Examining the Structure of Concepts: Using Interactions Between Items

Dirk J. M. Smits,[1] **Paul De Boeck,**[1] **and Machteld Hoskens**[2]
K. U. Leuven, Belgium[1]**, CTB/McGraw-Hill, Monterey, CA**[2]

A framework is presented for modeling the relational structure of concepts using item response theory (IRT) models with interactions between the items, so-called models with local item dependency (LID). The proposed approach works for unidimensional and multidimensional concepts. For the relational structure of a concept to be analyzed, two types of items are used: items that directly refer to the concept and items that refer to the underlying components. The dependencies (the LIDs) are included in the model to analyze the mutual relations between the components and between the components and the concept. In a study on guilt, it was found that a unidimensional model complemented with situation-specific dependencies could explain the data that were gathered. Because of its flexibility, the approach is a promising tool for a structural analysis of concepts.
Index terms: item response theory, local item dependency, structural analysis, guilt.

Psychological concepts often contain different components. For example, Michel and Shoda (1995) conceive of personality as

> a stable system that mediates how the individual selects, construes and processes social information and generates social behaviors. . . . This theory assumes individual differences in the features of the situations that individuals select and the cognitive-affective mediating units (such as encodings and affects) that become activated and interact with and activate other mediating units (e.g. expectancies, goals, behavioral scripts and plans) in the personality system. (p. 246)

So, the concept of personality can be decomposed into subprocesses or components. Another way of looking at the concept of personality is less process oriented and mainly dimensional, as in, for example, the theory of the Big Five (see, e.g., Costa & McCrae, 1987, 1989; McCrae & Costa, 1997, 1999). In this theory, the personality of a person is described as a position on each of the five basic dimensions: Extraversion, Neuroticism, Openness to Experience, Agreeableness, and Conscientiousness. These dimensions are orthogonal simple-structure factors. Each of these dimensions can in turn be decomposed into what are called facets. For example, for Conscientiousness, the facets are competence, order, dutifulness, achievement striving, self-discipline, and deliberation (Costa & McCrae, 1995; Costa, McCrae, & Dye, 1991). Many other psychological concepts can be decomposed into more basic aspects in a similar way, including emotion concepts. In appraisal theory, for example, different emotions are supposed to be built on different patterns of appraisals and action tendencies, sometimes completed with other aspects such as bodily feedback (e.g., Ellsworth & Smith, 1988; Frijda, 1986, 1993; Frijda, Kuipers, & ter Schure, 1989; Frijda & Zeelenberg, 2001;

Applied Psychological Measurement, Vol. 27 No. 6, November 2003, 415–439
DOI: 10.1177/0146621603259277

© 2003 Sage Publications

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Izard, 1993; Omdahl, 1995; Reisenszein & Hofmann, 1993; Roseman, Antoniou, & Jose, 1996; Roseman & Smith, 2001; Scherer, 1993, 1997, 1999; Smith & Lazarus, 1993). In this theory, appraisals and action tendencies can be viewed as basic processes of emotions.

By decomposing psychological concepts into components, the structure of these components can be investigated. A common way to unravel the structure of such concepts is by using a multidimensional analysis. The most popular technique used for this purpose has been factor analysis, as in the Big Five theory.

If components are derived from factor analysis of items concerning a given trait, then the factors can be understood as ways in which individuals differ in how they show the underlying trait. The factors can refer to a different kind of behavior or to the same kind of behavior in different types of situations (e.g., Ortony & Turner, 1990).

Factor analysis is not the only technique with which to decompose concepts. Embretson (1980, 1984), for example, developed the multicomponent latent trait model (MLTM). In the MLTM, the probability of success on an item is modeled as the product of the probabilities of success on items referring to different subprocesses or components. Embretson's approach is different from factor analysis, but, like factor analysis, it is a multidimensional technique in that each of the components is a source of individual differences.

Even when the existing individual differences are not multidimensional, a concept may still be decomposable into more basic components. Different processes can be necessary for a behavior to arise, without these processes showing specific individual differences. For example, solving a mathematical problem, such as $3 \cdot (4 + 5)$, requires two different operations ($4 + 5$; $3 \cdot 9$), which both may be based on the same ability. An example from a totally different domain is that appraisals of a certain situation—for example, the situation being appraised as blocking a goal, as due to others, and as unfair—all are associated with a certain emotion, for example, anger (Ellsworth & Smith, 1988; Fitness & Fletcher, 1993; Frijda, 1986, 1993; Frijda et al., 1989; Ortony, Clore, & Collins, 1988; Scherer, 1993), without these appraisals being based on specific sources of individual differences. In principle, the individual differences in the various appraisals underlying an emotion can all be based on the same underlying person characteristic (e.g., trait anger). This means that a concept, which contains different components, can be unidimensional and is not necessarily multidimensional. Unidimensionality is not being proposed as the most plausible structure, but it is a possibility one may want to consider.

In this article, the focus is on an approach based on item response theory (IRT), one that is especially appropriate for relational concepts. Relational concepts, in this study, are concepts with several components and with a possibly complicated pattern of relations between the components and the global concept. These relations will be conceived of as dependencies between the components, as well as between the components and the global concept, beyond the effect of the one dimension or the multiple dimensions that reflect the global concept; see Hoskens and De Boeck (1997, 2001) for the unidimensional and the multidimensional cases, respectively. These dependencies are called local item dependencies because they are not explained by the global underlying dimensions (e.g., one general underlying trait). Most often, these local item dependencies are treated as problematic because they complicate the parsimony of a simpler model. It is argued in this study that local item dependencies can tell one about the structure of a relational concept and that they can be used to test the validity of psychological theories without explicitly including additional dimensions. This can be done by specifying different theories about the relations between components and the concept and translating these relations into IRT models with local item dependencies, so that the theories can be tested through the corresponding models.

Local item dependencies imply that subgroups of items will show higher or lower intercorrelations than can be expected based on the underlying dimension(s), as defined by the person

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

parameters. It is possible to capture such dependencies by fixed-effects parameters (constant over all persons) so that there is no need to add person parameters to the model (see the section on the models). A distinction is made in this article between multidimensionality as defined by the number of person parameters and multidimensionality as captured by fixed-effects dependency parameters. Two clear advantages are seen with a local item dependency approach. The first advantage is theoretical and concerns the flexibility and fine-grained nature of dependency models. The patterns of interitem dependency that are dictated by a theory can be quite complex. Including local item dependency parameters is a flexible way to translate a theory into a model without augmenting the number of parameters too much. With local item dependency models, it is possible to specify in a direct way all kinds of networks of interitem relations, as well as networks that can hardly be specified by including more person parameters. The second advantage is practical. It can be cumbersome to estimate models with a high number of person parameters with the extant computer programs—especially within a maximum likelihood framework (e.g., SAS Institute, 1999a, 1999b)—whereas it is rather easy to estimate local item dependency models with the same programs. However, the second disadvantage can be circumvented by applying Markov chain Monte Carlo (MCMC) techniques. These techniques have been used successfully with a high number of respondents (2,000 or more) and up to seven dimensions (Beguin & Glas, 2001; Segall, 2001).

First, the approach will be explained, and second, an application is described with data collected about feelings of guilt and components of these feelings. The same approach can be followed for other kinds of feelings but also for cognitive abilities, with the components referring to more elementary cognitive processes.

Modeling the Relational Component Structure Using Interactions

The approach to be presented is a general one. Neither the models nor the design for the data it requires are new. It is the aim of this study to present and illustrate the application of both (models and design) as an approach to test psychological theories in the test data and as a way of studying the internal validity of a test or a questionnaire. The models it is based on are IRT models; more specifically, they are models for local item dependencies (Hoskens & De Boeck, 1997, 2001; Jannarone, 1986; Kelderman, 1984; Kempf, 1977; Thissen & Steinberg, 1988; Thissen, Steinberg, & Mooney, 1989; Tuerlinckx & De Boeck, 2001, 2002; Wilson & Adams, 1995; Yen, 1993).

The design it requires implies two kinds of items: *component items* and *composite items* (Embretson, 1981, 1984). Component items are items for a single component that is assumed to underlie the concept, whereas composite items are items for the total concept. The test or questionnaire consists of families of items with the two types of items. An item family contains one composite item and several component items, all based on a common item stem (in cognitive tasks) or a common situation (in an inventory on emotions). The application in this study is based on a questionnaire for situational guilt feelings; each item family is associated with one situation (the common stimulus) and comprises four items: three component items, with each referring to a different component of situational guilt feelings in the given situation, and one composite item that refers to the guilt feeling itself. The guilt components studied are norm violation, worrying, and a tendency to restitute. Therefore, the component items for each situation are as follows:

- Do you feel like having violated a moral, an ethic, a religious, and/or a personal code in this situation? (Norm violation)
- Do you worry about what you did or failed to do in this situation? (Worrying)

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

- Do you want to do something to restitute for what you did or failed to do in this situation? (Tendency to restitute)

The composite item is the following:

- Do you feel guilty about what you did or failed to do in this situation? (Guilt feelings)

Together, these four items constitute the item family for the situation in question. The first three questions are based on a literature review on guilt (Barrett, 1995; Baumeister, Stillwell, & Heatherton, 1994, 1995; Caprara, Barbaranelli, Pastorelli, Cermak, & Rosza, 2001; Frijda, 1986; Gilbert, Pehl, & Allan, 1994; Izard, 1978; Lindsay-Hartz, De Riviera, & Mascolo, 1995; Smith & Lazarus, 1993; Tangney, 1995; Wicker, Payne, & Morgan, 1983) and on two pilot studies mentioned in the Application section.

It is the aim of this study to show how local item dependency models for data from a test design as explained can be used to compare, in a flexible way, various theoretically meaningful patterns of relations between component items and between component items and composite items. The flexibility concerns the specification of the model as well as its estimation. The patterns of relations can be quite complicated, without serious consequences for the estimation, because the number of person parameters does not increase.

The Model

As a starting point, the Rasch model (Rasch, 1960) is taken for binary data. In this model, the probability of a response x_{vi} to an item $i (i = 1, \ldots, I)$ by person $v (v = 1, \ldots, V)$ can be written as in equation (1):

$$P\,(X_{vi} = x_{vi}\,|\theta_v, \beta_i\,) = \frac{e^{[x_{vi}(\theta_v - \beta_i)]}}{1 + e^{(\theta_v - \beta_i)}}. \tag{1}$$

In equation (1), θ_v represents the person parameter or latent trait value of person v, and β_i represents the item parameter of item i, also called the item difficulty. Note that the Rasch model assumes equal discrimination of all items. This is not a necessary restriction for the models that will be discussed; see Hoskens and De Boeck (1997) for models with heterogeneous item discrimination. For the interpretation that is required in this study, a reparameterization is needed, with $\beta_i - \theta_v$ instead of $\theta_v - \beta_i$, so that the signs need to be reversed. After a reversal of the signs, and taking into account the guilt context, θ_v can be interpreted as the person's threshold for experiencing the three appraisals and guilt. The β_i can be interpreted as the inducing power from a situation with respect to the corresponding appraisal or guilt.

The Rasch model relies, among other assumptions, on the assumption of conditional independence or *local stochastic independence* (LSI). This assumption means that the dependence between the responses of an individual is solely attributed to the underlying trait, without the responses of the other items containing any additional information for the probability of responses to the item in question, so that equation (2) holds:

$$P(X_{v1} = x_{v1}, \ldots, x_{vI}\,|\theta_v\,) = \prod_{i=1}^{I} P(X_{vi} = x_{vi}\,|\theta_v). \tag{2}$$

If equation (2) does not hold, it is said that there is local item dependency (LID) because after partialling out the latent trait, covariances between the items do remain. It should be noted that LID is always defined in terms of a given model. Because the assumption of LSI is often too strong, LID has attracted some attention in the literature. What is called LID can be dealt with in several

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Table 1

Model for Constant Pairwise Interaction

Response Pattern (x_{vi}, x_{vj})	Adjusted Formula[a]
(0,0)	$1/v(\theta)$
(0,1)	$\exp(\theta_v - \beta_j)/v(\theta)$
(1,0)	$\exp(\theta_v - \beta_i)/v(\theta)$
(1,1)	$\exp[(\theta_v - \beta_i) + (\theta_v - \beta_j) - \beta_{ij}]/v(\theta)$

a. β_{ij} is the interaction parameter for the corresponding item pair, and $v(\theta) = 1 + \exp(\theta_v - \beta_i) + \exp(\theta_v - \beta_j) + \exp(2\theta_v - \beta_i - \beta_j - \beta_{ij})$.

ways (e.g., Andrich, 1985; Bradlow, Wainer, & Wang, 1999; Chen & Thissen, 1997; Hoskens & De Boeck, 1997, 2001; Jannarone, 1986; Kelderman, 1984; Kempf, 1977; Thissen & Steinberg, 1988; Thissen et al., 1989; Tuerlinckx & De Boeck, 2001, 2002; Wilson & Adams, 1995; Yen, 1993).

A major concern has been how to deal with LID so that the measurement quality is preserved while using models without LID parameters. An efficient solution is to group dependent items in a testlet, so that the number of items correct defines categories of the superitem that corresponds to the testlet (Andrich, 1985; Thissen & Steinberg, 1988; Thissen et al., 1989; Wilson & Adams, 1995; Yen, 1993).

A somewhat different approach that is less focused on measurement but concentrates on modeling instead is model extension. A prominent example of this approach is the model of Bradlow et al. (1999), in which random effects, and therefore new dimensions, are added to capture the dependencies. An alternative for this approach is to use fixed LID parameters for the dependent items (Jannarone, 1986; Kelderman, 1984). The latter approach will be followed in this study because, first, the number of dimensions needs to be restricted and, second, because the random-effects approach may require an extra dimension per local dependency between item responses (for all item pairs with LID). However, for other applications, the random-effects approach may be the one to be preferred.

When following this fixed-effects LID approach, the interaction between items (used here as another term for dependency) can be constant or dimension dependent. Following the terminology used by Hoskens and De Boeck (1997), constant interaction is dependency that is constant over all participants, independent of their position on the latent trait, whereas dimension-dependent interaction depends on the position of a person on the latent trait(s). For reasons of simplicity, this study will concentrate on constant LID. Dimension-dependent models have also been tested by the authors but without success because they did not explain the data any better. This means that there were no individual differences in the degree of LIDs.

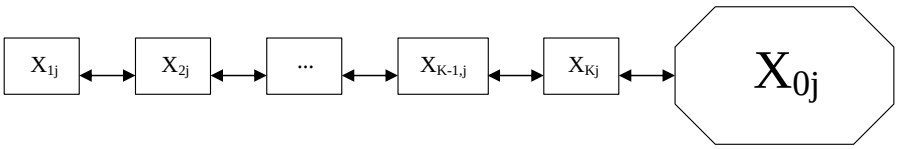
Table 1 shows the basic model formulation for the case when there is constant interaction between a pair of items i and j.

It is now easy to see that when β_{ij} is negative, the probability of observing the response pattern (1,1) increases, and when β_{ij} is positive, the probability decreases, in comparison to the probability of the same event under the Rasch model. A negative vale of β_{ij} indicates a positive interaction, whereas a positive value of β_{ij} indicates a negative interaction.

The implication of this interaction model is that the item parameters β_i and β_j are difficult to interpret because they are no longer pure reflections of the difficulty but dependent on the interaction as well. This may be a reason to prefer an alternative approach (see earlier discussion). However, for the reasons explained earlier, the fixed-effects LID approach will be pursued here.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Figure 1
A Linear-Sequence Structure



Note. The size of the interactions between consecutive components and the concept can be different within and over item families (item families are denoted with index j).

Model Estimation

All the models presented by Hoskens and De Boeck (1997) and the ones presented below can be estimated with existing IRT programs such as CONQUEST (Wu, Adams, & Wilson, 1997), LOGIMO (Kelderman & Steen, 1993), or MULTILOG (Thissen, 1988). The appendix of Hoskens and De Boeck describes how the models can be estimated using these programs. It is also possible to use SAS V8, PROC NLMIXED (Wolfinger, 1999) for the estimations (Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003).

Testing the Fit of the Model

If two models are nested, a likelihood ratio test can be used to compare the fit of both models. When the models are not nested (the different structures to be presented are not all nested one into the other), Akaike's information criterion (AIC) (Akaike, 1977) can be used. The AIC is a measure of lack of fit. A model has a better fit than another does if the AIC of the first model is lower than the AIC of the second. The index contains a penalty for the number of parameters added: It equals the likelihood ratio value plus twice the number of parameters estimated. To further investigate the goodness of fit, a bootstrap method (Efron & Tibshirani, 1993) will be used, as explained in the Estimation section.

Specific Interaction Structures

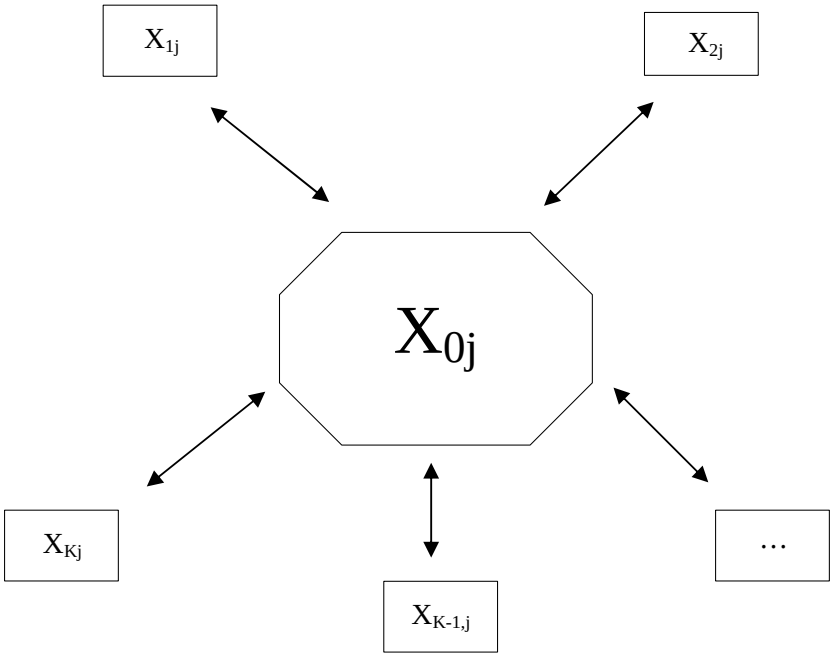
Consider a questionnaire with the following structure: J item families (index $j = 1, \ldots, J$), each with K componential items ($k = 1, \ldots, K$) and a global or composite item ($k = 0$). The items will be denoted with a double subscript jk, X_{jk}. For example, X_{jk} with $k \neq 0$ refers to the componential item k from family j, whereas X_{jk} with $k = 0$ refers to the composite item from family j. Persons are denoted with an index $v (v = 1, \ldots, V)$, so that X_{vjk} is the response of person v to item k from family j.

Different types of dependency patterns, also called interaction structures, may exist. Four structures are described here: a linear-sequence structure, a star structure, a cluster structure, and an item family structure. The dependencies are be represented with arrows, with each arrow representing one dependency. The first three structures are meant to apply within each of the item families, but their degree may differ depending on the item family.

In a *linear-sequence structure*, the components have an order, so that they interact only with adjacent items. Suppose further that the composite item reflects an end point in the process and that it interacts only with the "last" component. The result is a linear-sequence structure as represented in Figure 1.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Figure 2
A Star Structure



Note. The size of the interactions of the components with the concept can be different within and over item families (item families are denoted with index *j*).

The ordering can be based on an order in time, but it may reflect as well a chain-like overlap structure between the components and the end result, without any reference to an order in time. Note that when a separate person parameter or random effect would be used to model the dependencies to obtain a similar model, as many person parameters as the number of adjacent pairs times the number of item families would be needed. An analogous consideration applies to the following types of structures.

In a *star structure*, each component item interacts only with the composite item. The corresponding component structure is represented in Figure 2.

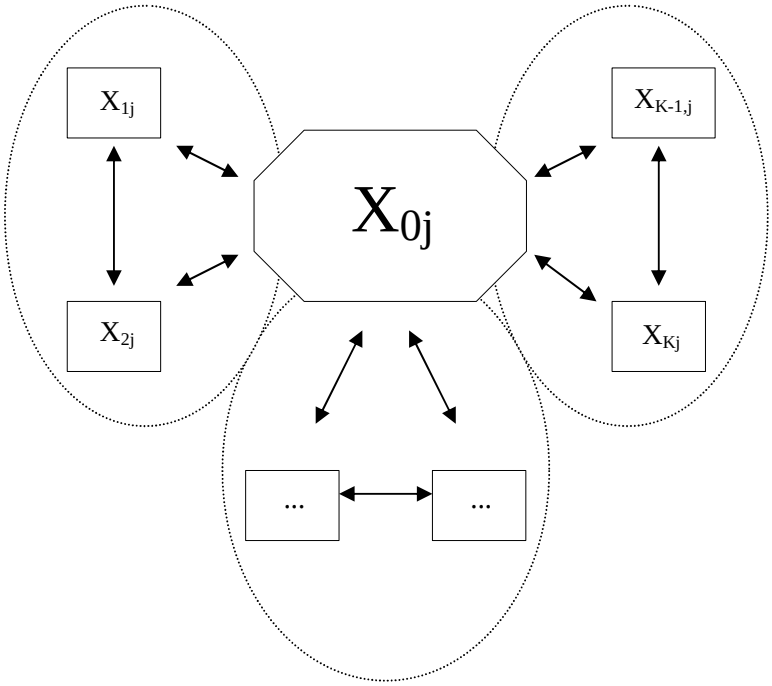
In a *cluster structure*, the items are structured within clusters, and the interaction occurs between all pairs of component items belonging to the same cluster, but there are no interactions between clusters. It is further assumed that the composite item is the only element of overlap between all clusters. An example of the cluster structure is represented in Figure 3. An extension of this structure, which will not be considered here because it does not yield better results, is one with higher order interactions within the clusters.

In the *item family structure*, pairwise interactions occur between all items of the same item family. In Figure 4, an item family structure is shown. Higher order interactions are not considered for the same reason as for the cluster structure.

The different structures each reflect a different psychology of the phenomenon under investigation. In this application, the phenomenon is feeling guilty. A linear-sequence structure suggests

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Figure 3
Cluster Structure



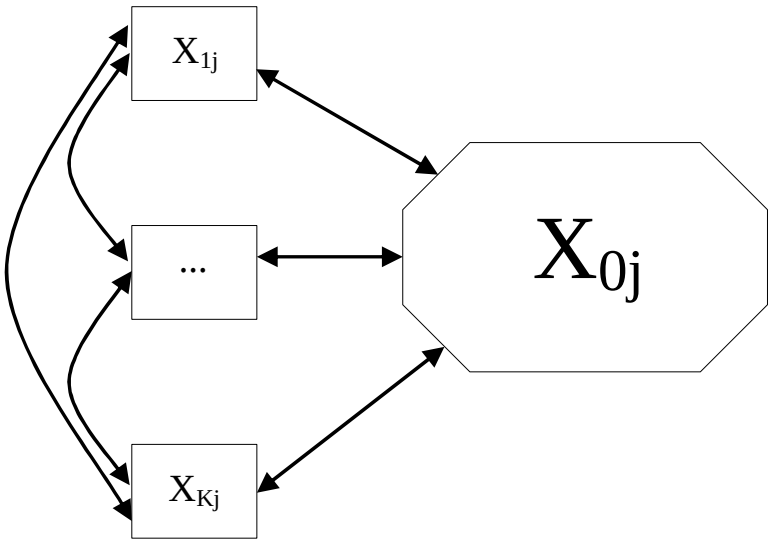
Note. The size of the interactions within each cluster between components and the concept can be different within and over item families (item families are denoted with index *j*).

that feeling guilty is an end product of the components norm violation, worrying, and tendency to restitute, and these components are ordered in a linear way in how they affect feeling guilty. For example, authors such as Frijda (1986) and Frijda et al. (1989) argue that appraisals precede action tendencies and that action tendencies are experienced before or together with the emotion. This theory could lead to a structure in which the feeling of having violated a norm (an appraisal) precedes worrying (a covert act), whereas they both precede the tendency to restitute (an overt act), and guilt is the end product, corresponding with a linear-sequence structure. Note that not only a psychological order can lead to a linear-sequence structure but that also other orders (e.g., order of presentation in the questionnaire) can lead to a similar structure.

A star structure implies that the components each interact independently with the feeling and not with one another (unless through the underlying trait). A cluster structure could mean that the dependency is organized into a cluster for the appraisals (feeling of having violated a norm) and one for action tendencies (worrying and the tendency to restitute), each complemented with feeling guilty. This distinction between appraisals and action tendencies is primarily based on the work of Frijda (1986) and Frijda et al. (1989), who state that appraisal and action tendencies can be separate from emotion processes. Finally, an item family structure would suggest that guilt feelings and their components show situational specificity to some extent because each item family is defined on the basis of a different situation. The basis for the dependency is the shared situation. To approximate the corresponding structure using person parameters would imply that, in addition to the general

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Figure 4
Item Family Structure



Note. The size of the interactions between the components and the concept is equal within an item family but can be different over item families (item families are denoted with index *j*).

dimension, as many dimensions are defined (and estimated) as there are item families. In this application, there are 10 item families, but there may be more in other cases.

Various other kinds of structures may exist, but they will not be described here. As mentioned earlier, in principle, higher order interactions (e.g., triple interactions) are also possible, but because of their complexity and because they do not yield a better fit, they will not be considered further here.

Application: Modeling of Guilt Feelings

Data and Preliminary Analysis

Ten situations have been selected for the application on guilt feelings in the following way: In a first pilot study, a sample consisting of 46 (20 males and 26 females) 18-year-old subjects was asked to describe three situations they felt guilty about, each stemming from a different domain of life: (a) work or study situation, (b) personal relationships, and (c) leisure time. To use the descriptions in this study, all information about responses from the person in the situation was deleted, and only the information about the situation was retained. Subsequently, 10 stories were selected using the following criteria: understandability, equal representation of the three domains of life, variation in content and assumed guilt-inductive power, conformity with the environment of 18-year-old people, and equal representation of stories stemming from males or females. The selected situations are listed in the appendix. This article uses the same numbers as used in the appendix and a keyword to refer to the situations.

Based on a second pilot study in which 12 judges rated the 10 selected situations on three appraisal components (self-responsibility, norm violation, and negative self-evaluation), two of these components were omitted. The first component, self-responsibility, was omitted because the ratings were

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Table 2

Eigenvalues for the First 12 Principal Components of the Nondichotomized and the Dichotomized Data

Principal Component	Nondichotomized Data	Dichotomized Data
1	9.58	6.78
2	3.30	2.99
3	2.47	2.31
4	2.33	2.26
5	2.26	2.00
6	1.91	1.79
7	1.81	1.61
8	1.71	1.59
9	1.62	1.53
10	1.59	1.42
11	1.04	1.13
12	1.01	1.01

fairly constant over all the judges. Apparently, for this set of situations, self-responsibility can be considered a rather objective appraisal primarily based on the situation descriptions and not to be modeled as based on an individual sensitivity, as is assumed in the models used here. Second, the third component, negative self-evaluation, was also not retained because over situations, it showed an extremely high correlation (.98) with norm violation, so that the two aspects could not be differentiated. Because norm violation seems more important from the literature on guilt, the negative self-evaluation appraisal was omitted. In sum, only one appraisal component will be included, next to two action tendency components: worrying and tendency to restitute.

The data analyzed here are from a much larger third study with 10 situations and its four associated questions (one item family per situation). The questionnaire was completed by 268 high school students age 18 (130 males and 138 females) who answered on a 4-point scale (0 = *no*, 1 = *not likely*, 2 = *likely*, and 3 = *yes*) whether the corresponding appraisals, action tendency, or guilt feeling would apply to them in the described situation. The data can be dichotomized in a natural way by recoding 0 and 1 (*no* or *not likely*) into 0 and 2 and 3 (*likely* and *yes*) into 1. The internal consistency of these data, as measured with Cronbach's alpha, was equal to .91 before dichotomization and to .87 after dichotomization.

To find out whether a model with one general latent trait complemented with LIDs could fit the data, the authors did a principal components analysis (PCA). For the nondichotomized and the dichotomized data, the eigenvalues for the first 12 principal components are given in Table 2. From both PCAs, it may be concluded that there is a dominant first component. Therefore, a model that has only one person parameter but is supplemented with LIDs will be given first.

Furthermore, the DIMTEST program (Stout, 1987; Stout, Nandakumar, Junker, Chang, & Steidinger, 1992; Stout, Douglas, Junker, & Roussos, 1993) was applied as an explicit test for the lack of unidimensionality in the responses to all items (10×4 in total). The DIMTEST statistics assess whether one group of items is dimensionality distinct from another group of items. Therefore, the first group of items has to be as unidimensional as possible and seem dimensionally distinct from the rest of the items. To partition the item set, the authors have followed a method that is provided by the program based on a factor analysis. An unrotated principal axis factor analysis using

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

tetrachoric correlations between the items was performed as a first step of the program. Based on this analysis, the program selected 10 items that had the highest loadings on the second factor to obtain a group of items that is as unidimensional as possible and, at the same time, seems dimensionally distinct from the rest of the items. The DIMTEST statistics compare these 10 items to the other 30 items with respect to dimensionality. Of the examinees, 85% were used for calculating the DIMTEST statitistics: 40 examinees were removed from the calculation of the DIMTEST statistics as they led to sparse data cells. The index T measures the departure from unidimensionality (Stout, 1987). The value for the T statistic equals .663 ($p = .254$), and the value for the more powerful T' statistic, based on refinements of the DIMTEST procedure made by Nandakumar and Stout (1993), equals .814 ($p = .208$), meaning that the data here can be considered unidimensional in the sense as defined by DIMTEST.

In the following, various instantiations of the types of dependency models presented earlier will be formulated. They are each based on a different hypothesis, and they all have only one underlying latent trait. After an analysis based on these models and the selection of a best model, the PCA results will be presented further to see whether they are in agreement with the selected dependency model. Another way of proceeding would be to follow the specific suggestions from the PCA. However, this would be a purely exploratory approach, whereas the authors want to show the potential of the LID approach to compare theories that imply different interaction structures. For a PCA to be a good exploratory tool to indicate LIDs, the LIDs need to be strong enough in terms of explained variance and so that clear item structures can be delineated from the PCA. Because the aim is to illustrate a theory-based approach, the authors will continue with LIDs that are defined a priori to compare their goodness of fit.

Modeling

Five different relational structures are investigated. The first is a baseline structure without any interaction: the *main-effects model*, which is an independence model, if abstraction is made of the one underlying latent trait. The second structure is a *linear-sequence structure*. It is based on a sequential hypothesis of guilt feelings, with one component following the other and with guilt feelings as the end product. In this structure, each component interacts with the subsequent component, and only the last component interacts with guilt. The linear-sequence structure can have different variants depending on the sequence of the components. The expected order is that the appraisal (norm violation) comes before the action tendencies and that the action tendency for a covert act (worrying) precedes the action tendency for an overt act (restitution) and that the feeling of guilt follows. This order is indicated as *N*-*W*-*R*-*G* in Table 4. However, all possible orders with guilt in the last position were tried. The third structure is a *star structure*. It is based on the hypothesis of a convergent but independent activation of guilt from the various components. Each component interacts with guilt feelings but not with the other components. The fourth structure is a *cluster structure*. It is based on the hypothesis that components of a similar kind interact. This is a modification of the previous structure so that components interact not only with guilt feelings but also with components of a similar kind. To define similarity of components, these are grouped into appraisals and action tendencies (see, e.g., Frijda, 1986). As norm violation is the only appraisal among the three components, norm violation would interact only with guilt feelings, so that norm violation and guilt feelings form one cluster. The remaining two components are action tendencies: worrying and a tendency to restitute. They are assumed to interact with one another and with guilt feelings, so that together with guilt feelings, they constitute the second cluster. Finally, the fifth structure is an *item family structure*, with pairwise interactions between all items belonging to the same item family. This structure is based on the hypothesis that guilt has a partially situation-specific meaning, possibly varying in degree depending on the situation. This model has equal interaction

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

parameters within each item family as all items in the family refer to the same situation, but the interaction parameters may differ from situation to situation.

Two models were used as control models as a way to test for alternative and methodological explanations. First, one of the linear-sequence structures corresponds to the order of presentation of the items. If this model holds, the order of presentation is a candidate for explaining the interactions, and the result must be interpreted as an artifact. This model is called the *order-of-presentation model*. Second, if the respondents want to be consistent for responses related to the same situation, one may expect an item family structure, but there is no reason why the consistency should differ from situation to situation while being equal within a situation. Therefore, a model will be tested with only one interaction parameter for all interactions within item families. If this model fits, the LIDs could be due to a consistency style that is induced by the structure of the questionnaire. This model is called the *situational consistency model*. These two alternative models were added to test plausible alternative sources of item dependency that are not caused by guilt-related processes.

In Figure 5, a representation of the various interaction models is given. Only one variant of the linear-sequence structure is shown, but the other variants are estimated as well. For the item family structure, only the first two item families are shown. In the picture of the item family structure, equal types of arrows refer to equal interaction parameter values. For the other models, the interaction parameter can also differ depending on the item family.

Six dependency models have been described, but note that the order-of-presentation model is a particular variant of the linear-sequence structure model and that the situational consistency model is a special case of the item family structure model, so that there are in fact only four basic types of models. In the remainder of this section, the parameterization of the four basic types of models will be explained, based on Hoskens and De Boeck (1997). Writing the probabilities for the various models, the denominator is always the sum over the numerators for all possible response patterns of the item family. For all models applied to the data, it is assumed that the items have equal discriminations.

The first interaction model is the *linear-sequence structure*. This model contains three pairwise interaction effects per item family: two between component items and one between a component item and the composite item. Using the example from Figure 5, one can obtain the parameterization for one item family, as shown in the upper part of Table 3. The table shows the formulas for three response patterns. The index j denotes the situation the items are associated with or, in other words, the item family, and the index k for the components is given the values N, W, R, and G (N for norm violation, W for worrying, R for tendency to restitute, and G for guilt); β_{NWj}, β_{WRj}, and β_{RGj} represent the interaction parameters. As mentioned, in Table 3, only the parameterization for the linear-sequence structure for three response patterns is given as a way of presenting the model. The parameterization of the other response patterns is straightforward using the following rules:

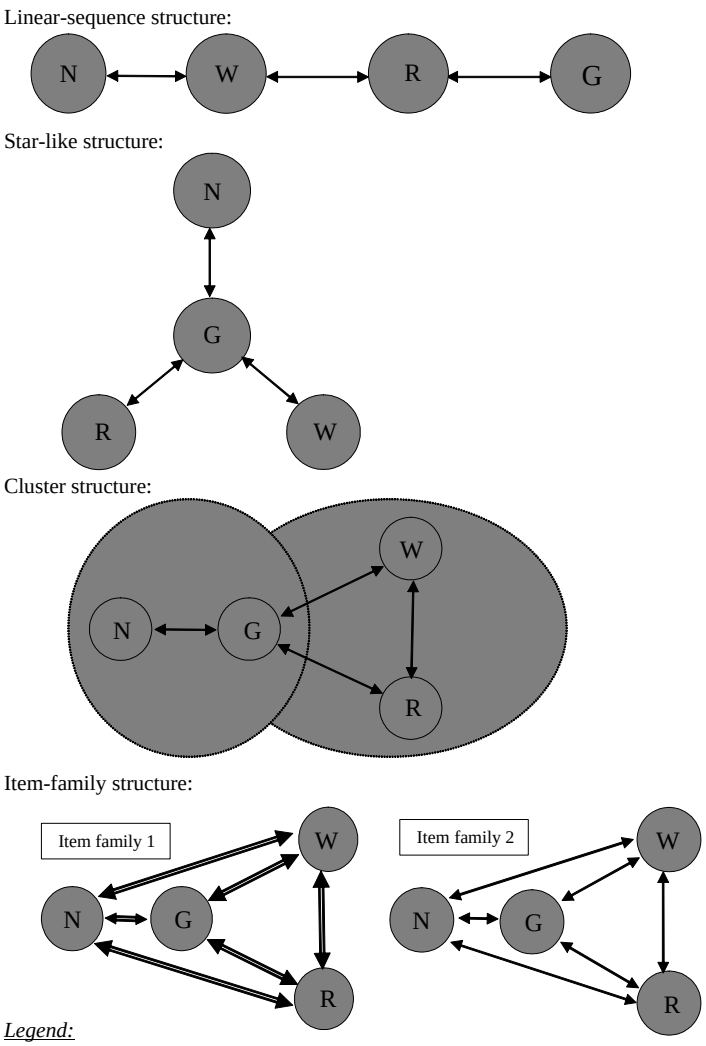
1. The item parameter is included in the numerator if, for that item, the response is 1.
2. A (pairwise) interaction parameter is included in the numerator if, for the items involved in the interaction, the item responses are both 1 and an interaction is assumed between the two items.
3. The denominator $v(\theta)$ is the sum of all different terms appearing in the numerators.

These rules hold also for the structures to be presented in the following.

The second interaction model is the *star structure*. Also, this model contains three pairwise interaction effects per item family: one between each component and guilt feelings. The parameterization of this model is exemplified in the second part of Table 3 for three response

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Figure 5
The Interaction Models for Guilt



patterns. The principle for constructing the numerators and the denominators is the same as for the linear-sequence structure model, but the interactions are different.

The third interaction model is the *cluster structure*. This model has four pairwise interaction parameters per item family: one for norm violation and guilt feelings and three for all pairs of worrying, tendency to restitute, and guilt feelings. The parameterization is exemplified for four response patterns in the third part of Table 3.

The fourth interaction model is the *item family structure*. The model has one interaction parameter per item family, the same for all item pairs within the item family. Its parameterization is illustrated in the fourth part of Table 3 for six response patterns.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Table 3

Parameterization of Dependency Models

Model	Norm (N) Violation	Worrying (W)	Restitution (R)	Guilt (G)	Parameterization
Linear-sequence structure (the N-W-R-G sequence)	1	1	0	0	$\exp(2\theta_v - \beta_{Nj} - \beta_{Wj} - \beta_{NWj})/v(\theta)$
	0	1	1	0	$\exp(2\theta_v - \beta_{Wj} - \beta_{Rj} - \beta_{WRj})/v(\theta)$
	0	0	1	1	$\exp(2\theta_v - \beta_{Rj} - \beta_{Gj} - \beta_{RGj})/v(\theta)$
Star structure	1	0	0	1	$\exp(2\theta_v - \beta_{Nj} - \beta_{Gj} - \beta_{NGj})/v(\theta)$
	0	1	0	1	$\exp(2\theta_v - \beta_{Wj} - \beta_{Gj} - \beta_{WGj})/v(\theta)$
	0	0	1	1	$\exp(2\theta_v - \beta_{Rj} - \beta_{Gj} - \beta_{RGj})/v(\theta)$
Cluster structure	1	0	0	1	$\exp(2\theta_v - \beta_{Nj} - \beta_{Gj} - \beta_{NGj})/v(\theta)$
	0	1	0	1	$\exp(2\theta_v - \beta_{Wj} - \beta_{Gj} - \beta_{WGj})/v(\theta)$
	0	0	1	1	$\exp(2\theta_v - \beta_{Rj} - \beta_{Gj} - \beta_{RGj})/v(\theta)$
	0	1	1	0	$\exp(2\theta_v - \beta_{Wj} - \beta_{Rj} - \beta_{WRj})/v(\theta)$
Item-family structure	1	1	0	0	$\exp(2\theta_v - \beta_{Nj} - \beta_{Wj} - \beta_{INTj})/v(\theta)$
	1	0	1	0	$\exp(2\theta_v - \beta_{Nj} - \beta_{Rj} - \beta_{INTj})/v(\theta)$
	1	0	0	1	$\exp(2\theta_v - \beta_{Nj} - \beta_{Gj} - \beta_{INTj})/v(\theta)$
	0	1	1	0	$\exp(2\theta_v - \beta_{Wj} - \beta_{Rj} - \beta_{INTj})/v(\theta)$
	0	1	0	1	$\exp(2\theta_v - \beta_{Wj} - \beta_{Gj} - \beta_{INTj})/v(\theta)$
	0	0	1	1	$\exp(2\theta_v - \beta_{Rj} - \beta_{Gj} - \beta_{INTj})/v(\theta)$

For the models that appear promising, a restricted version will be estimated as well: one with each kind of interaction parameter being constant over all situations (item families) to test whether a common parameter value can be generalized over situations. For the item family structure model, the restricted variant equals the situational consistency model.

Estimation

All the models are estimated with Conquest (Wu et al., 1997). Conquest uses an expectation-maximization algorithm (Dempster, Laird, & Rubin, 1977) following the approach of Bock and Aitkin (1981). The integrals are approximated numerically using a quadrature method with 20 quadrature points in the interval −6 to 6. Furthermore, it is assumed that the person parameters are normally distributed over persons (marginal maximum likelihood) (Baker, 1992). Individual person parameter estimates are obtained with empirical Bayes estimation.

The main-effects model will be used as a reference. The AIC (Akaike, 1977) will be used as a relative measure of fit. To further investigate the fit of the best-fitting models, the interitem correlations of the data will be compared to the interitem correlations as expected from those models. Therefore, a bootstrap methodology will be used (Efron & Tibshirani, 1993): Based on the parameter estimates obtained under the models, 500 new data sets for each model will be generated. These replicated data are used to derive the 95% confidence interval for each interitem correlation. The empirical pairwise interitem correlations will be compared with these 95% confidence intervals to see which model leads to a similar pattern of interitem correlations as the data in this study. The proportion of empirical interitem correlations within the 95% confidence intervals will be used as a goodness-of-fit measure. This measure will be derived for the set of all correlations and for the set of within-situation correlations.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Table 4

Fit of the Various Models: Main-Effects Model, Linear-Sequence Structure, Star Structure, Cluster Structure, Item Family Structure, and Two Control Models

Model	–2 Log-Likelihood	Number of Parameters	Akaike Information Criterion
No interactions	10545.7	41	10627.7
Linear-sequence structures			
N-*W*-*R*-*G*	9805.8	71	9947.8
N-*R*-*W*-*G*	9609.8	71	9751.8
W-*N*-*R*-*G*	9860.4	71	10002.4
W-*R*-*N*-*G*	9756.4	71	9898.4
R-*W*-*N*-*G*	9611.0	71	9753.0
R-*W*-*N*-*G*	9702.0	71	9844.0
Star structure	9481.1	71	9623.1
Cluster structure	9365.5	81	9527.5
Cluster structure with interactions constant over situations*	9432.8	45	9522.8
Item family structure*	9329.3	51	9431.3
Control models			
Order of presentation model		See *N*-*W*-R-*G* above	
Situational consistency model	9366.5	42	9450.5

Note. Selected models are marked with an asterisk. *N* = norm violation; *W* = worrying; *R* = restitution; *G* = guilt.

Results

In Table 4, the results for the different LID models are summarized. The three best-fitting models are indicated with an asterisk.

The main-effects model (no LIDs) fits the data clearly worse than all other models. All the linear-sequence structures do clearly better, and among these, the one with the sequence reflecting the order of presentation (*N*-*W*-*R*) is certainly not the best, so that order of presentation can be ruled out as a basis for the LID structure. As a sequence for the components, the *N*-*R*-*W* order (norm violation, tendency to restitute, and worrying) seems the best. The star structure does slightly better than the linear-sequence structure, but the cluster structure outperforms both. Therefore, the restricted variant with all interaction parameters constant over situations was also tested. The AIC value of this restricted variant is lower than the one of the nonrestricted version. This model corresponds with a within-cluster interaction structure that can be generalized over the situations.

However, the best-fitting model is the one with an item family structure. Hence, the interactions can probably be attributed to a situational specificity of guilt feelings, as all items in one item family share the same situation. The control model with only one interaction parameter that is equal for all situations yields a higher AIC value than its nonrestricted version mentioned above, but the difference is only minor. From these results, the cluster structure and the item family structure seem reasonable structures for the concept of guilt.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Table 5

Percentage of Empirical Interitem Correlations Covered by the 95% Confidence Interval as Constructed Based on the Bootstrap Methodology

Model	All Correlations	Within-Situation Correlations
Cluster structure	87	65
Item family structure	89	92
Situational consistency model	87	70

Table 6

Estimated Values for the Item and Interaction Parameters and Their Standard Error (for the Item Family Structure)

	Norm Violation		Worrying		Tendency to Restitute		Guilt		Interactions	
Item Family	β	SE	β	SE	β	SE	β	SE	β	SE
1. Breakup	1.82	.13	–.06	.15	1.10	.13	.70	.14	–.85	.03
2. Trumpet	3.73	.23	2.59	.18	2.75	.18	2.80	.19	–1.33	.05
3. Shoes	2.41	.14	1.12	.13	1.86	.13	1.34	.13	–1.05	.03
4. Movie	2.06	.13	1.54	.14	1.54	.14	1.09	.14	–1.16	.03
5. Discussion	1.93	.16	1.20	.18	.25	.21	.89	.19	–1.32	.03
6. Secret	0.60	.23	1.48	.19	1.96	.17	.42	.24	–1.46	.03
7. Youth movement	1.95	.15	.66	.18	2.82	.14	.85	.18	–1.49	.03
8. Pen pal	1.92	.14	2.18	.14	1.55	.14	1.78	.14	–1.23	.03
9. Jacket	2.85	.13	.43	.17	–1.17	.26	1.18	.15	–1.13	.03
10. Homework	2.44	.15	2.84	.16	2.05	.14	2.84	.16	–1.35	.03

To further investigate the goodness of fit of these three models, the previously presented bootstrap methodology was followed. The percentages of the empirical correlations covered by the 95% confidence intervals, as constructed for the clusture structure, the item family structure, and the situational consistency structure, are given in Table 5. Clearly, the item family structure is superior in explaining the correlations. It is not only superior, but the percentages are sufficiently high to conclude that the model has a reasonable goodness of fit. Note that the within-situation correlations form only 8% of the total number of correlations, so that they do not have a strong impact on the percentage of all correlations that fall within the confidence interval.

Many other models were fitted to the data as well, including models with higher order interactions and with dimension-dependent interactions (see earlier), but by far not any of these did better than the item family structure model.

The values for the item parameters and the interaction parameters of the item family structure are given in Table 6.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

For an interpretation of the values, it should be noted that the mean of the person parameters is set to zero for reasons of identification. Table 6 shows the parameter estimates of the component items (norm violation, worrying, and tendency to restitute) and of the composite items (guilt). Note that the lower the value of β, the higher the inductive power of the situation. The higher the estimated value, the less the corresponding component or the less guilt is elicited by the situation involved. The interaction parameters are all negative, meaning that there is a positive interaction between the items. The lowest guilt-inducing situations (column 8, Table 6) are Situations 2 (trumpet) and 10 (homework), whereas the highest guilt-inducing situations are Situations 1 (breakup) and 6 (secret), which are both about close relationships, whereas Situations 2 and 10 are not. As explained earlier, the item parameters cannot safely be interpreted as difficulties because they depend on the interaction. In case the item parameters correspond with the item difficulties, a perfect negative correlation with the proportions of 1-responses is expected. As the item parameters of the item family structure correlate only $-.78$ with the proportions of 1-responses, one can conclude that, in general, they resemble item difficulties, but the correspondence is not perfect. Treating them as item difficulties could be misleading.

To interpret the interactions in terms of item characteristic curves (ICCs), the ICCs of the main-effects model (Rasch model) have been compared with those of the item family structure (see Figure 6). The ICCs show the probability of giving a 1-response as a function of the latent trait. In the left panel, the ICCs of the item family structure are shown for all four items per item family. In the right panel, the corresponding ICCs of the same items are plotted for the Rasch model. A thicker line is used for the ICC of a composite item.

Note that the ICCs of the item family structure are based on the sum of probabilities for different response patterns. For example, the probability of giving a 1-response to the first item is, according to the item family structure, equal to the sum of the probabilities of all response patterns that contain a 1-response for Item 1, so $\mathrm{P}(\mathrm{Y}_{v11} = 1|\theta) = \mathrm{P}(1000|\theta) + \mathrm{P}(1100|\theta) + \mathrm{P}(1010|\theta) + \mathrm{P}(1001|\theta) + \mathrm{P}(1110|\theta) + \mathrm{P}(1011|\theta) + \mathrm{P}(1101|\theta) + \mathrm{P}(1111|\theta)$, where $\mathrm{P}(1000|\theta)$ is $\mathrm{P}(\mathrm{Y}_{vN1} = 1, \mathrm{Y}_{vW1} = 0, \mathrm{Y}_{vR1} = 0, \mathrm{Y}_{vG1} = 0|\theta)$, and so forth.

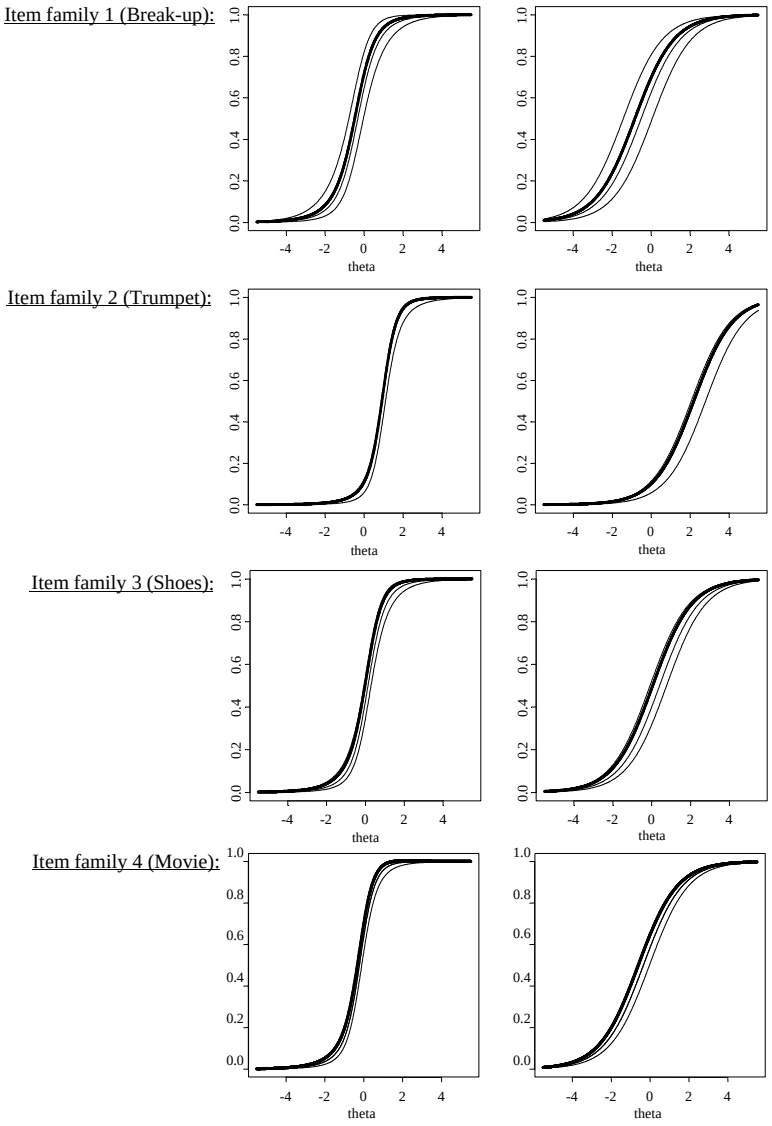
Comparing the ICCs in both panels, one can see that the ICCs for the item family structure are steeper than the ICCs for the Rasch model. Note that the steeper ICCs do not follow from a different θ scale in both models. The reason for the steeper ICCs is that all interactions are positive. For items that show no interaction, the ICCs are actually equally steep. Positive interaction has a positive effect on the steepness of the ICCs of the interacting items, and negative interaction has a negative effect (Tuerlinckx & De Boeck, 2001; Yen, 1993). The item family structure allows that the slopes differ depending on the situation, although the item weights are equal over all items. It is shown by Tuerlinckx and De Boeck (2001) that one can approach dependencies quite well with a model without dependencies but with differing item weights, as in the two-parameter logistic model (2PLM), although the ICCs deviate slightly from the logistic form. This may explain why, in an analysis of the data in this study, item weights were needed if no LID parameters were included. However, not taking LID into account leads to biased parameter estimates for the discrimination parameters and the item parameters (Thissen et al., 1989; Tuerlinckx & De Boeck, 1999; Yen, 1993). More important, although such a model may have a good fit, it cannot reveal the theoretical interitem dependencies that need to be studied.

It may be expected that as a consequence of estimating a model with LID parameters, the variance of the person parameter is reduced. From Figure 7, it is clear that the scale of the person parameters shrinks when moving from the Rasch model to the item family structure. However, the relative position of the persons remains the same: the correlation between the person parameters as estimated with the Rasch model and the person parameters as estimated with the item family structure is .99.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Figure 6

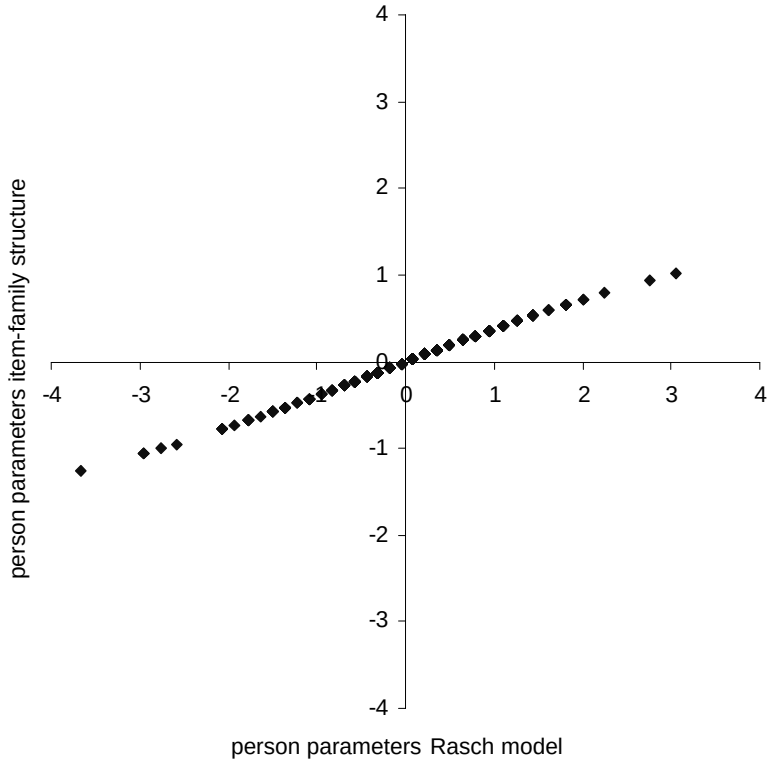
Item Characteristic Curves (ICCs) for the Items Per Item Family for the Item Family Structure (Left Column) and the Main-Effects Models (the Rasch Model) (Right Column)



As mentioned earlier, from a PCA, one dominant principal component was obtained. The loadings of the items on this component were all positive and varied from .63 to .27 for the nondichotomized data and from .58 to .15 for the dichotomized data. On the basis of a scree test, one could choose either a 1-component solution or a 10-components solution (see eigenvalues in Table 2). After a varimax rotation of the 10 components, the components can be interpreted as situation components, as all items of one item family load primarily on one and the same principal component. The highest cross-loadings were equal to .26 and .25 for the nondichotomized and the dichotomized

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Figure 7
Person Parameters of the Rasch Model Versus Person
Parameters of the Item Family Structure (Empirical Bayes Estimates)



data, respectively, whereas the corresponding loadings of the items on their situational component varied from .86 to .62 and from .84 to .50, respectively. A large majority of the cross-loadings was smaller than .1.

As in the item family structure, the dominant unrotated component can be interpreted as a "guilt dimension" on which all items have positive loadings. The situation components correspond to the LID within item families. A correlation of .60 was found between the LID parameter within each item family and the average loading of the four corresponding dichotomized items on the corresponding situation component (after a varimax rotation). Although a PCA on binary items is problematic, a moderately good approximation of the selected LID structure was obtained. Although the item family structure translates easily into a PCA structure, this is not necessarily the case, for example, because the number of PCA components would be higher or there would be too much overlap between the PCA components. It is an empirical result that the PCA could have told one what the kind of structure was, but this is not by definition so.

Discussion and Conclusions

Using a unidimensional model complemented with patterns of LID, it was found that two theory-based kinds of structures were clearly better in capturing the dependencies: a cluster structure with appraisals separated from action tendencies and an item family structure with pairwise dependencies

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

between all items that share the same situations. The item family structure was the best choice of the two. The cluster structure only had a slightly higher AIC value, and it did equally well in explaining the interitem correlations overall but failed in explaining the correlations within the same situation. In a similar way, the control model for the item family structure, the model with an equal LID parameter for all item families, is inferior to the item family structure because this structure also explains less well the correlations within the same situation.

The item family structure implies that sensitivity to guilt cannot be perfectly generalized over situations because of the situational specificity implied in the structure. It seems important to understand not only the abstract notion of guilt but also its specific situational appearances. The fine-grained IRT approach with LID models combines the global and the specific. The item family structure presents an opportunity to model situational specificity while still being based on a general latent trait. The LID model captures the specifics without detracting from the global view. The multidimensional equivalent of unidimensionality complemented with LIDs would be a structure with a large number of dimensions, one for the general latent trait and also as many as there are situations (for the item family structure), in addition to having even more for the other dependency patterns. Such multidimensional models were not tested for two reasons: First, the DIMTEST analysis showed that there is only one dominant dimension present in the data. Second and more important, there does not exist a multidimensional alternative for every structure investigated in this study, so the comparison with multidimensional IRT models could only be partial and would detract from the generality of the approach. However, when all local dependency structures can be easily formulated in terms of specific random effects (dimensions), the multidimensional approach using a Bayesian approach would be preferred.

Testing LID models, as done in this study, is a way of unraveling a concept into its underlying processes while allowing for situational specificity. It is also a way of investigating the internal validity of the questionnaire. Different psychological theories seem plausible for the data gathered here. If the responses are in agreement with one of these theories, two types of conclusions can be drawn. First, the theory is supported by the data, so that insight is gained in guilt phenomena. Second, evidence for the internal validity of the test is found, as the responses are in agreement with a psychological theory. In this case, the conclusion is that the structure of guilt is situation specific to some extent.

However, it cannot be excluded that the structure found in this study stems from response consistency within situations; in that case, LID modeling is also an efficient way of dealing with the phenomenon in question, without expanding the explicit dimensionality of the model. Especially for applications with many situations, LID modeling is a way to keep the number of person parameters low.

The approach followed in this study is similar to a structural equation modeling (SEM) (Du Toit, Du Toit, Jöreskog, & Sörbom, 1999; Everitt & Dunn, 1991) approach with correlated error terms for what were called the LIDs here. However, classical SEM requires aggregates of items for a successful modeling (Marsh & O'Neill, 1984), so that a less microscopic view will be obtained. Using SEM for binary data, one can model the mean structure and the covariance structure of the data (Muthén & Muthén, 1998-2001). Therefore, tetrachoric correlations are used, which are based on a normal distribution underlying the binary response. As such, they correspond to normal ogive IRT models that are similar to logistic IRT models used here.

The item family structure is closely related to the "testlet model" of Bradlow et al. (1999) and to the bifactor model of Gibbons and Hedeker (1992). These models are attractive alternatives for item sets with a clearly clustered structure, as it turned out to be the case in the results presented here. However, when the dependency pattern is more complicated or when a rather simple structure has to be compared with more complicated ones, as in this study, it may be appealing to choose an

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

approach that does not increase the number of random effects. The approach followed in this study corresponds to what is called the conditional approach in the statistical literature as an alternative for the random-effects approach and the marginal approach (Diggle, Heagerty, Liang, & Zeger, 2002; Fahrmeir & Tutz, 2001).

In the kind of LID models used here, any pattern of interitem dependencies can be specified, as well as types of patterns that deviate from the type that can be explained by common underlying sources, unless one would want to define such a source for every interdependent pair. It turned out that for the data in this study, the pattern that was supported from the data is one that can be explained from a common underlying source for each situation, but this is an empirical finding and not a necessity.

The price to pay using LID models is that the item parameters are more difficult to interpret. According to the Rasch model, an item parameter corresponds to the value on the person parameter scale in which a person has a .5 probability of answering the item correctly. However, due to the LIDs, this interpretation is not valid any more because the probability also depends on the responses to other items. Because of this interpretational difference, one cannot compare item parameters of both models. However, one can compare the ICCs of both models as in Figure 6. From this comparison, it may be concluded that all ICCs of the item family structure are steeper and closer to one another within an item family. This is due to the positive dependence between items referring to the same situation. Second, taking the .5 probability as the location of the ICCs, for some item families, one can see a shift to the left of the ICCs of the item family structure in comparison to the Rasch model (Item Families 2 and 10), whereas for some other item families, one can see a shift to the right (Item Families 5, 6, and 7). Situations 2 and 10 are the weakest when it comes to inducing guilt (column 8, Table 6), whereas Situations 5, 6, and 7 are among the strongest. The effect is less clear in the other strong situation (Item Family 1). This shift of the ICCs means that the guilt-inducing power of a situation is overestimated by the Rasch model for strong guilt situations and underestimated for weak guilt situations.

To conclude, LID models can be used independent of the dimensionality (number of person parameters), in that LID parameters can always be added. LID modeling is an easy way of restricting the number of person parameters or random effects, taking into account dependencies beyond those from latent traits that are explicitly incorporated in the model. Two important advantages of the LID approach are its flexibility in the formulation of all kinds of theory-based dependencies and the easy way of estimating these dependencies. For the domain of emotion research, the LID approach can help to clarify the fine-grained structure of emotions and how they are related to appraisals and action tendencies.

Appendix

The 10 descriptions selected in the first study are listed as follows (translated from Dutch to English):

1. You have been dating for some time a person you are not really in love with. When you break up, you find out that he or she was in love with you (and was taking the relationship very seriously). The breakup hurts him or her considerably. (Breakup)
2. You have been a member of a brass band for some years now. As a result, you learned to play trumpet for free. Now that you're skilled enough, you leave the band because you don't like the members of the band any more. (Trumpet)
3. During the holidays, you are working as a salesperson in a clothing and shoe store. One day, a mother with four children enters the store. One of the kids wants Samson shoes (Samson

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

is a popular doll featured in a Belgian TV series for children). The mother leaves the child with you while she goes to look for clothes for the other children. The child tries on different types and sizes of shoes, but after a while the child gets tired of fitting the shoes and refuses to continue. She picks a pair she has not tried on before, and you sell this pair to the mother afterwards. The next day, the mother wants to return the shoes because they do not fit. Your boss takes back the shoes and reimburses the mother. The shoes have been worn, however, and they are dirty. Because of this, they cannot be sold anymore. Your boss says that it doesn't matter and that everyone is capable of mistaking the size of shoes. (Shoes)

4. A not so close friend asks you if you want to join him or her to go to the movies. You tell him or her that you don't feel like it and want to spend a quiet evening at home. That evening you do go out with a closer friend. (Movie)
5. During a discussion, you make a stinging remark toward one of your friends. You notice that it hurts him or her, but you pretend not to see it. (Discussion)
6. A friend tells you something in confidence and adds that he or she would not like you to spread it around. Later, you do tell it to someone else. (Secret)
7. You are a member of a youth movement. One day the group leaders hang a rope between two trees, so you can glide from one tree to another. Jokingly, some other members make the stop of the pulley unclear. You see them doing it, but you do not help them. The following member, who wants to glide to the other tree, did not see that the stop was made unclear. You do not warn him or her. Halfway, the person falls from the rope, and he passes out. (Youth movement)
8. You have a pen pal. You get bored of writing with him or her, and suddenly, you stop corresponding with the person. After a year and a half, he or she writes you again and again, but you do not respond. (Pen pal)
9. You borrowed a jacket from a friend to wear when you go out. At the party, you leave the jacket on a chair. When you are about to leave, you notice the jacket has disappeared. In all probability, it has been stolen. (Jacket)
10. One evening, you do not feel like doing your homework. The following day, you copy the assignment of a friend who clearly has gone through a lot of trouble finishing it. You get a good grade for your assignment, the same grade as your friend. (Homework)

References

Akaike, H. (1977). On entropy maximization principle. In P. R. Krisnaiah (Ed.), *Applications of statistics* (pp. 27-41). Amsterdam: North Holland.

Andrich, D. (1985). A latent-trait model for items with response dependencies: Implications for test construction and analysis. In S. E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics* (pp. 245-275). New York: Academic Press.

Baker, F. B. (1992). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.

Barrett, K. C. (1995). A functionalist approach to shame and guilt. In J. P. Tangney & K. W. Fischer (Eds.), *Self conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 25-63). New York: Guilford.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin, 115*, 243-267.

Baumeister, R. F., Stillwell, A. M., & Heatherton, T. F. (1995). Interpersonal aspects of guilt: Evidence from narrative studies. In J. P. Tangney & K. W. Fischer (Eds.), *Self conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 255-273). New York: Guilford.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Beguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*, 541-562.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, *46*, 443-459.

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

Caprara, G. V., Barbaranelli, C., Pastorelli, C. Cermak, & Rosza, S. (2001). Facing guilt: Role of negative affectivity, need for reparation, and fear of punishment in leading to prosocial behaviour and aggression. *European Journal of Personality*, *15*, 219-237.

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

Costa, P. T., & McCrae, R. R. (1987). Validation of the five factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81-90.

Costa, P. T., & McCrae, R. R. (1989). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. *Multivariate Behavioral Research, 24*, 107-124.

Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*, 21-50.

Costa, P. T., McCrae, R. R., & Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences*, *12*, 887-898.

Dempster, A. P., Laird, N., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1-38.

Diggle, P. J., Heagerty, P. J., Liang, K., & Zeger, S. L. (2002). *Analysis of longitudinal data*. Oxford, UK: Oxford University Press.

Du Toit, S., Du Toit, M., Jöreskog, K. G., & Sörbom, D. (1999). *Interactive LISREL: User's guide*. Chicago: Scientific Software International.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapmann & Hall.

Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion*, *12*, 271-302.

Embretson (Whitely), S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, *45*, 479-494.

Embretson (Whitely), S. E. (1981). Measuring aptitude processes with multicomponent latent trait models. *Journal of Educational Measurement*, *18*, 67-84.

Embretson (Whitely), S. E. (1984). A general latent trait model for response processes. *Psychometrika*, *49*, 175-186.

Everitt, B. S., & Dunn, G. (1991). Covariance structure models. In B. S. Everitt & G. Dunn (Eds.), *Applied multivariate data analysis* (pp. 257-275). London: Edward Arnold.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York: Springer Verlag.

Fitness, J., & Fletcher, G. J. O. (1993). Love, hate, anger and jealousy in close relationships: A prototype and cognitive appraisal analysis. *Journal of Personality and Social Psychology*, *65*, 942-958.

Frijda, N. H. (1986). *The emotions*. Cambridge, UK: Cambridge University Press.

Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition and Emotion*, *7*, 357-387.

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*, 212-228.

Frijda, N. H., & Zeelenberg, M. (2001). Appraisal: What is the dependent? In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 141-155). Oxford, UK: Oxford University Press.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*, 423-436.

Gilbert, P., Pehl, J., & Allan, S. (1994). The phenomenology of shame and guilt: An empirical investigation. *British Journal of Medical Psychology*, *67*, 23-36.

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 IRT models for polytomous items. *Applied Psychological Measurement*, *25*, 19-37.

Izard, C. E. (1978). Guilt, conscience and morality. In C. E. Izard (Ed.), *Human emotions* (pp. 421-452). New York: Plenum.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and non cognitive processes. *Psychological Review*, *100*, 68-90.

Jannarone, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika*, *51*, 357-373.

Kelderman, H. (1984). Loglinear Rasch models. *Psychometrika*, *49*, 223-245.

Kelderman, H., & Steen, R. (1993). *LOGIMO* [Computer program]. Groningen, The Netherlands: PROgramma.

Kempf, W. F. (1977). Dynamic models for the measurement of "traits" in social behavior. In W. F. Kempf & B. H. Repp (Eds.), *Mathematical models for social psychology* (pp. 14-58). Bern, Switzerland: Hans Huber.

Lindsay-Hartz, J., De Riviera, J., & Mascolo, M. F. (1995). Differentiating guilt and shame and their effects on motivation. In J. P. Tangney & K. W. Fischer (Eds.), *Self conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 174-300). New York: Guilford.

Marsh, H. W., & O'Neill, R. (1984). Self description questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. *Journal of Educational Measurement*, *21*, 153-174.

McCrae, R. R., & Costa, T. P. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*, 509-516.

McCrae, R. R., & Costa, T. P. (1999). A five factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139-153). New York: Guilford.

Michel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics and invariance in personality structure. *Psychological Review*, *102*, 246-268.

Muthén, L., & Muthén, B. (1998-2001). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

Nandakumar, R., & Stout, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics*, *18*, 41-68.

Omdahl, B. L. (1995). Cognitive appraisal theories of emotions. In B. L. Omdahl (Ed.), *Cognitive appraisal, emotion, and empathy* (pp. 37-90). Englewood Cliffs, NJ: Lawrence Erlbaum.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. New York: Cambridge University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about emotions? *Psychological Review*, *97*, 315-331.

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

Reisenzein, R., & Hofmann, T. (1993). Discriminating emotions from appraisal-relevant situational information: Baseline data for structural model of cognitive appraisals. *Cognition and Emotion*, *7*, 271-293.

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

Roseman, I. J., Antoniou, A. A., & Jose, P. E. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and Emotion*, *10*, 241-277.

Roseman, I. J., & Smith, C. A. (2001). Appraisal theory: Overview, assumptions, varieties, controversies. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 3-19). Oxford, UK: Oxford University Press.

SAS Institute. (1999a). *SAS OnlineDocTM version 8*. Cary, NC: Author.

SAS Institute. (1999b). *SAS System V8 for Windows*. Cary, NC: Author.

Scherer, K. R. (1993). Studying the emotion-antecedent appraisal process: An expert system approach. *Cognition and Emotion*, *7*, 325-355.

Scherer, K. R. (1997). Profiles of emotion antecedent appraisal: Tesing theoretical predictions across cultures. *Cognition and Emotion*, *11*, 113-150.

Scherer, K. R. (1999). On the sequential nature of appraisal processes: Indirect evidence from a recognition task. *Cognition and Emotion*, *13*, 763-793.

Segall, D. O. (2001). General ability measurement: An application of multidimensional item response theory. *Psychometrika*, *66*, 79-97.

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion*, *7*, 233-269.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, *52*, 589-617.

Stout, W., Douglas, J., Junker, B., & Roussos, L. (1993). *DIMTEST manual*. Urbana-Champaign, IL: William Stout Institute for Measurement.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Stout, W., Nandakumar, R., Junker, B., Chang, H., & Steidinger, D. (1992). DIMTEST: A Fortran program for assessing dimensionality of binary item responses. *Applied Psychological Measurement*, *16*, 236.

Tangney, J. P. (1995). Shame and guilt in interpersonal relationships. In J. P. Tangney & K. W. Fischer (Eds.), *Self conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 114-139). New York: Guilford.

Thissen, D. (1988). *Multilog* [Computer program]. Mooresville, IN: Scientific Software.

Thissen, D., & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin*, *104*, 385-395.

Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement*, *26*, 247-260.

Tuerlinckx, F., & De Boeck, P. (1999). Distinguishing constant and dimension-dependent interaction: A simulation study. *Applied Psychological Measurement*, *23*, 299-307.

Tuerlinckx, F., & De Boeck, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters. *Psychological Methods*, *6*, 181-195.

Tuerlinckx, F., & De Boeck, P. (2002). Measuring needs with the thematic apperception test: A psychometric study. *Journal of Personality and Social Pychology*, *82*, 448-461.

Wicker, F. W., Payne, G. C., & Morgan, R. D. (1983). Participant descriptions of guilt and shame. *Motivation and Emotion*, *7*, 25-39.

Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika*, *60*, 181-198.

Wolfinger, R. (1999). *Fitting nonlinear models with the new NLMIXED procedure* (SUGI 24 Conference Proceedings, Paper 287). Cary, NC: SAS Institute, Inc.

Wu, M., Adams, R., & Wilson, M. (1997). *Conquest* [Computer program]. Melbourne: Australian Council for Educational Research.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*, 187-213.

Acknowledgment

The research is supported by a GOA 2000/2 grant from the K. U. Leuven.

Author's Address

Address correspondence to Dirk J. M. Smits, K. U. Leuven, Department of Psychology (H.C.I.V.), Tiensestraat 102, B-3000 Leuven, Belgium; e-mail: Dirk.Smits@psy.kuleuven.ac.be.

Downloaded from http://apm.sagepub.com at Lmog Ku Leuven Campusbiblio on April 11, 2008
© 2003 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.