PSYCHOMETRIKA—VOL. 69, NO. 2, 191–216
JUNE 2004

LOCALLY DEPENDENT LATENT TRAIT MODEL FOR POLYTOMOUS RESPONSES WITH APPLICATION TO INVENTORY OF HOSTILITY

EDWARD H. IP

WAKE FOREST UNIVERSITY

YUCHUNG J. WANG

RUTGERS UNIVERSITY

PAUL DE BOECK
MICHEL MEULDERS

K.U. LEUVEN

Psychological tests often involve item clusters that are designed to solicit responses to behavioral stimuli. The dependency between individual responses within clusters beyond that which can be explained by the underlying trait sometimes reveals structures that are of substantive interest. The paper describes two general classes of models for this type of locally dependent responses. Specifically, the models include a generalized log-linear representation and a hybrid parameterization model for polytomous data. A compact matrix notation designed to succinctly represent the system of complex multivariate polytomous responses is presented. The matrix representation creates the necessary formulation for the locally dependent kernel for polytomous item responses. Using polytomous data from an inventory of hostility, we provide illustrations as to how the locally dependent models can be used in psychological measurement.

Key words: generalized log-linear model, hybrid kernel, combination dependency models, EM algorithm, social inhibition.

1. Introduction

The item response theory (IRT) has proved to be a powerful tool in modeling individual responses to behavioral stimuli. Within an IRT framework, individual differences in a specific construct (e.g., intelligence quotient, ability, attitude, or propensity) are captured by the relative positions of the individuals' scores on a common continuous scale of the latent trait. Since the pioneering work of Lord and Novick (1968) and Fischer (1974), IRT has been extended to cover a wide range of psychological, cognitive, and educational applications. One area in which IRT has been extended is its focus on the development of locally dependent models. Two broad approaches to local dependency (LD) can be distinguished. In the first approach, LD is viewed as being unintentionally created by the *design* of a test. Subsequently, LD is treated as a nuisance factor that is peripheral to the purpose of the study. For example, in a reading comprehension test different items may share a common stem—a common reading passage. As a consequence, the usual assumptions regarding local independence—that item responses are conditionally independent given the latent trait—do not necessarily apply. While one still wants to estimate the underlying latent trait, the residual dependency not captured by the latent trait is considered a nuisance factor, which, if not taken into account in the measurement model, might lead to serious distortion. In particular, it might result in "double counting" of information contained in multiple responses that are related to a common stem.

Requests for reprints should be sent to Edward Ip, Department of Public Health Sciences, Wake Forest University School of Medicine, MRI Building, Winston-Salem, NC 27157. Email: eip@wfubmc.edu

0033-3123/2004-2/2001-0945-A $00.75/0

© 2004 The Psychometric Society

In the literature, model-based methods have been proposed to estimate LD effects (Bradlow, Wainer, & Wang, 1999; Hoskens & De Boeck, 1997; Ip, 2000, 2001, 2002; Scott & Ip, 2002). Gibbons and Hedekker (1992) proposed a bi-factor item response model that permits conditional dependence within identified subsets of items. Nonmodel-based methods—notably item bundle (Rosenbaum, 1988); testlet (Wainer & Kiely, 1987); indexes for estimating effects of LD (DETECT, Zhang & Stout, 1999; HCA/CCPROX, Roussos, Stout, & Marden, 1998); and other LD managing strategies (Yen, 1993)—have also been extensively developed to address the problem. DIMTEST (Stout, 1987), which is based on hypothesis testing, has been widely used for detecting LD as well.

On the other hand, one may have an interest in *substantive* phenomena that implies an alternative type of dependency that deserves to be carefully analyzed. For example, Jannarone (1986) and Hoskens and De Boeck (1995, 2001) discussed dependency models to capture the cognitive components underlying a task as an alternative for componential LD modeling (Embretson, 1984; Fischer, 1973, 1994). This treatment of LD does not preclude test designs that use clustered items. In fact, the application we will use in the subsequent discussions contains this type of data.

The following psychological example can be used to illustrate the two approaches to treating LD. Endler and Hunt (1968) constructed a Self-Report (S-R) inventory of hostility containing 23 situations, with multiple items associated with each situation. For every situation (e.g., "you miss your train because the clerk has given you the wrong information"), a set of three item pairs follows a common design. Each pair refers to a hostile behavior, and each behavior is presented in two response variants: wanting to show the behavior and actually showing it. The hostile behaviors and variants are: "wanting to curse" and "curse," "wanting to scold" and "scold," "wanting to shout" and "shout." For each situation, the participants were asked to rate on a 3-point scale the degree to which they would display the response.

The first treatment of LD would imply that while one wants to measure hostility, the dependency between these items may distort the results if the possible LD effect—presumably arising from a common situation or from common wording within each pair of items—is not taken into account. In other words, dependency is a nuisance factor.

However, it may be expected that beyond the underlying trait indicating general hostility there exists dependency within item pairs due to a general inclination toward a specific behavior (e.g., want to and do curse) that stimulates simultaneously the desire for and the manifestation of the behavior. Figure 1 shows the item responses of a sample of 40 individuals to a sample of 6 item clusters from a data set collected by Vansteelandt (1999). Item responses to "want to curse" and "want to scold" are paired with "curse" and "scold," respectively, as are "want to shout" and "shout." In addition to a consistent pattern of general aggressiveness within individuals, Figure 1 suggests that item pairs also display substantial dependency. Therefore, directly modeling LD provides an opportunity to investigate the effect (if any) of inclinations.

The dependency structure is further complicated when one considers whether or not another party is involved in a situation. When another party is involved (e.g., in the train situation), social inhibitions may counteract the inclination to exhibit hostile behavior. In fact, in many situations in the inventory another party can be blamed for an adverse outcome, while in other cases it is clearly the individual's own fault that something went wrong.

Designing flexible and viable LD models is important for analyzing this kind of data, especially when the purpose is to empirically support or refute related psychological theories. For example, it can be hypothesized that when another party is involved, social inhibition has a divergent effect on the desire to exhibit hostility and the actual hostile behavior. As a result, negative dependency may exist between the "want to" and "do" item responses (e.g., want to curse and do curse) for situations involving a third party. Likewise, it is possible that dependency exists between the "do" item responses and manifests itself in the following way. Inhibited individuals

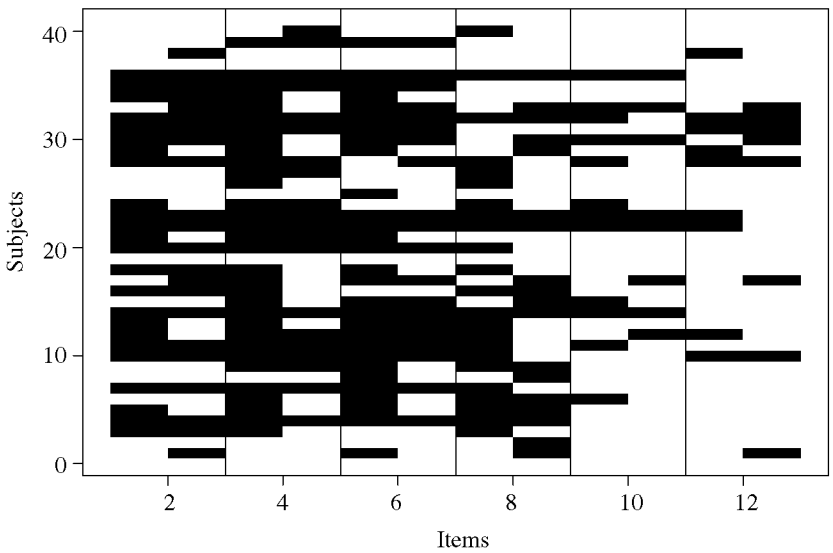


FIGURE 1.
Visualizing polytomous responses to a sample of item clusters for hostility inventory. Each line represents the responses in grade scale (1 is lightest and 3 is darkest). Clusters are separated by vertical lines. The items are in the context of, respectively, *wtsh, sh*; *wtsc, sc*; *wtc, c*; *wtsc, sc*; *wtsh, sh*, where want to shout = *wtsh*, shout = *sh*, want to scold = *wtsc*, scold = *scold*, want to curse = *wtc*, and curse = *c*.

may feel that they want to engage in aggressive behavior, but they do not. When uninhibited individuals, on the other hand, feel hostile, they actually exhibit aggressiveness. Because of this, an analysis of the dependency structure in item responses can both identify the underlying dynamics and gauge their strengths. Specifically, one may conjecture that the dynamics vary from one person to another—an interesting issue that we will discuss later in this paper.

Technically, one approach to modeling LD is to modify the so-called IRT kernels, which are the conditional distributions that appear in observed probability mixtures for IRT models (see Jannanore, 1986). Consider an example of a cluster of two trichotomous items with responses (Y_1, Y_2), where each variable may take any one of the three possible levels 1, 2, 3. The IRT kernel is $P(Y_1 = y_1, Y_2 = y_2|\theta)$, where θ is the latent trait. Instead of using a locally independent IRT kernel, whereby

$$P(Y_1 = y_1, Y_2 = y_2|\theta) = P(Y_1 = y_1|\theta)P(Y_2 = y_2|\theta),$$

one could replace the right-hand side by a more general form of distribution that is able to capture the dependency structure of the joint responses. Note that this item cluster (Y_1, Y_2) generates a total of nine possible categorical joint response patterns $(1, 1), (2, 1), \ldots, (3, 3)$. Because of its capacity to handle general categorical responses, the log-linear model (Bishop, Fienberg, & Holland, 1975) logically provides a proven means of representing the interactions of various orders among joint responses from an item cluster. Cox (1972) and Zhao and Prentice (1990) developed computationally efficient forms of the log-linear model for *binary* responses. Laird (1991) and Holland (1990) called them the generalized log-linear models (GLLM). Using the GLLM as a basis, Ip (2002) presented a class of LD kernels for binary responses.

This paper is motivated by the need to (1) extend LD kernels to accommodate polytomous responses, and (2) apply the results from (1) to psychological problems of substantive interests. In psychological and educational testing, polytomous responses are becoming increasingly common. The development of theories and methods for polytomous LD kernels would suggest a

nontrivial extension to the binary case of Ip (2002). One of the foremost requirements of an extension to polytomous responses is a generalization of the binary GLLM such as in Zhao and Prentice (1990). Because the LD kernel is based on the theory of exponential family, the result from the generalization must be able to conform to an exponential family representation. Furthermore, the order of complexity of the formulation now greatly increases from binary to polytomous responses. Interactions are now present both between categories within an item and *between items*. Given the already potentially high dimensionality because of the multivariate nature of the problem—the number of interactions increases exponentially as the cluster size increases—a succinct and interpretable parameterization of the polytomous GLLM is therefore critical.

In this paper, we propose a compact and highly interpretable matrix system for representing the extended GLLM, which now describes a system of multivariate polytomous responses. The representation is an extension of results in Wang (1986). A special case of the extension is discussed in Ip and Wang (2003). Because of the possible high dimensionality of the dependency structure, evaluation and interpretation of parameters is important to applications that involve large item clusters. We introduce a system of differential and integral operators to enable users to evaluate and interpret parameters.

The extended GLLM forms the basis for the development of two broad classes of LD kernels—the polytomous canonical and the hybrid. The polytomous canonical kernel leads to an extension of the constant-combination dependency model and the dimension-dependent-combination dependency model of Hoskens and de Boeck (1997). The canonical kernel also provides a formal statistical framework for these models. Alternately, the polytomous hybrid model, as noted previously, extends the binary case in Ip (2002).

The remainder of this paper is organized as follows. After describing the general systems of matrix representation and the operators, we develop the two classes of kernels, canonical and hybrid. Then we demonstrate how the two resulting families of LD models can be used in the context of analyzing an inventory of hostility. This is finally followed by a discussion.

2. Compact Representation of Multivariate Polytomous Responses

Matrix Formulation

The structure of multivariate polytomous response data can be specified by a lattice. An example would be a two-dimensional polytomous response in which the first and second variables Y_1 and Y_2 both have three ordered levels, which are $\Lambda_1 = (1, 2, 3)$, and $\Lambda_2 = (a, b, c)$, respec-

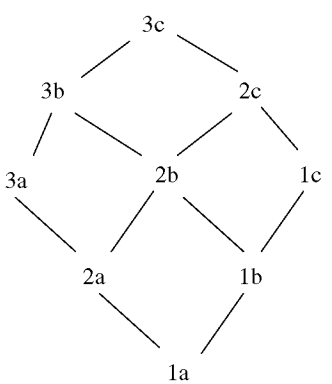


FIGURE 2.
Dominance structure of two responses that are respectively ordered as (3-2-1) and (c-b-a).

tively. Assume that a partial ordering $\prec$ exists on the two item responses so that $1 \prec 2 \prec 3$, and $a \prec b \prec c$, each of which now forms a chain. Then, the Cartesian product $\Lambda_1 \times \Lambda_2$ forms a lattice under $\prec$ (Aigner, 1979, chapter 4). The lattice is shown in Figure 2.

In general, let $Y = (Y_1, \ldots, Y_I)$ be a vector of polytomous responses. Each variable Y_i, $i = 1, \ldots, I$ may take a value in the ordered category set $\Lambda_i = \{C_i(1), \ldots, C_i(K_i)\}$, where K_i denotes the number of possible response categories for item i. We assume without loss of generality that $C_i(1) \prec C_i(2) \prec \cdots \prec C_i(K_i)$ for every i. An item-level dominance matrix $B = (b_{uv})$ describes the ordered relationship between categories: $b_{uv} = 1$ if $C_i(u) \prec C_i(v)$ and 0 otherwise. The concept extends to multiple items and is best illustrated by an example.

Consider an asymmetric 2×3 example. A dominance matrix $A = (a_{uv})$ is derived from the six possible response patterns $(1a, 2a, 1b, 2b, 1c, 2c)$. It is straightforward to show that A is the tensor product of the item-level dominance matrices B_1 and B_2 of the chains $2-1$, and $c-b-a$, respectively. Specifically,

$$B_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, B_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

$$A = B_2 \otimes B_1 = \begin{pmatrix} B_1 & 0 & 0 \\ B_1 & B_1 & 0 \\ B_1 & B_1 & B_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

The matrix A represents the dominance relationship among all realizations of response patterns. For example, the "1" at the intersection of the fourth row and third column indicates that in the pattern $(1a, 2a, 1b, 2b, 1c, 2c)$ the fourth element, $2b$, dominates the third, $1b$.

In general, it can be shown by induction that the dominance matrix A can be expressed as the tensor product of the item-level dominance matrices of Λ_i, $i = 1, \ldots, I$. That is, $A = B_I \otimes B_{I-1} \otimes \cdots \otimes B_1$. Under lattice theory, the invertibility of B_i is guaranteed by a Möbius inversion (Aigner, 1979, chapter 4). Indeed, the inverse of B_i is given by

$$B_i^{-1} = \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & . & . & \\ & & -1 & 1 & \\ & & & -1 & 1 \end{pmatrix}. \tag{1}$$

Therefore A is also invertible, and $A^{-1} = B_I^{-1} \otimes \cdots \otimes B_1^{-1}$. The matrix A^{-1} plays an important role in the subsequent development of the theory in that it defines the contrasts of log cell probabilities, and we shall refer to it as a contrast matrix.

Now, suppose all possible realizations of $Y = (Y_1, \ldots, Y_I)$ are represented by a vector $\mathbf{X}$ of indicator functions, so that $X_{k_1 \ldots k_I} = 1$ if the response is $(C_1(k_1), \ldots, C_I(k_I))$ and 0 otherwise. For simplicity, responses hereafter are labeled as $(k_1, \ldots, k_I)$. Also, let the probability for each realization of response pattern, $P(X_{k_1 \ldots k_I} = 1)$ be denoted by $\pi_{k_1 \ldots k_I}$, $k_i = 1, \ldots, K_i$, $i = 1, \ldots, I$. Then the multinomial distribution of the response has the density function $h(x) \exp(x^T \log \pi)$, with $\sum \pi_k = 1$. The vectors x and π are arranged in lexicographical order so that the first index changes fastest.

Apparently, the multinomial distribution for all possible realizations of response patterns is not suitable for describing dependency. Using a matrix transformation of the multinomial, one

can obtain a representation of the dependency structure through contrasts of logarithms of the cell probabilities. Based on the dominance matrix A, the following transformation of the kernel $x^T \log \pi$ of the multinomial distribution gives:

$$x^T \log \pi = x^T A A^{-1} \log \pi = s^T \omega, \tag{2}$$

where $s = A^T x$, and $\omega = A^{-1} \log \pi$. Here, both s and ω are vectors of length $K_1 \times \cdots \times K_I$. Consequently, the log-likelihood equation after the transformation (2) is given by

$$\ell(\omega) = s^T \omega - \kappa(\omega) + H(s), \tag{3}$$

where $\kappa(\omega)$ is the cumulant generating function, and H is some function of s. Equation (3) is called a GLLM of canonical representation of polytomous responses. The following two examples illustrate the re-parameterization of three-dimensional dichotomous and two-dimensional polytomous responses into the canonical representation.

Example 1. ($2 \times 2 \times 2$ Ordered Response) Let $x = (x_{111}, x_{211}, x_{121}, x_{221}, x_{112}, x_{212}, x_{122}, x_{222})^T$. The dominance matrix A is given by

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

and thus

$$s = A^T x = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} x_{111} \\ x_{211} \\ x_{121} \\ x_{221} \\ x_{112} \\ x_{212} \\ x_{122} \\ x_{222} \end{pmatrix}$$

$$= (x_{+++}, x_{2++}, x_{+2+}, x_{22+}, x_{++2}, x_{2+2}, x_{+22}, x_{222})^T,$$

where $+$ indicates summation over the subscript. Note that the vector s can be expressed as functions of the item responses y_i, $i = 1, 2, 3$, where responses 1, 2 are respectively recoded into 0, 1:

$$(1, y_1, y_2, y_1 y_2, y_3, y_1 y_3, y_2 y_3, y_1 y_2 y_3)^T.$$

Furthermore,

$$\omega = A^{-1} \log \pi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ -1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \log \pi_{111} \\ \log \pi_{211} \\ \log \pi_{121} \\ \log \pi_{221} \\ \log \pi_{112} \\ \log \pi_{212} \\ \log \pi_{122} \\ \log \pi_{222} \end{pmatrix}$$

$$= (\omega_0, \omega_1, \omega_2, \omega_{12}, \omega_3, \omega_{13}, \omega_{23}, \omega_{123})^T.$$

The components of the vector ω can be grouped into four categories: the zeroth-order term $\omega_0 = \log \pi_{111}$; first-order terms such as $\omega_1 = \log(\pi_{211}/\pi_{111})$, which is the conditional logit of Y_1 given $Y_2 = Y_3 = 1$; second-order interactions such as $\omega_{12} = \log(\pi_{221}\pi_{111})/(\pi_{211}\pi_{121})$, which is the conditional local log odds ratio between Y_1 and Y_2 given $Y_3 = 1$; and, finally, the third-order interaction term $\omega_{123} = \log[(\pi_{222}\pi_{112})/(\pi_{212}\pi_{122})] - \log[(\pi_{221}\pi_{111})/(\pi_{211}\pi_{121})]$, which indicates the "step" between the conditional log odds ratios at two different levels of the conditioning variable Y_3.

As a result of the transformation using the A matrix, the joint distribution can be expressed as:

$$\log P(\mathbf{Y} = \mathbf{y}) = \sum_{i=1}^{3} y_i \omega_i + \sum_{i<j} y_i y_j \omega_{ij} + y_1 y_2 y_3 \omega_{123} - \kappa(\omega). \tag{4}$$

Equation (4) is the familiar form of GLLM of order 3 for dichotomous responses (Zhao & Prentice, 1990).

Example 2. (3×3 Ordered Response) The vector x is $(x_{11}, x_{21}, x_{31}, x_{12}, x_{22}, x_{32}, x_{13}, x_{23}, x_{33})^T$. The corresponding vector s is given by $(s_{k_1k_2})$, where $s_{k_1k_2} = I(Y_1 \geq k_1, Y_2 \geq k_2)$, $k_1, k_2 = 1, 2, 3$, and $I(.)$ is the indicator function. Therefore, the vector s indicates whether each component of a response pattern is higher than or equal to a specific category. Table 1 shows the structure of the vector s, which comprises indicator functions of rectangular regions of the table. Each region always contains the "corner" cell x_{33}. The contrast matrix, on the other hand, produces the vector ω:

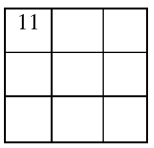
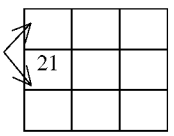
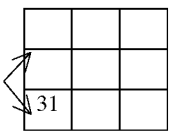
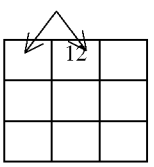
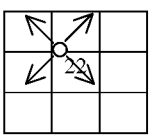
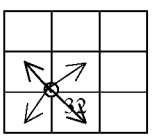
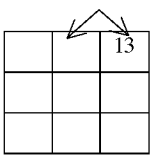
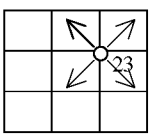
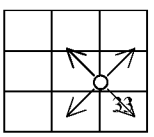
$$\omega = A^{-1} \log \pi$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \log \pi_{11} \\ \log \pi_{21} \\ \log \pi_{31} \\ \log \pi_{12} \\ \log \pi_{22} \\ \log \pi_{32} \\ \log \pi_{13} \\ \log \pi_{23} \\ \log \pi_{33} \end{pmatrix}.$$

The last column in Table 1 shows the entire vector of contrasts of log probabilities in ω. Note in Table 1 that the third column shows a compact form for representing interaction terms in ω based on the log-linear notation (Bishop, Fienberg, and Holland, 1975). For example, $\omega_{12(23)}$ shows the local log odds ratio at the level $Y_1 = 2$ and $Y_2 = 3$. The general notation and interpretation of ω will be discussed in detail in the next subsection.

Briefly, the contrast matrix A^{-1} summarizes the necessary contrasts that are used to define various orders of local interactions (local in the sense that comparison is made between adjacent categories that could be either within or between items). The latter possibility is especially intriguing, because to the best of our knowledge there had not existed a literature on modeling interaction between levels of responses between two polytomous items in the IRT context. Therefore, the method that we developed should provide a useful tool for researchers interested in examining granular item interactions.

TABLE 1.
3×3 example of canonical parameter and s statistic. Each 3×3 table represents a possible set of item responses for a pair of items with one side of the table representing one item and the other side representing the other.

	s_{11}*	ω_0**	$\log \pi_{11}$
	s_{21}	$\omega_{1(2)}$	$\log \frac{\pi_{21}}{\pi_{11}}$
	s_{31}	$\omega_{1(3)}$	$\log \frac{\pi_{31}}{\pi_{21}}$
	s_{12}	$\omega_{2(2)}$	$\log \frac{\pi_{12}}{\pi_{11}}$
	s_{22}	$\omega_{12(22)}$	$\log \frac{\pi_{22}\pi_{11}}{\pi_{21}\pi_{12}}$
	s_{32}	$\omega_{12(32)}$	$\log \frac{\pi_{21}\pi_{32}}{\pi_{31}\pi_{22}}$
	s_{13}	$\omega_{2(3)}$	$\log \frac{\pi_{13}}{\pi_{12}}$
	s_{23}	$\omega_{12(23)}$	$\log \frac{\pi_{12}\pi_{23}}{\pi_{22}\pi_{13}}$
	s_{33}	$\omega_{12(33)}$	$\log \frac{\pi_{22}\pi_{33}}{\pi_{32}\pi_{23}}$

*s is the indicator function of the collection of cells indicated by the shaded area.
**The last two columns indicate the notation and values of the canonical parameters, and their roles are visualized by arrows in the first column.

The joint distribution of the multivariate polytomous responses can be cast in a form analogous to the GLLM (4). Let $Z_{i(k_i)} = I(Y_i \geq k_i)$, $i = 1, 2$, $k_i = 2, 3$. Then,

$$\log P(\mathbf{Y} = \mathbf{y}) = \sum_{k_i=2}^{3} \sum_{i=1}^{2} z_{i(k_i)} \omega_{i(k_i)} + \sum_{k_1=2}^{3} \sum_{k_2=2}^{3} z_{1(k_1)} z_{2(k_2)} \omega_{12(k_1 k_2)} - k(\omega). \tag{5}$$

Difference and Integral Operators

The dominance matrix A and the contrast matrix A^{-1} are often quite large and inconvenient to use. We provide an alternative approach to deriving ω. The interaction terms of ω are generated as mixed derivatives of the log density with respect to categories. Following Whittaker (1990, chapter 2), we define a difference operator

$$\nabla_i g(k_1, \ldots, k_i, \ldots, k_I) = g(k_1, \ldots, k_i, \ldots, k_I) - g(k_1, \ldots, k_i - 1, \ldots, k_I),$$

for $1 < k_i \leq K_i$, and $\nabla_i g(k_1, \ldots, k_I) = g(k_1, \ldots, k_I)$ for $k_i = 1$. Define

$$\nabla_j \nabla_i g = \nabla_j (\nabla_i g), \quad \text{and} \quad \nabla_i (g_1 \pm g_2) = \nabla_i g_1 \pm \nabla_i g_2.$$

It is straightforward to verify that the operator satisfies the commutative law $\nabla_j \nabla_i g = \nabla_i \nabla_j g$. For simplicity, write the mixed derivative $\nabla_i \nabla_j g$ as $\nabla_{ij} g$. Note that $\nabla_{ij} g = \nabla_i g$ if $k_j = 1$.

Now, consider the function $g(k_1, \ldots, k_I) = \log \pi_{k_1 \cdots k_I}$. The application of the difference operator to $\log \pi_{k_1 \cdots k_I}$ produces conditional contrasts of various orders.

Example 3. (3×3 ordered responses (continued).) Three sets of interaction terms are produced by the application of ∇_{12} to $\log \pi_{k_1 k_2}$, $k_1, k_2 = 1, 2, 3$: the zeroth-order term, $\nabla_{12} \log \pi_{11} = \pi_{11} = \omega_0$; first-order terms such as $\nabla_{12} \log \pi_{21} = \nabla_1 \log \pi_{21} = \log(\pi_{21}/\pi_{11}) = \omega_{1(2)}$, the adjacent-categories logit; and second-order terms such as $\nabla_{12} \log \pi_{22} = \log(\pi_{11}\pi_{22}/\pi_{12}\pi_{21}) = \omega_{12(22)}$, the local log odds ratio at (2, 2). Table 1 shows how each cell is mapped to the corresponding ω, with the arrows in the first column indicating cells that are involved in the derivative, the second column indicating the position of the cell, and the last column indicating the result of the operator ∇_{12} on the specific cell.

Note that in the $2 \times 2 \times 2$ example, the third-order interaction term is given by $\nabla_{123} \log \pi_{222} = \omega_{123}$.

The generalization to the 3×3 example requires the application of $\nabla_{1 \cdots I}$ to the cells of a multivariate response. Formally, let $\nabla^i = (\nabla_i)$ be a K_i-vector of ∇_i, $i = 1, \ldots, I$. For example, for 3×3 ordered responses, $\nabla^1 = (\nabla_1, \nabla_1, \nabla_1)^T$. The vector $\nabla = \nabla^I \otimes \cdots \otimes \nabla^1$ is therefore of length $K = K_1 \times \cdots \times K_I$. Because of the associative and commutative properties of the difference operators, each component of ∇ is a copy of $\nabla_{1 \cdots I}$. Define a Hamadan product $\cdot$ between a vector of operators $\nabla = (\nabla_1, \ldots, \nabla_I)^T$ and a vector of functions $g = (g_1, \ldots, g_I)^T$ as $\nabla \cdot g = (\nabla_1 g_1, \ldots, \nabla_I g_I)^T$. We have the following lemma.

Lemma 1. $\omega = A^{-1} \log \pi = \nabla \cdot \log \pi$.

The proof is provided in the appendix. The canonical parameter ω derived from the difference operator can always be expressed in terms of conditional probabilities. For the 3×3 example, $\omega_{1(2)} = \log \pi_{21}/\pi_{11}$ is the conditional logit $P(Y_1 = 2 | Y_2 = 1)/P(Y_1 = 1 | Y_2 = 1)$. In summary, the difference operator ∇ defines a set of interaction parameters that consists of the conditional adjacent-categories logits, the conditional log odds ratios, and the conditional log ratios of odds ratios, and so on. Therefore, Lemma 1 provides a means to identify and interpret the potentially large number of $K_1 \times \cdots \times K_I$ interaction terms of various orders.

While A^{-1} is viewed as a differential operator, the matrix A can be logically construed as an integral operator. Define the addition operator as

$$\triangle_i g(x_1, \ldots, x_i, \ldots, x_I) = \sum_{x_t = x_i}^{K_i} g(x_1, \ldots, x_t, \ldots, x_I),$$

where x_t is in the ith position, $\triangle_{ij} g = \triangle_i(\triangle_j g)$, and $\triangle_i(g_1 \pm g_2) = \triangle_i g_1 \pm \triangle_i g_2$. Further, let $\triangle = \triangle^I \otimes \triangle^{I-1} \otimes \cdots \otimes \triangle^1$, where $\triangle^i = (\triangle_i)$ is a K_i-vector of $\triangle_i$. It can then be proved that

Lemma 2. $s = A^T x = \triangle \cdot x$.

To clarify notation and facilitate the evaluation of the interactions, we use two sets of subscripts. The first set of indexes in ω indicates the items for which differentiations between response levels are made, and the second set, within parenthesis, indicates the levels at which differentiations take place. For example, in the 3×3 case (Table 1), $\omega_{1(2)}$ differentiates at level 2 for item 1 but not item 2. The indexes for items that are not differentiated are not included for the sake of brevity.

3. Model Development and Estimation

The lattice parameterization for a multivariate polytomous response such as (5) forms the basis for the polytomous LD kernel of the proposed IRT models. In the IRT context, each of the canonical parameters ω in (5) is in fact some function of the latent trait θ and will be referred to as functional parameter $\omega(\theta)$. In other words, IRT postulates an individual-specific, unobserved, underlying trait θ such that the probability $\pi_{k_1 \cdots k_I}$ of observing a response pattern $(k_1, \ldots, k_I)$ is a function of θ. Accordingly, $\omega(\theta) = A^{-1} \log \pi(\theta)$ is also a function of θ. In an LD model, θ can be interpreted as the dominant factor that explains the correlation among item responses from the same individual.

Two types of LD models are distinguished: (1) canonical models that directly specify response functions in terms of $\omega(\theta)$ in the IRT kernel, and (2) hybrid models that apply a mixed parameterization of the IRT kernel to transform purely canonical representation into partially marginal probabilities $\mu(\theta)$ and partially canonical interactions $\omega(\theta)$. The marginal probabilities $\mu(\theta)$ appear in the form $P(Y_i \geq k_i|\theta)$, which, of course, is the familiar item response function (IRF) of item i higher than or at level k_i in conventional IRT models. For both types of LD kernels, the manifest probability of subject j, $j = 1, \ldots, J$, with response vector $\mathbf{Y} = \mathbf{y}$ is given by the integral of the kernel

$$p(\mathbf{y}_j) = \int p(\mathbf{y}_j|\theta) dF(\theta), \tag{6}$$

where θ is distributed as $F(\theta)$. The joint likelihood is therefore given by

$$L(\theta|\mathbf{y}) = \prod_{j=1}^{J} p(\mathbf{y}_j).$$

Canonical Kernels

Several models for the canonical kernel can be distinguished. These models are extensions to the locally dependent models of Hoskens and De Boeck (1997, 2001). The LD models of Hoskens and De Boeck are based on extensions to the partial credit model (Masters, 1982). We use the 3×3 trichotomous item pair example to illustrate these models.

The first model is the constant-combination dependency model (CCDM), in which one defines the parameters referring to within-item transitions, such as from (1, 2) to (1, 3), to be random-effects parameters, each being an unweighted function of the latent trait θ:

$$\omega_{i(k_i)} = \theta - \beta_{i(k_i)}, \tag{7}$$

for the ith item, $i = 1, 2$, and $k_i = 2, 3$. In Table 1, these components correspond to the "edge" elements, 21, 31, 12, 13, of the 3×3 table. The remaining $\omega(\theta)$, or the association component, is specified by a corresponding constant term of γ. For example,

$$\omega_{12(k_1k_2)}(\theta) = \gamma_{k_1k_2}, \tag{8}$$

$k_1, k_2 = 2, 3$. In Table 1, these association terms correspond to the "inside" elements of the 3×3 table. The CCDM model specification (7)–(8) implies that individuals differ by within-item transitions (as from (1, 2) to (1, 3)), but not by combined transitions (as from (1, 2) to (2,3)). Each of the latter transitions is subject independent—that is, CCDM specifies that LD between items is not a source of individual differences.

In the second model, the dimension-dependent-combination dependency model (DDCDM), the parameters are defined to be random effects, each being an unweighted function of θ. Specifically, the within-item transitions (edge components) follow (7), and instead of following (8), the remaining ω (inside components) follow the form:

$$\omega_{12(k_1,k_2)}(\theta) = \theta - \gamma_{k_1k_2}. \tag{9}$$

In DDCDM, which is now specified by equations (7) and (9), individuals differ by the transitions within items as well as by combined transitions, but in each case the latent trait plays a similar role. The LD between the items in DDCDM depends upon the underlying trait, and therefore it may vary from person to person.

The third model is a straightforward extension of DDCDM in which each of the ω modeled θ is assigned an item-specific weight λ:

$$\omega_{i(k_i)} = \lambda_i\theta - \beta_{i(k_i)} \tag{10}$$

and

$$\omega_{12(k_1k_2)}(\theta) = \lambda_{12}\theta - \gamma_{k_1k_2}, \tag{11}$$

$k_1, k_2 = 2, 3$. The extension (10) applies to CCDM, as well. For CCDM, however, the association for the 2-parameter logistic (2PL) version is still given by (8). The 2PL DDCDM in (10) and (11) allows for different impacts of the underlying latent trait on the transition.

A fourth model uses two separate latent traits, one for each of the edge components $\omega_{i(k_i)}$ and the inside components $\omega_{ij(k_ik_j)}$. An example would be:

$$\omega_{i(k_i)}(\theta_1) = \theta_1 - \beta_{i(k_i)}, \tag{12}$$

and

$$\omega_{12(k_1k_2)}(\theta_2) = \theta_2 - \gamma_{k_1k_2}. \tag{13}$$

This kind of model can be used to investigate hypotheses that two distinct psychological traits drive the response and the association between responses, respectively (Hoskens & De Boeck, 2001).

All of the canonical models that are presented in this paper can be estimated with the recently developed PROC NLMIXED procedure of SAS (SAS Institute, Inc., 1999). The procedure PROC NLMIXED allows for MML estimates of the item parameters and can be programmed to provide empirical Bayes estimates of the person parameters. The MML approach involves the maximization of the marginal likelihood formed from (6). In particular, PROC NLMIXED only allows for the specification of the prior distribution of θ as normal (e.g., $N(0, \sigma^2)$). Various options are available in PROC NLMIXED for approximating the integrated likelihood and for maximizing the approximation to the likelihood. In our data analysis, we used a nonadaptive Gaussian quadrature method to approximate the likelihood (Pinheiro & Bates, 1995) and a Newton–Raphson procedure to maximize the approximate likelihood.

Hybrid Kernels

Following Ip (2002), the hybrid kernel for polytomous responses exploits the theory of mixed parameterization (Barndorff-Nielsen, 1978). Mixed parameterization of the log-likelihood function (3) proceeds as follows: Let the mean parameter of ω be defined by the mapping $\mu(\omega) = Es = \partial\kappa(\omega)/\partial\omega$, where μ is a one-to-one, continuously differentiable mapping. Consider a partition of the canonical parameter ω into $(\omega^{(1)}, \omega^{(2)})$, and denote the corresponding partition of the mean parameter μ by $(\mu^{(1)}, \mu^{(2)})$. When the joint distribution (3) is parameterized in terms of the hybrid parameter $(\mu^{(1)}, \omega^{(2)})$, the component $\mu^{(1)}$ is called the mean component, and $\omega^{(2)}$ is called the canonical component. It can be proved that $(\mu^{(1)}, \omega^{(2)})$ uniquely determines the joint distribution (Barndorff-Nielsen, 1978, p. 121).

The formalization of model specification using the lattice language enables us to directly apply inferential separation properties of $(\mu^{(1)}, \omega^{(2)})$ developed by Barndorff-Nielsen (1978) and Barndorff-Nielsen and Koudou (1995). One of the most important results of mixed parameterization is that $\mu^{(1)}$ and $\omega^{(2)}$ are variationally independent. That is, the parameter space of $\mu^{(1)}$ is not restricted in any way by the values of $\omega^{(2)}$ and vice versa. This property is useful because without it, the arbitrary value of $(\mu^{(1)}, \omega^{(2)})$ may yield negative probabilities. Furthermore, $\mu^{(1)}$ and $\omega^{(2)}$ are orthogonal, which leads to certain simplifications in computing the variance-covariance matrix of $(\mu^{(1)}, \omega^{(2)})$.

We use the trichotomous item pair to illustrate mixed parameterization. First, group the "edge" components in Table 1—namely, $\omega_{1(2)}$, $\omega_{1(3)}$, $\omega_{2(2)}$, and $\omega_{2(3)}$ with $\omega^{(1)}$ and the remaining "inside" components; $\omega_{12(22)}$, $\omega_{12(32)}$, $\omega_{12(23)}$, and $\omega_{12(33)}$ with $\omega^{(2)}$. Then we have

$$\begin{aligned}\mu^{(1)}(\theta) &= (Es_{21}, Es_{31}, Es_{12}, Es_{13}) \\ &= (P(Y_1 \geq 2|\theta), P(Y_1 \geq 3|\theta), P(Y_2 \geq 2|\theta), P(Y_2 \geq 3|\theta)).\end{aligned}$$

One way to specify the hybrid LD kernel is to separately model the mean (marginal) component $\mu^{(1)}(\theta)$, and canonical component $\omega^{(2)}(\theta)$. The marginal component $\mu^{(1)}(\theta)$ may be specified as a graded response model (Samejima, 1969):

$$\text{logit}\{\mu_{i(k_i)}(\theta)\} = \text{logit}\{P(Y_i \geq k_i|\theta)\} = a_i\theta - b_{i(k_i)}, \tag{14}$$

where $i = 1, 2$, $k_i = 2, 3$, and $\text{logit}(x) = \log[x/(1-x)]$. The canonical component $\omega^{(2)}(\theta)$ may follow any association model used for CCDM or DDCDM, such as (11).

The model (14), when juxtaposed to the 2PL version of DDCDM (10), reveals certain similarities. However, the following differences should be recognized.

Remark 1. Unlike the canonical representation, the interpretation of parameters in the mean component of the hybrid kernel does not change with the size of the cluster I. This property is referred to as semi-reproducibility (Ip, 2002). As a result, a and b in (14) can be interpreted as

discrimination and difficulty parameters. Other potential benefits in using the hybrid kernel are discussed in Ip (2002).

Remark 2. No closed-form equation similar to (5) exists for the hybrid kernel (Fitzmaurice, Laird, & Rotnitzky, 1993).

Remark 3. Because of the cumulative nature of the matrix transformation $s = A^T x$, the mean parameters $\mu = Es$ are cumulative probabilities. This is in contrast to ω, which contains conditional logits and local log odds ratios between adjacent probabilities. Both cumulative and adjacent probabilities (logits) have been used in the literature for models for polytomous item responses of a single dimension, most notably in the context of graded response models and partial credit models, respectively. In brief, the model based on the hybrid kernel can be regarded as a multivariate extension of the graded response model, and that based on the canonical kernel can be considered an extension of the partial credit model. An in-depth investigation of their relationships is the topic of a separate paper now being prepared.

Estimation of the hybrid kernel for the polytomous LD model is based upon a generalized version of the EM algorithm (Dempster, Laird, & Rubin, 1977). The EM algorithm updates parameters by clusters and requires alternate maximization of log-likelihood with respect to item parameters (a, b) and the cluster parameters (λ, γ). In the appendix, we provide some details of the algorithm beyond that which is described in Ip (2002).

4. Data Analysis

Data and Cluster Structure

As an illustration for the canonical and hybrid models, we analyze data that were collected by Vansteelandt (1999) in a study of individual differences in hostile behavior. In this study, 316 Dutch-speaking persons indicated, on a 3-point scale, the extent to which they would display each of 21 hostile responses in each of 23 frustrating situations (1 = not, 2 = limited, 3 = strong). Situations and responses were selected from an S-R inventory of hostility (Endler & Hunt, 1968).

To illustrate our approach to modeling LD, we study the dependencies between a subset of six responses in a set of four situations. The six responses (curse, want to curse, scold, want to scold, shout, want to shout) consist of three behaviors (curse, scold, and shout), each of which is presented in two variants (want to and actually do). This particular set of six responses was selected because it could be hypothesized on the basis of their substantive content that they cluster into sets of locally dependent responses. For instance, responses could be dependent because they are based on the same type of behavior (e.g., curse and want to curse), because they refer to showing aggression openly in a verbal manner (curse, scold, shout), or because they refer to only wanting to show aggression (want to curse, want to scold, want to shout). The four situations consisted of two pairs that express substantively different reasons for becoming frustrated. The first pair ("the grocery store closes just as you are about to enter," and "you use your last 10 cents to call a friend and the operator disconnects you because you don't have enough money") are situations in which the person is to blame himself or herself for the frustrating occurrence. In contrast, with the second pair of situations ("you are waiting at the bus stop and the bus fails to stop for you," and "you miss your train because the clerk has given you the wrong information") a person could blame someone else for their becoming frustrated.

One way to proceed with analyzing the data is to model dependencies between all possible $6 \times \frac{5}{2} = 15$ pairs of responses. Alternatively, one can derive a hypothesis about the cluster structure on the basis of a substantive theory. We adopt the latter approach and propose an explanation,

based on a theory of social inhibition, for a particular cluster structure. Specifically, for a situation in which one can blame oneself, we hypothesize that dependency exists between wanting to display an aggressive behavior and actually showing the behavior, because in a "self-blame" situation the aggression is directed toward oneself and therefore no social inhibition occurs. As a result, for self-blame situations we form clusters for the want and the do variant of each type of behavior (want to curse, curse; want to scold, scold; want to shout, shout), assuming that people tend to do what they want to do.

On the other hand, for others-to-blame situations, it is hypothesized that other factors such as social inhibition should make a direct expression of aggression less likely when it is directed toward others. Behaviors pertaining to openly showing aggression (curse, scold) or wanting to show aggression (want to curse, want to scold) that are directed toward others are grouped together, suggesting a divergence between wanting to do and actually doing. At the same time, shouting (in Dutch, the verb is "het uitschreeuwen," which means a "shouting out" that is not necessarily directed towards others but represents a way of expressing feeling) is not inhibited when one is inclined to shout. Therefore (want to shout, shout) forms the remaining cluster. Table 2 summarizes the cluster structure that is used for subsequent LD modeling.

Canonical Analysis

In this case of two trichotomous items (within an item cluster), the probability of a response pattern can be expressed as:

$$P(Y_1 = y_1, Y_2 = y_2|\theta) = \exp[s_{11}\omega_0 + s_{21}\omega_{1(2)} + s_{31}\omega_{1(3)} + s_{12}\omega_{2(2)} + s_{22}\omega_{12(22)} + s_{32}\omega_{12(32)} + s_{13}\omega_{2(3)} + s_{23}\omega_{12(23)} + s_{33}\omega_{12(33)}]/v(\omega),$$

where $s_{y_1,y_2} = I(Y_1 \geq y_1, Y_2 \geq y_2)$, $v(\omega)$ is the normalizing constant, and the argument θ in $\omega(\theta)$ is abbreviated. See Table 1.

In order to investigate the dependencies between pairs of responses, we compared several models, the parameterizations of which are summarized in Table 3. They include (1) LIM, a locally independent model; (2) CCDM, specified by (7) and (8); (3) DDCDM, specified by (7) and (9); and (4) two-dimensional (2D)-DDCDM, specified by (12) and (13).

In order to validate the dependency structure, we include in Table 3 the results of a bootstrap experiment that uses correlation matrix, a device that was suggested by a referee. For each model, we generated 2,000 parametric bootstrapped response data sets using the estimated parameters. Then for each bootstrapped data set, we computed the correlations of $C_2^6 = 15$ item pairs for each of the four situations. Accordingly, there were $15 \times 4 = 60$ correlations—$3 \times 4 = 12$ from clustered pairs and the remaining 48 from nonclustered pairs. We report in Table 3 the proportions of observed correlations for both clustered and nonclustered pairs that fall within

TABLE 2.
Cluster structure of situations and items

Situations	Cluster 1	Cluster 2	Cluster 3
Self-blame			
Grocery store	want to curse	want to scold	want to shout
Phone call	curse	scold	shout
Others-to-blame			
Wait for bus	want to curse	curse	want to shout
Miss train	want to scold	scold	shout

TABLE 3.
Parameterization, number of parameters (Npar), AIC and BIC of models fitted to S-R data

		LIM	CCDM	DDCDM	2D-DDCDMotb
Parameterization	ω_0	0	0	0	0
	$\omega_{1(2)}$	$\theta - \beta_{1(2)}$	$\theta - \beta_{1(2)}$	$\theta - \beta_{1(2)}$	$\theta_1 - \beta_{1(2)}$
	$\omega_{1(3)}$	$\theta - \beta_{1(3)}$	$\theta - \beta_{1(3)}$	$\theta - \beta_{1(3)}$	$\theta_1 - \beta_{1(3)}$
	$\omega_{2(2)}$	$\theta - \beta_{2(2)}$	$\theta - \beta_{2(2)}$	$\theta - \beta_{2(2)}$	$\theta_1 - \beta_{2(2)}$
	$\omega_{2(3)}$	$\theta - \beta_{2(3)}$	$\theta - \beta_{2(3)}$	$\theta - \beta_{2(3)}$	$\theta_1 - \beta_{2(3)}$
	$\omega_{12(22)}$	0	$-\gamma_{22}$	$\lambda\theta - \gamma_{22}$	$\theta_2 - \gamma_{22}$
	$\omega_{12(23)}$	0	$-\gamma_{23}$	$\lambda\theta - \gamma_{23}$	$\theta_2 - \gamma_{23}$
	$\omega_{12(32)}$	0	$-\gamma_{32}$	$\lambda\theta - \gamma_{32}$	$\theta_2 - \gamma_{32}$
	$\omega_{12(33)}$	0	$-\gamma_{33}$	$\lambda\theta - \gamma_{33}$	$\theta_2 - \gamma_{33}$
Correlation	Pclustered*	$\frac{1}{12}$	$\frac{12}{12}$	$\frac{12}{12}$	$\frac{12}{12}$
	Pothers	$\frac{40}{48}$	$\frac{38}{48}$	$\frac{40}{48}$	$\frac{37}{48}$
Measure	Npar**	49	97	109	99
	AIC	12743	12260	12268	12356
	BIC	12927	12624	12677	12728

*Pclustered (Pothers) indicates the proportion of observed correlations between responses from clustered (nonclustered) item pairs that fall within a 95% bootstrap confidence interval.

**Npar includes the parameter σ^2 specified in the prior. For item parameters in LIM, there are 4 (situations) $\times$3 (clusters) $\times$2 (items per cluster) $\times$2 (logits for each item) = 48. In CCDM, we added to LIM another 4 (situations) $\times$3 (clusters) $\times$4 (interactions per cluster) = 48, bringing the total to 96. In DDCDM, each cluster has an identical weight parameter λ. Therefore, we added to CCDM another 4 (situations) $\times$3 (clusters) = 12, bringing the total to 108. For 2D-DDCDMotb, the number of parameters is (number of parameters in CCDM) plus 2 (from the variance of the second latent trait and the covariance).

the respective 95 percent bootstrap confidence intervals. The results strongly support the use of locally dependent models over the locally independent model.

Table 3 also summarizes the information about the parsimoniousness and the fit of the different models. The table shows the number of parameters and some well-known likelihood-based information criteria, namely AIC (Akaike, 1973) and BIC (Schwarz, 1978), for model choice. The information criteria AIC and BIC both yield the same conclusion with regard to which models fit better: both the CCDM and DDCDM fit better than LIM. This confirms the need for modeling dependencies between responses within clusters. Furthermore, the fact that CCDM has both a lower AIC and BIC suggests that dependency is not dimension-dependent.

For the best-fitting model, CCDM, the interaction parameters $\omega_{12(22)}$, $\omega_{12(23)}$, $\omega_{12(32)}$, and $\omega_{12(33)}$ were in the ranges [0.47, 2.4], [-1.73, 10.74], [-1.33, 1.55], and [0.44, 3.06], respectively, and they had median values of 1.7, -1.33, -0.07, and 1.86. Hence, interaction terms $\omega_{12(22)}$ and $\omega_{12(33)}$ (see Table 1) are always positive, whereas interactions $\omega_{12(23)}$, $\omega_{12(32)}$ can attain both positive and negative values. These consistent patterns observed throughout the clusters imply that subjects who have the same general aggressiveness tend to indicate that, across the various situations, they would display related behaviors to the same extent (e.g., strong-strong) rather than to different extents (e.g., strong-limited).

The facts that the CCDM fits better than the LIM and that the interaction terms are consistent across situations lend support to the social inhibition theory. The above empirical findings suggest that, for self-blame situations, being inclined to exhibit a certain type of aggressive behavior is not socially inhibited and thus may lead to actually exhibiting that behavior. For situations in which another party could be blamed, showing (curse, scold) and only being inclined to show aggressive behaviors (want to curse, want to scold) directed toward others are positively related. Furthermore, in others-to-blame situations, being inclined to "shout" provokes actually "shout-

ing," because this behavior does not necessarily imply aggression toward others. In addition, because the DDCDM does not fit better than the CCDM, the dependencies between behaviors do not seem to vary with the general aggressiveness of a person.

In order to further investigate the social inhibition phenomenon, we conducted two types of additional analyses. First, we validated our hypothesized social inhibition theory to give rise to the clustering structure specified in Table 2. We studied several competing structures for clustering responses by examining the fit of their respective CCDM models. In the first structure, the clusters for the three behaviors of cursing, scolding, and shouting were symmetric. The second and third were control structures, in which cursing and scolding were each treated the way shouting was treated in the hypothesized model. Specifically, the following three alternative structures were investigated.

1. CCDMa, a structure that clusters "do" and "want" responses of each type of behavior. The clusters for CCDMa are: want to curse, curse; want to scold, scold; and want to shout, shout.
2. CCDMb, a structure for which the clusters are: want to scold, want to shout; scold, shout; and want to curse, curse.
3. CCDMc, a structure with the following clusters: want to curse, want to shout; curse, shout; and want to scold, scold.

Fitting the CCDM to these alternative clusters of responses produced the following goodness-of-fit statistics: AIC equals 12430, 12475, and 12524, and BIC equals 12794, 12839, and 12889, respectively, for CCDMa, CCDMb, and CCDMc. Compared with the fitted CCDM model in Table 3 (AIC = 12260, BIC = 12624), all three alternative models performed worse, which lent support to the social inhibition theory. In particular, when we performed a principal components analysis of the responses per situation, the PCA result further validated the clustering structure of the responses that we hypothesized and defined in Table 2. We also found in PCA analysis that there did not exist a general component. This suggests that the plausible within-situation dependency, if present, is not general but rather differentiated among the specific types of behavior or among the variants of wanting and doing.

We conducted a second study to investigate whether there is evidence for a separate latent trait that arises from social inhibition. It can be hypothesized that individual differences in social inhibition constitute a second latent trait that is distinct from the general tendency to exhibit aggression. Accordingly, we estimated a model that assumes that interactions between responses within clusters arise from a second latent trait of social inhibition (Equation (13)). The model, labeled 2D-DDCDMotb (Table 3), distinguishes between the traits to exhibit aggression θ_1 and to suppress aggression directed toward others θ_2. The results in Table 3 suggest that the model 2D-DDCDMotb fits better than the alternative models CCDMa, CCDMb, and CCDMc. The 2D-DDCDMotb model, however, has AIC and BIC values higher than the more parsimonious CCDM. Conclusions concerning individual differences in social inhibition therefore are not definitive and need further investigation.

A sample of SAS codes for a simplified structure for the above procedure, with explanatory notes, is provided in the appendix.

Hybrid Analysis

We propose a family of hybrid models to analyze the hostility data: a marginal constant-combination dependency model (MCCDM) and a marginal dimension-dependent-combination dependency model (MDDCDM). The uniqe difference between the hybrid and the corresponding canonical models is their specification of the "edge" elements of the responses. For example, in

TABLE 4.
Number of parameters and AIC for different types of locally dependent hybrid kernel models

Model	Number of parameters	AIC
LIM	48	12743
MCCDM	96	11677
MDDCDM	108	11817
2PL-LIM	72	13091
2PL-MCCDM	120	12084
2PL-MDDCDM	132	12160

MCCDM, the marginal component $P(Y_i \geq k_i|\theta)$ follows (14), while the association component follows the CCDM equations (7) and (8).

Subsequently, the hostility data were analyzed using the following hybrid models: (1) LIM, a locally independent model, (2) MCCDM, a marginal version of CCDM, with $a_i = 1$ for all i in (14), and (3) MDDCDM, a marginal version of DDCDM, also with $a_i = 1$ in (14).

For MDDCDM, we found that out of the 12 clusters, only one showed a slightly significant value of λ at the 0.05 level. This finding suggests that LD does not vary with the latent trait. An analysis based on AIC (Table 4) validates this finding. The value of AIC is lower for MCCDM than for MDDCDM. Because AIC values are determined up to a constant, we shifted AIC values so that the value for the locally independent model coincides with that in Table 3. Despite the very different formulations and interpretations of the canonical and hybrid models, we found that both of them strongly suggest the existence of LD, which is evidence to reject the locally independent model. Also, neither support the hypothesis that the dependency is a function of the latent trait of aggressiveness.

Table 4 shows that the MCCDM is the best among the several models. For this model, the range of the association term γ is $[-3.20, 2.17]$ with a mean standard deviation of 0.47. Figure 3

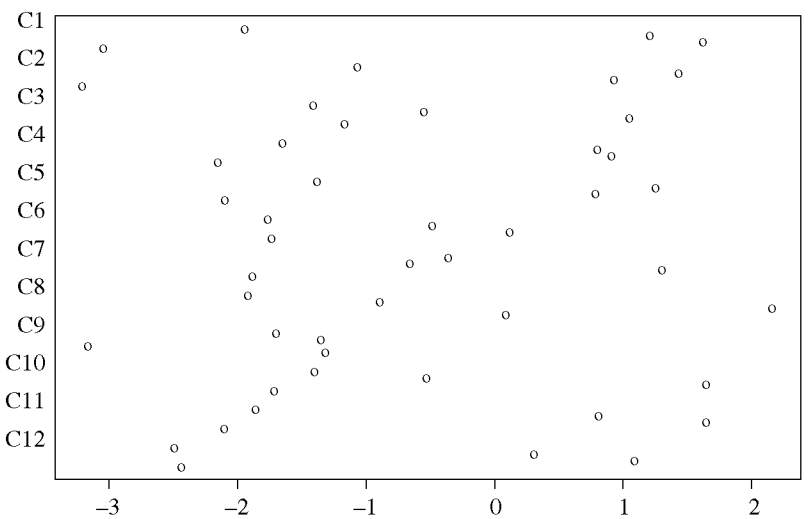


FIGURE 3.
Dot chart showing the distribution of the association parameter for the canonical component for the first cluster C1 through the last cluster C12.

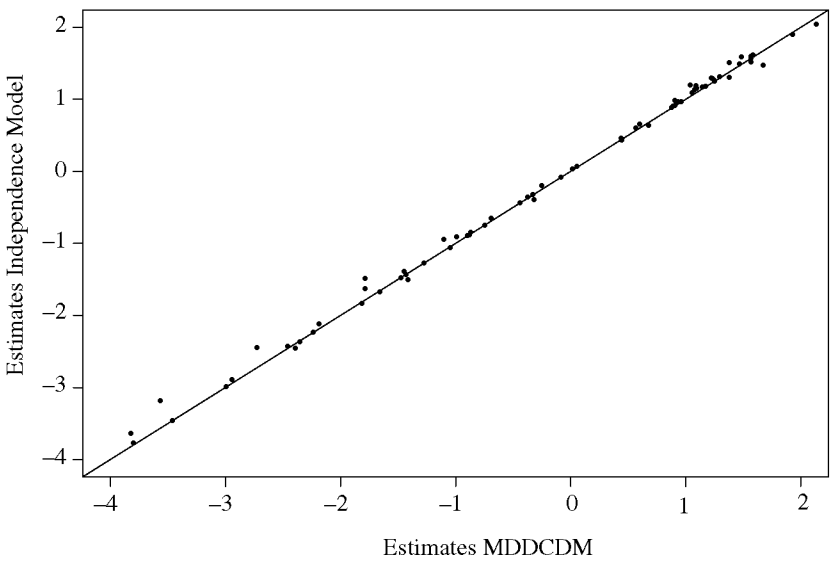


FIGURE 4.
Scatterplot showing the corresponding item parameters in the 2PL independence model and the 2PL version of the hybrid model MDDCDM in which the association parameters are assumed not to be a function of latent trait.

is a dot chart that shows the heterogeneity of values of interaction terms within clusters. The dot chart suggests that a more parsimonious model, such as one restricting the interactions within a cluster to the same value, may not be appropriate. Furthermore, the dot chart shows that there is a significantly larger number of negative values, indicating that there generally exists a positive association between item responses within clusters.

We conducted experiments that used 2PL versions in the marginal models, and the results are summarized in Table 4. The goodness-of-fit statistics in terms of AIC do not seem to provide support for the use of 2PL models.

We also found that the item parameters remain rather stable across the three models. Figure 4 shows the values of the item parameters a and b estimated under the 2PL versions of the locally independent model 2PL-LIM and the 2PL-MCCDM. The results from other models look rather similar. This confirms previous findings that estimates of item parameters are generally not adversely biased by LD (e.g., Ip, 2001; Junker, 1991). Note that the a and b parameters do have identical interpretation—for example, a is the discrimination parameter, and b is the difficulty parameter—across the models. This property can be further exploited if auxillary information about item difficulty or discrimination becomes available. Unlike the canonical models, the additional information can be readily accommodated into the hybrid model without too much effort.

The interpretability of the hybrid models comes with a price. Its computational overhead is higher than models derived from a canonical kernel, and no commercial software is immediately available for model estimation. The above analyses for the hybrid model were conducted using a customized program written in Splus (Becker, Chambers, & Wilks, 1988). The program, which took approximately one minute to complete one iteration of the EM algorithm, is rather lengthy and is not included here because of space limitations. Interested readers could contact the corresponding author and ask for a copy.

5. Discussion

This paper contains several contributions. First, we demonstrate the use of a novel and compact notation of a matrix to represent a system of multivariate polytomous responses. The representation allows for the direct modeling of local dependency. Second, we establish a formal statistical framework that leads to two general classes of LD latent trait models, canonical and hybrid, for polytomous responses. We believe that these models significantly expand the scope of IRT to cover item responses that are clustered, by design or otherwise. Finally, we illustrate the utility of these models through a psychological data set from the inventory of hostility. The application shows the capacity of LD models beyond their simple accounting of dependency as a nuisance factor.

It can be argued that using a multivariate probit model can accomplish the same goal of modeling LD polytomous responses. A multivariate probit model is able to use orthant probabilities formed from integrating specific regions of a multivariate normal distribution to specify the joint distribution of multivariate polytomous responses. However, the resulting probit models may not have sufficient flexibility for extensive modeling efforts. For example, the associations between categories are often indirectly specified in terms of the covariance structure of the underlying Gaussian distribution. In addition, constructing models for high-order interactions among responses using a multivariate probit approach is not a trivial task. Preferably, one would want to use a general form for representing multivariate categorical data so that interactions within or between items can be extensively studied—an approach that we adopt in this paper. Furthermore, unlike the multivariate probit model, the proposed method separates the model for the joint distribution of the responses from the link function (e.g., probit, logistic), which allows greater flexibility in the specification of the link function.

An alternative way to model LD among polytomous items within a cluster is to use random effects models in such a way that items from the same cluster have a common random effect, the variance of which may or may not vary across clusters (Bradlow, Wainer, & Wang, 1999; Scott & Ip, 2002). Such models are most appropriate for cases in which LD is considered a nuisance because random effects models might not be sufficiently flexible to specify characteristics of substantive interest—for example, when LD increases with the latent trait, or when item responses are negatively correlated given the latent trait.

Broadly speaking, a multidimensional model could also serve the purpose of being a meaningful model for LD. The factor scores of additional dimensions can be construed as measuring underlying traits not necessarily captured by a unidimensional model. However, as Bejar (1983, p. 31) argued, unidimensionality has a primary role in most psychological measurement applications. As long as items function in unison—that is, response and performance on each item are affected by the same process and in the same form—unidimensionality will hold. Keeves (1997) and Bond and Fox (2001, p.103) echoed this view and provided empirical evidence that supports the unidimensionality assumption in several longitudinal and developmental studies. The argument implies that unidimensional models are still relevant to a wide range of tests that might appear to have a multidimensional nature by inspection (Keeves, 1997). For example, in a test of several domain areas, unidimensional models can be constructed for each individual domain area. Although the LD model proposed in this paper is not necessarily unidimensional, for many practical purposes an underlying unidimensional latent trait is sufficient. Accordingly, we designed LD models for the item-person interaction that are beyond the explanation of a single latent trait. We also demonstrated how two conceptually discernable latent traits, not necessarily formulated under the framework of multi-dimensional IRT, can be incorporated into separate components of a LD model.

While it is true that, from a purely technical perspective, the questions of whether LD is a nuisance factor or of substantive interest are not important, there is an important difference

between the emphases of the two views. Treating LD as a nuisance would imply that dependency needs to be accounted for only to the extent that it affects the estimation of quantities of primary interest (e.g., item parameter, individual traits). In such cases, a crude model for LD (e.g., Ip, 2000) suffices. However, when LD itself is of interest, as in the cases we presented in this paper, then carefully constructed, viable, interpretable, and extensible models for LD would become a critical component of the modeling activities. For example, when collateral information for LD (e.g., item cluster attributes; see Fischer, 1994) is available, it can be included as input to regression-type linear logit models to replace equations such as (8)–(11). A research project along these lines is currently in progress.

Appendix A: Proof of Lemmas

The lemmas in the subsection *Difference and Integral Operators* can be proved by mathematical induction on J in

$$A^{-1}\log\pi = B_J^{-1}\otimes\cdots\otimes B_1^{-1}\log\pi = \nabla\cdot\log\pi.$$

Denote the k_jth row of (B_j^{-1}) by $(B_{jk_j}^{-1})^T$. Except when $k_j = 1$, $B_{jk_j}^{-1}$ is a vector that has -1 and $+1$, respectively, at its (k_j-1)th and (k_j)th positions (Equation (1)), and 0 otherwise. When $k_j = 1$, it is a unit coordinate vector with a 1 at its first position. It suffices to show that

$$\nabla_J\cdots\nabla_1\log\pi_{k_1\cdots k_J} = \left(B_{Jk_J}^{-1}\otimes\cdots\otimes B_{1k_1}^{-1}\right)^T\log\pi, \tag{A15}$$

for $k_j = 1,\ldots,K_j$, $j = 1,\ldots,J$. When $J = 1$, the case when $k_1 = 1$ is trivial, both sides of (A15) are $\log\pi_1$. When

$$K_1 \geq k_1 > 1, \text{LHS} = \nabla^1\log\pi_{k_1} = -\log\pi_{k_1-1} + \log\pi_{k_1} = \left(B_{1k_1}^{-1}\right)^T\log\pi = \text{RHS}.$$

Assume that (A15) is true when $J = m$. The vector $\log\pi$ for $J = m+1$ can be constructed by $\text{vec}((\log\pi)_t)$, $t = 1,\ldots,K_{m+1}$, where $(\log\pi)_t$ is the probability vector for $J = m$, with an additional index t for its $(m+1)$th dimension.

Consider the two cases in which $k_{m+1} = 1$ and $K_{m+1} \geq k_{m+1} > 1$. For the former, we observe that the first element in ∇^{m+1} is 1 and $B_{m+1,1}^{-1}$ is a unit coordinate vector with a 1 at the first position. Therefore, LHS in (A15) is $(1\ \nabla_m\cdots\nabla_1)\log\pi_{k_1\cdots 1}$, while

$$\text{RHS} = \left(B_{mk_m}^{-1}\otimes\cdots\otimes B_{1k_1}^{-1}\right)^T(\log\pi)_1.$$

The equality is established by the induction assumption. For the latter case, when $K_{m+1} \geq k_{m+1} > 1$,

$$\begin{aligned}
\nabla_{(m+1)}\cdots\nabla_1\log\pi_{k_1\cdots k_{m+1}} &= -\nabla_m\cdots\nabla_1\log\pi_{k_1\cdots(k_{m+1}-1)} + \nabla_m\cdots\nabla_1\log\pi_{k_1\cdots k_{m+1}}\\
&= -\left(B_{mk_m}^{-1}\otimes\cdots\otimes B_{1k_1}^{-1}\right)^T(\log\pi)_{k_{m+1}-1}\\
&\quad + \left(B_{mk_m}^{-1}\otimes\cdots\otimes B_{1k_1}^{-1}\right)^T(\log\pi)_{k_{m+1}}\\
&= \left(B_{(m+1)k_{m+1}}^{-1}\otimes\cdots\otimes B_{1k_1}^{-1}\right)^T\log\pi.
\end{aligned}$$

Appendix B: Estimation of Hybrid Kernel

This section outlines some specific and important details required for marginal maximum likelihood estimation beyond what is described in Ip (2002). From (6), for simplicity we write

$dF(\theta) = g(\boldsymbol{\theta}|\eta)d\boldsymbol{\theta}$, where $g(\boldsymbol{\theta}|\eta)$ is the prior density of latent traits with parameters η, and we assume that η is known. Assuming independence between subjects, the log-likelihood of jointly observing the collection of response patterns $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_J)$ from J subjects is

$$L = \sum_{j=1}^{J} \log p(\mathbf{y}_j). \tag{A16}$$

Let ξ denote the Q item parameters of one-way IRFs, and let α denote R cluster parameters that specify the association component.

The key to solving the log-likelihood equation (A16) is to obtain derivatives of the kernel with respect to item and cluster parameters and the posterior distribution of θ, because

$$\frac{\partial L}{\partial \xi} = \sum_{j=1}^{J} \int \frac{\partial}{\partial \xi}[\log p(\mathbf{y}_j|\theta, \xi, \alpha)][P(\theta|\mathbf{Y}_j, \xi, \alpha)]d\theta, \tag{A17}$$

where $P(\theta|\mathbf{Y}, \xi, \alpha)$ is the posterior density of θ given $(\mathbf{Y}_j, \xi, \alpha)$ (Baker, 1992, p.175). A similar equation can be obtained for the derivative with respect to α.

To obtain the derivatives, we first compute the covariance function of $s^\star$, the reduced set of s that deletes the element 1. Consider the corresponding reduced matrix $A^\star$ given by

$$A^T = \begin{pmatrix} \mathbf{1} \\ A^{\star T} \end{pmatrix},$$

where $\mathbf{1}$ is the row vector of 1's. Denote the covariance matrix $\mathrm{Cov}(s^\star) = A^\star \mathrm{Cov}(\mathbf{y}_j)A^{\star T}$ by V_j, where $\mathrm{Cov}(\mathbf{y}_j) = \mathrm{diag}(\pi_j) - \pi_j \pi_j^T$. Furthermore, denote the log kernel of the jth student $\log p(\mathbf{y}_j|\theta, \xi, \alpha)$ by ℓ_j. Suppose the hybrid kernel is $(\phi^{(1)}, \tau^{(2)})$, where $\phi^{(1)} = \mu^{(1)}$ and $\tau^{(2)} = \omega^{(2)}$.

The following equations of the derivative are essential to solving the maximum likelihood estimates (Fitzmaurice & Laird, 1993; Ip, 2002).

$$\frac{\partial \ell_j}{\partial \xi} = \frac{\partial \phi_j^{(1)}}{\partial \xi} V_{j11}^{-1}\left(s_j^{(1)} - \phi_j^{(1)}\right), \tag{A18}$$

$$\frac{\partial \ell_j}{\partial \alpha} = \frac{\partial \tau_j^{(2)}}{\partial \alpha}\left(s_j^{(2)} - \mu_j^{(2)} - V_{j21}V_{j11}^{-1}\left(s_j^{(1)} - \phi_j^{(1)}\right)\right), \tag{A19}$$

where

$$V_j = \begin{pmatrix} V_{j11} & V_{j21}^T \\ V_{j21} & V_{j22} \end{pmatrix}.$$

The Hessians are given by

$$\frac{\partial^2 \ell_j}{\partial \xi^T \partial \xi} = -\left(\frac{\partial \phi_j^{(1)}}{\partial \xi}\right) V_{j11}^{-1} \left(\frac{\partial \phi_j^{(1)}}{\partial \xi}\right)^T,$$

$$\frac{\partial^2 \ell_j}{\partial \alpha^T \partial \alpha} = -\left(\frac{\partial \tau_j^{(2)}}{\partial \alpha}\right) \left(V_{j22} - V_{j21}V_{j11}^{-1}V_{j21}^T\right) \left(\frac{\partial \tau_j^{(2)}}{\partial \alpha}\right)^T.$$

We give an example of the derivatives from the 2PL version of model MDDCDM, which is described in the subsection of this paper on hybrid analysis. For simplicity, ignore the subscript for cluster. The item parameters $\xi = (a_1, b_{1(2)}, b_{1(3)}, a_2, b_{2(2)}, b_{2(3)})$, and the cluster parameters $\alpha = (\lambda_{12}, \gamma_{22}, \gamma_{23}, \gamma_{32}, \gamma_{33})$ are, respectively, vectors of length 6 and 5. Then

$$\frac{\partial \phi^{(1)}}{\partial \xi} = \begin{pmatrix} \mu_{1(2)}\bar{\mu}_{1(2)}\theta & \mu_{1(3)}\bar{\mu}_{1(3)}\theta & 0 & 0 \\ \mu_{1(2)}\bar{\mu}_{1(2)} & 0 & 0 & 0 \\ 0 & \mu_{1(3)}\bar{\mu}_{1(3)} & 0 & 0 \\ 0 & 0 & \mu_{2(2)}\bar{\mu}_{2(2)}\theta & \mu_{2(3)}\bar{\mu}_{2(3)}\theta \\ 0 & 0 & \mu_{2(2)}\bar{\mu}_{2(2)} & 0 \\ 0 & 0 & 0 & \mu_{2(3)}\bar{\mu}_{2(3)} \end{pmatrix},$$

$$\frac{\partial \omega^{(2)}}{\partial \alpha} = \begin{pmatrix} \theta & \theta & \theta & \theta \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

where $\bar{\mu} = 1 - \mu$.

Note that $\phi_j^{(1)}$, $\tau_j^{(2)}$, V_{j11}, and V_{j21} are all functions of θ. This implies that the computation of joint probability for each subject needs to be evaluated individually in such a way that the marginal distribution fits the mean component and the set of remaining interactions fits the canonical component. The procedure could be carried out by an iterative proportional fitting (IPF) algorithm (Deming & Stephan, 1940). The IPF starts with appropriate odds ratios and fits cell probabilities to given margins. Odds ratios are invariant to row and column adjustments in IPF iterations. In the example of a trichotomous item pair, the four odds ratios (σ_1, σ_2, σ_3, σ_4), which specify the inside component in Table 1, are set in accordance with Table A1.

Because θ is not known, the derivative in (A17) is replaced by their conditional expectation, given the observed responses and some provisional value of α. Similarly, the derivative for α can be evaluated. An EM algorithm solves the maximum likelihood estimation by alternating between an M-step that maximizes with respect to ξ and α one cluster at a time, and an E-step that computes the conditional expectations. Within the M-step, one alternates between maximization with respect to ξ and α until a convergence criterion is satisfied.

TABLE A1.
Starting value for iterative proportional fitting

$\sigma_1\sigma_2\sigma_3\sigma_4$	$\sigma_3\sigma_4$	1
$\sigma_2\sigma_4$	σ_4	1
1	1	1

Appendix C: Sample SAS Codes for CCDM with Two Clusters Each Containing One Pair of Trichotomous Items

```
1. data ccdm;
2. infile 'c:\data.txt';
3. input person cat disc i1 i2
         cb12_1 cb13_1 cb22_1 cb23_1 cb12_2 cb13_2 cb22_2 cb23_2
         cg22_1 cg23_1 cg32_1 cg33_1 cg22_2 cg23_2 cg32_2 cg33_2;
```

```
4. run;
5. proc sort data=ccdm;
6. by person;
7. run;

8. PROC NLMIXED data=ccdm noad technique=newrap qpoints=20;
9. parms b12_1=0 b13_1=0 b22_1=0 b23_1=0 g22_1=0 g23_1=0 g32_1=0 g33_1=0
         b12_2=0 b13_2=0 b22_2=0 b23_2=0 g22_2=0 g23_2=0 g32_2=0 g33_2=0
         sigma=1;
10. num1=disc*sigma*th
         +b12_1*cb12_1+b13_1*cb13_1+b22_1*cb22_1+b23_1*cb23_1
         +g22_1*cg22_1+g23_1*cg23_1+g32_1*cg32_1+g33_1*cg33_1;
11. denom1=1+exp(sigma*th-b22_1)+exp(2*sigma*th-b22_1-b23_1)
           +exp(sigma*th-b12_1)+exp(2*sigma*th-b12_1-b22_1-g22_1)
           +exp(3*sigma*th-b22_1-b23_1-b12_1-g22_1-g23_1)
           +exp(2*sigma*th-b12_1-b13_1)+exp(3*sigma*th-b12_1-b13_1-b22_1
           -g22_1-g32_1)+exp(4*sigma*th-b12_1-b13_1-b22_1-b23_1-g22_1
           -g23_1-g32_1-g33_1);
12. num2=disc*sigma*th
         +b12_2*cb12_2+b13_2*cb13_2+b22_2*cb22_2+b23_2*cb23_2
         +g22_2*cg22_2+g23_2*cg23_2+g32_2*cg32_2+g33_2*cg33_2;
13. denom2=1+exp(sigma*th-b22_2)+exp(2*sigma*th-b22_2-b23_2)
           +exp(sigma*th-b12_2)+exp(2*sigma*th-b12_2-b22_2-g22_2)
           +exp(3*sigma*th-b22_2-b23_2-b12_2-g22_2-g23_2)
           +exp(2*sigma*th-b12_2-b13_2)+exp(3*sigma*th-b12_2-b13_2-b22_2
           -g22_2-g32_2)+exp(4*sigma*th-b12_2-b13_2-b22_2-b23_2-g22_2
           -g23_2-g32_2-g33_2);
14. loglik= i1*(num1-log(denom1))+i2*(num2-log(denom2));
15. model cat~general(loglik);
16. random th~normal(0,1) subject=person;
17. run;
```

Explanatory Notes for SAS Codes

Input

Line 1-2 : Read observations from file "data.txt"

An example of data file

```
Person cat disc i1 i2 cb12_1 ... cg33_2;
1      3   2    1  0     ...
1      9   4    0  1
2      1   0    1  0
2      2   1    0  1
...
316    5   2    1  0
316    1   0    0  1
```

Line 3: Specify 21 variables (for 2 pairs of trichotomous items in CCDM): (1) "person" = person ID, (2) "cat" = observed scores of persons on clusters (1..9), (3) "disc" = weight for latent

variable, (4–5) "i1" and "i2" = indicators whether observation is from cluster 1 or 2, 1 = yes, 0 = no, (6–21) 16 cluster-specific design variables. CB and CG code beta and gamma parameters of the CCDM (Table 3), respectively.

The design variables CB and CG are defined as follows:

Y1	Y2	CAT	CB12	CB13	CB22	CB23	CG22	CG23	CG32	CG33
1	1	1	0	0	0	0	0	0	0	0
1	2	2	0	0	-1	0	0	0	0	0
1	3	3	0	0	-1	-1	0	0	0	0
2	1	4	-1	0	0	0	0	0	0	0
2	2	5	-1	0	-1	0	-1	0	0	0
2	3	6	-1	0	-1	-1	-1	-1	0	0
3	1	7	-1	-1	0	0	0	0	0	0
3	2	8	-1	-1	-1	0	-1	0	-1	0
3	3	9	-1	-1	-1	-1	-1	-1	-1	-1

Example of a response function derived from the design variables:

$$P(cat = 5) = \frac{\exp(2\theta - \beta_{1(2)} - \beta_{2(2)} - \gamma_{22})}{\nu(\theta)}$$

where the numerator corresponds to the 5th row of the design matrix, and the denominator is the sum of numerators for all categories. The logits in the numerator are given in Table 3.

Group

line 5–7: Group observations by person.

Analyze

Line 8: Specify options.

Line 9 : Assign initial values to model parameters. Note sigma is the SD of theta ("th").

Line 10–15: Compute the general form of loglikelihood.

Line 16: Specify theta ("th") as normal with mean 0 and SD 1, so `sigma*th~N(0,sigma^2)`.

Remark: The run time of one analysis of 12 clusters was approximately 1 hour on a Pentium IV 1500mHz machine.

References

Aigner, M. (1979). *Combinatorial Theory*. Berlin: Springer-Verlag.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 271–281). Budapest, Hungary: Academiai Kiado.

Baker, F.B. (1992). *Item Response Theory*. New York: Marcel Dekker.

Barndorff-Nielsen, O.E. (1978). *Information and Exponential Families in Statistical Theory*. New York: John Wiley.

Barndorff-Nielsen, O.E., & Koudou, A.E. (1995). Cuts in natural exponential family. *Theory of Probability and Its Applications, 40*, 361–372.

Becker, R.A., Chambers, J.M., & Wilks, A.R. (1988). *The New S Language*. New York: Chapman & Hall.

Bishop, Y., Fienberg, S.E., & Holland, P. (1975). *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.

Bejar, I.I. (1983). *Achievement Testing: Recent Advances*. Beverly Hills, CA: Sage.

Bond, T., & Fox, C.M. (2001). *Applying the Rasch model: Fundamental Measurements in the Human Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Bradlow, E., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*, 153–168.
Cox, D.R. (1972). The analysis of multivariate binary data. *Applied Statistics, 21*, 113–120.
Deming, W.E., & Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics, 11*, 427–444.
Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Ser. B, 39*, 1–38.
Embretson (Whitely), S.E. (1984). A general latent model for response processes. *Psychometrika, 49*, 175–186.
Endler, N.S., & Hunt, J.M. (1968). S-R inventories of hostility and comparisons of the proportions of variance from persons, behaviors, and situations for hostility and anxiousness. *Journal of Personality and Social Psychology, 9*, 309–315.
Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359–374.
Fischer, G.H. (1974). *Introduction to the Theory of Psychological Tests* (in German). Bern, Switzerland: Huber.
Fischer, G.H. (1994). The linear logistic test model. In G.H. Fischer & I. W. Molenaar (Eds.), *Rasch Models: Foundation, Recent Developments, and Applications* (pp. 131–156). New York: Springer-Verlag.
Fitzmaurice, G.M., & Laird, N.M. (1993). A likelihood-based method for analyzing longitudinal binary responses. *Biometrika, 80*, 141–151.
Fitzmaurice, G.M., Laird, N.M., & Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statistical Science, 8*, 284–309.
Gibbons, R.D., & Hedekker, D.R. (1992). Full-information item bi-factor analysis. *Psychometrika, 57*, 423–436.
Holland, P. (1990). The Dutch identity: A new tool for the study of item response models. *Psychometrika, 55*, 5–18.
Hoskens, M., & De Boeck, P. (1995). Componential IRT models for polytomous items. *Journal of Educational Measurement, 32*, 364–384.
Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological Methods, 2*, 261–277.
Hoskens, M., & De Boeck, P. (2001). Multidimensional componential item response models for polytomous items. *Applied Psychological Measurement, 25*, 19–37.
Ip, E.H. (2000). Adjusting for information inflation due to local dependency in moderately large item clusters. *Psychometrika, 65*, 73–91.
Ip, E.H. (2001). Testing for local dependency in dichotomous and polytomous item response models. *Psychometrika, 66*, 109–132.
Ip, E.H. (2002). Locally dependent latent trait model and the Dutch identity revisited. *Psychometrika, 67*, 367–386.
Ip, E.H., & Wang, Y. J. (2003). A strategy for designing telescoping models for analyzing multiway contingency tables using mixed parameters. *Sociological Methods and Research, 31*, 291–324.
Jannarone, R.J. (1986). Conjunctive item response theory kernels. *Psychometrika, 51*, 357–373.
Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56*, 255–278.
Keeves, J. P. (1997, March). *International Practice in Rasch Measurement, with Particular Reference to Longitudinal Research Studies*. Invited paper presented at the Annual Meeting of the Rasch Measurement Special Interest Group, American Educational Research Association, Chicago.
Laird, N.M. (1991). Topics in likelihood-based methods for longitudinal data analysis. *Statistica Sinica, 1*, 33–50.
Lord, F.M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.
Pinheiro, J.C., & Bates D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics, 4*, 12–35.
Rosenbaum, P.R. (1988). Item bundles. *Psychometrika, 53*, 349–359.
Roussos, L.A., Stout, W.F., & Marden, J. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement, 35*, 1–30.
Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, No. 17.
SAS Institute, Inc. (1999). SAS OnlineDoc (Version 8) [software manual on CD-ROM]. Cary, NC: SAS Institute, Inc.
Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6*, 461–464.
Scott, S., & Ip, E.H. (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.
Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensinonality. *Psychometrika, 52*, 589–617.
Vansteelandt, K. (1999). A formal model for the competency-demand hypothesis. *European Journal of Personality, 13*, 429–442.
Wainer, H., & Kiely, G.L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*, 185–201.
Wang, Y.J. (1986). Ordered-dependent parameterization of multinomial distributions. *Scandinavian Journal of Statistics, 13*, 199–205.
Whittaker, J. (1990). *Graphical Models in Applied Mathematical Multivariate Statistics*. West Sussex, UK: Wiley.
Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item independence. *Journal of Educational Measurement, 30*, 187–213.

Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, *64*, 213–249.

Zhao, L.P., & Prentice, R.L. (1990). Correlated binary regression using a generalized quadratic model. *Biometrics*, *77*, 642–648.

Manuscript received 26 OCT 2001
Final version received 4 APR 2003