PSYCHOMETRIKA—VOL. 72, NO. 3, 393–411
SEPTEMBER 2007
DOI: 10.1007/S11336-007-9005-4

COPULA FUNCTIONS FOR RESIDUAL DEPENDENCY

JOHAN BRAEKEN, FRANCIS TUERLINCKX, AND PAUL DE BOECK

UNIVERSITY OF LEUVEN

Most item response theory models are not robust to violations of conditional independence. However, several modeling approaches (e.g., conditioning on other responses, additional random effects) exist that try to incorporate local item dependencies, but they have some drawbacks such as the nonreproducibility of marginal probabilities and resulting interpretation problems. In this paper, a new class of models making use of copulas to deal with local item dependencies is introduced. These models belong to the bigger class of marginal models in which margins and association structure are modeled separately. It is shown how this approach overcomes some of the problems associated with other local item dependency models.

Key words: item response theory, local item dependency, copula.

1. Introduction

In most measurement situations test items are developed with the intention to provide unique information about the ability or skill being measured. Items are, however, often related through a common theme, stimulus material, or formulation. Examples are legion; reading comprehension items linked to the same reading passage and multistage items, for instance, are widespread. Responses on these items will partially show dependency due to their "artificial" common ground and not only because they relate to the skill or ability intended to be measured (Ferrara, Huynh, & Michaels, 1999; Yen, 1993). If one wants to get a "pure" assessment of this skill or ability, measurement models should take this noise, often labeled as local item dependency, into account.

A well-known measurement model used in different areas of research in psychology and educational measurement is the Rasch model (Rasch, 1960). For a person p ($p = 1, \ldots, P$) and item i ($i = 1, \ldots, I$), a binary random variable Y_{pi} is defined and the probability of a realization y_{pi} equals

$$\Pr(Y_{pi} = y_{pi}|\theta_p) = \frac{\exp(y_{pi}(\theta_p - \beta_i))}{1 + \exp(\theta_p - \beta_i)},$$

where θ_p is the proficiency of person p and β_i is the difficulty of the item i. For the moment we will assume that $Y_{pi} = 1$ corresponds to a correct response and $Y_{pi} = 0$ to an incorrect one (but see the second application in section 5).

For the remainder of the paper, it will be convenient to define the Rasch model as a latent threshold model in which an underlying (latent) continuous variable X_{pi} is logistically distributed with mean $\theta_p - \beta_i$, a scale parameter equal to one, and a threshold parameter set at 0 (see, e.g., De Boeck & Wilson, 2004). More formally, this can be written as follows: $X_{pi} = \theta_p - \beta_i + \varepsilon_{pi}$, with ε_{pi} following a standard logistic distribution (i.e., location zero and

The authors wish to thank Yuri Goegebeur and Taoufik Bouezmarni for their helpful suggestions and comments. We are also indebted to the reviewers of this paper, their generous comments and remarks greatly improved the setup and clarity of the presented material. Preparation of this manuscript was supported in part by the Fund for Scientific Research Flanders (FWO) Grant G.0148.04 and by the K.U. Leuven Research Council Grant GOA/2005/04.

Requests for reprints should be sent to Johan Braeken, Research Group Quantitative and Personality Psychology, Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium. E-mail: johan.braeken@psy.kuleuven.be

© 2007 The Psychometric Society

scale one) such that ε_{pi} can be called a latent residual for person p and item i. Furthermore, it can be deduced that

$$\Pr(Y_{pi} = 1|\theta_p) = \Pr(X_{pi} > 0|\theta_p) = \Pr(\varepsilon_{pi} > -\theta_p + \beta_i) = 1 - F(-\theta_p + \beta_i)$$
$$= 1 - \frac{\exp(-\theta_p + \beta_i)}{1 + \exp(-\theta_p + \beta_i)} = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}.$$

Before turning to some basic characteristics of the Rasch model, we will first introduce some necessary notation. In general, we will make a notational distinction between a random variable (written in capitals; e.g., Y_{pi} or X_{pi}) and its realization (e.g., y_{pi} or x_{pi}). An exception to this rule pertains to random variables denoted with Greek letters (e.g., ε_{pi}) for which the same notation will be used to denote random variable and realization (a standard convention in the IRT literature). The cumulative distribution function (cdf) of a random variable X evaluated for an arbitrary value x will be denoted as $F_X \equiv F_X(x)$. The joint cdf of a vector of random variables $\boldsymbol{X} = (X_1, \ldots, X_I)$ will be referred to as $F_{\boldsymbol{X}} \equiv F_X(x_1, \ldots, x_I)$. In some situations, it is necessary to express explicitly the conditioning on the latent trait θ (e.g., $F_{X|\theta}(x|\theta)$).

A basic assumption of the Rasch model is local stochastic or conditional independence:

$$\Pr(\boldsymbol{Y}_p = \boldsymbol{y}_p|\theta_p) = \prod_{i=1}^{I} \Pr(Y_{pi} = y_{pi}|\theta_p),$$

where $\boldsymbol{Y}_p$ is the random vector of responses for person p on the set of I items (and $\boldsymbol{y}_p$ is the vector of corresponding realizations). Note that in the latent threshold model formulation above, the conditional assumption is equivalent to uncorrelated logistic error terms ε_{pi} across items such that $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \ldots, \varepsilon_{pI}) = \prod_{i=1}^{I} F_{\varepsilon_p i}(\varepsilon_{pi})$, where $F_{\boldsymbol{\varepsilon}_p}$ is the joint distribution function of the latent residuals for person p.

The Rasch model, and more in general most item response theory models, are not robust to violations of local stochastic independence. Violations of the assumption are called local item dependencies or residual dependencies (both terms will be used interchangeably in this paper). Local item dependencies can seriously affect the estimation of the model parameters (see, e.g., Chen & Thissen, 1997; Sireci, Thissen, & Wainer, 1991; Tuerlinckx & De Boeck, 2001a; Yen, 1984, 1993). To illustrate this intuitively, consider the extreme case wherein several slightly differently phrased questions are used in one test, this is almost equivalent to asking a single question. This redundancy situation will lead to an inflation in or "double counting of" information (Ip, Wang, De Boeck, & Meulders, 2004). The effects of this redundancy are, for instance, the underestimation of the standard error of a person's proficiency parameter θ_p (Junker, 1991) and a positively biased test information function. For the effect on the item discrimination parameter in a two-parameter logistic model (2PLM), see Masters (1988). In general, when an item response model that assumes local independence is used for a test that suffers from local item dependencies, this can result in biased estimates for both person and item parameters, and the test information function and other diagnostics that assume conditional independence.

Once local item dependency problems have been detected, there are several possible approaches to model them; for an overview see Tuerlinckx and De Boeck (2004). Two typical problems that are associated with some of the most popular models for local item dependencies are nonreproducibility and the impossibility of interpreting β_i as the difficulty of item i. To illustrate these problems, we take the constant combination interaction model (CCI) of Hoskens and De Boeck (1997) as a starting point and, for simplicity, only the case with two items (1 and 2) showing residual dependency is considered. The joint probability of response (y_{p1}, y_{p2}) then equals

$$\Pr(Y_{p1} = y_{p1}, Y_{p2} = y_{p2}|\theta_p) = \frac{\exp(y_{p1}(\theta_p - \beta_1) + y_{p2}(\theta_p - \beta_2) + y_{p1}y_{p2}\lambda)}{1 + \exp(\theta_p - \beta_1) + \exp(\theta_p - \beta_2) + \exp(2\theta_p - \beta_1 - \beta_2 + \lambda)}.$$

Note that conditional on θ_p, the parameter λ expresses the conditional log odds ratio for the item pair given the responses on all other items, hence the CCI model only reduces to the familiar Rasch model in case λ equals zero (see Tuerlinckx & De Boeck, 2004, pp. 303–304, for the interpretation of higher-order associations between three or more items). Calculating the probability of answering the first item correctly gives (the marginal item characteristic curve or ICC for item 1):

$$\begin{aligned}\Pr(Y_{p1} = 1|\theta_p) &= \sum_{y_{p2}=0}^{1} \Pr(Y_{p1} = 1, Y_{p2} = y_{p2}|\theta_p) \\ &= \frac{\exp(\theta_p - \beta_1) + \exp(2\theta_p - \beta_1 - \beta_2 + \lambda)}{1 + \exp(\theta_p - \beta_1) + \exp(\theta_p - \beta_2) + \exp(2\theta_p - \beta_1 - \beta_2 + \lambda)}. \qquad (1)\end{aligned}$$

It can be seen that, when $\lambda \neq 0$ (hence residual dependency between items 1 and 2 exists), the marginal probability of responding correctly to item 1 is not a Rasch model anymore (i.e., equation (1) is not a logistic function). Therefore, it is said that the marginals are not reproducible.[1] Moreover, the parameter β_1 looses its natural interpretation of difficulty parameter because it is not simply the location on the latent trait for which the probability of responding correctly is 0.5 (see also Ip, 2002). Both problems are illustrated in the panels of Figure 1 where the item characteristic curve of item 1 derived from the CCI model (see equation (1)) is plotted for a certain configuration of parameters. The item characteristic Rasch curve is plotted for comparison reasons in each panel by a dotted line. It can be seen that the probability of responding correctly to item 1 also depends on λ and β_2 (cf., equation (1)). The same kinds of problems also occur in random-effect testlet models (Bradlow, Wainer, & Wang, 1999) and other types of conditional models (e.g., the dynamic Rasch model of Verhelst & Glas, 1993).

A class of models that does not suffer from the aforementioned problems are the so-called marginal models (see, e.g., Molenberghs & Verbeke, 2005). In these models, the univariate margins and the dependence structure are modeled separately. In consequence, these models will not suffer of the problem of nonreproducible marginals. Examples proposed in the literature are the Bahadur–Ip model (Bahadur, 1961; Ip, 2000, 2001) and the hybrid kernel approach developed by Ip (2002) that can be related to the work of Holland (1990) and generalized log-linear models (Cox, 1972; Zhao & Prentice, 1990). If we want to make use of logistic marginals, as is the case for the Rasch model, a logical approach would be to use a multivariate random effect logit model analogous to the multivariate random effect probit model. Unfortunately, there is no single or generally accepted multivariate logistic distribution (see, e.g., Kotz, Balakrishnan, & Johnson, 2000), analogous to the multivariate normal distribution, which can be used as a starting point for introducing dependencies between the latent residuals. However, with copula functions this paper will introduce a convenient tool to model dependency and construct multivariate distributions. The residual dependencies will be taken into account by these copulas without changing anything to the marginal model part (i.e., item characteristic curve and interpretation of the parameters). A residual dependency model that preserves Rasch logistic margins will be used as an obvious and prototypical working example of the copula approach.

[1]Note that we employ the concept of reproducibility here in an intuitive fashion. More formal definitions can be found in Ip (2002) and in Fitzmaurice, Laird, and Rotnitzky (1993).

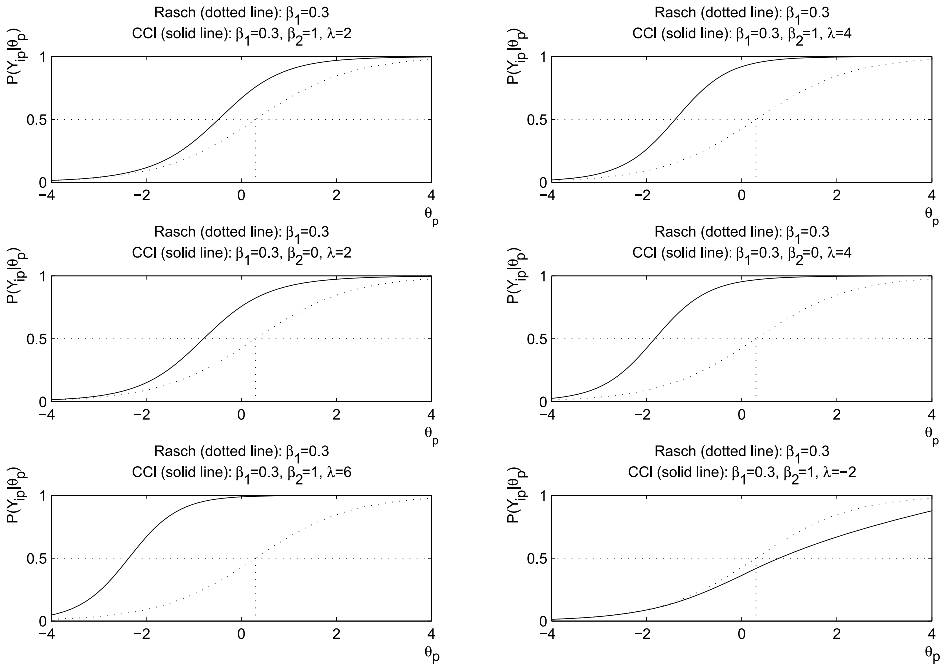


FIGURE 1.
Nonreproducibility of the constant combination interaction model.

The remainder of the paper is organized as follows. First, some basic theory with respect to copulas will be discussed. We will continue with the formulation of a copula model for residual dependencies that has Rasch logistic margins and is quite straightforward in its approach and interpretation. Then this copula model will be applied to two data sets containing residual dependencies. Finally, the paper is closed with a discussion.

2. Copula Theory

In order that this paper should be reasonably self-contained, we begin with introducing copula functions as a mathematical concept. A more thorough overview can be found in the reference works by Joe (1997) or by Nelsen (1999). In mathematics, a copula (Latin for link or tie) defines a function that relates, or "couples," a multivariate distribution function to its univariate margins.

It is said that an R-dimensional copula is a function $C : [0, 1]^R \to [0, 1]$ with the following properties:

1. For every vector $\boldsymbol{u} \in [0, 1]^R$, $C(\boldsymbol{u})$ is increasing in each component u_r with $r \in 1, 2, \ldots, R$.
2. For every vector $\boldsymbol{u} \in [0, 1]^R$, $C(\boldsymbol{u}) = 0$ if at least one coordinate of the vector is 0 and $C(\boldsymbol{u}) = u_r$ if all the coordinates of the vector are equal to one except the rth one.
3. For every $\boldsymbol{a}, \boldsymbol{b} \in [0, 1]^R$ with $\forall r \in \{1, 2, \ldots, R\}$, $a_r \le b_r$, given a hypercube $\boldsymbol{B} = [\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_R, b_R]$ whose vertices lie in the domain of C, and with $V_C(\boldsymbol{B}) \ge 0$. The volume $V_C(\boldsymbol{B})$ is defined as

$$V_C(\boldsymbol{B}) = \sum_{k_1=1}^{2} \sum_{k_2=1}^{2} \cdots \sum_{k_R=1}^{2} (-1)^{k_1+k_2+\cdots+k_R} C\big(d_1^{(k_1)}, d_2^{(k_2)}, \ldots, d_R^{(k_R)}\big), \tag{2}$$

where $d_r^{(1)} = a_r$ and $d_r^{(2)} = b_r$ $(r = 1, \ldots, R)$.

An important result from the theory of copulas is Sklar's theorem (Sklar, 1959). The theorem states that for any R-dimensional distribution function $F_{\boldsymbol{X}}$ with univariate margins $F_{X_1}, F_{X_2}, \ldots, F_{X_R}$ there exists a copula function C such that this multivariate distribution $F_{\boldsymbol{X}}$ can be represented as a function of its margins through this copula: $F_{\boldsymbol{X}} = C(F_{X_1}, F_{X_2}, \ldots, F_{X_R})$. The key idea is that any existing multivariate distribution can be reformulated according to this copula presentation and, conversely, based upon any kind of univariate margins, a joint (i.e., multivariate) distribution can be constructed by means of a copula function. Note that we will make use of exactly this part in our modeling approach to construct a multivariate distribution using copulas, such that the univariate margins of this newly constructed multivariate distribution are still the original univariate distributions we started from. Making use of the second property of the copula definition, it can easily be deduced that the univariate marginal distribution for X_r equals $C(1, \ldots, 1, F_{X_r}, 1, \ldots, 1) = F_{X_r}$. Hence, in this way, an association between the R random variables is allowed while preserving the univariate margins.

For each joint distribution with margins $F_{X_1}, F_{X_2}, \ldots, F_{X_R}$, and constructed by means of a copula, the following applies:

$$W(F_{X_1}, \ldots, F_{X_R}) \leq C(F_{X_1}, \ldots, F_{X_R}) \leq M(F_{X_1}, \ldots, F_{X_R}),$$

where $W(F_{X_1}, \ldots, F_{X_R}) = \max(F_{X_1} + \cdots + F_{X_R} - R + 1, 0)$, and $M(F_{X_1}, \ldots, F_{X_R}) = \min(F_{X_1}, \ldots, F_{X_R})$. The functions $W(F_{X_1}, \ldots, F_{X_R})$ and $M(F_{X_1}, \ldots, F_{X_R})$ correspond to the Fréchet–Hoeffding lower and upper bounds (Fréchet, 1951; Hoeffding, 1940) and define the maximum negative and positive dependency of a joint distribution that can be obtained given fixed margins. These bounds can be used to indicate the range of dependency a copula function can capture.

A wide variety of possible copula functions exists which allow for fitting a wide range of dependency types (see, e.g., Joe, 1993; Nelsen, 1999). Our modeling approach will focus on the class of Archimedean copulas (Genest & MacKay, 1986; Joe, 1993; Nelsen, 1999). Archimedean copulas have a simple structure and can be written as

$$C(u_1, \ldots, u_R) = \psi^{-1}\left(\sum_{r=1}^{R} \psi(u_r)\right),$$

where $\psi : [0, 1] \to [0, \infty]$ is a continuous strictly decreasing function, called the generator function, such that $\psi(0) = \infty$ and $\psi(1) = 0$, and ψ^{-1} is completely monotonic on $[0, \infty)$, such that $(-1)^k (d^k/dt^k)\psi^{-1}(t) \geq 0$ $\forall t \in [0, \infty)$ and $k \in \mathbb{N}$. Archimedean copulas have nice symmetry properties, as there are permutation symmetry, $C(u_1, u_2) = C(u_2, u_1)$, and associativity, $C(u_1, u_2, u_3) = C(u_1, C(u_2, u_3)) = C(C(u_1, u_2), u_3)$; which makes them especially attractive for modeling symmetrically dependent data. Furthermore, notice that the independence case can also be rewritten in a convenient way under an Archimedean copula representation:

$$F_{\boldsymbol{X}} = \prod_{r=1}^{R} F_{X_r} = C(F_{X_1}, \ldots, F_{X_R}) = \exp\left(-\sum_{r=1}^{R} [-\log(F_{X_r})]\right),$$

with $\psi(u) = -\log(u)$ and $\psi^{-1}(t) = \exp(-t)$. This copula is also known as the product copula, denoted by Π (Frees & Valdez, 1998; Nelsen, 1999).

Two Archimedean copula families are presented in this paper:

1. The Frank copula (Frank, 1979):

$$C(F_{X_1}, F_{X_2}, \dots, F_{X_R}) = \frac{-1}{\delta} \log\left(1 - \frac{\prod_{r=1}^{R}(1 - \exp(-\delta F_{X_r}))}{\prod_{r=1}^{R-1}(1 - \exp(-\delta))}\right),$$

   with for $R = 2$: $\delta \in \mathbb{R}/\{0\}$, if $\delta \to -\infty$ then $C \to W$, if $\delta \to 0$ then $C \to \Pi$, if $\delta \to \infty$ then $C \to M$; and for $R > 2$: $\delta > 0$, if $\delta \to 0$ then $C \to \Pi$, if $\delta \to \infty$ then $C \to M$.
2. Cook–Johnson copula (Clayton, 1978; Cook & Johnson, 1981):

$$C(F_{X_1}, F_{X_2}, \dots, F_{X_R}) = \left(\sum_{r=1}^{R} F_{X_r}^{-\delta} - R + 1\right)^{-1/\delta},$$

   with $\delta > 0$, if $\delta \to 0$ then $C \to \Pi$, if $\delta \to \infty$ then $C \to M$.

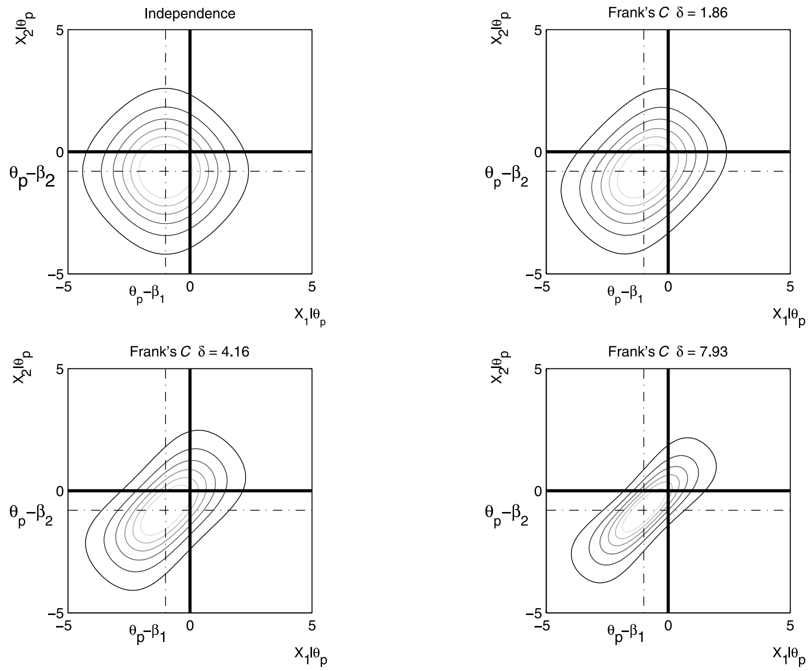
Both copula families will be used extensively throughout this paper (we discuss below why these two functional forms were chosen). The parameter δ that defines each of these Archimedean copulas has an interpretation of an association parameter. For instance, when the value of δ increases, the dependence captured by the copula gets closer to the theoretical maximum positive dependence for $F_{\mathbf{X}}$ (i.e., M, the Fréchet–Hoeffding upper bound).

When we consider standard logistic marginals, the Frank and the Cook–Johnson copulas lead to different joint distributions. In Figure 2 we illustrate the degree and kind of dependence by depicting for the two-item case the contour lines of several latent bivariate distributions $F_{\mathbf{X}_p|\theta_p}$ with standard logistic margins $F_{X_{p1}|\theta_p}$ and $F_{X_{p2}|\theta_p}$, constructed by means of a Frank copula (upper four panels) and a Cook–Johnson copula (lower four panels) for varying values of the association parameter δ. Notice that the copulas can capture a broad range of dependency and differ in the type of dependence they induce; for instance, the Cook–Johnson copula has a prominent lower tail (i.e., more formally, $C(u, \dots, u)/u$ converges to a constant c in [0, 1] as $u \to 0$; Joe, 1993; Nelsen, 1999), while the Frank copula leads to a similar kind of dependence in both tails.

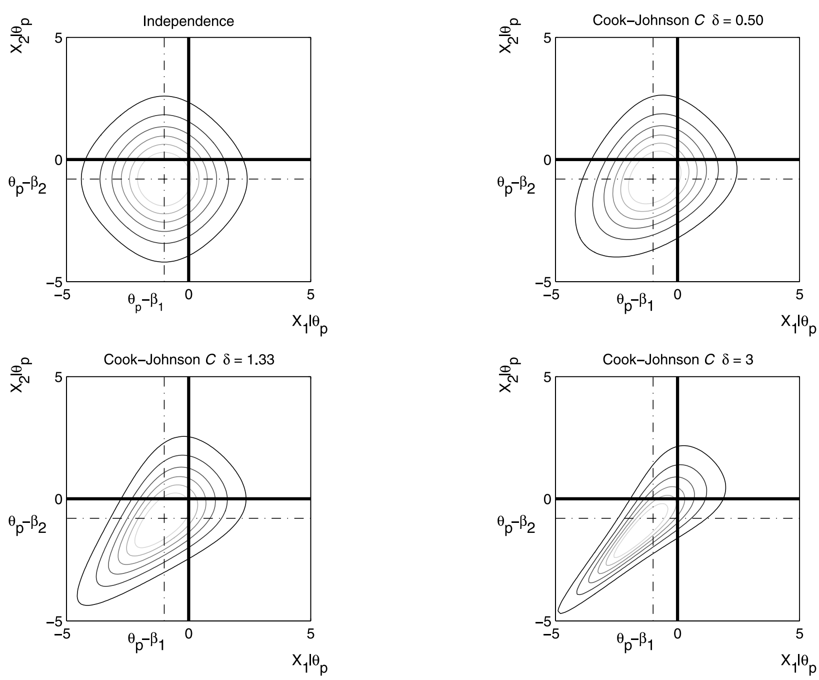
3. The Copula Model for Residual Dependencies

With θ_p being the proficiency of person p $(p = 1, \dots, P)$, β_i being the difficulty of the item i $(i = 1, \dots, I)$, latent variable $X_{pi} = \theta_p - \beta_i + \varepsilon_{pi}$, and observed binary variable $Y_{pi} = I(X_{pi} > 0) = I(\varepsilon_{pi} > -\theta_p + \beta_i)$, the vector $\boldsymbol{\varepsilon}_p = (\varepsilon_{p1}, \dots, \varepsilon_{pI})^T$ is assumed to have an I-variate distribution with standard logistic univariate margins $F_{\varepsilon_{pi}}$. The conditional independence assumption of the Rasch model implies that the logistic error terms ε_{pi} are uncorrelated across items, such that $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \dots, \varepsilon_{pI}) = \prod_{i=1}^{I} F_{\varepsilon_{pi}}(\varepsilon_{pi})$. However, when residual dependencies within different subsets of items are detected, a different dependency structure for $\boldsymbol{\varepsilon}_p = (\varepsilon_{p1}, \dots, \varepsilon_{pI})^T$ is required and, consequently, a more appropriate I-variate distribution $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \dots, \varepsilon_{pI})$ is needed. The marginal distributions $F_{\varepsilon_{pi}}$ are univariate logistic distributions because we want the marginal probabilities of a correct response to be a Rasch model. The unknown I-variate joint distribution $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \dots, \varepsilon_{pI})$ will be constructed from these margins by means of copula functions.

Consider S disjoint subsets of $\{1, \dots, I\}$ denoted as $J_1, \dots, J_S$, where J_s has cardinality I_s. The vector of error terms $\boldsymbol{\varepsilon}_p$ is similarly divided into subsets $\boldsymbol{\varepsilon}_p^{(1)}, \dots, \boldsymbol{\varepsilon}_p^{(S)}$ where $\boldsymbol{\varepsilon}_p^{(s)} = (\varepsilon_{pi},\ i \in J_s)$. The different subsets are independent, and the variables in a subset $\boldsymbol{\varepsilon}_p^{(s)}$ are assumed exchangeable. Subsets of items can be chosen based upon diagnostic tests for residual dependency, context of items, or substantive theory.



(a) The Frank copula



(b) The Cook–Johnson copula

FIGURE 2.
Bivariate logistic density contour plots for different copula models.

The joint probability of the response vector of a person p is

$$\Pr(\boldsymbol{Y}_p|\theta_p) = \prod_{s=1}^{S} \Pr(Y_{pi} = y_{pi}, i \in J_s|\theta_p),$$

with the joint probability of responses on item subset J_s equal to

$$\Pr(Y_{pi} = y_{pi}, i \in J_s|\theta_p) = \Pr\big(d_{pi}^{(1)} < \varepsilon_{pi} \le d_{pi}^{(2)}, i \in J_s|\theta_p\big),$$

where for $y_{pi} = 1$, $d_{pi}^{(1)} = -\theta_p + \beta_i$ and $d_{pi}^{(2)} = \infty$, and for $y_{pi} = 0$, $d_{pi}^{(1)} = -\infty$ and $d_{pi}^{(2)} = -\theta_p + \beta_i$. If the cardinality of subset J_s is larger than one ($I_s > 1$), $\Pr(d_{pi}^{(1)} < \varepsilon_{pi} \le d_{pi}^{(2)}, i \in J_s|\theta_p)$ is evaluated from the copula $C_S(\cdot; \delta_s)$ for $(\varepsilon_{pi}, i \in J_s)$ as

$$\begin{aligned}
&\Pr\big(d_{pi}^{(1)} < \varepsilon_{pi} \le d_{pi}^{(2)}, i \in J_s|\theta_p\big) \\
&= \sum_{k_1=1}^{2} \cdots \sum_{k_{I_s}=1}^{2} (-1)^{k_1+\cdots+k_{I_s}} C_s\big(F_{\varepsilon_{p1}}(d_{p1}^{(k_1)}), \ldots, F_{\varepsilon_{pi}}(d_{pi}^{(k_i)}), \ldots, F_{\varepsilon_{pI_s}}(d_{pI_s}^{(k_{I_s})})\big).
\end{aligned}$$

For clarity, assume $I = 2$ and $J_s = \{1, 2\}$, the equations above simplify as follows for a $(0, 0)$-response:

$$\begin{aligned}
&\Pr(Y_{p1} = 0, Y_{p2} = 0|\theta_p) \\
&\quad = \Pr\big(d_{pi}^{(1)} < \varepsilon_{pi} \le d_{pi}^{(2)}, i \in J_s|\theta_p\big) \\
&\quad = \Pr\big(d_{p1}^{(1)} < \varepsilon_{p1} \le d_{p1}^{(2)}, d_{p2}^{(1)} < \varepsilon_{p2} \le d_{p2}^{(2)}|\theta_p\big) \\
&\quad = \Pr(-\infty < \varepsilon_{p1} \le -\theta_p + \beta_1, -\infty < \varepsilon_{p2} \le -\theta_p + \beta_2) \\
&\quad = C_s\big(F_{\varepsilon_{p1}}(-\infty), F_{\varepsilon_{p2}}(-\infty)\big) - C_s\big(F_{\varepsilon_{p1}}(-\theta_p + \beta_1), F_{\varepsilon_{p2}}(-\infty)\big) \\
&\quad\quad - C_s\big(F_{\varepsilon_{p1}}(-\infty), F_{\varepsilon_{p2}}(-\theta_p + \beta_2)\big) + C_s\big(F_{\varepsilon_{p1}}(-\theta_p + \beta_1), F_{\varepsilon_{p2}}(-\theta_p + \beta_2)\big) \\
&\quad = 0 - 0 - 0 + C_s\big(F_{\varepsilon_{p1}}(-\theta_p + \beta_1), F_{\varepsilon_{p2}}(-\theta_p + \beta_2)\big) \\
&\quad = C_s\big(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p)\big).
\end{aligned}$$

The last equality follows from the definition of X_{pi}. The other probabilities are then

$$\begin{aligned}
\Pr(Y_{p1} = 1, Y_{p2} = 1|\theta_p) &= 1 - F_{X_{p1}|\theta_p}(0|\theta_p) - F_{X_{p2}|\theta_p}(0|\theta_p) \\
&\quad + C_s\big(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p)\big), \\
\Pr(Y_{p1} = 1, Y_{p2} = 0|\theta_p) &= F_{X_{p2}|\theta_p}(0|\theta_p) - C_s\big(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p)\big), \\
\Pr(Y_{p1} = 0, Y_{p2} = 1|\theta_p) &= F_{X_{p1}|\theta_p}(0|\theta_p) - C_s\big(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p)\big).
\end{aligned}$$

Figure 2 offers an intuitive insight into these calculations by presenting the contour lines of bivariate logistic densities constructed by means of both the Frank copula and the Cook–Johnson copula. Each contour plot is divided into quadrants made up by the solid lines drawn at the latent thresholds (the dashed lines indicating the marginal means $\theta_p - \beta_i$). In order to calculate the joint probabilities from the joint distribution functions of the latent random variables $X_{p1}|\theta_p$ and $X_{p2}|\theta_p$, the volume under the density for the corresponding quadrant needs to be calculated (see, e.g., Mood, Graybill, & Boes, 1974, and see also equation (2)).

Once the partitioning of items into subsets is given, and the copula families are given, the following set of parameters has to be estimated: the item parameters β_i $(i = 1, \dots, I)$, the distributional parameters of the latent trait (the person's proficiency θ_p is, generally but not necessarily, assumed to be normally distributed with mean zero and unknown variance σ^2), and the association parameters $\delta_1, \dots, \delta_S$. The regular Rasch model arises as a special case when $S = I$ and each subset J_s has size 1; or when the different C are assumed to be the product copula Π (equivalent to independence). As an example, consider a test with $I = 7$ items where items 1 and 2 exhibit some symmetric residual dependence, the set of items 3 to 5 also form a dependent subset (independent of the first) and items 6 and 7 do not show any violation of the general conditional independence assumption and are independent of the first two subsets. Thus, $\{1, \dots, 7\}$ is partitioned as $J_1 = \{1, 2\}$, $J_2 = \{3, 4, 5\}$, $J_3 = \{6\}$, and $J_4 = \{7\}$. The proposed I-variate distribution for the error component vector $\boldsymbol{\varepsilon}_p = (\varepsilon_{p1}, \dots, \varepsilon_{pI})^T$ is then $F_{\boldsymbol{\varepsilon}_p}(\varepsilon_{p1}, \dots, \varepsilon_{pI}) = C_1(F_{\varepsilon_{p1}}(\varepsilon_{p1}), F_{\varepsilon_{p2}}(\varepsilon_{p2})) \times C_2(F_{\varepsilon_{p3}}(\varepsilon_{p3}), F_{\varepsilon_{p4}}(\varepsilon_{p4}), F_{\varepsilon_{p5}}(\varepsilon_{p5})) \times F_{\varepsilon_{p6}}(\varepsilon_{p6}) \times F_{\varepsilon_{p7}}(\varepsilon_{p7})$.

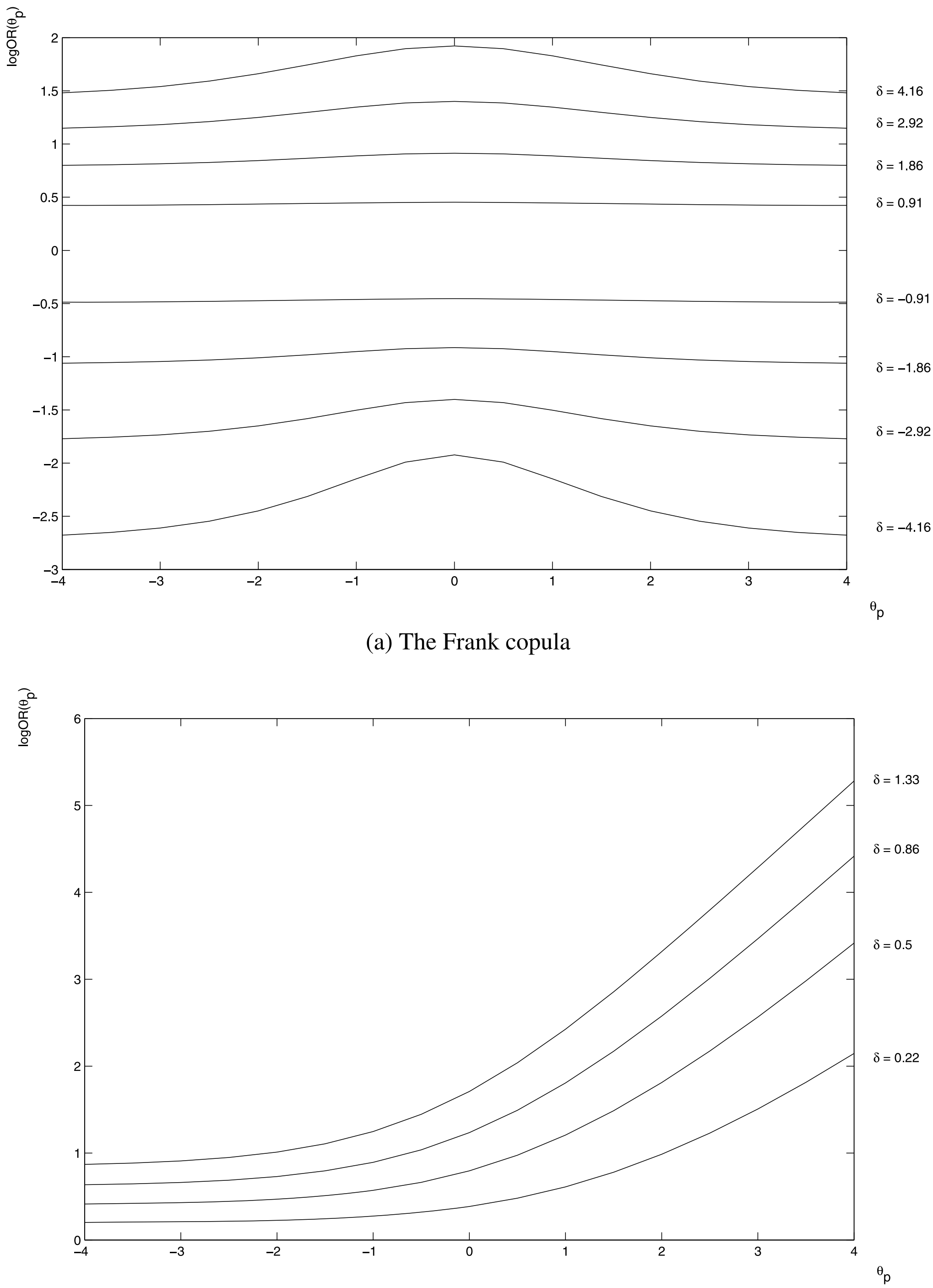
Because a copula model is a form of marginal modeling, a broad range of association structures (by means of different copula functions) for the subsets showing residual dependency can be compared without fundamentally changing the base model of the marginal probabilities and thus preserving the individual item characteristic curves of the items. In the previous example, C_1 could be either the Frank copula or the Cook–Johnson copula, and the same holds for C_2.

In order to illustrate that the introduction of the copula can take residual dependencies into account, the odds ratio (conditional on θ_p) for items 1 and 2 involved in the copula C can be computed as follows:

$$\begin{aligned} \mathrm{OR}(\theta_p) &= \frac{\Pr(Y_{p1} = 1, Y_{p2} = 1|\theta_p)\Pr(Y_{p1} = 0, Y_{p2} = 0|\theta_p)}{\Pr(Y_{p1} = 1, Y_{p2} = 0|\theta_p)\Pr(Y_{p1} = 0, Y_{p2} = 1|\theta_p)} \\ &= \frac{(1 - F_{X_{p1}|\theta_p}(0|\theta_p) - F_{X_{p2}|\theta_p}(0|\theta_p) + C)C}{(F_{X_{p2}|\theta_p}(0|\theta_p) - C)(F_{X_{p1}|\theta_p}(0|\theta_p) - C)}, \end{aligned}$$

with $C = C(F_{X_{p1}|\theta_p}(0|\theta_p), F_{X_{p2}|\theta_p}(0|\theta_p))$. Using the Frank copula, the value of the log odds ratio is then computed for several values of δ and for θ_p ranging from -4 to 4 and the result is shown in the upper panel of Figure 3. For the Cook–Johnson copula, the same procedure is followed and this result is shown in the lower panel of the same figure. For ease of demonstration the two margins were set equal to one another, with difficulty parameters β_1 and β_2 equal to zero, so that the log odds ratio (conditional on θ_p) was only a function of the copula's association parameter and the marginal probabilities as determined by θ_p. From both panels it can be seen that when the value of the copula parameter δ rises the log odds ratio also increases, indicating the copula parameter's function as an association measure.

In both the upper and lower panels of Figure 3, it appears that for a fixed value of the copula parameter δ, there is a dependency between the log odds ratio and the value of the latent trait. The Frank copula shows a more static residual dependence between the two items (the log odds ratio is more or less constant, unless for large values of δ). On the contrary, the Cook–Johnson copula has a dimensional residual dependence structure because the log odds ratio increases as θ_p increases (for an example of a residual dependency structure that is dimension dependent see, e.g., Bell, Pattison, & Withers (1988), and also the second application in this paper). Our attention is restricted in this paper to the Frank and the Cook–Johnson copulas because they show such distinct and interesting patterns of association. Different models for static and dimensional dependence structures have also been proposed by Hoskens and De Boeck (1997) in their framework of conditional residual dependency models (Hoskens & De Boeck (1997), use the labels "constant combination" and "dimension-dependent" interaction models). Notice that the dimensional residual dependence induced by the Cook–Johnson copula is positively accelerated, whereas in the dimension-dependent interaction model this can be either linearly increasing or decreasing.



(a) The Frank copula

(b) The Cook–Johnson copula

FIGURE 3.
The conditional log odds ratio given different values of the copula parameter.

4. Statistical Inference in the Copula Model

We followed a marginal maximum likelihood (MML) approach (see De Boeck & Wilson, 2004) to fit the Rasch copula model. Over all persons, and items, the likelihood to be optimized is

$$\prod_{p=1}^{P} \int_{\theta_p} \prod_{s=1}^{S} \big[\Pr\big(d_{pi}^{(1)} < \varepsilon_{pi} \le d_{pi}^{(2)}, i \in J_s | \theta_p\big)\big] \phi(\theta_p | \sigma^2)\, \mathrm{d}\theta_p.$$

Usually the log is taken for numerical reasons. The negative of this loglikelihood is minimized using a quasi-Newton algorithm and the approximation of the intractable integral with respect to the distribution of θ_p is carried out with a Gauss–Hermite quadrature. In extensive simulations and applications we have found that usually 21 nodes are sufficient for a good recovery and to obtain stable results.

With respect to model checking and selection, the tools usually applied in nonlinear mixed models are available (Wald, score, and likelihood ratio tests). The independence model (i.e., a Rasch model) is nested within the copula model for both the Frank copula and the Cook–Johnson copula. Therefore, a test of the null hypothesis that the association parameter equals a particular value (0 for both the Frank and the Cook–Johnson copulas) is a comparison between the Rasch model and a certain copula model.

Note that for the Frank copula, Meester (1991) showed that once R exceeds 2, the lower bound of the copula parameter δ needs to be adapted in a function of R. For any R this adapted lower bound is always strictly less than zero, thus not restricting the positive association range and technically leaving the position of the independence point in the interior of the parameter space.[2] However, for the Cook–Johnson copula the independence case lies on the boundary of the parameter space since δ cannot be lower than zero. Therefore the appropriate reference distribution for the likelihood ratio test statistic comparing the independence and the Cook–Johnson model is not a chi-square with 1 degrees of freedom, but a mixture of two chi-square distributions, with 0 and 1 degree of freedom, respectively. When more copulas and, consequently, more association parameters are involved, deducing the appropriate mixture of chi-square distributions to function as reference distribution may get very complicated (see, e.g., Self & Liang, 1987). Therefore, we will rely in all cases on the traditional reference distribution, which yields a more conservative test of the null hypothesis.

Attention has to be given to the selection of a particular copula function. The Frank and the Cook–Johnson copulas do not have a nested relation and, consequently, the selection is best based on methods such as the AIC (Akaike, 1973) or BIC (Schwarz, 1978).

5. Applications

In this section we will apply the Rasch copula model to two data sets. One data set originates from ability testing and the other from a research on verbal aggression. In both cases, the presence of residual dependencies is to be expected based on the design of the study.

[2]The range of negative dependency in the case that $R > 2$ is rather limited and attaching a meaningful interpretation to negative dependency between three or more items is not very likely. Hence, we do not consider the use of the Frank copula for negative residual dependency between more than two items for reasons of clarity.

TABLE 1.
Rasch and Rasch Copula models for the small reading test data.

	Model	AIC	Association parameter(s) (SE between parentheses)
1	Rasch (i.e., $\Pi{:}J_1 = \{1,2,3,4,5,6\}$)	3157	—
2	$\text{Frank}{:}J_1 = \{1,6\}$, $\text{Frank}{:}J_2 = \{4,5\}$, $\Pi{:}J_3 = \{2,3\}$	3108	$\delta_1 = 3.56(0.62)$, $\delta_2 = 2.53(0.73)$
3	$\text{Frank}{:}J_1 = \{1,6\}$, $\text{C–J}{:}J_2 = \{4,5\}$, $\Pi{:}J_3 = \{2,3\}$	3113	$\delta_1 = 3.57(0.62)$, $\delta_2 = 0.39(0.16)$
4	$\text{C–J}{:}J_1 = \{1,6\}$, $\text{Frank}{:}J_2 = \{4,5\}$, $\Pi{:}J_3 = \{2,3\}$	3111	$\delta_1 = 0.99(0.22)$, $\delta_2 = 2.50(0.73)$
5	$\text{C–J}{:}J_1 = \{1,6\}$, $\text{C–J}{:}J_2 = \{4,5\}$, $\Pi{:}J_3 = \{2,3\}$	3116	$\delta_1 = 0.99(0.23)$, $\delta_2 = 0.39(0.16)$
6	$\text{Frank}{:}J_1 = \{1,4,5,6\}$, $\Pi{:}J_2 = \{2,3\}$	3140	$\delta_1 = 1.40(0.10)$
7	$\text{C–J}{:}J_1 = \{1,4,5,6\}$, $\Pi{:}J_2 = \{2,3\}$	3148	$\delta_1 = 0.26(0.09)$

5.1. Application 1: Small Reading Test

A group of high-school students interested in studying law in college ($P = 441$) answered six multiple choice questions about a text on the president and the separation of powers in the United States of America. The answers were recoded in right and false for ease of demonstration and scored 1 and 0, respectively. Various types of tools have been developed to detect local item or residual dependencies (see, e.g., Chen & Thissen, 1997; Holland & Rosenbaum, 1986; Rosenbaum, 1984; Yen, 1984; or, for a recent comparison, see Tate, 2003). In previous analyses (Tuerlinckx & De Boeck, 2001b) it is shown that two pairs of items showed residual dependencies: items 1 and 6, and items 4 and 5. Here we fit a series of models including the simple Rasch model and possible residual dependence models using the Frank and the Cook–Johnson copulas in different combinations. The fits of these models are assessed using the appropriate tools outlined above: the likelihood ratio test and/or the AIC. The list of estimated models and results on the estimated association parameters are given in Table 1.

From the results in Table 1 it can be seen that, according to the AIC, the most appropriate model is that copula model with a separate Frank copula for each item subset (Model 2; $\text{Frank}{:}J_1\{1,6\}$, $\text{Frank}{:}J_2\{4,5\}$, and $\Pi{:}J_3\{2,3\}$). If we compare Model 2 to the regular Rasch model (Model 1; $\Pi{:}J_1\{1,2,3,4,5,6\}$), there is overwhelming evidence against the null hypothesis that both association parameters are equal to zero ($LRT = 49$, $df = 2$, $p < 0.0001$). Both association parameters differ significantly from zero ($p < 0.01$ and 0.05, for δ_1 and δ_2, respectively).

We also looked at the Mantel–Haenszel (MH) test as a diagnostic tool (Mantel & Haenszel, 1959) for detecting residual dependencies as proposed by Rosenbaum (1984), Ip (2001), Scott and Ip (2002). The main idea is to test for equal odds ratio between groups of persons and the group division is based upon the estimated latent trait of θ_p in the model under investigation. Significant deviations from the hypothesis of equal odds ratios will give an indication of the presence of residual dependency between pairs of items while controlling for the dimension θ_p. Applying the Rasch model to the reading data set gives significant MH tests for item pairs $\{1,6\}$ and $\{4,5\}$: $z = 6.38$ ($p < .0001$) and $z = 2.33$ ($p = .03$), respectively. When we repeat this procedure after taking into account the residual dependencies through a copula model (i.e., Model 2), the MH tests for both item pairs become nonsignificant: $z = 1.37$ ($p = .16$) for item pair $\{1,6\}$ and $z = -0.61$ ($p = .33$) for item pair $\{4,5\}$. Thus, it appears that the copula model has taken into account the existing residual dependencies.

The other parameter estimates for Models 1 and 2 are given in detail in Table 2. It can be seen that the estimates of the item parameters do not change very much between the Rasch and the Rasch Copula model: There is a very slight tendency for the copula estimates to be shrunk more toward zero than the Rasch model estimates. However, the estimate of the standard deviation of the person ability distribution in the Rasch Copula model is considerably smaller than the

TABLE 2.
Parameter estimates for the Rasch model and the Rasch Copula model, and the constant combination interaction model, and a combination of the previous.

Model parameter	Rasch estimates (SE)		Copula estimates (SE)		CCI estimates (SE)		Copula and CCI estimates (SE)	
β_1	−0.36	(0.12)	−0.34	(0.13)	0.34	(0.16)	−0.34	(0.12)
β_2	0.99	(0.15)	0.94	(0.12)	0.89	(0.12)	0.92	(0.13)
β_3	0.01	(0.11)	0.00	(0.12)	−0.01	(0.11)	−0.01	(0.12)
β_4	−1.38	(0.14)	−1.32	(0.16)	−0.41	(0.23)	−0.52	(0.23)
β_5	−1.50	(0.13)	−1.44	(0.16)	−0.58	(0.22)	−0.68	(0.23)
β_6	−0.02	(0.13)	−0.03	(0.13)	0.83	(0.19)	−0.03	(0.12)
σ_θ	1.28	(0.09)	1.12	(0.10)	0.98	(0.09)	1.07	(0.09)
Frank: $J_1 = \{1, 6\}$; δ_1	—	—	3.56	(0.62)	—	—	3.62	(0.62)
Frank: $J_2 = \{4, 5\}$; δ_2	—	—	2.53	(0.73)	—	—	—	—
CCI: $J_1 = \{1, 6\}$; λ_1	—	—	1.47	(0.24)	—	—	—	—
CCI: $J_2 = \{4, 5\}$; λ_2	—	—	—	—	1.26	(0.27)	1.16	(0.27)
AIC	3157		3108		3104		3102	

equivalent estimate under the Rasch model. Thus, the location of the items on the continuum is hardly affected by introducing the copulas, but the scale of the distribution for the θ_p is shrunken somewhat. Therefore, after taking into account the residual dependencies, there is less differentiation among the examinees. Because of the large association parameter estimates, together with the lower AIC value for the copula model, we may conclude that not taking into account the residual dependencies leads to a wrong inference about the population distribution of the person abilities. Consequently, the predictions of the individual θ_p may be less accurate under the Rasch model.

For reasons of comparison we also fitted the interaction models of Hoskens and De Boeck (1997) to this data set and combinations of both this conditional approach and the copula approach. The parameter estimates and AIC for the best fitting model of both these model classes are shown in Table 2. A constant combination interaction for both item subsets was assessed as the best fitting of the interaction models. Compared to its copula counterpart it showed slightly better fit, but also slightly larger standard errors for the item parameters; and, as expected, the item location parameters were affected by applying the CCI model, and shifted along the latent dimension. Both interaction parameters differ significantly from zero ($p < 0.01$ and 0.05, for λ_1 and λ_2, respectively). The best fitting model, however, was a hybrid model with the Frank copula for one item subset, and a constant combination interaction for the second item subset. Based on these fitting results it is hard to make a conclusive statement about which modeling approach should be preferred for this data set; this does not come as a surprise for us considering the moderate sample size and, especially, given the small subset sizes.

5.2. *Application 2: Verbal Aggression Data*

As a second example, we analyze the data from a behavioral questionnaire (Vansteelandt, 2000) on verbal aggression. The questionnaire consists of 24 items and all items refer to verbally aggressive reactions in a frustrating situation, and it was administered from 316 persons. Four different situations were used to construct the 24 items: “A bus fails to stop for me”, “I miss a train because a clerk gave me faulty information”, “The grocery store closes just as I am about to enter”, and “The operator disconnects me when I had used up my last 10 cents for a call”. The description of the situation is followed by a statement about the behavioral mode of a verbal aggressive reaction (“I would want to” or “I would”) and then three verbal aggressive reactions:

cursing, scolding, and shouting (the questionnaire was translated from Dutch, and shouting refers to an expressive aggressive reaction in Dutch). The original items had three response categories ("yes", "perhaps", and "no") but we dichotomized them ("yes" and "perhaps" were scored 1 and "no" was scored 0; see also De Boeck and Wilson (2004) for an extensive study of this data set).

Because the 24 items are clustered in four groups of six items due to the common situations, we may expect residual dependencies among the items that are related to a similar situation. Making use of graphical exploratory techniques, Tuerlinckx and De Boeck (2004) showed that there are indeed residual dependencies in this data set. In order to take the residual dependencies originating from the same situation into account, we fitted a model with four copulas, one for each situation (each containing six items), and Rasch marginals. Again we have fitted a series of models with different functional form for the copula. The best fitting Rasch Copula model has a Frank copula for situation 1 (bus fails to stop, Frank:$J_1\{1, 2, 3, 13, 14, 15\}$) and situation 2 (train miss, Frank:$J_2\{4, 5, 6, 16, 17, 18\}$), and a Cook–Johnson copula for situation 3 (store closes, C–J:$J_3\{7, 8, 9, 19, 20, 21\}$) and situation 4 (disconnection, C–J:$J_4\{10, 11, 12, 22, 23, 24\}$). The results are presented in Table 3. From the results in Table 3, one can see that the Rasch Copula model outperforms the regular Rasch model. Note that while the interpretation of the individual item parameters remains the same in the Rasch Copula model (i.e., all β's can be seen as the points on the latent continuum where the probability of a yes response is 0.50), the model clearly has a significant better fit than the regular Rasch model ($LRT = 197$, $df = 4$, $p < 0.0001$). All four association parameters, δ_1, δ_2, δ_3, and δ_4, differ significantly from zero ($p < 0.05$).

Contrary to the first example, the four copulas in the model do not all have the same functional form: For the first two situations the Frank copula was judged to be the appropriate one and for the last two situations we have chosen Cook–Johnson. As a consequence, the verbal aggressive reactions in the former situations show static local item dependence (i.e., the log odds ratio does not depend on the latent verbal aggression θ_p; see Figure 3) and the latter dimensional local item dependence (i.e., the log odds ratio increases as the latent verbal aggression θ_p increases; see Figure 3). Note that this is why we do not report the Mantel–Haenszel test statistics here, as the procedure assumes constant odds ratio over groups contrary to what we find in this data set. Although the residual dependencies are basically considered as a nuisance in this paper, we find here a case in which an interesting observation from a substantive point of view occurs. The first two situations are situations where someone else is to blame (the bus driver, the clerk), while the last two are self-blame situations. It appears that highly verbally aggressive people want to and/or tend to show the whole range of verbally aggressive reactions in the self-to-blame situations more than can be expected based solely from their general degree of verbal aggressiveness. In contrast, persons with a lower θ_p-value do not exhibit a lot of residual dependency in self-to-blame situations. For other-to-blame situations, the log odds ratio is basically independent of the verbal aggressiveness θ_p.

The data set also contained information about the gender of the participants and notice that in both the regular Rasch model and the Rasch Copula model the latent verbal aggression θ_p is regressed on gender (gender is dummy coded in a variable G_p which takes a value 1 for males and 0 for females). This resulted in the following model for the margins: latent variable $X_{pi} = \theta_p + g_p\gamma - \beta_i + \varepsilon_{pi}$, with observed binary variable $Y_{pi} = (X_{pi} > 0) = I(\varepsilon_{pi} > -\theta_p - g_p\gamma + \beta_i)$. Incorporating this additional covariate information only changes the formulation of the marginal model part, but not the model of the residual dependence structure as formulated by the copulas. Another advantage of the copula approach is that the effect can still be interpreted using the (log) odds ratio, as is the case for Rasch models with latent regression of θ_p (see De Boeck & Wilson, 2004, for more examples). Accordingly, this model can be labeled a "latent regression Rasch" Copula model. For the verbal aggression data, γ the estimated effect of gender under the copula model is 0.30 on the logit scale, with a standard error of 0.19. From this we can conclude that males are not significantly more inclined to verbal aggression than females. The estimated odds ratio is 1.35 with a 95% confidence interval ranging from 0.93 to 1.96. Notice that this hardly differs with the estimate of gender under the regular Rasch model; here the estimated odds ratio is 1.42 with a 95% confidence interval ranging from 0.98 to 2.06.

Again, for reasons of comparison, we fitted the interaction models by Hoskens and De Boeck (1997) on this data set. Given six items in a subset, this would lead to 15 (6!/(4!2!) second-order interaction terms, 20 (6!/(3!3!)) third-order interaction terms, 15 (6!/(4!2!) forth order interaction terms, and 6 (6!/(5!1!) fifth-order interaction terms. In other words, conditioning on a large amount of items can lead to a quite extensive model that might need to be restricted in some way to remain practical. To be able to compare the copula approach with an equally parsimonious model, two types of constrained models were chosen. In the first type of constrained model

TABLE 3.
Rasch and Rasch Copula models, and a constrained constant combination pairwise interaction model for the verbal aggression data.

Model parameter	Rasch estimate (SE)		Rasch Copula estimate (SE)		Restricted CCI estimate (SE)	
β_1	−1.13	(0.17)	−1.15	(0.17)	−0.04	(0.19)
β_2	−0.47	(0.16)	−0.50	(0.16)	0.68	(0.19)
β_3	0.01	(0.16)	−0.01	(0.16)	1.22	(0.20)
β_4	−1.66	(0.18)	−1.67	(0.18)	−0.59	(0.19)
β_5	−0.62	(0.18)	−0.64	(0.16)	0.58	(0.19)
β_6	0.08	(0.16)	0.06	(0.16)	1.37	(0.21)
β_7	−0.44	(0.16)	−0.42	(0.16)	0.15	(0.15)
β_8	0.78	(0.16)	0.66	(0.16)	1.50	(0.18)
β_9	1.61	(0.17)	1.53	(0.17)	2.44	(0.21)
β_{10}	−0.99	(0.17)	−0.95	(0.16)	−0.09	(0.17)
β_{11}	0.44	(0.16)	0.37	(0.16)	1.59	(0.19)
β_{12}	1.13	(0.17)	1.09	(0.16)	2.41	(0.22)
β_{13}	−1.13	(0.17)	−1.13	(0.17)	−0.04	(0.19)
β_{14}	−0.30	(0.16)	−0.33	(0.16)	0.88	(0.20)
β_{15}	0.96	(0.17)	0.93	(0.16)	2.25	(0.23)
β_{16}	−0.78	(0.16)	−0.79	(0.15)	0.40	(0.19)
β_{17}	0.15	(0.16)	0.10	(0.16)	1.45	(0.21)
β_{18}	1.57	(0.18)	1.52	(0.17)	3.02	(0.26)
β_{19}	0.30	(0.16)	0.27	(0.15)	0.97	(0.16)
β_{20}	1.59	(0.18)	1.46	(0.17)	2.42	(0.21)
β_{21}	3.06	(0.24)	2.91	(0.23)	4.04	(0.29)
β_{22}	−0.62	(0.16)	−0.64	(0.16)	0.34	(0.17)
β_{23}	0.47	(0.16)	0.41	(0.15)	1.63	(0.19)
β_{24}	2.09	(0.19)	1.99	(0.18)	3.51	(0.26)
γ	0.35	(0.19)	0.30	(0.19)	0.19	(0.12)
σ_θ	1.37	(0.07)	1.31	(0.07)	0.80	(0.06)
Frank: $J_1\{1, 2, 3, 13, 14, 15\}$; δ_1	—	—	0.84	(0.25)	—	—
Frank: $J_2\{4, 5, 6, 16, 17, 18\}$; δ_2	—	—	1.08	(0.26)	—	—
C–J: $J_3\{7, 8, 9, 19, 20, 21\}$; δ_3	—	—	0.50	(0.09)	—	—
C–J: $J_4\{10, 11, 12, 22, 23, 24\}$; δ_4	—	—	0.57	(0.09)	—	—
CCpwI: $J_1\{1, 2, 3, 13, 14, 15\}$; λ_1	—	—	—	—	0.42	(0.05)
CCpwI: $J_2\{4, 5, 6, 16, 17, 18\}$; λ_2	—	—	—	—	0.47	(0.05)
CCpwI: $J_3\{7, 8, 9, 19, 20, 21\}$; λ_3	—	—	—	—	0.49	(0.06)
CCpwI: $J_4\{10, 11, 12, 22, 23, 24\}$; λ_4	—	—	—	—	0.53	(0.05)
AIC	8123		7928		7970	

the interaction terms for an item subset were restricted to be all equal; in the second type of constrained model the second-order interactions for an item subset were restricted to be equal and all the other higher-order interactions were restricted to zero. The latter type of models will be referred to as pairwise interaction models. Both models result in one interaction parameter for each item subset, in correspondence with one copula parameter for each item subset in the copula approach. Both constant combinations as dimension dependent interactions were considered. The pairwise interaction model with a constant combination interaction (CcpwI) for each item subset was the best fitting of the possible interaction models; its parameter estimates are displayed in Table 3. Notice how the item parameter estimates are not comparable anymore to the regular Rasch or Rasch Copula models, and that they have slightly larger standard errors. The estimates of the interaction parameters are low, but all significant. The estimate of the standard deviation of the person ability distribution in the Rasch Copula model is considerably smaller than the equivalent estimate under the Rasch model, or the Rasch Copula model. The estimated effect of gender γ under the model is 0.19 on the logit scale, with a standard error of 0.12. The estimated odds ratio is 1.21 with a 95% confidence interval ranging from 0.95 to 1.53. This is slightly smaller than the corresponding estimates for the Rasch and Rasch Copula models, but still within both their 95% confidence interval. All four interaction parameters, $\lambda_1, \lambda_2, \lambda_3$, and λ_4, differ significantly from zero ($p < 0.01$). The overall goodness-of-fit as indicated by the AIC for this interaction model is 7970, an improvement compared to the regular Rasch model (AIC = 8123), but underperforming compared to the Rasch Copula model (AIC = 7928). Overall, it appears that large item subsets favor the copula approach with regard to both practical application and goodness-of-fit, whereas the conditional approach bumps into its boundaries.

6. Discussion

In this paper we have introduced the use of copulas for modeling residual dependencies. The basic idea is that we start out with a specific continuous latent response process (conditional on the random effects). Instead of assuming that, given the random effects, these latent response processes are independent, they are coupled through a copula function. As a consequence, the discretized observed random variables (i.e., responses to the items) derived from coupled latent processes will exhibit residual dependency. In this way, the residual dependencies can be taken into account without altering the marginal item characteristic curves (which were typically Rasch models in our illustration of the approach).

It has been shown that the proposed copula model overcomes some of the problems in existing models for residual dependencies. The model has the property of reproducibility because the univariate marginals are Rasch models and the item parameters can still be interpreted as difficulty parameters. This reproducibility property (see also Liang, Zeger, & Qaqish, 1992), or what McCullagh calls "upward compatibility" (McCullagh, 1989), means that we can remain within a well-known and conceptually attractive framework as offered by Rasch ICCs when applying a Rasch Copula model.

The different dependence structures of the varying copula functions enable the modeling of static as well as dimensional residual dependencies. The general class of Archimedean copulas (with Frank and Cook–Johnson as special cases) is limited to the case in which the dependency is symmetric. Therefore, Archimedean copulas are less suited to model serial dependence as in learning phenomena.

Building on the basic model proposed in this paper, extensions to other more complicated models can easily be implemented. For instance, the univariate margins do not need to be restricted to the Rasch model: One can easily insert the 2PLM or 3PLM, and both within-subset constant and varying covariates can easily be added. Moreover, the method can be applied to response variables other than binary. Instead of dichotomizing the latent continuous response variable, as has been done throughout this paper, we can also use two or more cutoffs and, consequently, end up with a polytomous observed random variable (if the underlying distribution is logistic, the model for the categorical data is a version of the graded response model; Samejima, 1969). The copula approach can easily be extended beyond binary variables to polytomous variables, in which case the number of possible interval points $d_{pi}^{(k)}$ in the distribution function of equation (2) increases with increasing categories. An advantage of the copula approach is that even for these polytomous items, taking residual dependencies into account means that only a few parameters have to be added. However, note that the treatment of residual dependencies in polytomous models is much more complex than in the binary case. In the latter case, one has only a single association or odds ratio (conditional on θ_p), but when considering the residual dependencies between two items with M_1 and M_2 categories, respectively, there are $(M_1 - 1)(M_2 - 1)$ (global or local) odds ratios. For example, one might hypothesize that in some cases the residual association becomes stronger with higher categories. The potential of the copula approach to accommodate for these kinds of phenomena will be part of the focus of future work.

The flexibility of the copula IRT models brings up a model selection issue. In practice, one has to choose the most adequate model for the marginal probabilities and find out which copula function shows the best representation of the residual dependence structure. Because the copula IRT model is fully likelihood-based, this practical issue is resolved by using the traditional likelihood based model selection tools (e.g., likelihood ratio tests, AIC, BIC). The construction of specific diagnostic tools within the IRT framework for assessing the fit of both the margins and the copula may be desirable and is an open field where much work can be done.

This rather straightforward and parsimonious copula approach can make use of existing and known IRT models for the marginals, without changing anything about the interpretation of the marginal model part. Adding or removing a residual dependent item or item set also does not influence the marginal model part and the interpretation of the model parameters as, for instance, would be the case when applying a conditional type of residual dependency model (in which the marginal probability model of an individual residual dependent item is dependent on other than the individual item parameter; see equation (1)). Obviously these characteristics make it a lot easier to explain these types of copula models to practitioners that have a basic and conceptual understanding of item response theory, but are less familiar with mathematical statistics and more complicated models that are available to approach the problem of local item dependency.

Another alternative approach would be to fix the margins to be normally distributed instead of following a logistic distribution, leading to multivariate random effect probit models (Ashford & Sowden, 1970). This would allow the introduction of residual dependencies by estimating the variance–covariance matrix of the multivariate normal distribution. Note that it can be shown that this approach fits in the copula modeling framework: the multivariate normal distribution can be reformulated as a joint distribution constructed by means of a so-called Gaussian copula with univariate normally distributed margins. However, the computational burden of a multivariate probit model is quite high because, to compute the joint probability for a certain response pattern, a multivariate normal distribution function has to be computed (implying multiple integrals that need to be approximated). In contrast, the copula approach presented in this paper avoids these extra integrals, and is not restricted to a linear dependency structure as implied by the use of the multivariate normal distribution. Thus the copula approach allows for a greater variety of models for the margins and models, i.e., copulas, for the dependency structure.

Although based upon modeling possibilities, technical statistical properties, and interpretability, the copula IRT model is clearly distinguished from other local item dependency models, general differences and similarities between the application of copula IRT models and their conditional (or random effects) counterparts are worthwhile to receive further investigation.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csáki (Eds.), *2nd international symposium on information theory* (pp. 267–281). Armenia, USSR: Tsahkadsov.

Ashford, J.R., & Sowden, R.R. (1970). Multivariate probit analysis. *Biometrics, 26*, 535–546.

Bahadur, R. (1961). A representation of the joint distribution of responses to n dichotomous items. In H. Solomon (Ed.), *Studies in item analysis and prediction* (pp. 158–168). Palo Alto, CA: Standford University Press.

Bell, R.C., Pattison, P.E., & Withers, G.P. (1988). Conditional independence in a clustered item test. *Applied Psychological Measurement, 12*, 15–26.

Bradlow, E.T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*, 153–168.

Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265–289.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika, 65*, 141–151.

Cook, R.D., & Johnson, M.E. (1981). A family of distributions to modeling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society, Series B, 43*, 210–218.

Cox, D.R. (1972). The analysis of multivariate binary data. *Applied Statistics, 21*, 113–120.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

Ferrara, S., Huynh, H., & Michaels, H. (1999). Contextual explanations of local dependence in item clusters in a large-scale hands-on science performance assessment. *Journal of Educational Measurement, 36*, 119–140.

Fitzmaurice, G.M., Laird, N.M., & Rotnitzky, A.G. (1993). Regression models for discrete longitudinal responses. *Statistical Science, 8*, 284–309.

Frank, M.J. (1979). On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequationes Mathematica, 19*, 194–226.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université Lyon, Série 3, 14*, 53–77.

Frees, A.W., & Valdez, E.A. (1998). Understanding relationships using copulas. *Actuarial Research Clearing House, 1*, 5–45.

Genest, C., & MacKay, J. (1986). Copules archimédiennes et familles de lois bi-dimensionelles dont les marges sont donées. *Canadian Journal of Statistics, 14*, 145–159.

Hoeffding, W. (1940). Masstabinvariante Korrelations-Theorie. *Schriften des Matematischen Instituts und des Instituts für angewandte Mathematik der Universität Berlin (5:3*, pp. 179–223). [Reprinted as *Scale-invariant correlation theory in the Collected Works of Wassily Hoeffding*, N.I. Fischer, & P.K. Sen (Eds.), New York: Springer].

Holland, P.W. (1990). The Dutch identity: A new tool for the study of item response models. *Psychometrika, 55*, 5–18.

Holland, P.W., & Rosenbaum, P.R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics, 14*, 1523–1543.

Hoskens, M., & De Boeck, P. (1997). A parametric model for local item dependencies among test items. *Psychological Methods, 2*, 261–277.

Ip, E. (2000). Adjusting for information inflation due to local dependence in moderately large item clusters. *Psychometrika, 65*, 73–91.

Ip, E. (2001). Testing for local dependence in dichotomous and polutomous item response models. *Psychometrika, 66*, 109–132.

Ip, E. (2002). Locally dependent latent trait model and the Dutch identity revisited. *Psychometrika, 67*, 367–386.

Ip, E., Wang, Y.J., De Boeck, P., & Meulders, M. (2004). Locally dependent latent trait models for polytomous responses. *Psychometrika, 69*, 191–216.

Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis, 46*, 262–282.

Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman & Hall.

Junker, B.W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56*, 255–278.

Kotz, S., Balakrishnan, N., & Johnson, N. (2000). *Continuous multivariate distributions* (Vol. 1). New York: Wiley.

Liang, K.-Y., Zeger, S.L., & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B, 54*, 3–40.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute, 22*, 719–748.

Masters, G.N. (1988). Item discrimination: When more is worse. *Journal of Educational Measurement, 25*, 15–29.

McCullagh, P. (1989). Models for discrete multivariate responses. *Bulletin of the International Statistics Institute, 53*, 407–418.

Meester, S.G. (1991). *Methods for clustered categorical data*. Unpublished doctoral dissertation. University of Waterloo, Canada.

Molenberghs, G., & Verbeke, G. (2005). *Models for discrete longitudinal data*. New York: Springer.

Mood, A.M., Graybill, F.A., & Boes, D.C. (1974). *Introduction to the theory of statistics*. New York: McGraw-Hill.

Nelsen, R.B. (1999). *An introduction to copulas*. New York: Springer.

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

Rosenbaum, P.R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*, 425–435.
Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 7*.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
Scott, S.L., & Ip, E. (2002). Empirical Bayes and item clustering effects in a latent variable hierarchical model: A case study from the national assessment of educational progress. *Journal of the American Statistical Association, 97*, 409–419.
Self, G.H., & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association, 82*, 605–610.
Sireci, S.G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*, 237–247.
Sklar, A. (1959). Fonctions de répartition à n dimension et leurs marges. *Publications Statistiques Université de Paris, 8*, 229–231.
Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27*, 159–203.
Tuerlinckx, F., & De Boeck, P. (2001a). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods, 6*, 181–195.
Tuerlinckx, F., & De Boeck, P. (2001b). Non-modeled item interactions lead to distorted discrimination parameters: A case study. *Methods of Psychological Research, 6*. [Retrieved May 20, 2005 from http://www.mpr-online.de/issue14/art3/Tuerlinckx.pdf].
Tuerlinckx, F., & De Boeck, P. (2004). Models for residual dependencies. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 289–316). New York: Springer.
Vansteelandt, K. (2000). *Formal models for contextualized personality psychology*. Unpublished doctoral dissertation. K.U. Leuven, Belgium.
Verhelst, N.D., & Glas, C.A.W. (1993). A dynamic generalization of the Rasch model. *Psychometrika, 58*, 395–415.
Yen, W.M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125–145.
Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187–213.
Zhao, L.P., & Prentice, R.L. (1990). Correlated binary regression using a generalized quadratic model. *Biometrics, 77*, 642–648.

Manuscript received 2 JUN 2006
Final version received 22 DEC 2006
Published Online Date: 8 JUN 2007