Published in final edited form as:

Br J Math Stat Psychol. 2010 May ; 63(Pt 2): 395–416. doi:10.1348/000711009X466835.

Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models

Ip Edward Haksing*

Wake Forest University, Winston-Salem, North California, USA

Abstract

Multidimensionality is a core concept in the measurement and analysis of psychological data. In personality assessment, for example, constructs are mostly theoretically defined as unidimensional, yet responses collected from the real world are almost always determined by multiple factors. Significant research efforts have concentrated on the use of simulated studies to evaluate the robustness of unidimensional item response models when applied to multidimensional data with a dominant dimension. In contrast, in the present paper, I report the result from a theoretical investigation that a multidimensional item response model is empirically indistinguishable from a locally dependent unidimensional model, of which the single dimension represents the actual construct of interest. A practical implication of this result is that multidimensional response data do not automatically require the use of multidimensional models. Circumstances under which the alternative approach of locally dependent unidimensional models may be useful are discussed.

1. Introduction

In contrast to biomedical and other physical measurements, which usually focus on a single and relatively well-defined construct, testing and measurement in psychology and education inherently require a multitude of items to operationalize and quantify a construct of interest that is often neither crisp nor unambiguously defined. The classical test theory quickly found its limits for handling the increasingly heterogeneous test designs and item structures. As a result, item response theory (IRT; Lord, 1980; Rasch, 1966) has fittingly emerged as a contemporary tool of choice for measurement, and to a certain extent for explanation, in psychological and educational testing (De Boeck & Wilson, 2004; Embretson & Reise, 2000).

Partly because of its simplicity and mathematical elegance, unidimensional IRT has historically been predominantly used across psychological and educational research. Unidimensional IRT in its basic form, however, has many limitations. It assumes that each item within a test measures the same construct (the unidimensionality assumption), and also that item responses, given the latent construct, are conditionally independent (the local item independence assumption). While test designers generally strive to create tests that target a

*Correspondence should be addressed to Dr Edward Haksing Ip, Medical Center Boulevard, WC23, Winston-Salem, NC 27157, USA (eip@wfubmc.edu).

single construct, in practice it is rare to find a test that is purely unidimensional, at least to the extent that the test possesses sufficient ‘substantive breadth’ (Cattell, 1966; Reise, Morizot, & Hays, 2007) to be useful. The local independence assumption has also been found to be too stringent in many testing situations (Yen, 1993).

Motivated by a broad range of applications, the heavy reliance of psychometric research on traditional IRT has changed significantly over the past two decades. Specifically, considerable advances have been made along many fronts, two of which are particularly pertinent to this paper. First, IRT models have been greatly expanded to relax the stringent assumption of local independence (Bradlow, Wainer, & Wang, 1999; Braeken, Tuerlinckx, & De Boeck, 2007; Douglas, Kim, Habing, & Gao, 1998; Hoskens & De Boeck, 1997; Ip, 2000, 2002; Ip, Smits, & De Boeck, 2009; Ip, Wang, De Boeck, & Meulders, 2004; Jannarone, 1986; Rosenbaum, 1988; Scott & Ip, 2002; Stout, 1990; Wang & Wilson, 2005; Wilson & Adams, 1995). Second, accompanied by the arrival of software such as NOHARM, TESTFACT, ConQuest, and Mplus, methods for fitting multidimensional IRT (MIRT) models to response data have become better developed (Bock, Gibbons, & Muraki, 1988; Gibbons & Hedeker, 1992; McDonald, 1985; Reckase, 1997; Reckase & McKinley, 1991; Samejima, 1974; Segall, 1996).

These two literatures have largely evolved independently of one another, and justifiably so. While unidimensionality and local independence are conceptually related, they are unequivocally distinct mathematical entities. To illustrate the distinction between multidimensionality and local independence, consider a test that is deemed unidimensional in its content, and yet is designed in such a way that a current response is dependent upon earlier responses – for example, when there is a learning effect; see the dynamic model proposed by Verhelst and Glas (1993). The test would be unidimensional but not locally independent. Conversely, a test can possess two dimensions, and yet its items could be locally independent given *both* of the latent traits representing the respective dimensions.

In this paper, I report results that show a direct connection between the two bodies of research. It is shown that an MIRT model is empirically indistinguishable (to be formally defined later) from a locally dependent, unidimensional item response model. In layman’s language, if an analyst is only given the response data matrix but not access to the source of the data, he or she cannot tell from the distributions of the response data alone whether the data have been generated from a locally dependent unidimensional model or from an MIRT model. A formal mathematical relation between the two models is presented in this paper.

2. Background

To see the practical implications of the empirical indistinguishability results, one needs to understand the precursors in the current literature regarding the MIRT and the locally dependent IRT. The starting-point for discussing the precursors is the recognition that unidimensionality is more of an abstract ideal than a reality. Achievement and psychological tests, from a validity perspective, are almost always multidimensional. The balance between dimensionality and validity was acknowledged in work on factor methods dating from as early as the 1920s and 1930s (Holzinger & Swineford, 1937; Kelley, 1928; Spearman,

1933). Kelley (1928, chap. 1) maintained that the designation of a trait as a category of mental life requires the inclusion of all measurements that are ‘definable and verifiable’. Humphreys (1986) highlighted the tension between unidimensionality and validity by going as far as to suggest that tests should be deliberately constructed to include numerous minor factors in addition to the dominant dimension. In personality assessment, Ozer (2001) contended that it is ‘exceedingly difficult’ to achieve structural validity of unidimensionality because ‘most constructs are theoretically defined as unidimensional, but item responses, as individual behaviours in their own right, are usually multiply determined’. In fact, it is hard to argue that truly valid unidimensional tests exist in any subject matter area. Therefore, it may even be fair to assert that (to the credit of Milton Friedman) multidimensionality is *always* and *everywhere* a validity phenomenon.

There are several extant approaches to resolve this validity-versus-unidimensionality dilemma. The first strategy is to use unidimensional IRT as an ‘approximation’ model for item responses that are deemed not strictly unidimensional. A substantial literature exists in addressing the ‘what can go wrong?’ question through simulation experiments (Ackerman, 1989; Ansley & Forsyth, 1985; Drasgow & Parson, 1983; Folk & Green, 1989; Harrison, 1986; Junker & Stout, 1994; Kim, 1994; Kirisci, Hsu, & Yu, 2001; Reckase, 1979; Reckase, Carlson, Ackerman, & Spray, 1986; Spencer, 2004; Walker & Beretvas, 2003; Way, Ansley, & Forsyth, 1988).

As summarized by Gibbons, Immekus, and Bock (2007), two important findings appeared to emerge from this literature. If there is a predominant general factor in the data, and if the dimensions beyond that major dimension are relatively small, the presence of multidimensionality has little effect on item parameter estimates and the associated ability estimates. If, on the other hand, the data are multidimensional with strong factors beyond the first one, unidimensional parameterization results in parameter and ability estimates that are drawn towards the strongest factor in the set of item responses (this tendency is ameliorated to some extent if the factors are highly correlated). The ability estimate tends to be a weighted composite of the measures from each individual dimension. For a critical review, see Goldstein and Wood (1989).

The second approach to the validity-versus-unidimensionality dilemma is to first determine the dimension of a test – empirically or relying on expert knowledge (McDonald, 1981, 1985) – and to judiciously select an MIRT model for fitting the response data at hand. As MIRTs are not created equal, different variants of MIRT can be considered. For example, if an item within a test can only be loaded on one dimension, then one can use a so-called between-item MIRT model (Adam, Wilson, & Wang, 1997). As opposed to the standard IRT model represented in Figure 1a, Figure 1b shows a between-item MIRT model. The leftmost four items in Figure 1b belong to one dimension (represented by a latent variable, which is depicted as an oval in the graph), while the remaining two belong to another distinct dimension. The two dimensions can be correlated (indicated by the double-headed arrow). Alternatively, one can fit a bifactor model (Gibbons & Hedeker, 1992) in which a general factor underlies all items, and two or more group factors. Figure 1c shows the structure of a bifactor model, of which each item has at most two dimensions – a generic factor and one of many group factors that correspond to specified mutually exclusive subsets of items (here

the terms ‘dimension’ and ‘factor’ are used interchangeably). This kind of item-level bifactor pattern (Muthén, 1989) is especially useful for tests that contain a general underlying factor (e.g. general reading ability) and clearly identifiable domains (e.g. reading to achieve a special purpose such as information gathering).

A third approach, which admittedly is ‘the road less travelled’, is to fit a locally dependent unidimensional IRT model to the data. The argument for this approach follows from the observation that when there exist identifiable domains within a test, items will be locally dependent within a domain, but locally independent between domains. Figure 1d shows the structure of a locally dependent unidimensional model that corresponds to the bifactor model in Figure 1c. In a locally dependent model, the conditional covariance matrix of the item responses is non-diagonal given the general factor, and it can be subject to further modelling. Presumably, because of a lack of understanding of how locally dependent IRT models function, this approach is not commonly adopted for analysing potentially multidimensional data.

Simply (and graphically) put, the results reported in the present paper state that (i) one cannot distinguish, at least empirically, the model represented by Figure 1c and that represented by Figure 1d, and (ii) the strength of the local dependencies (double-headed arrows in Figure 1d) can be delineated. This is directly relevant to all of the three strategies above. First, in situations in which the first strategy is employed, one can use the numerical result (ii) to explore the impact of multidimensionality on the parameters of unidimensional models. Second, and more importantly, the finding (i) could be used to inform the second strategy and to provide theoretical justification for the third. I will elaborate these points in Section 7. Let us now begin the formal derivation by first describing the necessary mathematical set-up.

3. Multidimensional item response model

Following Reckase (1997), a basic form of the compensatory MIRT model is given by

$$P(Y_{ij} = 1 | \underline{\Theta}_j) = \frac{\exp(\underline{a}_i^T \underline{\Theta}_j - d_i)}{1 + \exp(\underline{a}_i^T \underline{\Theta}_j - d_i)}, \quad (1)$$

where Y_{ij} is the binary response of person j to item i , $\underline{a}_i$ is a vector of item parameters, $\underline{\Theta}_j$ is a vector of latent traits of dimension $q - 2$, and d_i is a parameter related to the difficulty of the item. Note that in contrast to the usual convention, the negative sign is used for d_i so that it can later be compared to a locally dependent IRT model. A more general form of the MIRT model can be expressed as $P(Y_{ij} = 1 | \underline{\Theta}_j) = g^{-1}(\underline{a}_i^T \underline{\Theta}_j - d_i)$. The function g^{-1} is often referred to as an inverse link function (McCullagh & Nelder, 1989). The focus is on the probit link (i.e. $g^{-1} = \Phi$, where Φ is the standard normal cumulative distribution) and the logit link $g^{-1}(u) = \exp(u) / [1 + \exp(u)]$.

For the purpose of illustration, consider a two-dimensional IRT binary response model with a probit link:

$$P(Y_{ij} = 1 | \underline{\Theta}_j) = \Phi(a_{i1}\theta_{j1} + a_{i2}\theta_{j2} - d_i), \quad (2)$$

where $\underline{\Theta}_j = (\theta_{j1}, \theta_{j2})$, a_{i1} and a_{i2} are item discrimination parameters along dimensions 1 and 2 respectively, and d_i is the item difficulty parameter. The strength with which an item measures each dimension can be summarized by the angular direction $\cos^{-1}(a_{i1}/\sqrt{a_{i1}^2 + a_{i2}^2})$. If the angle is less than 45° , then the item measures θ_1 better than it measures θ_2 . Furthermore, assume that the latent score vector follows a bivariate normal distribution:

$$\underline{\Theta}_j = (\theta_{j1}, \theta_{j2}) \sim N(\underline{0}, \Sigma), \quad (3)$$

where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

with a further assumption that $\sigma_1 > 0$ and $\sigma_2 > 0$. For ease of description, let θ_1 be the dimension of interest (hence the assumption $\sigma_1 > 0$). The other dimension θ_2 in the model is treated as a nuisance dimension. Clearly, the model becomes unidimensional when $\sigma_2 = 0$. The distinction between θ_1 and θ_2 is arbitrary, and as we shall see later, the mathematical derivation does not necessitate such a distinction. As a result of (3), the two dimensions are allowed to attain different variances and be correlated with correlation coefficient ρ . Constraints are generally required to maintain identifiability of the model (e.g. σ_1 and σ_2 fixed at specific values; or correlation between dimensions fixed). However, for the purpose of mathematical derivation of the main results, identifiability constraints are not necessary and will therefore not be enforced. The manifest probability is given by integrating out the so-called kernel of the probability distribution – in this case, $\Phi(a_{i1}\theta_{j1} + a_{i2}\theta_{j2} - d_i)$:

$$P(Y_{ij} = 1) = \int \Phi(a_{i1}\theta_{j1} + a_{i2}\theta_{j2} - d_i) \phi(\underline{\Theta}_j) d\underline{\Theta}_j, \quad (4)$$

where $\phi(\cdot)$ denotes the density function of the normal distribution.

4. Locally dependent unidimensional model

The local item dependence (LID) unidimensional item response model used for the purpose of this paper extends the formulation of locally dependent models described in Ip (2002) and Ip *et al.* (2004), and follows the so-called population-averaged approach in the statistics literature (Liang & Zeger, 1986). The population-averaged approach focuses on the marginal

expectation of outcome variables across the population. Recently, Braeken *et al.* (2007) developed a copula approach that is similar in spirit. The model is specified by the following three components:

LID1. The unidimensional kernel of each item response given the subject's latent trait. Often known as the item response function (IRF) in the IRT literature, this is the conditional mean $\mu^*(\theta) = E(Y_{ij}|\theta_j)$ of the response Y_{ij} given θ_j (Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003). A commonly used kernel, which I shall follow in this paper, takes the form of the logistic function:

$$\mu_{ij}^*(\theta) = E(Y_{ij}|\theta_j) = P(Y_{ij} = 1|\theta_j) = \frac{\exp[a_i^*(\theta_j - b_i^*)]}{1 + \exp[a_i^*(\theta_j - b_i^*)]}, \quad (5)$$

where a^* , b^* are the item discrimination and difficulty parameters, respectively.

LID2. The conditional variance function of each item response given θ_j , which is assumed to be some function of the conditional mean:

$$\text{Var}(Y_{ij}|\theta_j) = v(\mu_{ij}^*(\theta_j)). \quad (6)$$

LID3. The residual pairwise associations among the item responses after the effect of the latent trait have been partialled out. This can be specified as pairwise conditional correlations or odds ratios among the set of responses given θ (see McDonald, 1981; Stout *et al.*, 1996). For locally independent IRT, the residual correlation is identically zero.

The specification in condition LID3 allows genuine deviation from the standard local item independence assumption made in IRT. As such, LID3 distinguishes from Stout's notion of essential independence assumption (Stout, 1990), which assumes that the averaged correlation is necessarily zero. By design, the locally dependent unidimensional model specified in conditions (LID1)–(LID3) does not specify the full joint distribution of responses given θ . Because the number of association terms grows exponentially with the number of item responses, it is actually advantageous to avoid the explicit specification of higher-order association (e.g. three-way association between three responses given θ) by following the principle of the marginal model approach (e.g. Fitzmaurice, Laird, & Ware, 2004, p. 319).

5. Main results

5.1. Empirical indistinguishability

Our goal is to show that an MIRT model is 'equivalent' to a locally dependent unidimensional model that is specified by (LID1)–(LID3). To be more precise about what is meant by the term 'equivalent', I provide an operational definition. Suppose a random vector $\underline{Y} = (Y_i)$, $i = 1, \dots, I$ (possibly multidimensional), is generated from reference model Ψ_R .

Denote the corresponding mean and covariance functions of the reference model, assuming that they both exist, by $E_{\Psi_R}(\underline{Y})$ and $\text{Cov}_{\Psi_R}(\underline{Y})$, respectively. In the context of latent-variable modelling, these two quantities are, respectively, called the manifest mean and manifest covariance. Alternatively, consider a comparison model Ψ_A for which both a mean function $E_{\Psi_A}(\underline{Y})$ and a covariance function $\text{Cov}_{\Psi_A}(\underline{Y})$ exist. Then the models Ψ_R and Ψ_A are called weakly empirically indistinguishable (or empirically indistinguishable for short) if their respective manifest mean and covariance functions are identical:

$$E_{\Psi_R}(\underline{Y}) = E_{\Psi_A}(\underline{Y}) \quad (7a)$$

and

$$\text{Cov}_{\Psi_R}(\underline{Y}) = \text{Cov}_{\Psi_A}(\underline{Y}). \quad (7b)$$

Equations (7a) and (7b) represent a weak form of equivalence because they only require equality of the first two moments of the two distributions. One can also call this weak form of equivalence second-order empirical indistinguishability as it concerns only the first two moments. It is noteworthy that in basic item response models, only the first conditional moment is considered. The inclusion of second-order moments sets the stage for models embracing local dependencies.

In the present context, I use the MIRT model as the reference model Ψ_R and the locally dependent unidimensional model as a comparison model Ψ_A . The following key lemma suggests a sufficient condition for establishing empirical indistinguishability between the reference and comparison models.

Lemma 1. Denote the dimension of interest by θ_1 in the MIRT model Ψ_R , and denote the latent trait in the comparison locally dependent unidimensional model Ψ_A also by θ_1 . The mean and covariance are, respectively, denoted by E_* and Cov_* , where $*$ denotes either the reference (R) or the comparison (A) model. The marginal distributions for θ_1 under Ψ_A and Ψ_R are denoted, respectively, by $p_A(\theta_1)$ and $p_R(\theta_1)$. The following conditions are sufficient for Ψ_R and Ψ_A to be (weakly) empirically indistinguishable. For all $\underline{Y}$, θ_1 ,

$$E_{\Psi_R}(\underline{Y}|\theta_1) = E_{\Psi_A}(\underline{Y}|\theta_1), \quad (8a)$$

$$\text{Cov}_{\Psi_R}(\underline{Y}|\theta_1) = \text{Cov}_{\Psi_A}(\underline{Y}|\theta_1), \text{ and } \quad (8b)$$

$$p_R(\theta_1) = p_A(\theta_1). \quad (8c)$$

Proof.

$$E_{\Psi_R}(\underline{Y}) = E_R[E_{\Psi_R}(\underline{Y}|\theta_1)] = E_A[E_{\Psi_A}(\underline{Y}|\theta_1)] = E_{\Psi_A}(\underline{Y}), \quad (9)$$

$$\begin{aligned} \text{Cov}_{\Psi_R}(\underline{Y}) &= E_R \text{Cov}_{\Psi_R}(\underline{Y}|\theta_1) + \text{Cov}_R E_{\Psi_R}(\underline{Y}|\theta_1) \\ &= E_R \text{Cov}_{\Psi_A}(\underline{Y}|\theta_1) + \text{Cov}_R E_{\Psi_A}(\underline{Y}|\theta_1) \quad \text{from conditions (8a), (8b)} \\ &= E_A \text{Cov}_{\Psi_A}(\underline{Y}|\theta_1) + \text{Cov}_{\Psi_R} E_{\Psi_A}(\underline{Y}|\theta_1) \quad \text{from conditions (8c)} \\ &= \text{Cov}_{\Psi_A}(\underline{Y}). \end{aligned} \quad (10)$$

Note that $\text{Cov}_{\Psi_A}(\underline{Y}|\theta_1)$ and $E_{\Psi_A}(\underline{Y}|\theta_1)$ are both functions of θ_1 in the second line of (10).

Thus, the expectations or covariances over the distribution of θ_1 for the two models are equivalent according to condition (8c). The logic applies to (9) as well.

Lemma 2. Given a two-dimensional MIRT model with the logit link in equation (1), there exists an empirically indistinguishable unidimensional locally dependent model that is characterized by (LID1)–(LID3). Specifically, (i) the IRF specified by LID1 is given by

$$\mu_i^*(\theta_j) = P(Y_{ij} = 1 | \theta_j) = \frac{\exp(a_i^* \theta_j - d_i^*)}{1 + \exp(a_i^* \theta_j - d_i^*)}, \quad (11)$$

where

$$a_i^* = \lambda_{\text{logit}} \left(a_{i1} + \frac{a_{i2} \rho \sigma_2}{\sigma_1} \right), \quad d_i^* = \lambda_{\text{logit}} d_i, \quad (12)$$

with $\lambda_{\text{logit}} = [k^2 a_2^2 (1 - \rho^2) \sigma_2^2 + 1]^{-1/2}$, $k = 16\sqrt{3}/(15\pi) = 0.588$, and (ii) the covariance function specified in LID2 and LID3 is given by the equation

$$\text{Cov}(\underline{Y}|\theta_1) = \text{Cov}[E(\underline{Y}|\theta_1, \theta_2)] + E[\text{Cov}(\underline{Y}|\theta_1, \theta_2)]. \quad (13)$$

Both terms on the right-hand side of (13) can be evaluated via numerical integration.

Corollary 1. Approximately, by the Taylor expansion, the conditional variance function (13) can be explicitly derived:

$$\begin{aligned}\text{Var}(Y_i|\theta_1) &= v_i(\theta_1) \\ &\approx \frac{\exp(a_{i1}\theta_1 - d_i)}{[1 + \exp(a_{i1}\theta_1 - d_i)]^2} + \frac{a_{i2}^2\sigma_2^2(1 - \rho^2)[\exp(a_{i1}\theta_1 - d_i)]^2}{[1 + \exp(a_{i1}\theta_1 - d_i)]^4} \\ &\quad + \frac{a_{i2}\rho\theta_1\sigma_2}{\sigma_1}h'(a_{i1}\theta_1 - d_i) \\ &= p_i^*q_i^*[1 + k_1p_i^*q_i^* + k_2\theta_1(q_i^* - p_i^*)],\end{aligned}\quad (14)$$

where $p_i^* = \exp(a_{i1}\theta_1 - d_i)/[1 + \exp(a_{i1}\theta_1 - d_i)]$, $q_i^* = 1 - p_i^*$, $k_1 = a_{i2}^2\sigma_2^2(1 - \rho^2)$, $k_2 = a_{i2}\rho\sigma_2/\sigma_1$, $\sigma_1 > 0, \sigma_2 \geq 0$, $h'(u) = \exp(u) - [\exp(u)]^2/[1 + \exp(u)]^3$, whereas the conditional correlation between item u and item v ($u \neq v$) is given by

$$\text{corr}(Y_u, Y_v|\theta_1) = \frac{s_{uv}(\theta_1)}{\sqrt{v_u(\theta_1)v_v(\theta_1)}}, \quad (15)$$

where $s_{uv}(\theta_1)$ is given by

$$\begin{aligned}s_{uv}(\theta_1) &= \frac{a_{u2}a_{v2}\sigma_2^2(1 - \rho^2)[\exp(a_{u1}\theta_1 + a_{v1}\theta_1 - d_u - d_v)]}{[1 + \exp(a_{u1}\theta_1 - d_u)]^2[1 + \exp(a_{v1}\theta_1 - d_v)]^2} \\ &= a_{u2}a_{v2}\sigma_2^2(1 - \rho^2)p_u^*q_u^*p_v^*q_v^* \quad \text{if } u \neq v.\end{aligned}\quad (16)$$

The proofs of Lemma 2 and Corollary 1 are provided in Appendices A and B.

5.2. The general case of multiple minor traits

I further extend the results to include the case for multiple minor traits in which the nuisance dimensions are denoted by the $(q - 1)$ -vector $\underline{\Theta}_2$. To set up notation, define $\underline{\Theta}^T = (\underline{\Theta}_1, \underline{\Theta}_2^T)^T$ and assume that $\underline{\Theta} \sim N(\underline{0}, \Sigma)$, where the $q \times q$ covariance matrix Σ can be further partitioned into

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \underline{\gamma}^T \\ \underline{\gamma} & \Sigma_2 \end{pmatrix}, \quad (17)$$

where σ_1^2 is the variance of the dimension of interest, $\underline{\gamma}^T$ is a covariance vector of length $q-1$, and Σ_2 is a $(q-1) \times (q-1)$ covariance matrix. Further, let $\underline{a}_i = (a_{i1}, a_{i2}^T)^T$. We have the following corollary to Lemma 2.

Corollary 2. Given a q -dimensional ($q > 2$) MIRT model with the logit link in equation (1), there exists an empirically indistinguishable unidimensional LID model with the kernel function

$$P(Y_{ij} = 1 | \theta_j) = \frac{\exp(a_i^* \theta_j - d_i^*)}{1 + \exp(a_i^* \theta_j - d_i^*)}, \quad (18)$$

where $a_i^* = \lambda_{\text{logit}}[a_{i1} + (1/\sigma_1)a_2^T \underline{\xi}]$ and $d_i^* = \lambda_{\text{logit}} d_i$; $\underline{\xi} = (\xi_m)$, $m = 2, \dots, q$, $\xi_m = \varphi_{1m} \sigma_m$,

$$\lambda_{\text{logit}} = \frac{1}{\sqrt{1 + k^2 \underline{a}_2^T \bar{\Sigma} \underline{a}_2}}, \quad (19)$$

$$\bar{\Sigma} = (\psi_{mn}),$$

$$\psi_{mn} = \begin{cases} \sigma_{m+1}^2 (1 - \varphi_{1m+1}^2), & \text{if } m = n, \\ \sigma_{m+1} \sigma_{n+1} (\varphi_{(m+1)(n+1)} - \varphi_{1m+1} \varphi_{1n+1}), & \text{if } m \neq n, \end{cases}$$

$m, n = 1, \dots, q-1$, where φ_{rs} denotes the population correlation between the s th and the r th dimension, and $k = 16\sqrt{3}/(15\pi)$. A proof of this result is given in Appendix C.

Because the matrix $\bar{\Sigma}$, which is the conditional covariance matrix of $\underline{\Theta}_2$ given θ_1 , is non-negative definite, λ_{logit} is positive and less than one. A tighter bound for λ_{logit} is given by the Raleigh quotient bounds (Abadir & Magnus, 2005, p. 344):

$$1 - \frac{1}{2} \|\underline{a}_2\|^2 \lambda_{q-1} \approx \frac{1}{(1 + \|\underline{a}_2\|^2 \lambda_{q-1})^{1/2}} \leq \lambda_{\text{logit}} \leq \frac{1}{(1 + \|\underline{a}_2\|^2 \lambda_1)^{1/2}} \approx 1 - \frac{1}{2} \|\underline{a}_2\|^2 \lambda_1, \quad (20)$$

where $\lambda_1 \ \lambda_2 \ \dots \ \lambda_{q-1}$ are the (positive) eigenvalues of the matrix $\bar{\Sigma}$, and $\|\cdot\|$ denotes the Euclidean norm. The approximation holds if the minor dimensions are relatively weak (i.e. $\|\underline{a}_2\|^2 \lambda_1$ is small).

As a generalization of (13), the $I \times I$ covariance matrix conditional on the trait of interest θ_1 is given by

$$\text{Cov}(\underline{Y}|\theta_1) = \text{Cov}[E(\underline{Y}|\theta_1, \underline{\Theta}_2)] + E[\text{Cov}(\underline{Y}|\theta_1, \underline{\Theta}_2)], \quad (21)$$

assuming that both terms on the right-hand side of (21) exist. Each of the terms $\text{Cov}[E(\underline{Y}|\theta_1, \underline{\Theta}_2)]$ and $E[\text{Cov}(\underline{Y}|\theta_1, \underline{\Theta}_2)]$ can be computed via numerical integration. For example, the term $E[\text{Cov}(\underline{Y}|\theta_1, \underline{\Theta}_2)] = \text{diag}\{v_{ii}(\theta_1)\}$ can be computed through term-by-term numerical integration:

$$v_{ii}(\theta_1) = \int \frac{\exp(a_{i1}\theta_1 + a_{i2}^T \underline{\Theta}_2 - d_i)}{[1 + \exp(a_{i1}\theta_1 + a_{i2}^T \underline{\Theta}_2 - d_i)]^2} \phi(\underline{\Theta}_2|\theta_1) d\underline{\Theta}_2. \quad (22)$$

The $I \times I$ matrix $\text{Cov}[E(\underline{Y}|\theta_1, \underline{\Theta}_2)]$ generally contains non-zero off-diagonal elements, which can be thought of as reflecting the LID that is being induced by the nuisance dimensions in the MIRT model. Closed-form approximations for the covariance are possible through the use of techniques such as multivariate Taylor expansion, but they will not be further elaborated here.

6. Numerical results

The quality of the approximation in Corollary 1 is evaluated through a comparison of the approximated solution and numerical integration. I conducted comparisons across a broad array of conditions, some of which will be described below. The results showed that the approximations are accurate under mild conditions, but they are not necessarily highly precise across the range of latent traits under more extreme conditions. Space limitations preclude the reporting of all of the comparison results, but Table 1 summarizes four scenarios that were selected to demonstrate the quality of the approximation on a pair of latent traits: (a) standard condition, under which the variance of the dimension of interest is larger than the minor dimension, the correlation between the two dimensions is moderate, and the item discrimination of the dominant dimension is also higher; (b), (c), and (d) are similar to (a) but with the following respective differences: a very high correlation exists between the two dimensions; there is lower discrimination in the dimension of interest; and there are comparable variances in the two dimensions.

Figure 2 shows the IRF of the unidimensional model. As in all subsequent graphs, the solid line in the graph in Figure 2 is obtained through numerical integration. The curve using (11) and (12) is virtually indistinguishable from the curve obtained via numerical integration, and is not shown.

Figures 3 and 4 show the approximations of the two components of the covariance functions in Corollary 1. The Taylor approximation works well in scenarios (a) and (b), but the approximation shows discrepancies from the curve obtained via numerical integration under scenarios (c) and (d), in which either the discrimination of the minor dimension is higher or its variance becomes dominant. The result is not surprising because the Taylor expansions in

(14) and (15) are obtained from linear approximations of the respective functions about the point $\theta_2 = 0$ (see Appendix B), and because their accuracies begin to deteriorate when the linear relations are extrapolated too far out. The term $h'(\cdot)$ in (14) can become especially problematic under scenarios (c) and (d) because it can turn negative. In my experience the approximation actually improves somewhat if the term $h'(\cdot)$ is set to zero when it takes on negative values (Figure 5).

7. Discussion

It sounds like an oxymoron, but by showing that MIRT is empirically indistinguishable from a locally dependent unidimensional model, a salient message that comes out of the theoretical investigation is that multidimensionality does not necessitate the use of multidimensional models.

One circumstance under which locally independent IRT can be useful is when multiple diffused, minor dimensions deemed not to be of substantive interest pervade the entire test. Robust analytic results may not be available (e.g. poor fit to IRT), and MIRT may produce too complex a model that is beyond meaningful interpretation (e.g. 10 or more dimensions are required). In the context of a latent-class model, Reboussin, Ip, and Wolfson (2008) showed that using a locally dependent model could meaningfully improve model fit and successfully solve the so-called misspecification-versus-interpretation dilemma, which refers to the tension between fitting too few (but substantively interpretable) latent classes, leading to model misspecification, and fitting too many, leading to spurious and hard-to-interpret latent classes. It is reasonable to think that the lessons learned there are germane to the circumstance described here.

Curiously, the empirical indistinguishability result in Lemma 2 implies a different approach to ‘composite dimension’ estimation (i.e. fitted to an IRT model and settled with a composite estimate of multiple dimensions as an approximate solution). According to Lemma 2, the minor dimensions can be treated as a nuisance factor such that one can conduct appropriate inference on the ‘purified’ major dimension (i.e. the dimension of interest). From a measurement perspective, obtaining a purified measure that is independent of the content of the items (Bollen & Lennox, 1991) is appealing, because a ‘contaminated’ (composite) factor creates a measurement dilemma, which is that the estimated score is test-specific and its interpretation requires the test itself as a referent. As a reviewer of this paper pointed out, a composite would change depending upon the relative contribution of content facets, and thus the IRT invariance property would not make sense. For example, the unidimensional ability estimate for a quantitative reasoning test that involves a verbal component would lack a global interpretation because it is a function of the extent to which verbal ability is required in the specific test. By using Lemma 2 as a basis for ‘purifying’ the contaminated construct in order to strictly obtain an estimate of the construct of interest (θ_1 in our notation), the interpretation of ability will be invariant across tests. Some work has been done in this direction (e.g. Ip, Goetghebeur, Molenberghs, & De Boeck, 2006).

Yet another implication of the main result is the potential use of a locally dependent, unidimensional model to expand the existing IRT-and MIRT-based methods. Consider the

following example from the National Assessment of Educational Progress (NAEP) on reading comprehension. Scott and Ip (2002) described a between-item MIRT model. The model also accounts for the testlet effect (Wainer, Bradlow, & Wang, 2007).¹ A testlet is a collection of clustered items that are all related to a common theme. Figure 6a shows the graphical representation of the model, which is structurally equivalent to a bifactor model embedded within a between-item MIRT. Here, two reading domains – reading for information and reading for literary experience – are shown. Figure 6a also shows one testlet in the reading for information domain, within which a subset of items clustered around a reading paragraph (a real example of an article about catching blue crabs by George Frame is used in Figure 6). The potential local dependencies between items within a reading paragraph are often not of substantive interest and considered a nuisance factor.

Figure 6b shows an alternative model in which the locally dependent IRT and the between-item MIRT models can be used in conjunction when testlets exist within a domain. A similar hierarchical factor structure can be exemplified by self-reported symptom-assessment data collected from patients with brain tumours (e.g. Rijmen, Ip, Rapp, & Shaw, 2008). In addition to a general underlying factor suggesting overall symptom severity, the symptom items can be partitioned according to bifactor groups (domains) such as memory problems, speech problems, and non-somatic symptoms. The memory domain may further contain a testlet of items that are all related to short-term memory recall. For non-hierarchical data structures, a hybrid model of local dependency and bifactor/MIRT models may serve such data well.

It should be pointed out that sometimes the local dependency itself may be of substantive interest. It is conceivable that within depressive patients the conditional correlation between two depressive symptoms converges with the presence of co-morbidity, and accordingly the correlation could provide insight into possible interventions. From a modelling perspective, the (residual) association, or local dependency, can be directly related to explanatory factors. Ip *et al.* (2009) report an application of such models to aggressive behaviour data. Moreover, negative correlation between items (e.g. between positive mood and negative mood in quality-of-life assessment), which cannot be directly modelled through the use of a second factor, as evidenced from (16), can be captured through a general locally dependent IRT. Programs developed in PROC NLMIXED (SAS, Inc., Cary, NC, USA) for estimating locally dependent models can be found in Ip *et al.* (2004).

I would further make one technical remark about the main results of the present paper. While a locally dependent unidimensional model that is empirically indistinguishable from an MIRT model always exists, it is not unique. It is clear that marginalizing (2) over the minor dimension θ_2 of the MIRT (see also equation (A4)) would produce yet another empirically indistinguishable solution. Generally, the results in (11), (14), and (15) are not symmetric about the dimensions represented by θ_1 and θ_2 .

Although the testlet item response model (Wainer *et al.*, 2007) has been commonly used to accommodate local dependency, it has been shown that it is equivalent to a constrained bifactor model (Li, Bolt, & Fu, 2006). This paper treats the testlet model – at least technically – as being more similar to the bifactor model than to the locally dependent IRT model described in Section 6. Some other random effects-based testlet models (e.g. Wilson & Adams, 1995) are treated similarly.

In conclusion, IRT-based measurement and analytic methods in psychology are perpetually challenged by the increasingly complex test designs emanating from the proliferation of new applications, such as those recently arising in psychopathology (Meijer & Baneke, 2004; Sharp, Goodyer, & Croudace, 2006), exercise science (Rejeski, Ip, Katula, & White, 2006), personality inventory (Reise & Cook, 2010), and self-report health-related psycho-behavioural outcomes (Reeve, Hayes, Chang, & Perfetto, 2007; Reise *et al.*, 2007). It is my hope that the theoretical results reported here will further the understanding of how different IRT-based models function, and enhance the capacity of current psychometric tools to tackle these practical challenges.

Acknowledgements

This work is supported by National Science Foundation grant SES-0719354. The author would like to thank Dr Steve Reise for providing valuable suggestions that led to improvements in the presentation of the paper, and Dr Cheng-Der Fu for his comments and suggestions.

Appendix A: Proof of Lemma 2

In this and the following appendices, the key proof steps are outlined. We use boldface to indicate random variables when the distinction between a random variable and its realization is necessary.

Conditions (8b) and (8c) of Lemma 1 are satisfied by definition. For condition (8a), the specific form of the IRF, $E(Y|\theta_1)$, follows from applying the conditional expectation theorem (Williams, 1991, p. 88) to the conditional expectation of the MIRT model:

$$E(Y|\theta_1) = E[E(Y|\theta_1, \theta_2)]. \quad (A1)$$

The manifest probability, starting with a two-dimensional MIRT model, is given by

$$\begin{aligned} P(Y = 1) &= \iint P(Y = 1|\theta_1, \theta_2) f(\theta_1, \theta_2) d\theta_1 d\theta_2 \\ &= \iint E(Y = 1|\theta_1, \theta_2) f(\theta_1, \theta_2) d\theta_1 d\theta_2 \\ &= \int \left[\int E(Y = 1|\theta_1, \theta_2) f(\theta_2|\theta_1) d\theta_2 \right] f(\theta_1) d\theta_1 \\ &= \int \{E[E(Y|\theta_1, \theta_2)]\} f(\theta_1) d\theta_1, \end{aligned} \quad (A2)$$

where $f(\cdot)$ represents the density function. The two-dimensional kernel $E(Y|\theta_1, \theta_2)$ is equivalent to $E(Y|\theta_1)$, and our goal is to compute these two functions. Mathematically, it is easier to first derive our results with a probit link:

$$P(Y = 1) = \int \left[\int \Phi(a_1\theta_1 + a_2\theta_2 - d)\phi(\theta_2|\theta_1)d\theta_2 \right] \phi(\theta_1)d\theta_1. \quad (A3)$$

The two-dimensional conditional probit kernel is the inside integral in (A3) and is given by

$$\kappa(\theta_1) = \int \Phi(a_1\theta_1 + a_2\theta_2 - d)\phi_{\theta_1}(\theta_2)d\theta_2 = \int \Phi(a_2\theta_2 + \delta_{\theta_1})\phi_{\theta_1}(\theta_2)d\theta_2, \quad (A4)$$

where $\delta_{\theta_1} = a_1\theta_1 - d$. Let W denote a random variable that follows the standard normal distribution. It follows from (A4) that

$$\begin{aligned} \kappa(\theta_1) &= \int P(W \leq a_2\theta_2 + \delta_{\theta_1} | \theta_1 = \theta_1) \phi_{\theta_1}(\theta_2) d\theta_2 \\ &= \int P(W - a_2\theta_2 \leq \delta_{\theta_1} | \theta_1 = \theta_1) \phi_{\theta_1}(\theta_2) d\theta_2 \quad (A5) \\ &= E[P(W - a_2\theta_2 \leq \delta_{\theta_1} | \theta_1 = \theta_1)] \\ &= P(W - a_2\theta_2 \leq \delta_{\theta_1}). \end{aligned}$$

The variable $W - a_2\theta_2$ is also normally distributed as Φ_s , which has mean $-a_2\rho\theta_1\sigma_2/\sigma_1$ and variance $a_2^2(1 - \rho^2)\sigma_2^2 + 1$. Therefore, the kernel can now be re-expressed as

$$\begin{aligned} \kappa(\theta_1) &= \Phi_s(\delta_{\theta_1}) \\ &= \Phi\left(\lambda_{\text{probit}}\left(\delta_{\theta_1} + a_2\rho\theta_1\sigma_2/\sigma_1\right)\right) \quad (A6) \\ &= \Phi\left(\lambda_{\text{probit}}\left(a_1\theta_1 - d + a_2\rho\theta_1\sigma_2/\sigma_1\right)\right), \end{aligned}$$

where $\lambda_{\text{probit}} = [a_2^2(1 - \rho^2)\sigma_2^2 + 1]^{-1/2}$, the scaling factor that transforms Φ_s into the standard normal distribution for the probit link (Caffo, An, & Rohde, 2007; Gilmour, Anderson, & Rae, 1985; Heagerty & Zeger, 2000; Zeger, Liang, & Albert, 1988).

The scale factor for the logit link is given by (Johnson & Kotz, 1970, p. 6)

$$\lambda_{\text{logit}} = [k^2 a_2^2 (1 - \rho^2) \sigma_2^2 + 1]^{-1/2}, \quad (A7)$$

where $k = 16\sqrt{3}/(15\pi) = 0.588$. This approximation is known to be of sufficiently high-quality for most practical purposes (Demidenko, 2004, p. 334).

Appendix B: Proof of Corollary 1

Using the logit link function $g^{-1}(u)$, $\text{Cov}[E(\underline{Y}|\underline{\theta}_1, \underline{\theta}_2)]$ can be expressed as the $I \times I$ matrix that takes the form $\text{Cov}[g^{-1}(a_{i1}\theta_1 + a_{i2}\theta_2 - d_i)] = \text{Cov}[g^{-1}(\delta_{i\theta_1} + a_{i2}\theta_2)]$. Using the Taylor expansion about the point $\theta_2 = 0$ gives

$$g^{-1}(\delta_{i\theta_1} + a_{i2}\theta_2) = g^{-1}(\delta_{i\theta_1}) + \left[\frac{\partial g^{-1}(\delta_{i\theta_1} + a_{i2}\theta_2)}{\partial \theta_2} \right]_{\theta_2=0} \times \theta_2 + O(\theta_2^2), \quad (\text{B1})$$

where $g^{-1}(u) = \exp(u)/[1 + \exp(u)]$ and

$$\frac{\partial g^{-1}(u)}{\partial u} = h(u) = \frac{\exp(u)}{[1 + \exp(u)]^2}. \quad (\text{B2})$$

Thus, ignoring the second- and higher-order terms $O(\theta_2^2)$ in (B4a) and (B4b), the covariance matrix $\text{Cov}[E(\underline{Y}|\underline{\theta}_1, \underline{\theta}_2)]$ with the covariance function taken with respect to $\underline{\theta}_1$ given that $\underline{\theta}_1 = \theta_1$ is given by

$$\mathbf{S} = \left[\frac{\partial g^{-1}(\delta_{i\theta_1} + a_{i2}\theta_2)}{\partial \theta_2} \right]_{\theta_2=0} \text{Cov}(\underline{\theta}_2|\underline{\theta}_1) \left[\frac{\partial g^{-1}(\delta_{i\theta_1} + a_{i2}\theta_2)}{\partial \theta_2} \right]_{\theta_2=0}^T, \quad (\text{B3})$$

where $\text{Cov}(\underline{\theta}_2|\underline{\theta}_1) = \sigma_2^2(1 - \rho^2)$. The entries (s_{uv}) in $\mathbf{S}$ therefore are given by the expression

$$s_{uv}(\theta_1) = \frac{a_{u2}^2 \sigma_2^2 (1 - \rho^2) [\exp(a_{u1}\theta_1 - d_u)]^2}{[1 + \exp(a_{u1}\theta_1 - d_u)]^4}, \quad \text{if } u = v \quad (\text{B4a})$$

$$= \frac{a_{u2} a_{v2} \sigma_2^2 (1 - \rho^2) [\exp(a_{u1}\theta_1 + a_{v1}\theta_1 - d_u - d_v)]}{[1 + \exp(a_{u1}\theta_1 - d_u)]^2 [1 + \exp(a_{v1}\theta_1 - d_v)]^2}, \quad \text{if } u \neq v. \quad (\text{B4b})$$

The covariance term $\text{Cov}(\underline{Y}|\underline{\theta}_1, \underline{\theta}_2)$ in MIRT is a diagonal matrix in which the i th element is given by $p_i q_i$, $p_i = \exp(a_{i1}\theta_1 + a_{i2}\theta_1 - d_i)/[1 + \exp(a_{i1}\theta_1 + a_{i2}\theta_1 - d_i)]$ and $q_i = 1 - p_i$.

Accordingly, the conditional expectation of the i th element with respect to the distribution of θ_2 given that $\theta_1 = \theta_1$ is given by

$$E[\text{Var}(Y_i | \theta_1, \theta_2)] = \int \frac{\exp(a_{i1}\theta_1 + a_{i2}\theta_2 - d_i)}{[1 + \exp(a_{i1}\theta_1 + a_{i2}\theta_2 - d_i)]^2} \phi_{\theta_1}(\theta_2) d\theta_2. \quad (\text{B5})$$

The conditional variance function $p_{ij} q_{ij}$ takes the form $h(u)$ in (B2). A Taylor expansion of this function of $h(\delta_{i\theta_1} + a_{i2}\theta_2)$ at $\theta_2 = 0$ leads to the expression

$$\begin{aligned} E[\text{Cov}(Y_i | \theta_1, \theta_2)] &= \int \left\{ h(\delta_{i\theta_1}) + \left[\frac{\partial h(\delta_{i\theta_1} + a_{i2}\theta_2)}{\partial \theta_2} \right]_{\theta_2=0} \times \theta_2 + O(\theta_2^2) \right\} \phi_{\theta_1}(\theta_2) d\theta_2 \\ &\approx \frac{\exp(a_{i1}\theta_1 - d_i)}{[1 + \exp(a_{i1}\theta_1 - d_i)]^2} + \left[\frac{\partial h(\delta_{i\theta_1} + a_{i2}\theta_2)}{\partial \theta_2} \right]_{\theta_2=0} \int \theta_2 \phi_{\theta_1}(\theta_2) d\theta_2, \end{aligned} \quad (\text{B6})$$

where the integral is the conditional mean of θ_2 given that $\theta_1 = \theta_1$, which is given by $\rho\theta_1\sigma_2/\sigma_1 > 0$, σ_1 and $\sigma_2 \geq 0$. Furthermore, the derivative of the function $h(u)$ is given by

$$h'(u) = \frac{\exp(u) - [\exp(u)]^2}{[1 + \exp(u)]^3}. \quad (\text{B7})$$

Therefore,

$$E[\text{Cov}(Y | \theta_1, \theta_2)] \approx \text{diag} \left\{ \frac{\exp(a_{i1}\theta_1 - d_i)}{[1 + \exp(a_{i1}\theta_1 - d_i)]^2} + \frac{a_{i2}\rho\theta_1\sigma_2}{\sigma_1} h'(a_{i1}\theta_1 - d_i) \right\}. \quad (\text{B8})$$

Appendix C: Proof of Corollary 2

Consider the MIRT ($q > 2$) model with probit link function:

$$E(Y = 1 | \underline{\Theta}) = p(Y = 1 | \theta_1) = \Phi(\underline{a}^T \underline{\Theta} - d). \quad (\text{C1})$$

The kernel can be expressed as:

$$\begin{aligned}
 \kappa(\theta_1) = E(Y|\theta_1) &= \int \Phi(a_1\theta_1 + \underline{a}_2^T \underline{\Theta}_2 - d) \phi_{\theta_1}(\underline{\Theta}_2) d\underline{\Theta}_2 \\
 &= \int \Phi(\delta_{\theta_1} + \underline{a}_2^T \underline{\Theta}_2) \phi_{\theta_1}(\underline{\Theta}_2) d\underline{\Theta}_2 = E[P(Z < \delta_{\theta_1} | \theta_1)],
 \end{aligned} \tag{C2}$$

where Z is normally distributed with mean $-\underline{a}_2^T \underline{\eta}$ and variance $1 + \underline{a}_2^T \bar{\Sigma} \underline{a}_2$, where $\underline{\eta} = (\rho_{1m} \sigma_m \theta_1 / \sigma_1)$, $m = 2, \dots, q$, is the conditional mean vector of $\underline{\Theta}_2 | \theta_1$, and $\bar{\Sigma} = \Sigma_2 - (1/\sigma_1^2) \gamma \gamma^T$ is its conditional covariance. This leads to the following unidimensional kernel corresponding to its multidimensional counterpart in (C1):

$$\kappa(\theta_1) = E(Y|\theta_1) = \Phi \left(\frac{\delta_{\theta_1} + \underline{a}_2^T \underline{\eta}}{\sqrt{1 + \underline{a}_2^T \bar{\Sigma} \underline{a}_2}} \right). \tag{C3}$$

When a logit link is used, the scale factor $\sqrt{1 + \underline{a}_2^T \bar{\Sigma} \underline{a}_2}$ needs to be modified to $\sqrt{1 + k^2 \underline{a}_2^T \bar{\Sigma} \underline{a}_2}$ where $k = 16\sqrt{3}/(15\pi)$.

References

- Abadir KM, & Magnus JR (2005). Matrix algebra Cambridge: Cambridge University Press.
- Ackerman TA (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement*, 13, 113–127.
- Adam RJ, Wilson M, & Wang W-C (1997). The multidimensional random coefficient multinomial logit model. *Applied Psychological Measurement*, 21, 1–23.
- Ansley TM, & Forsyth RA (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9, 39–48.
- Bock RD, Gibbons RD, & Muraki E (1988). Full-information factor analysis. *Applied Psychological Measurement*, 12, 261–280.
- Bollen KA, & Lennox R (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 100, 305–314.
- Bradlow E, Wainer H, & Wang X (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153–168.
- Braeken J, Tuerlinckx F, & De Boeck P (2007). Copulas for residual dependencies. *Psychometrika*, 72, 393–411.
- Caffo B, An M, & Rohde C (2007). Flexible random intercept model for binary outcomes using mixture of normals. *Computational Statistics and Data Analysis*, 51, 5220–5235. [PubMed: 18628822]
- Cattell RB (1966). Psychological theory and scientific method. In Cattell RB (Ed.), *Handbook of multivariate experimental psychology* (pp. 1–18). Chicago: Rand McNally.
- De Boeck P, & Wilson M (2004). *Explanatory item response models* New York: Springer.
- Demidenko E (2004). *Mixed models: Theory and applications* Hoboken, NJ: Wiley.
- Douglas J, Kim HR, Habing B, & Gao F (1998). Investigating local dependence with conditional covariance functions. *Journal of Educational and Behavioral Statistics*, 23, 129–151.

- Drasgow F, & Parson CK (1983). Application of unidimensional item response theory to multidimensional data. *Applied Psychological Measurement*, 7, 189–199.
- Embretson SE, & Reise SP (2000). *Item response theory for psychologists* Mahwah, NJ: Erlbaum.
- Fitzmaurice GM, Laird NM, & Ware JH (2004). *Applied longitudinal analysis* Hoboken, NJ: Wiley.
- Folk VG, & Green BF (1989). Adaptive estimation when the unidimensionality assumption of IRT is violated. *Applied Psychological Measurement*, 13, 373–389.
- Gibbons RD, & Hedeker DR (1992). Full-information item bi-factor analysis. *Psychometrika*, 57, 423–436.
- Gibbons RD, Immekus JC, & Bock RD (2007). *The added value of multidimensional IRT models. Multidimensional and hierarchical modeling monograph 1* Chicago: Center for Health Statistics, University of Illinois.
- Gilmour AR, Anderson RD, & Rae AL (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika*, 72, 593–599.
- Goldstein H, & Wood R (1989). Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology*, 42, 139–167.
- Harrison DA (1986). Robustness of parameter estimation to violations to the unidimensionality assumption. *Journal of Educational Statistics*, 11, 91–115.
- Heagerty PJ, & Zeger SL (2000). Marginalized multilevel models and likelihood inference. *Statistical Science*, 15, 1–26.
- Holzinger KJ, & Swineford F (1937). The bi-factor method. *Psychometrika*, 2, 41–54.
- Hoskens M, & De Boeck P (1997). A parametric model for local dependence among test items. *Psychological Methods*, 2, 261–277.
- Humphreys LG (1986). An analysis and evaluation of test and item bias in the predictive context. *Journal of Applied Psychology*, 71, 327–333.
- Ip EH (2000). Adjusting for information inflation due to local dependency in moderately large item clusters. *Psychometrika*, 65, 73–91.
- Ip EH (2002). Locally dependent latent trait model and the Dutch identity revisited. *Psychometrika*, 67, 367–386.
- Ip EH, Goetghebeur Y, Molenberghs G, & De Boeck P (2006). All unidimensional models are wrong, but some are useful: Functional unidimensionality and methods of estimation. Paper presented at the 71st Meeting of the Psychometric Society, 14–17 June, Montreal, Canada.
- Ip EH, Smits D, & De Boeck P (2009). Locally dependent linear logistic test model with person covariates. *Applied Psychological Measurement*, 33(7), 555–569. 10.1177/0146621608326424
- Ip EH, Wang Y, De Boeck P, & Meulders M (2004). Locally dependent latent trait model for polytomous responses with application to inventory of hostility. *Psychometrika*, 69, 191–216.
- Jannarone RJ (1986). Conjunctive item response theory kernels. *Psychometrika*, 51, 357–373.
- Johnson NL, & Kotz S (1970). *Continuous univariate distributions* (Vol. 1). New York: Wiley.
- Junker BW, & Stout WF (1994). Robustness of ability estimation when multiple traits are present with one trait dominant. In Laveault D, Zumbo BD, Gessaroli ME, & Boss MW (Eds.), *Modern theories of measurement: Problems and issues* (pp. 31–61). Ottawa, Canada: University of Ottawa.
- Kelley TL (1928). *Crossroads in the mind of man: A study of differentiable mental abilities* Stanford, CA: Stanford University Press.
- Kim H (1994). *New techniques for the dimensionality assessment of standardized test data*. Doctoral dissertation, Department of Statistics, University of Illinois, Urbana-Champaign.
- Kirisci L, Hsu T, & Yu L (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*, 25, 146–162.
- Li Y, Bolt DM, & Fu J (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30, 3–21.
- Liang KY, & Zeger SL (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121–130. [PubMed: 3719049]
- Lord FM (1980). *Applications of item response theory to practical testing problems* Mahwah, NJ: Erlbaum.

- McCullagh P, & Nelder JA (1989). Generalized linear models (2nd ed.). London: Chapman & Hall.
- McDonald RP (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100–117.
- McDonald RP (1985). Unidimensional and multidimensional models for item response theory. In Weiss DJ (Ed.), *Proceedings of the 1982 item response theory and computerized adaptive testing conference* (pp. 127–148). Minneapolis: University of Minnesota.
- Meijer RR, & Baneke JJ (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods*, 9, 354–368. [PubMed: 15355153]
- Muthén BO (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54, 557–585.
- Ozer D (2001). Four principles of personality assessment. In Pervin LA & John OP (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 671–688). New York: Guilford Press.
- Rasch G (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 19, 49–57. [PubMed: 5939145]
- Reboussin B, Ip EH, & Wolfson M (2008). Locally dependent latent class models with covariates: An application to underage drinking in the United States. *Journal of the Royal Statistical Society A*, 171, 877–897.
- Reckase MD (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207–230.
- Reckase MD (1997). A linear logistic multidimensional model for dichotomous item response data. In van der Linden WJ & Hambleton RK (Eds.), *Handbook of item response theory* (pp. 271–286). New York: Springer.
- Reckase MD, Carlson JE, Ackerman TA, & Spray JA (1986). The interpretation of unidimensional IRT parameters when estimated from multidimensional data. Paper presented at the Annual Meeting of the Psychometric Society, Toronto.
- Reckase MD, & McKinley RL (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 14, 361–373.
- Reeve BB, Hays RD, Chang C, & Perfetto EM (2007). Applying item response theory to enhance health outcomes assessment. *Quality of Life Research*, 16, 1–3. [PubMed: 17033892]
- Reise SP, & Cook KF (2010). Item response theory and the unidimensionality assumption: Toward a bifactor future Manuscript submitted for publication.
- Reise SP, Morizot J, & Hays RD (2007). The role of bifactor models in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16, 19–31. [PubMed: 17479357]
- Rejeski J, Ip EH, Katula J, & White L (2006). Older adults' desire for physical competence. *Medicine and Science in Sports and Exercise*, 38, 100–105. [PubMed: 16394960]
- Rijmen F, Ip EH, Rapp S, & Shaw E (2008). Qualitative longitudinal analysis of symptoms in patients with primary or metastatic brain tumors. *Journal of the Royal Statistical Society 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. [PubMed: 12924814]
- Rosenbaum PR (1988). Item bundles. *Psychometrika*, 53, 349–359.
- Samejima F (1974). Normal ogive model for the continuous response level in the multidimensional latent space. *Psychometrika*, 39, 111–121.
- Scott S, & Ip EH (2002). Empirical Bayes and item clustering effects in latent variable hierarchical models: A case study from the National Assessment of Educational Progress. *Journal of the American Statistical Association*, 97, 409–419.
- Segall DO (1996). Multidimensional adaptive testing. *Psychometrika*, 61, 331–354.
- Sharp C, Goodyer IM, & Croudace TJ (2006). The Short Mood and Feelings Questionnaire (SMFQ): A unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of Abnormal Child Psychology*, 34, 379–391. [PubMed: 16649000]
- Spearman C (1933). The factor theory and its troubles. III. Misrepresentation of the theory. *Journal of Educational Psychology*, 24, 591–601.

- Spencer SG (2004). The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated. Doctoral dissertation, Department of Instructional Psychology and Technology, Brigham Young University, Provo, UT.
- Stout W (1990). A new item response theory modeling approach with applications to unidimensional assessment and ability estimation. *Psychometrika*, 55, 293–326.
- Stout W, Habing B, Douglas J, Kim HR, Roussos L, & Zhang J (1996). Conditional covariance based nonparametric multidimensional assessment. *Applied Psychological Measurement*, 20, 331–354.
- Verhelst ND, & Glas GAW (1993). A dynamic generalization of the Rasch model. *Psychometrika*, 58, 391–415.
- Wainer H, Bradlow ET, & Wang X (2007). *Testlet response theory and its applications* Cambridge: Cambridge University Press.
- Walker CM, & Beretvas SN (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement*, 40, 255–275.
- Wang W, & Wilson M (2005). Exploring local item dependence using a random-effects facet model. *Applied Psychological Measurement*, 29, 296–318.
- Way WD, Ansley TN, & Forsyth RA (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12, 239–252.
- Williams D (1991). *Probability with martingales* Cambridge: Cambridge University Press.
- Wilson M, & Adams RJ (1995). Rasch models for item bundles. *Psychometrika*, 60, 181–198.
- Yen WM (1993). Scaling performance assessments: Strategies for managing local item independence. *Journal of Educational Measurement*, 30, 187–213.
- Zeger SL, Liang KY, & Albert P (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44, 1049–1060. [PubMed: 3233245]

Figure 1.
 (a) Locally independent unidimensional model. (b) Between-item MIRT model. (c) Bifactor model. (d) Locally dependent unidimensional model. Square represents item response, and oval represents latent factor.

Figure 2.

Comparison of approximation of IRF through equation (19) and through numerical integration under four scenarios (a)–(d) in Table 1. Dotted lines represent the approximated IRF, and solid lines represent the IRF from numerical integration.

Figure 3.

Comparison of approximation of expected variance through equation (B8) and through numerical integration under four scenarios (a)–(d) in Table 1. Dashed lines represent the approximated expected variance, and solid lines represent expected variance from numerical integration.

Figure 4.

Comparison of approximation of variance of expected value through equation (B4a) and through numerical integration under four scenarios (a)–(d) in Table 1. Dashed lines represent the approximated expected variance and solid lines represent expected variance from numerical integration.

Figure 5.

Comparison of approximation of correlation between two items through equations (15) and (16) and through numerical integration under four scenarios (a)–(d) in Table 1. The two items are assumed to have identical item parameters. Dashed lines represent the approximated correlation, and solid lines represent correlation from numerical integration.

Figure 6.

(a) Bifactor (testlet) model for between-item MIRT in application to NAEP data. The testlet within the subscale reading for information is modelled through a random effect. Only one testlet is shown. In the actual test and the model in Scott and Ip (2002), there are multiple testlets. (b) Corresponding locally dependent model for between-item MIRT. The testlet within the subscale is modelled through specification of the conditional covariance matrix.

Table 1.

Different scenarios for showing quality of linear approximation between MIRT and locally dependent unidimensional models

Parameter	Scenarios			
	Standard (a)	High correlation (b)	Low a_1 , high a_2 (c)	Comparable variance (d)
a_1	2	2	0.5	2
a_2	1.5	1.5	2.0	1.5
d	1	1	1	1
σ_1	1	1	1	1
σ_2	0.5	0.5	0.5	1
ρ	0.4	0.9	0.4	0.4
Angular direction (deg)	36.7	36.7	75.6	36.7