} is the general dimension, and the dimensions pertaining to each measurement occasion are the specific dimensions. The corresponding moral graph is presented in Figure 5. Since no cycles have more than three nodes, the graph is already triangulated.

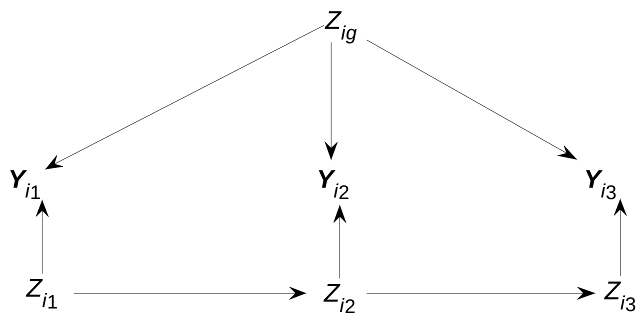

Figure 4. **Directed acyclic graph for a bifactor model with a first-order Markov structure between the specific dimensions.**

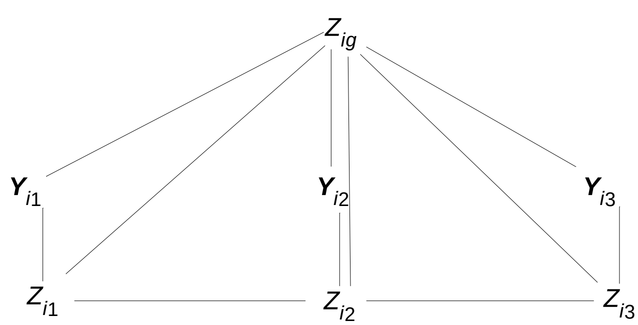

Figure 5. **Moral graph for a bifactor model with a first-order Markov structure between the specific dimensions.**

The maximal subsets of nodes that are all interconnected (the cliques) can be read directly from the triangulated graph in Figure 5. They are the sets $\{Z_{ig}, Z_{it}, Z_{it-1}\}$ for t = 2, …, T, and $\{Z_{ig}, Z_{it}, \mathbf{Y}_{it}\}$ for t =2, …, T. Hence, maximum likelihood estimation involves a sequence of integrations (summations) over three-dimensional latent spaces, which is computationally still feasible.

Instead of a bifactor structure, one could also specify a second-order model. In this model, the general dimension accounts for the additional associations between the occasion specific dimensions. The directed acyclic and moral (triangulated) graphs for such a model are presented in Figures 6 and 7. It is easily verified that the computational complexity of the second-order structure is of the same order as the computational complexity of the bifactor structure (i.e., requires a summation over three-dimensional latent spaces).

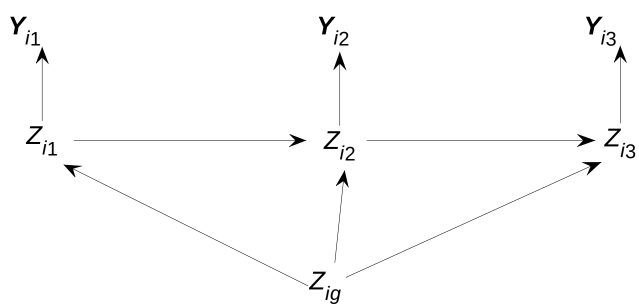

***Figure 6.* Directed acyclic graph for a second-order model with a first-order Markov structure between the specific dimensions.**

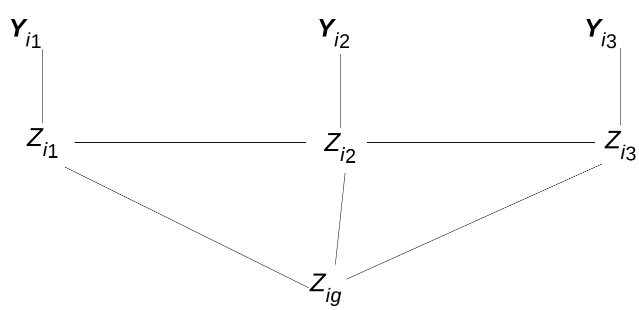

***Figure 7.* Moral graph for a second-order model with a first-order Markov structure between the specific dimensions.**

Bifactor Model Within Measurement Occasions—Markov Structure Model Across Occasions

Let's now turn to the situation where a multidimensional test structure exists within each measurement occasion.

First consider the situation where no prior dimensional structure is assumed for the items within each measurement occasion. In that case, each item depends conditionally on a vector of latent variables $\mathbf{Z}_{it} = \left(Z_{it1},...,Z_{itd},...,Z_{itD}\right)'$, where D is the number of dimensions for the IRT model within a measurement occasion. In a way completely analogous to the case of unidimensional within-occasion models, a first-order Markov structure can be added for each dimension to account for additional dependencies across measurement occasions between items measuring the same dimension. This may be a viable approach when only a couple dimensions D are involved at each measurement occasion, but obviously it does not scale up well with increasing within-measurement occasion multidimensionality.

Therefore, consider the case where one can make simplifying assumptions about the within-occasion dimensional structure. In particular, the case in which a bifactor or second-order structure can be assumed within each measurement occasion is focused upon. Figure 8 displays the directed acyclic graph for a multidimensional model with a bifactor structure within each measurement occasion and a first-order Markov structure defined on the general factor. The figure shows three measurement occasions (T = 3) and three specific dimensions within each measurement occasion (and hence D = 3 + 1 = 4). Indices refer respectively to person, measurement occasion, and dimension. Figure 9 shows the directed acyclic graph for a model with a second-order structure within each measurement occasion. The corresponding moral graphs are shown in Figures 10 and 11. Again, the graphs are already triangulated. For both models, no clique contains more than two latent variables, and hence both models are computationally tractable.

Bifactor Model Within Measurement Occasions—Bifactor Model With a Markov Structure Across Occasions

For the models discussed in the previous section, the first-order Markov structure for the general dimensions is assumed to account for all dependencies over time. Similar to the model presented in the section on unidimensional within-measurement occasion models, this structure can be complemented with a factor that is common to all measurement occasions. The result is a tri-factor model: D specific dimensions within each measurement occasion, T general dimensions across measurement occasions, and one overarching dimension Z_{iG} common to all items within and across measurement occasions.

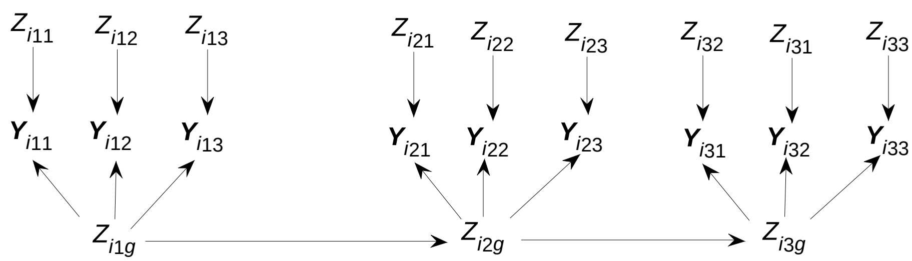

Figure 8. **Directed acyclic graph for a within-occasion bifactor model (*D* = 3 + 1) and a first-order Markov structure for the general dimension over time (*T* = 3). Indices refer respectively to person, measurement occasion, and dimension.**

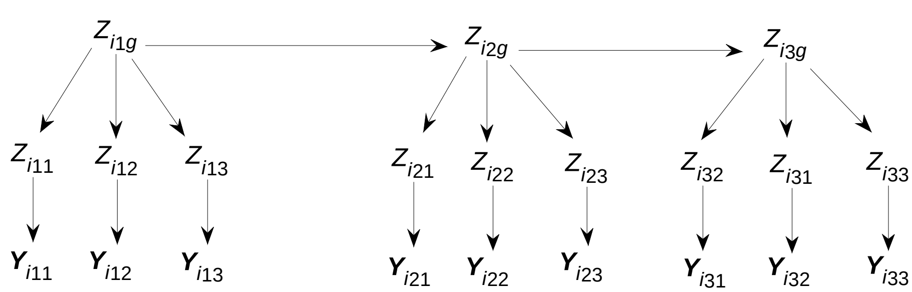

Figure 9. **Directed acyclic graph for a within-occasion second-order model (*D* = 3 + 1) and a first-order Markov structure for the general dimension over time (*T* = 3). Indices refer respectively to person, measurement occasion, and dimension.**

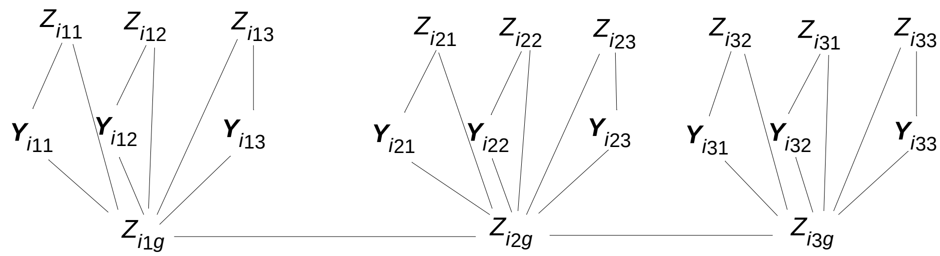

Figure 10. **Moral graph for a within-occasion bifactor model (*D* = 3 + 1) and a first-order Markov structure for the general dimension over time (*T* = 3). Indices refer respectively to person, measurement occasion, and dimension.**

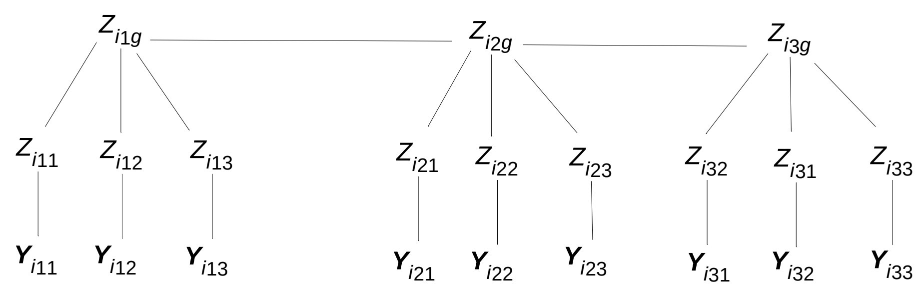

Figure 11. **Moral graph for a second-order model (*D* = 3 + 1) and a first-order Markov structure for the general dimension over time (*T* = 3). Indices refer respectively to person, measurement occasion, and dimension.**

Figures 12 and 13 represent the directed acyclic and moral (triangulated) graphs. It is remarkable that the computational complexity of this model is of the same order as the computational complexity of the unidimensional within-measurement occasion model with a combined bifactor and first-order Markov structure across measurement occasions. For both models, at most three latent variables appear in the same clique.

Again, a similar model could be specified with a higher-order rather than a bifactor structure. Such a model would be of the same computational complexity in that at most three latent variables appear in the same clique.

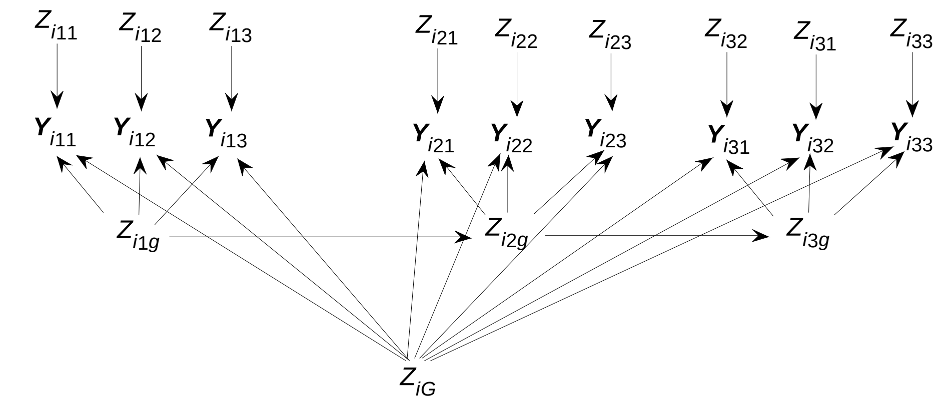

Figure 12. **Directed acyclic graph for a within-occasion bifactor model (*D* = 3 + 1), a first-order Markov structure for the general dimension over time (*T* = 3), and an overarching general dimension. Indices refer respectively to person, measurement occasion, and dimension.**

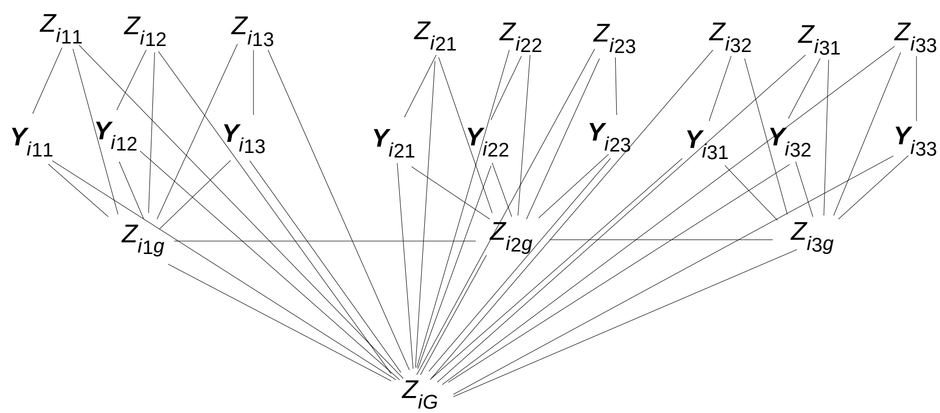

Figure 13. **Moral graph for a within-occasion bifactor model (*D* = 3 + 1), a first-order Markov structure for the general dimension over time (*T* = 3), and an overarching general dimension. Indices refer respectively to person, measurement occasion, and dimension.**

Bifactor Model Within Measurement Occasions—Markov Structure Model Across Occasions for Both General and Specific Dimensions

All models discussed up to now did not require the addition of edges to triangulate the moral graph. Therefore, let's specify a model that does require additional edges during triangulation to illustrate the concept. For this, consider the bifactor within-measurement occasion model with a Markov structure over time for the general dimension. Now, additional first-order Markov structures are assumed for each of the specific dimensions over time. Figure 14 presents the directed acyclic graph, and Figure 15, the moral graph. The moral graph contains cycles consisting of four nodes that are chordless, for example the cycle $Z_{i1g} - Z_{i2g} - Z_{i21} - Z_{i11} - Z_{i1g}$. The subgraph formed by these four nodes and their edges can be triangulated in two ways: adding an edge between Z_{i1g} and Z_{i21}, or adding an edge between Z_{i2g} and Z_{i11} . As mentioned before, a graph can often be triangulated in a variety of ways. For the current example, a heuristic triangulation algorithm that minimizes the total number of latent variables within a clique (Murphy, 2001) was used. This way, the computational complexity for an EM-algorithm carrying out local computations on the latent clique variables is kept at a minimum. The resulting triangulated graph is presented in Figure 16. Edges that were added during triangulation are displayed with dotted lines. It can be seen that an edge was added between Z_{i1g} and Z_{i21} to break the chordless cycle $Z_{i1g} - Z_{i2g} - Z_{i21} - Z_{i11} - Z_{i1g}$. The maximal number of latent variables in a clique is four when both $D = 3$ and $T = 3$. However, unlike all models discussed previously, this number increases with T. For $T = 6$, the largest number of latent variables in a clique was six using the same heuristic triangulation algorithm.

In contrast to all previously presented models, the model with a bifactor within-measurement occasion structure and a first-order Markov structure over time for all dimensions does not scale well with the number of measurement occasions.

The latest example also illustrates how it becomes increasingly complex to transform the directed acyclic graph into a triangulated moral graph by hand. Fortunately, these transformations can be carried out in an algorithmic way and hence carried out by a computer. All graph transformations in this paper were carried out using the Bayes Net Toolbox for Matlab (Murphy, 2001, 2007). In the toolbox, directed acyclic graphs are represented in a matrix, whose $(i, j)^{\text{th}}$ element equals 1 if there is an edge from node i to node j in the directed acyclic graph, and 0 otherwise. Upon specifying the directed acyclic graph in matrix form, one can readily obtain the moral graph, a triangulated graph, its cliques, and the corresponding junction tree using the Bayes net toolbox.

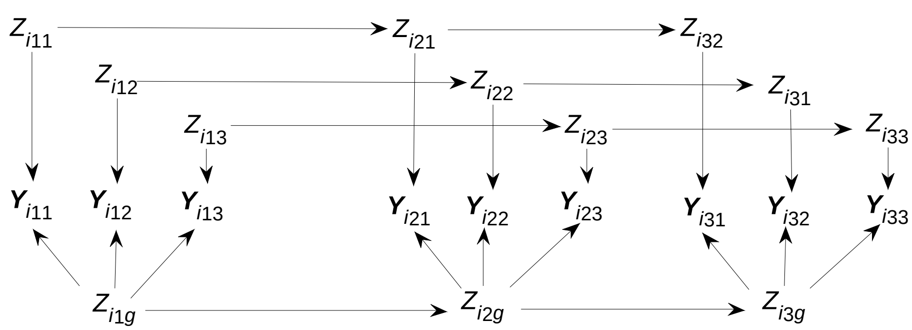

Figure 14. **Directed acyclic graph for a within-occasion bifactor model (*D* = 3 + 1), and first-order Markov structures for both the general and specific dimensions over time (*T* = 3).**

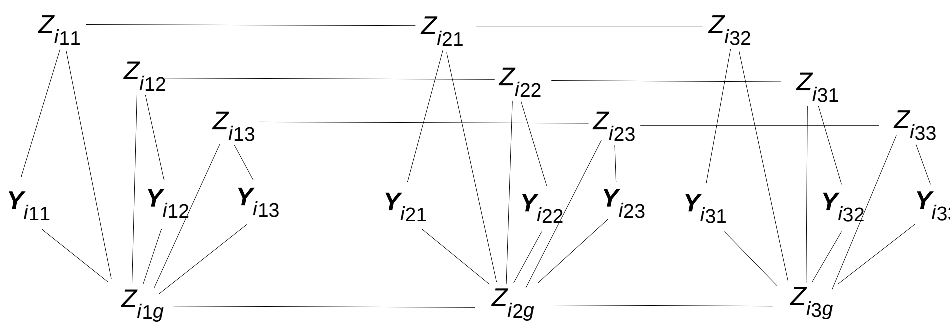

Figure 15. **Moral graph for a within-occasion bifactor model (*D* = 3 + 1) and first-order Markov structures for both the general and specific dimensions over time (*T* = 3).**

Concluding Remarks

In the first part of this paper, a data collection design was presented that was custom-tailored to the context of repeated measurements. The design was a combination of an equivalent groups design and a common-item design. Within each measurement occasion, items could be linked through the use of randomly equivalent groups. Different measurement occasions were linked through the use of common items. The common items were common across measurement occasions but never presented to the same individual twice. This presentation was possible because the design is incomplete at each measurement occasion.

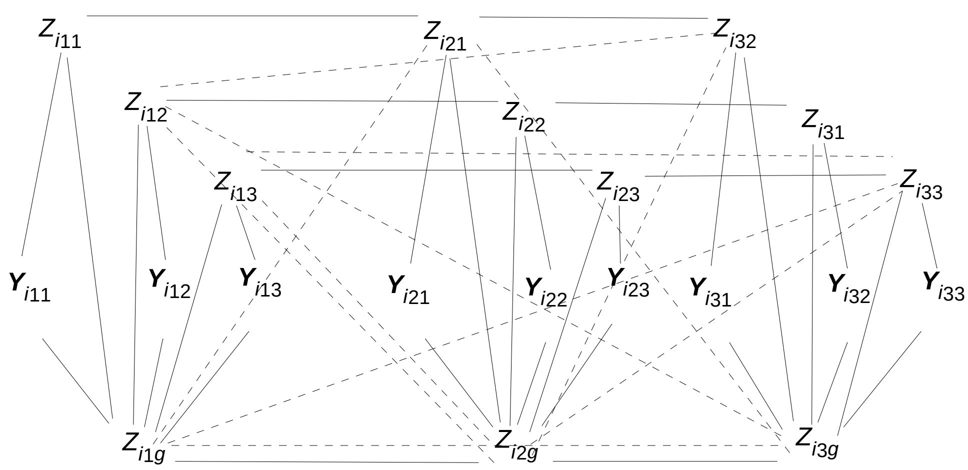

Figure 16. **Triangulated graph for a within-occasion bifactor model (*D* = 3 + 1) and first-order Markov structures for both the general and specific dimensions over time (*T* = 3). Dotted lines represent edges added during triangulation.**

This final section of the paper may be the appropriate place to reiterate that linking procedures all rely on a set of strong assumptions that may or may not be met in particular situations. A crucial assumption in the current context is that the construct one is measuring does not change across measurement occasions. The further apart these measurement occasions are, the less likely the assumption is realistic. Also, this assumption is less likely to be met for some constructs than for others. When these assumptions are not met, a meaningful scale cannot be constructed, no matter how carefully the data collection design is crafted.

The second and more elaborate part of the paper presented an introduction to the use of graphical models in statistical modeling. Graphs have been used for decades to visualize and communicate statistical models. However, the true value of graphical models relies on the fact that the graph representing a statistical model can be transformed in such a way that conditional independence relations implied by the statistical model are rendered explicit. The transformations on the graph are carried out in a completely algorithmic way. Hence, conditional independence relations can be obtained entirely automatically. No tedious algebraic manipulations of the probability function of a statistical model are involved.

Several multidimensional model structures were presented. Using the graphical model framework, it was shown that a high dimensional latent space does not necessarily imply that maximum likelihood (or Bayesian, for that matter) estimation procedures become computationally infeasible. Indeed, if one is willing to make certain conditional independence assumptions during model specification, these assumptions can be exploited to partition the high dimensional latent space into subspaces of lower dimensionality. Full-information maximum likelihood estimates can then be obtained by carrying out computations locally on these subsets.

The focus of this paper was primarily on deriving the sets of conditionally independent (latent) variables for various model structures using graphical models. In addition, through the use of junction trees, graphical model theory can also be used to construct an efficient EM-algorithm for a particular statistical model at hand. The algorithm is efficient in that posterior probabilities are computed in the E-step of the algorithm in a way that maximally exploits the conditional independence relations between them. Such an algorithm is presented by Rijmen, Vansteelandt, et al. (2008) and Rijmen (2009a).

References

Braun, H., & Wainer, H. (2007). Value-added modeling. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 867-892). Amsterdam, Netherlands: Elsevier.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics, 75,* 79–97.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems.* New York, NY: Springer.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London, England: Edward Arnold.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.

Jeon, M., & Rijmen, F. (2010, April). *Assessing differential item functioning for testlet-based tests using the bifactor model*. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika, 59*, 381–389.

Kjærulff, U. (1992). Optimal decomposition of probabilistic networks by simulated annealing. *Statistics and Computing, 2,* 7–17.

Martineau, J. A. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, *31*(1), 35–62.

Mislevy, R. J. (1985). *Recent developments in the factor analysis of categorical variables* (ETS Research Rep. No. RR-85-24). Princeton, NJ: ETS.

Murphy, K. (2001). The Bayes Net Toolbox for Matlab. *Computing Science and Statistics, 33,* 331–350.

Murphy, K. (2007). Bayes Net Toolbox for Matlab. Available from http://code.google.com/p/bnt/

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis method* (2nd ed.). Newbury Park, CA: Sage.

Rijmen, F. (2009a). *An efficient EM algorithm for multidimensional IRT models: Full information maximum likelihood estimation in limited time* (ETS Research Rep. No. (RR-09-03). Princeton, NJ: ETS.

Rijmen, F. (2009b). *Three multidimensional models for testlet based tests: Formal relations and an empirical comparison* (ETS Research Rep. No. RR-09-37). Princeton, NJ: ETS.

Rijmen, F. (2009c). *Horizontal and vertical linking in a longitudinal design* (ETS Research Memorandum RM-09-03). Princeton, NJ: ETS.

Rijmen, F. (in press). Formal relations and an empirical comparison between the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement.*

Rijmen, F., Ip, E. H., Rapp, S., & Shaw, E. G. (2008). Qualitative longitudinal analysis of symptoms in patients with primary and metastatic brain tumors. *Journal of the Royal Statistical Society, Series A, 171,* 739–753.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods, 8*, 185–205.

Rijmen, F., Vansteelandt, K., & De Boeck, P. (2008). Latent class models for diary method data: Parameter estimation by local computations. *Psychometrika, 73,* 167–182.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Scott, S. L. (2002). Bayesian methods for hidden Markov models, recursive computing in th21st century. *Journal of the American Statistical Association, 97,* 337–351.

Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology, 59,* 225–255.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York, NY: Springer-Verlag.

Yannakakis, M. (1981). Computing the minimum fill-in is NP-complete. *SIAM Journal on Algebraic and Discrete Methods, 2*, 77–79.