Multivariate Behavioral Research, *38* (2), 161-188
Copyright © 2003, Lawrence Erlbaum Associates, Inc.

A Componential IRT Model for Guilt

Dirk J. M. Smits and Paul De Boeck
K. U. Leuven, Department of Psychology

Componential IRT models are often used to investigate the process structure of a cognitive task in terms of its components. Although not yet used for emotions, these models are also helpful to investigate the process structure of emotions. We investigated the process structure of guilt with the OPLM-MIRID, a particular type of componential IRT model. Based on a first exploratory study, we selected 3 components that can be considered partial guilt responses. To test this structure, we collected 10 descriptions of guilt-inducing situations. For each situation, it was asked how guilty one would feel, and also three componential questions were presented, one for each of the three selected components. The inventory was administered to 270 high school students, 130 males and 140 females, all between 17 and 19 years old. The data were analyzed with the OPLM-MIRID. All 3 components were found to contribute (α =.05) to the global guilt response: norm violation, worrying about what one did, and a tendency to restitute. The same kind of modeling seems appropriate to investigate the structure of other emotions, and more in general to validate inventories with a componential design.

Many psychological concepts turn out to be multifaceted and complex – in other words, they are multi-componential. This seems true for intelligence as well as for personality. A common way of proceeding with such concepts is to unravel their structure with some form of multidimensional analysis. The most popular form is factor analysis, in which different factors are supposed to reflect the components of a concept. A recent example is the decomposition of the concepts associated with the Big Five into facets (Costa & McCrae, 1995, 1997; Costa, McCrae, & Dye, 1991). For instance, when a factor analysis is performed on the Conscientiousness items of the NEO-PI-R, six factors are found: Competence, Order, Dutifulness, Achievement Striving, Self-Discipline, and Deliberation (Costa & McCrae, 1989, 1997). Another example is the factor analysis of the Beck Depression Inventory-II. Two dimensions were found in self-reported depression:

I thank René Butter for his help with MIRID. We also thank the schools and the students for kindly participating to this study.

Correspondence concerning this article should be addressed to Dirk J. M. Smits, K. U. Leuven, Psychologisch Instituut, Dep. of Psychology (H.C.I.V.), Tiensestraat 102, B-3000 Leuven, Belgium. Ph: 003216/326133, Fax: 003216/325916, e-mail: Dirk.Smits@psy.kuleuven.ac.be

cognitive-affective symptoms, and somatic symptoms (Beck, Steer, & Brown, 1996; Osman et al., 1997).

Components derived on the basis of factor analysis of items for a given trait can be understood as ways in which individuals differ in how they display the underlying trait. Different dimensions refer to (a) different kinds of behavior or (b) the same kind of behavior (or behavior outcome) in different types of situations (e.g., Ortony & Turner, 1990). For example, when intelligence is studied, the dimensions refer to success (behavior outcome) in performing different kinds of tasks (different situations). We will call components that are derived from a multidimensional analysis *dimensional components*.

A different way of discerning components in a concept or a phenomenon is an analysis in terms of the processes that lead to a behavior or a behavior outcome. This approach has become quite popular for intelligence. A typical example is the componential approach to intelligence as elaborated by Sternberg (1977a, 1977b, 1978) for response times. He has developed cognitive-process models to predict the response time in intelligence tests from the time needed for the processes – also called components – underlying the response. For example, for analogy items, the processes or components he distinguished were encoding, inference, mapping, and application. Embretson (1980, 1984) has developed similar models for the probability of a correct response to the total task, with that probability being predicted from several componential probabilities each corresponding to a different process. The process outcomes are observed as subtask outcomes, each referring to one process. For example, when solving an analogy item (*father*: *mother*: *uncle*: ?), one first has to find a rule, an activity called rule generation (generation component), and one also has to evaluate the response generated from the rule (evaluation component). The corresponding first subtask would be "find the rule for …," with as a correct response being "the feminine of …"; the corresponding second subtask would be "the feminine of uncle is (a) sister, (b) aunt, (c) niece, and (d) cousin." Common features of this kind of componential approach are that intermediate responses are discerned from final responses and that the characteristics (response time, probability) of the final responses are explained from the corresponding characteristics of the intermediate responses. The intermediate processes are parts of solving the total cognitive task: the time taken for the final response is seen as the sum of the times taken for the intermediate processes (Sternberg, 1977b), and in Embretson's (1980, 1984) case, the probability of success at the total task is seen as the product of the success probabilities for the intermediate processes. Components that are derived from a process analysis will be called *process components*.

The process approach and the dimensional approach can be seen as complementary in two respects. First, the dimensional approach, using factor analysis or principal components analysis, can help to delineate domains of behaviors that then can be analyzed further with a process approach. For example, Inductive Reasoning is a first-order intelligence factor; consequently, one may use a process approach to study how inductive reasoning works (see, e.g., Sternberg, 1977b). Second, different processes may each refer to a different dimension, each process being a source of individual differences, and therefore each may show up as a factor in factor analyses. Although the dimensional approach by definition needs multidimensionality to find different components, the process approach does not. It makes perfect sense to analyze the processes behind a one-dimensional phenomenon, with the individual differences in the underlying processes all being based on one common trait, with the processes having a distinct nature and being distinguishable empirically – for example, because their response times add up to an observed response time that functions as a dependent variable.

Most applications of the process approach have been in the cognitive domain. It makes sense, however, to consider the same approach for emotion and personality. For example, there exist inventories for assessing guilt-proneness using items on feeling guilty (e.g., Ferguson & Crowley, 1997; Harder, 1995). On the other hand, emotions such as guilt have been conceptually analyzed from a process-componential approach (e.g., Gilbert, Pehl, & Allan, 1994; Wicker, Payne, & Morgan, 1983). The aim of our study was to investigate the process-componential structure of guilt and to demonstrate the applicability of the process approach to emotions. Our approach was very similar to the subtask methodology just explained. The participants of the study were given situations about which they had to answer several questions: the first kind of questions concerned components that may contribute to feeling guilty (the equivalents of subtasks), and the second kind concerned feeling guilty itself (the equivalent of the total task). The set of situations with the associated questions about feeling guilty and its components can be seen as a guilt-process inventory. Testing the componential structure is a way of examining the internal validity of the inventory. The type of internal validity we have in mind is what Embretson (1983) described as *construct representation*, referring to the validity of the model for the phenomena under study.

First, the process components of guilt will be described, subsequently we will explain the model, and why it makes sense to use it for guilt proneness. The kinds of components we will discuss are appraisals, covert reactions, and action tendencies, but not physiological reactions, primarily because of

methodological reasons: it is difficult to investigate physiological reactions with the self-report method we will use.

Components of Guilt

Emotions have been analyzed in terms of components, such as appraisals of the situation, or *emotivations* – a term coined by Roseman, Wiest, and Swartz (1994) to denote "emotional motives or things wanted while having a feeling" (see, e.g., Johnson-Laird & Oatley, 1989; Omdahl, 1995; Roseman, Antoniou & Jose, 1996; Roseman & Smith, 2001; Smith & Ellsworth, 1985; Smith & Lazarus, 1993) – covert reactions, and action tendencies (see, e.g., Frijda, Kuipers, & ter Schuure, 1989; Izard, 1993; Roseman, Wiest, & Swartz, 1994). From a review of the literature, five components of guilt were derived:

1. Guilt implies an *appraisal in terms of responsibility*. For example, Izard (1978) stated that guilt only appears in situations for which one feels personally responsible (see also Baumeister, Stillwell, & Heatherton, 1994, 1995; Frijda, 1986; Lindsay-Hartz, 1984; Lindsay-Hartz, DeRiviera, & Mascolo 1995; Smith & Lazarus, 1993; Wicker et al., 1983).

2. Guilt implies an *appraisal in terms of norm violation*. For example, Lindsay-Hartz (1984; Lindsay-Hartz et al., 1995) stated that a violation of a norm or the moral order precedes guilt. According to Lewis (1987), being able to make moral judgments is a precondition of guilt feelings (see also Ausubel, 1955; Barrett, 1995). Opinions differ as to how broad the notion of "norm" should be understood. Norms can be moral, religious, cultural, or personal. Most authors prefer a broad definition of norms when it comes to explaining the basis of guilt (Baumeister et al., 1995; Izard, 1978; Johnson-Laird & Oatley, 1989; Jones & Kugler, 1993; Jones, Kugler, & Adams, 1995; Tangney, 1995; Wicker et al., 1983).

3. Guilt implies a negative self-evaluation as a covert reaction of the type "I did something bad." The *negative self-evaluation relates to an act* and is not a definite disapproval of the entire self: "... we feel like a bad person, yet we know that while we did a bad thing, we are not really bad" (Lindsay-Hartz et al., 1995, p. 288; see also Ausubel, 1955; Barrett, 1995; Baumeister et al., 1994; Frijda, 1986; Gilbert et al., 1994; Lindsay-Hartz, 1984; Tangney, 1995; Wicker et al., 1983).

4. While feeling guilty, one's attention and inner thoughts are *covert, ruminative worrying reactions focused on the act* much more than on the self. Baumeister et al. (1994) defined guilt as related to a particular act, with a consequence that the act stays in one's mind while one reflects upon it (see also Barrett, 1995; Gilbert et al., 1994; Izard, 1978; Tangney, 1995).

5. Guilt implies the *emotivations and action tendencies related to restitution*. One is inclined to confess, to undo one's fault, to do after all what one ought to do and did not do, to apologize, to compensate, etcetera (Barrett, 1995; Baumeister et al., 1994, 1995; Lindsay-Hartz, 1984; Lindsay-Hartz et al., 1995; Tangney, 1995). For example, Baumeister et al. (1994) and Lindsay-Hartz (1984) stated that the motivation to make reparations or at least to apologize is part of feeling guilty.

We make an a priori distinction between objective appraisal components (1 and 2) and the other three components (3, 4, and 5). The first two components are directly related to the situation, as an interpretation of the situation. In the case of guilt, the situation is partly an event in which one is involved as an actual or potential actor. We assume that the first two components are objective features of the situation – so to speak, parts of the situation (Lindsay-Hartz, 1984; Lindsay-Hartz et al., 1995; Wicker et al., 1983), so that the appraisal is induced by the situation and not so much by a subjective appreciation. This assumption was checked in a first small, exploratory study, preceding the main study. Note that we do not claim at all that appraisals in general are objective and induced primarily by the situation, and not even that these two are always objective, but for the situations we studied, we believe they were.

The other three components can be understood as subjective reactions to the interpretation based on the first two components: "If a norm is violated, and if I am responsible for it, then the evaluation of what I did is negative, I keep thinking of what I did, and I am inclined to restitute what I did wrong." We do not insist on a strict distinction between subjective appraisals and emotivations and action tendencies. The exact labeling is less important than the role they play as guilt processes. Further, we do not claim to say anything about the order in time between the components. They all can play a role in guilt feelings at different moments. The only order that seems plausible is that the three subjective reactions (Components 3, 4 and 5) play their role later in time then the two objective interpretation or appraisal components (Components 1 and 2). The reason is that the first two components are considered features of the situation.

Although we have identified Component 3 as a reaction and not as an interpretation of the situation, we are not sure it can be differentiated empirically from Component 2. That in guilt-inducing situations a negative self-evaluation follows upon a personal norm violation seems highly plausible and is often documented in the literature (Barrett, 1995; Lindsay-Hartz, 1984; Lindsay-Hartz et al., 1995; Tangney, 1995; Wicker et al., 1983). Therefore, the differentiation between Component 2 and Component 3 was

also explored in the *exploratory study*. Their intercorrelation may not be too high in order to distinguish between both.

In the *main study*, we considered the Components 3, 4, and 5 the three subjective components. They may be considered part of guilt as partial guilt responses. The two appraisal components were left out of consideration in this second study as not being part of feeling guilty but as having an effect on guilt through the three partial responses. However, if Component 3 cannot be differentiated from Component 2 within the 10 situations we used, Component 2 is to be preferred, because it is more often cited in the literature as a guilt component (Baumeister et al, 1995; Izard, 1978; Johnson-Laird & Oatley, 1989; Jones & Kugler, 1993; Jones et al., 1995). Note that if these two components cannot be differentiated, then they are also of the same type (objective interpretations or subjective reactions to these interpretations).

It will be described now how we assume the more response-like components (3, 4 and 5) contribute to the global feeling of guilt. Basically, they are considered as part of the guilt response. What this means more specifically is explained now in three more specific assumptions.

First, for the partial responses as well as for the total response, the situations are assumed to have a *guilt-inducing power* based on the appraisals. The guilt-inducing power of a situation is seen as a compound that is built up from component-specific guilt-inducing powers. The hypothesis is that a situation makes one feels guiltier the more it makes one feel bad about what one did (or failed to do), the more one keeps thinking of the act (or absence of it), and the more one wants to restitute what one did wrong.

Second, the componential responses and the guilt response are conceived as stemming from the inductive situational power transgressing a *person's threshold*. The threshold represents the sensitivity of the person. In principle, a person's threshold can differ depending on the kind of response: three componential responses and one guilt response.

Third, it is assumed that all four kinds of responses (the three componential responses and the guilt response) show individual differences that are based upon one and the same kind of sensitivity, called guilt-proneness. The *unidimensionality* of guilt is not an essential assumption but a provisional assumption. Only if this assumption is rejected will a multidimensional view be taken. To represent individual differences, *persons* are thought of as having guilt thresholds. These thresholds can differ from person to person, and hence they reflect guilt-proneness. The lower the threshold, the higher is the guilt-proneness.

An inventory was constructed with a set of situations differing in degree of guilt induction. In our main study, four kinds of questions were presented to participants: (a) whether a negative self-evaluation would occur

(supposing that Component 3 can be differentiated from Component 2), (b) whether one would keep thinking of what one did, (c) whether one would be inclined to restitute what one did wrong, and (d) whether one would feel guilty. Normally, componential responses are covert, but in order to study their relation with guilt feelings, there is no other way then to make them overt, in this case by asking for self-reports. This view on guilt feelings was formalized into a psychometric model in order to test the assumptions explicitly.

The Model

The model we used to analyze the data and to test the componential structure of guilt is an adaptation of the MIRID (Model with Internal Restrictions on Item Difficulty) (Butter, De Boeck, & Verhelst, 1998). The adaptation was described in Butter (1994) and is very similar to the original model. The model has the following ingredients, which will be explained here for the case of guilt:

1. It is assumed that the *probability for a person to feel guilty in a situation* depends on the person's threshold for feeling guilty and on the situation's guilt-inducing power. The probability of feeling guilty is assumed to be a function of the difference between both, with chances being higher than 1 out of 2 if the guilt-inducing power exceeds the guilt threshold. More specifically, the logit of the probability of feeling guilty is modeled as follows:

$$\text{logit}\,(P_{ij}) = a_j\,(\beta_j - \theta_i) \tag{1}$$

with P_{ij} indicating the probability of person i ($i = 1, \ldots, I$) feeling guilty in situation j ($j = 1, \ldots, J$), and $\text{logit}\,(P_{ij}) = \ln[P_{ij} / (1 - P_{ij})]$; with σ_k denoting the guilt threshold of person i, called the *person parameter*; with β_j denoting the guilt-inducing power of situation j, called the *item parameter*; and with a_j being a weight for situation j, called the *discrimination value,* in order to allow situational differences in how much the guilt probability depends on the difference between β_j and θ_i. The a_j are not called parameters, because they were not estimated in the strict sense, see further.

2. It is assumed that the *total situational guilt-inducing power* is a weighted sum of contributions from different components, which can be expressed as a linear function:

$$\beta_j = \sum_{k=1}^{K} \sigma_k \beta_{jk} + \tau \tag{2}$$

with σ_k denoting the weight of the contribution of the component of type k

($k = 1, ..., K$); with β_{jk} denoting the kth type of contribution from situation j; and with τ denoting an additive scaling parameter. As a result, the β_j are no longer basic parameters as in Equation 1 of the model, but instead the β_{jk} and the σ_k are basic parameters.

3. The *probability of a componential response* is considered a function of the common underlying threshold θ_i and of the component-specific situational guilt-inducing power, represented in the parameter β_{jk}:

$$\text{logit}\,(P_{ijk}) = a_{jk}(\beta_{jk} - \theta_i) \tag{3}$$

with P_{ijk} indicating the probability of person i to show componential response k to situation j; with a_{jk} indicating the weight of the difference between the person threshold (θ_i) and the guilt-inducing power of situation j with respect to component k (β_{jk}). Note that the β_{jk} are situational parameters associated with the components. Consequently, the components involved in the model (and not just Component 1 and Component 2) may also be thought of as located in the situation and not in the person. This is true for the guilt-inducing powers, but not for the componential responses themselves. The componential responses are not situational but personal in two ways. First, the responses depend also on the θ_i, the personal threshold, so that individual differences will be displayed in the same situation. Second, the responses also depend on the a_{jk}. A lower discrimination weight is equivalent to a larger random part in the responses. In our design, the random part per component is not distinguishable from the person-by-situation interaction for the component in question. Hence, the lower the a_{jk} for a component, the more the componential responses depend on the particular pairing of a person and a situation.

Identification

Like in the Rasch model, the scale of the item parameters and the person parameter is identified only up to an additive constant: if another origin were chosen such that $\theta_i^* = \theta_i + c$, this could be compensated for by rescaling the β-parameters (β_j and β_{jk}). The β_{jk} have to be transformed into $\beta_{jk}^* = \beta_{jk} + c$, and the same holds for the β_j, $\beta_j^* = \beta_j + c$. Butter et al. (1998) described the consequences this has for the additive scaling parameter and for the component weights. The value of the new additive scaling parameter, τ^*, follows from

$$\beta_j^* = \sum_{k=1}^{K} \sigma_k \beta_{jk} + \tau + c = \sum_{k=1}^{K} \sigma_k (\beta_{jk}^*) + \tau^* \tag{4}$$

Solving this equation for τ^*, using the fact that $\beta_{jk}^* = \beta_{jk} + c$, gives :

$$\tau^* = \tau + c\left(1 - \sum_{k=1}^{K} \sigma_k\right). \quad (5)$$

Given that linear weights are invariant under translations of the scale (see also Butter et al., 1998), the component weights σ_k are identified and remain invariant under the scale transformations just described. To make the parameters identifiable, an identifiability constraint is needed. Although it seems to follow from Equation 5 that the identifiability constraint may be imposed on τ, this would not always solve the problem: if the sum of the weights σ_k is equal to 1, then the parameter τ becomes invariant under scale transformations, resulting in a special case of the model which is not identified in terms of its difficulties (β_j and β_{jk}) (Butter et al., 1998). Therefore, an identification constraint is imposed on the β_{jk}: the mean β_{jk} is fixed to zero.

The Discrimination Values in the OPLM and the OPLM-MIRID

The model just presented is the MIRID (Model with Internal Restrictions on Item Difficulties) as described by Butter et al. (1998), except for the a_j and the a_{jk} (discrimination values). The original MIRID has no discrimination values. With the a_j and the a_{jk}, the model is called the OPLM-MIRID (Butter, 1994).

The name 'OPLM-MIRID' can be clarified by explaining the OPLM (One Parameter Logistic Model; Verhelst & Glas, 1995). This model differs from the Rasch model only in that it allows for different but fixed discrimination values. Hence, in the OPLM and in the OPLM-MIRID, these discrimination values are integer constants instead of parameters to be estimated. They are fixed a priori to the estimation of the difficulty parameters (β), so that the items have only one parameter. Therefore, like for the original MIRID, a CML formulation (conditional maximum-likelihood; Molenaar, 1995) is possible, and basically the same estimation equations can be followed as described by Butter et al. (1998).

Following a CML approach (Fischer, 1977, 1983, 1995; Molenaar, 1995) means that the parameters are estimated by conditioning on the sum of the discrimination values of items the person agrees upon. In that way, given the sum-score, the probability of a specified response pattern can be estimated independently of the value of the person parameter (the guilt-proneness of

a person). The CML approach provides an excellent basis for statistical testing (Fischer, 1977, 1983, 1995).

To determine the values for the a_j and the a_{jk}, the same approach as implemented in the OPLM-program (Verhelst, Glas, & Verstralen, 1994) was followed. Based on a heuristic, the OPLM-program can suggest a reasonable set of discrimination values. This heuristic is based upon a least-squares procedure, under the assumption that the person parameters of respondents with the same unweighted sum score is approximately the same. Finally, the estimates of the discrimination indices are transformed into positive integers with a maximal range of 1 to 15. Given a user-specified geometric mean, OPSUG, a module in the OPLM program, suggests a set of discrimination values. The higher the chosen geometric mean, the larger the differences between the discrimination values of the items can be, and also the higher is the risk of capitalization on error. Therefore, the manual of the OPLM program (Verhelst et al., 1994) suggests to start with a value not higher than 3.

MIRID and LLTM

MIRID differs from the Linear-Logistic Test Model (LLTM; Fischer, 1973, 1983, 1996). The structure of the LLTM for the guilt-inducing power regarding the final response, given in Equation 6, is identical to the structure of Equation 2, and also, in the LLTM, item parameters are modeled as linear contributions, but so-called complexity factor values (q_{jk}) take the role of the β_{jk} in Equation 2. These complexity factor values are *a priori given*, and hence are constants, whereas the β_{jk} are to be estimated parameters. The matrix **Q** describes the situations a priori in terms of the components. The LLTM for the guilt-inducing powers regarding the final response is given in Equation 4:

$$\beta_j = \sum_{k=1}^{K} \sigma_k q_{jk} + \tau \tag{6}$$

For all values of j and for Component x, if $k = x$, $q_{jk} \neq 0$ and if $k \neq x$, $q_{jk} = 0$. In a similar way, formula Equation 6 applies to the componential guilt-inducing powers as well.

In the LLTM, the so-called componential items are not treated differently from the so-called total items, except for a difference in their **q**-vector: for componential items, for only one value of k, $q_{jk} \neq 0$, whereas for the other values of k, $q_{jk} = 0$. In contrast with the LLTM, the MIRID does not require any a priori knowledge about the *size* of the componential contributions q_{jk} (componential guilt-inducing powers), as in contrast to the

q_{jk}-values, the β_{jk}-values are parameters to be estimated. In other words, MIRID estimates the LLTM **Q**-matrix. The weights of the contribution (σ_k) are parameters in both models, and as such, they are to be estimated.

Applied to our guilt questionnaire, the LLTM has a serious drawback in that exact knowledge about the components is needed. For example, we do not know to which degree brooding (our fourth component) is elicited by each of the situations (size of componential contributions q_{jk}). The main advantage of the MIRID is that these values can be estimated from the data. In emotion research, this situation is common, because researchers do not have the required quantitative information about the components of emotions. Often the knowledge of researchers is restricted to qualitative knowledge, meaning that they have an idea about the components but do not know anything about the size of their contributions to the emotion or about the degree to which these components are elicited by situations. The size of the contribution can be estimated with the LLTM as the basic parameters, but the components themselves cannot.

Estimation and Testing

The parameters of the OPLM can be estimated with several IRT programs like the previously mentioned OPLM program of Verhelst, et al. (1994). For the estimation of parameters of the OPLM-MIRID, we used the MIRID program of Butter (1994). Given that this program has problems running on the current generation of computers, we also wrote a new Windows-oriented program (see Smits & De Boeck, 2002; Smits, De Boeck, Verhelst, & Butter, 2001) for estimating the item parameters under a CML formulation of the model via an iterative Newton-Raphson procedure (Gill, Murray, & Wright, 1981), which is similar to the approach Butter (1994) used in his program. Both programs lead to exactly the same results. When preferring an MML (Marginal Maximum-Likelihood) formulation of the models, one can estimate the parameters of the OPLM and the OPLM-MIRID with the PROC NLMIXED procedure of SAS V8 (1999), as explained in Rijmen, Tuerlinckx, De Boeck, and Kuppens (in press) and in Smits and De Boeck (2002). Using these approaches, it was found that for this application, all programs lead to similar results. Using PROC NLMIXED, a probit link function can be specified instead of a logit link function, so that the normal ogive version of the MIRID (NO-MIRID) can be estimated. After multiplying the component-item parameters and the additive scaling constant by 1.7, similar results are again obtained for all item parameters.

The check on the chosen a priori discrimination values is a goodness-of-fit test of the model taking the suggested a priori discrimination values into

account. The OPLM-MIRID was tested as follows: first, an OPLM was fitted to the data, and, because the OPLM-MIRID is a restriction of the OPLM, the fit of the OPLM-MIRID was then compared with the fit of the OPLM using a likelihood-ratio test (as explained in Butter et al., 1998).

The fit of the OPLM was determined with two statistics: the R1*c*-statistic and the DIMTEST procedure (Stout, 1987). Like the Martin-Löf test, the R1*c*-statistic was developed to have power against violations of monotone increasing item response functions (parallel if the discrimination values are equal, and proportional to the discrimination values if they differ a priori). It is a combination of Pearson chi-square tests of observed versus expected frequencies on the item level, and it has an asymptotic χ^2-distribution (Glas, 1988, 1989; Glas & Verhelst, 1995). As Glas (1981, 1988) has shown, this statistic is equivalent to the Martin-Löf *T*-statistic, but unlike the Martin-Löf *T*-statistic, the R1*c*-statistic fits the framework of the generalized Pearson statistics. Because the R1*c*-statistic is not very sensitive for violations of unidimensionality (Van den Wollenberg, 1982), we also used the DIMTEST procedure (Stout, 1987; Stout, Douglas, Junker, & Roussos, 1993; Stout, Nandakumar, Junker, Chang, & Steidinger, 1992) to test for unidimensionality. Only if the DIMTEST analysis did not reject the hypothesis of unidimensionality *and* if the R1*c*-statistic did not reject the OPLM with the chosen discrimination values, did we conclude that the OPLM showed a reasonable fit.

Hypotheses

In terms of our MIRID approach, there are two hypotheses at the basis of the study:

1. The same dimension θ_i – guilt-proneness – underlies all four kinds of responses: the three componential responses and the final response.

2. The guilt-inducing power (the item parameters for guilt) can be explained as a linear function of the guilt-inducing powers for the componential guilt responses (the item parameters for the components), as implied by the MIRID.

If the first hypothesis holds – that is, if the responses are unidimensional – then the total score of the inventory, including the responses to all four kinds of questions (weighted with the discrimination values a_{jk} or a_j), can be used as a measure of situational guilt-proneness. If the first hypothesis does not hold, then we will use a multidimensional variant of the MIRID (Butter, 1994), with one dimension for each kind of questions. If the second hypothesis holds (independently of the first) – that is, if the guilt-inducing powers can be decomposed into componential contributions – then the internal validity of the inventory will have been shown, given that the

responses would then agree with a theory of the basis for situational guilt feelings. An interesting feature of the psychometric modeling method followed is that it combines theory and measurement, given that the measurement model is also a process theory of guilt. Hence, the psychometric model is not only a tool for measuring and investigating the internal validity of an inventory, but also, more importantly, it is a formalization of a psychological theory.

In the following, first the collection of the guilt-inducing situations is described. The situations are needed to construct an inventory. Second, the exploratory study is described. It was set up to explore some properties of the Components 1 and 2, and to find out whether the Components 2 and 3 can be differentiated empirically. Third, the main study is described, which is set up to test the componential theory of guilt using the OPLM-MIRID. The guilt-inducing situations are used in both studies.

Collecting the Situations

Method

To collect the situations, a group of young people (17 to 19 years old) was asked to describe situations about which they had felt guilty. Because we wanted to make use of the situations in the inventories, it was important that their descriptions be as clear as possible. The descriptions should give a vivid and explicit description of what happened.

We used an open format with two parts: a common part and a specific part, referring to one of three different kinds of situations: (a) work or study situation, (b) personal relationships, and (c) leisure time.

The common part of the instructions was as follows: "We would like you to describe a situation you felt guilty about. To help you, we give you a set of questions, which can guide you describing the situations as completely as possible:

1. What happened?
2. Were other people involved?
3. Why did you feel guilty?
4. What were you thinking?
5. Were you thinking about yourself or about what you did?
6. When did you start feeling guilty?
7. How long did you feel guilty?
8. Why, in your opinion, did the feeling disappear, and did you do something to let the feeling disappear?"

The three specific questions were:

1. "Could you describe a situation related to your work or studies?"
2. "Could you describe a situation related to your personal relationships?"
3. "Could you describe a situation related to your free time?"

Under each of these three questions a half page was left to describe the requested situation.

Participants

Forty-six 17- to 19-year-old, high school students, 20 males and 26 females, were each given the task to describe three situations, one of each type.

Selection of the Situations

Ten stories were selected using the following six criteria: (a) understandability, (b) equal representation of each type of situation, (c) variation in content, (d) variation in assumed guilt-inducing power, (e) conformity with the environment of 18-year-old persons, and (f) equal representation of stories stemming from males of females.

In order to use the descriptions in our study, all information about responses from the person in the situation was deleted and only the information about the situation was retained. We asked for information about the reactions (see various questions) to make the task more natural for the respondents and to check whether and to what degree guilt was evoked. We subsequently omitted descriptions of reactions in order to avoid suggesting how other respondents would respond to the situation when used in an item in the inventory. The ten selected stories are listed in the Appendix, together with a keyword for each situation. Hereafter, we will use these keywords to refer to the situations.

Exploratory Study

This small study was set up to check whether Components 1 and 2 were mainly situational and whether Components 2 and 3 could be differentiated. We asked 12 judges to judge each of the 10 situations with respect to the three components on a two-point scale (1 = present, 0 = absent): responsibility (component 1), norm violation (component 2), and negative self-evaluation (component 3).

In order for an appraisal component to be almost exclusively situational instead of being also a person-dependent reaction to the situation, all persons should agree on their appraisal of the situations. This means that the intraclass correlation coefficient ICC(2, *k*) (Shrout & Fleiss, 1979), or the inter-judge reliability, needs to be very high over situations, and the absolute agreement must be very high as well.

To test the differentiation of Component 2 and 3, the correlation over situations was derived. If this correlation is close to 1, the two components are not sufficiently differentiated for the purpose of constructing a componential inventory.

The results concerning the situational character of Component 1 are as follows: The intraclass correlation coefficient for the first component was very high (.91), and the absolute agreement was also high (for six situations, the agreement between the judges was higher than 90%). Therefore, for our set of situations, responsibility can be considered a rather objective appraisal, primarily based on the situation descriptions. Consequently, responsibility will not be included as a component in the main study.

The Components 2 and 3 were less stable, meaning that the participants did not agree as much as for Component 1. The intraclass correlation coefficient was .64 for Component 2 and .73 for Component 3. The absolute agreement between the judges was also not very high (only in two situations was the percentage of agreement higher than 90%). This result is contrary to what we expected for Component 2. Because this component must be considered a personal reaction, it is a candidate for inclusion in the next study.

It turns out that Component 2 and Component 3 can hardly be differentiated. The correlation between the two components was as high as .98 ($p < 0.001$) using the means over persons. Therefore, in the main study, Component 3 will be omitted and replaced by Component 2. As mentioned earlier, this is because Component 2 is more often cited in the literature (Baumeister et al., 1995; Izard, 1978; Johnson-Laird & Oatley, 1989; Jones & Kugler, 1993; Jones et al., 1995).

Main Study

The structure of guilt was examined using an inventory based on the ten previously selected situations, each followed by four questions: one for each of the three components (2, 4 and 5) and one checking whether a respondent would feel guilty in the situation. The questions related to the components are called the *componential items*, and the guilt question is called the *composite item*.

It is common to use situation descriptions in an inventory (e.g., Ferguson & Crowley, 1997). It is also common to use componential questions along with a question about the (final) response under consideration. Componential questions are used in studies about the components of emotions (e.g., Frijda, Kuipers, & terSchure, 1989; Wicker et al., 1983), although not in an assessment context.

The inventory was administered to a group of persons and the data were analyzed with the OPLM-MIRID. The analysis allows us to test the two previously mentioned hypotheses: (a) unidimensionality over components, and (b) guilt-inducing power as a weighted sum of componential guilt-inducing powers.

Method

The instructions for the questionnaire were as follows:

"Please read the following stories and try to imagine you were the one this was happening to. After each story the following four questions are presented.

1. Do you feel like you violated a moral, an ethic, a religious and/or a personal code?
2. Do you worry about what you did or failed to do?
3. Do you want to do something to restitute what you did or failed to do?
4. Do you feel guilty about what you did or failed to do?

Answer by circling the number of your choice. '0' means 'no', '1' means 'not likely', '2' means 'likely', and '3' means 'yes'. Answer all questions as well as you can. The right answer is how you would feel in the given situation."

Subsequently the ten situations were presented, each followed by the four questions just mentioned.

Participants

The inventory was administered to a group of 270 students between 17 and 19 years old: 140 females and 130 males from three high schools. Each school distributed the inventories to students. The students were given some time during the day to fill in the inventories individually. Within 2 weeks, 268 completed inventories were returned (138 from females and 130 from males).

Descriptive Statistics

In Table 1, the mean and standard deviation are given for each item. As could be expected, the means of the items varied mainly between situations rather than between components. The intraclass correlation coefficients, ICC(3, k) (Shrout & Fleiss, 1979), between the components and over the situations, and the analogous coefficient between the situations and over the components were respectively equal to .92 and .20. This means that clearly more variance was associated with the kind of situation than with the kind of component.

Modeling

Because the MIRID (and the OPLM-MIRID) was formulated for binary data, the data were coded for yes ("0" and "1") and no ("2" and "3"). This was a natural dichotomization, given that "1" meant "not likely" and "2" meant "likely." The data were fitted using the MIRID program of Butter (1994; see also Butter et al., 1998). This program follows a CML approach to parameter estimation.

The unidimensional OPLM-MIRID can fit the data only if the OPLM (Verhelst & Glas, 1995; Verhelst et al., 1994) fits. The same discrimination values apply as for the OPLM, but the composite item parameters are restricted to be a linear combination of the componential item parameters. As mentioned

Table 1
Means and Standard Deviatons before Dichotomization

Situation	Component 2	Component 4	Component 5	Guilt item
Love affair	1.474 (1.040)	2.026 (.897)	1.769 (.943)	1.888 (.980)
Trumpet	.429 (.713)	.649 (.800)	.612 (.820)	.571 (.783)
Shoes	1.078 (1.004)	1.437 (.979)	1.265 (.987)	1.433 (1.035)
Film	1.478 (1.040)	1.649 (.966)	1.586 (1.030)	1.757 (.962)
Discussion	2.045 (.927)	2.239 (.832)	2.384 (.782)	2.265 (.794)
Confidence	2.500 (.757)	2.392 (.788)	2.369 (.853)	2.593 (.689)
Youth movement	2.097 (1.023)	2.328 (.864)	1.948 (1.076)	2.284 (.933)
Correspondent	1.366 (.983)	1.336 (.891)	1.470 (1.047)	1.437 (.940)
Jack	1.493 (1.166)	2.302 (.874)	2.634 (.682)	2.149 (1.013)
Homework	.843 (.997)	.791 (.941)	.910 (1.067)	.791 (1.010)

above, discrimination values are suggested prior to estimation of the model parameters. We chose a value of 2 for the geometric mean of the discrimination values, so that variation in the discrimination values was rather low.

The discrimination values of the OPLM are given in Table 2. It can be seen in Table 2 that the discrimination values varied primarily between the situations rather than between the components. When intraclass correlation coefficients ICC(3, k) (Shrout & Fleiss, 1979) were calculated between the components and over the situations, and the analogous coefficient between the situations and over the components, they turn out to be .62 and -.09. Some situations were more discriminative than others with respect to the guilt-proneness of the participants, whereas the different components and the composite items were about equally discriminative.

The OPLM was not rejected: the log-likelihood value was -4098.745 and the R1*c*-statistic was 117.105 (p = .48, df = 39). The assumption of unidimensionality was tested independently using the DIMTEST procedure (Stout, 1987; Stout et al., 1993; Stout et al., 1992). For calculating the DIMTEST statistics, 85.07% of the examinees were included. The value for the T-statistic, measuring unidimensionality, was .663 (p = .254), and the value for the more powerful T'-statistic, based on refinements of the DIMTEST-procedure (Nandakumar & Stout, 1993), was .814 (p = .208), meaning that our data can be considered unidimensional. A second indication of the unidimensionality of the components, and hence, of the

Table 2
Discrimination Values of the Component Items and the Composite Items in the OPLM and the OPLM-MIRID

Situation	Component 2	Component 4	Component 5	Guilt item
Love affair	1	2	2	2
Trumpet	2	1	1	1
Shoes	1	2	2	2
Film	3	3	2	2
Discussion	2	2	2	2
Confidence	2	2	2	2
Youth movement	3	3	3	3
Correspondent	3	2	2	3
Jack	1	2	2	1
Homework	3	3	3	3

inventory is the rather high Cronbach's alpha of 0.87. This is a high value, given that it is based on four rather different questions related to only 10 situations.

To check whether the OPLM-MIRID restrictions were reasonable for these data, a likelihood-ratio test was used to compare the fit of the OPLM-MIRID with the fit of the OPLM. Given that the OPLM-MIRID had a log-likelihood value of -4102.875, the value of the likelihood-ratio test statistic G was -2 [-4102.875 - (-4098.745)] = 8.26, whereas the critical $\chi^2(6)$ value was 12.6 ($\alpha = .05$). As a result, the OPLM-MIRID was not rejected and could be considered a reasonable model for the data.

The values for the item parameters of the componential items and their estimated standard errors are given in Table 3. A higher value within the same column means a higher componential contribution to the guilt-inducing power. Over columns, the componential weights also help to determine the contribution, given that each of the componential parameter values is multiplied by its corresponding weight.

For example, the componential contributions for copying another person's homework were all three lower than those for not preventing an accident during the activities of a youth movement group. By consequence, the young people in our sample were less likely to feel guilty for copying homework then for not having prevented an accident.

Table 3
Item Parameters of the Component Items (and their Standard Errors) and Reconstructed Item Parameters of the Composite Items as Estimataed with the OPLM-MIRID

Situation	Component 2	Component 4	Component 5	Composite item
Love affair	-.245 (.118)	.507 (.075)	.089 (.067)	.184
Trumpet	-1.536 (.115)	-1.971 (.157)	-2.062 (.172)	-2.242
Shoes	-.745 (.130)	-.060 (.062)	-.321 (.067)	-.396
Film	-.122 (.048)	.009 (.047)	.053 (.066)	-.091
Discussion	.563 (.078)	.775 (.081)	1.094 (.104)	.842
Confidence	1.272 (.116)	1.006 (.095)	.830 (.089)	1.073
Youth movement	.352 (.055)	.604 (.061)	.170 (.051)	.412
Correspondent	-.193 (.049)	-.279 (.058)	-.108 (.065)	-.327
Jack	-.077 (.123)	.814 (.087)	1.461 (.133)	.819
Homework	-.623 (.055)	-.718 (.053)	-.541 (.053)	-.821

The estimated weights or the importance of the components for the guilt-inducing power are given in Table 4, together with an estimation of their standard errors. The standard errors indicate that all components had a weight that differed significantly from zero. Component 4 (worrying about what one did or failed to do) seemed to be the component with the largest contribution.

The item parameters of the guilt items can be reconstructed from the estimated componential parameters of the OPLM-MIRID (in Table 3 and 4). For each situation, they are the weighted sum of the values for the three components (Table 3), using the weights of the components and the additive scaling parameter (Table 4). For example, the reconstructed item parameter (guilt-inducing power) of the first situation for the guilt item amounts to (.245 * -.245) + (.591 * .507) + (.300 * .089) - .082 = .184. All reconstructed item parameters (guilt-inducing power) are given in the last column of Table 4. Note that the sum of the componential weights is only slightly larger than 1.00, namely 1.136, and the intercept is nearly zero (-.082), so that the linear function approached a weighted average.

Comparing the reconstructed values of the guilt item parameters (10 in total), as estimated by the OPLM-MIRID, with the parameters of the guilt items, as estimated by the OPLM without the MIRID restrictions, yielded a correlation of .99.

We also fitted the 2PL model and its combination with the MIRID. For the component item parameters as well as for the item weights (discrimination values or parameters), the correspondence between the sets of parameters was very high. For the 2PL-MIRID and the OPLM-MIRID they were .99 for the component item parameters and .85 for the item weights (whereas for the OPLM, the weights were necessarily integer values). Most importantly, the estimates of the component weights were nearly the same in both approaches (.245, .591, and .300 for the OPLM-

Table 4
Weights of the Components and their Standard Errors, Estimated with the OPLM-MIRID

Basic Parameter	Value	SE-estimation
Component 2 (norm violation)	.245	.107
Component 4 (worrying)	.591	.118
Component 5 (restitution)	.300	.118
Additive scaling parameter	-.082	.033

MIRID, and .247, .545, and .315 for the 2PL-MIRID). Finally, the fit of the 2PL model was not significantly better than the fit of the OPLM: the value of the likelihood-ratio test statistic G was -2 [-5221.5 - (-5201)] = 41, and the fit of the 2PL-MIRID was not significantly better than the fit of the OPLM-MIRID: the value of the likelihood-ratio test statistic G was -2 [-5225.5 – (-5203.5)] = 44. The critical $\chi^2(39)$ value for both likelihood-ratio test statistics is 54.6 ($\alpha = .05$). Note that the log-likelihood values for the OPLM and the OPLM-MIRID (-5201,-5203.5) are lower than the previously mentioned log-likelihood values for the same models, as for the comparison with the 2PL models, we had to switch from the CML-framework to the MML-framework, which implies restrictions on the person distribution.

Discussion

Given the results in Table 4, Components 2, 4, and 5 must be seen as important components. They are not only important, but in this study they are sufficient to explain the guilt feelings. Together they give a full explanation of situational guilt feelings as measured with this questionnaire.

Comparing this structure with the evidence in the literature, we came to the following conclusions.

1. In our study, Component 1 is not needed to explain guilt, given that three other components are sufficient to explain the data. Component 1 probably has an effect only through the other components. There is no room left for responsibility to have an independent effect beyond norm violation, worrying, or an inclination to restitute. This view is shared by the authors who stated that Component 1 is part of the guilt-inducing situation and not part of the guilt feeling itself (Lindsay-Hartz, 1984; Lindsay-Hartz et al., 1995; Wicker et al., 1983).

2. The fact that the Components 2 and 3 could not be differentiated in the exploratory study contrasts with the view of Lindsay-Hartz (1984) and Lindsay-Hartz et al. (1995). In these articles, a very clear-cut distinction was made between the two components. In our study, there is a nearly perfect correlation between these components over situations, so that it did not make sense to distinguish between them. Perhaps our set of situations is suboptimal for differentiating these two components, given that it is limited to only 10 situations.

3. Component 4 is by far the most important component. This result is consistent with the results of Ferguson and Crowley (1997). These authors made the distinction between ruminative guilt and non-ruminative guilt. Ruminative guilt, as measured by the Test of Self-Conscious Affect–Modified [TOSCA-M] – an adaptation of the TOSCA-scale to which a

ruminative guilt response was added (Tangney et al., 1989) – was found to be related to other measures of guilt-proneness, such as the Personal Feelings Questionnaire – 2 [PFQ-2] and the Guilt Inventory [GI], whereas the non-ruminative guilt did not highly load on the latent construct of guilt-proneness. The importance of Component 4 is supported by this study.

4. We found no support in the literature for the configuration of exactly the three components of guilt we retained. The findings of Tangney (1995) and Izard (1978) are most closely related to those of our study. According to Tangney (1995), guilt comprises the Components 2 (norm violation), 3 (negative self evaluation, which is in our study enclosed through Component 2), 4 (covert actions with focus on the act), and 5 (emotivations and action tendencies related to the tendency to restitute). The difference with Izard (1978) is that in her view Component 1 is included. In sum, the view of Tangney (1995) is very much in line with our findings, although we could not differentiate between norm violation and negative self-evaluation (Components 2 and 3). The difference with Izard's (1978) view is also minor, but from our results we would consider feeling responsible a factor with an indirect effect on guilt through the other three components instead of having a direct effect on situational guilt feelings.

The components discussed here are seen as contributions to the situational guilt-inducing power. On the person's side, there is only one dimension responsible for the responses. In other words, the person's thresholds for the componential responses are the same as the thresholds for guilt responses. No evidence was found for multidimensionality. Our model is one in which different kinds of componential proneness are the same as guilt-proneness. This model supports the idea that the components are partial guilt responses rather than responses external to guilt but that effect guilt.

As can be seen in Table 3, the contribution of the components varies from situation to situation: for example, the *discussion* situation (hurting a friend in a discussion) and the *jack* situation (a jack you borrowed is stolen) have about the same guilt-inducing power (.842 and .819 respectively), but in the *discussion* situation the feeling of norm violation (Component 2) is stronger than in the *jack* situation, whereas the reverse is true for the inclination to restitute (Component 5). Feeling guilty in the *discussion* situation has more to do with the feeling that one has violated a norm than feeling guilty in the *jack* situation, whereas feeling guilty in the *jack* situation has more to do with wanting to restitute.

In modeling emotion data, the person parameter can be conceptualized as a latent trait as the basis for individual differences among persons. The value of each person on the latent trait can be estimated and then correlated

with other related measures – for example, to validate a questionnaire. An advantage of our approach is that, on the one hand, situational aspects and their structure can be studied, and, on the other hand, a latent trait estimate can be obtained. Thus, the current approach can be useful for different research questions, from research about the (situational) process structure of emotions to research about the relations between different latent traits, to questions about the validity of questionnaires and of the theory behind them.

We will look now at the modeling approach we have followed. An important question is whether there are any advantages to using this kind of formal model. In our opinion, there are at least three.

1. Using the MIRID (or OPLM-MIRID), it is possible explicitly to test the process structure of an emotion. By using such a model, one can look very closely at the underlying processes of feeling guilty, and one can determine which components are necessary and sufficient to explain these feelings. This reason is valid only as far as the self-report method can be trusted. For guilt, however, it is hard to find an alternative method.

2. In the approach we followed, inventory construction, hypothesis testing about the nature of an emotion, and measurement go hand in hand. Testing the model is testing a hypothesis about guilt, based on a particular design of the inventory, and testing the substantive model is also testing a measurement model.

3. MIRID offers a rich kind of information: in terms of regression analysis, one can state that not only the weights (σ_k) of the predictors can be estimated, but also the predictor values (β_{jk}). The latter are very useful in situations in which only a limited amount of information is available about the predictors – for example, just their presence or absence – whereas nothing is known about their degree or exact value. Further, a componential analysis is possible independent of whether the components are unidimensional or multidimensional, because the componential analysis is concentrated on the item parameters.

An important condition for the model to be relevant is that a *test design* (Embretson, 1985) is followed, with situational scenarios and componential questions for each scenario. Most inventories do not have such a systematic structure. A potential drawback of this type of design is its transparency. We can only trust that the transparency does not prevent the participants from responding honestly.

In this study we used only 10 situations to test our componential theory of guilt and to construct an inventory. One should realize that our sample of situations is small, as is the range of ages of persons in our study (17 to 19 years). These limitations as to the number of situations and the age of the participants prevent us from making strong claims for generalization about

the process structure of guilt. Nevertheless, the approach we have taken is a new and apparently promising way for examining and testing the process structure of emotions, combining an individual-differences approach and a measurement model with testing a general theory about the components of an emotion.

The componential approach we followed, complemented with a formal model, is one that can be used for other emotions as well. Whether a model with a unidimensional structure, a situational emotion-inducing power, and a personal threshold would also apply to other emotions may be a fruitful empirical question for future research to answer.

References

Ausubel, D. P. (1955). Relationships between shame and guilt. *Psychological Review, 62*, 378-390.

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 Press.

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 Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.

Butter, R. (1994). *Item response models with internal restrictions on item difficulty*. Unpublished doctoral dissertation, K. U. Leuven, Belgium.

Butter, R., De Boeck, P., & Verhelst, N. D. (1998). An item response model with internal restrictions on item difficulty. *Psychometrika, 63*, 1-17.

Costa, P. T., Jr. & 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., Jr. & 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., Jr. & McCrae, R. R. (1997). Conceptions and correlates of openness to experience. In R. Hogan & J. A. Johnson (Eds.), *Handbook of personality psychology* (pp. 825-847). San Diego: Academic Press.

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.

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

Embretson (Whitely), S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179-197.

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

Embretson, S. E. (Ed.) (1985). *Test design: Developments in psychology and psychometrics*. London: Academic Press.

Ferguson T. J. & Crowley S. L. (1997). Measure for measure: A multitrait-multimethod analysis of guilt and shame. *Journal of Personality Assessment, 69*, 425-441.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fischer, G. H. (1977). Linear logistic latent trait models: Theory and applications. In H. Spada & W. F. Kempf (Eds.), *Structural models of thinking and learning* (pp. 230-225). Bern, Switzerland: Huber.

Fischer, G. H. (1983). Logistic latent trait models with linear constraits. *Psychometrika, 48,* 3-26.

Fischer, G. H. (1995). The linear logistic test model. In G. H. Fischer & I. W. Molenaar (Eds.). *Rasch models: Foundations, recent developments and applications* (pp. 131-156). New York: Springer-Verlag.

Fischer, G. H. (1996). Unidimensional linear logistic Rasch models. In W. van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 225-244). New York: Springer-Verlag.

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

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.

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

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. London: Academic Press.

Glas, C. A. W. (1981). Het Raschmodel bij data in een onvolledig design (The Rasch model for data in an incomplete design). (PSM-Progress Reports 81-1). Utrecht, The Netherlands: Vakgroep PSM van de subfaculteit psychologie.

Glas, C. A. W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika, 53,* 525-546.

Glas, C. A. W. (1989). *Contributions to estimating and testing Rasch models*. Arnhem, The Netherlands: CITO.

Glas, C. A. W. & Verhelst, N. D. (1995). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 69-96). New York: Springer-Verlag.

Harder, D. W. (1995). Shame and guilt assessment, and relationships of shame- and guilt-proneness to psychopathology. In J. P. Tangney & K. W. Fischer (Eds.), *Self conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 368-392). New York: Guilford Press.

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

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

Johnson-Laird, P. N. & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion, 3*, 81-123.

Jones, W. H. & Kugler, K. (1993). Interpersonal correlates of the guilt inventory. *Journal of Personality Assessment, 61*, 246-258.

Jones, W. H., Kugler, K., & Adams, P. (1995). You always hurt the one you love: Guilt and transgressions against relationship partners. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 301-321). New York: Guilford Press.

Lewis, H. B. (1987). Introduction: Shame – The "sleeper" in psychopathology. In H. B. Lewis (Ed.), *The role of shame in symptom formation* (pp. 1-28). Hillsdale, NJ: Erlbaum.

Lindsay-Hartz, J. (1984). Contrasting experiences of shame and guilt. *American Behavioral Scientist, 27,* 689-704.

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 Press.

Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Eds.). *Rasch models: Foundations, recent developments and applications* (pp. 39-52). New York: Springer-Verlag.

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. *Cognitive appraisal, emotion, and empathy* (pp. 37-90). Hillsdale, NJ: Erlbaum.

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

Osman, A., Downs, W. R., Barrios, F. X., Kopper , B. A., Gutierrez, P. M., & Chiros, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory-II. *Journal of Psychopathology and Behavioral Assessment, 19*, 359-376.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (in press). A nonlinear mixed model framework for IRT models. *Psychological Methods*.

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, England: Oxford University Press.

Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviours and goals differentiate discrete emotions. *Journal of Personality and Social Psychology, 67,* 206-221.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Smith, C. A. & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48,* 813-838.

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

Smits, D. J. M. & De Boeck, P. (2002). *Estimation of item parameters under the MIRID: a program and a SAS-oriented approach*. Manuscript submitted for publication.

Smits, D. J. M., De Boeck, P., Verhelst, N., & Butter, R. (2001). *The MIRID Program 1.0* [Computer program and manual]. Leuven, Belgium: K. U. Leuven, Department of Psychology.

Sternberg, R. J. (1977a). Component processes in analogical reasoning. *Psychological Review, 84,* 353-378.

Sternberg, R. J. (1977b). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Lawrence Erlbaum.

Sternberg, R. J. (1978). Isolating the components of intelligence. *Intelligence, 2,* 117-128.

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*. Champaign, IL: The William Stout Institute for Measurement.

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 Press.

Tangney, J. P., Wagner P. E., & Gramzow, R. (1989). *Test of self-conscious affect*. Fairfax, VA: George Mason University.

Van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model. *Psychometrika, 47,* 123-140.

Verhelst, N. D. & Glas, C. A. W. (1995). One parameter logistic model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 215-238). New York: Springer-Verlag.

Verhelst, N. D, Glas, C. A. W., & Verstralen, H. H. F. M. (1994). *One parameter logistic model* [Computer program and manual]. Arnhem, The Netherlands: CITO.

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

Accepted September, 2002.

Appendix

The ten descriptions we selected in the first study are listed below (translated from Dutch to English), followed by the keyword used in the article:

1. During some time you were having a love affair, but you're not really in love with him/her. Making an end at this relation, you discover he/she supposed the relation was serious. He/she is very grieved. (*love affair*).

2. A few years ago, you started playing trumpet in a brass band. The schooling you needed is completely paid by the brass band. At the moment you're a good musician, you stop playing in the brass band, because you find yourself not fitting in the group musicians. (*trumpet*).

3. To earn some money during the holidays, you worked in a shoe and clothes shop as a salesperson. One day a mother with four children enters the shop. One of the kids wants Samsonshoes (Samson is a popular doll figuring in a the Belgian TV-series for children). The mother asks you to find fitting shoes for one child while she is busy with searching clothes for the other three children. You give the child different pairs of shoes of different sizes. After a while the child is tired of trying shoes. She doesn't want to

try any new pair of shoes, but instead she chooses a pair she hasn't tried. You sell this pair of shoes to the mother when she returns without saying that the child hasn't tried to fit them. The following day, the mother returns, and wants her money back because the pair you sold her doesn't fit the child. Your boss gives the mother her money back. The shoes themselves are no longer saleable after being used one day. According to the shopkeeper it doesn't matter, and he adds that everyone can be mistaken sometimes. (*shoes*).

4. A not so close friend asks you to accompany her going to a film. You answer you don't feel like going to a film, and want to stay home spending a quiet evening. That evening you go out with a better friend. (*film*).

5. While discussing, you hurt a friend by making a negative remark about him/her. You can see it hurts him/her, but you act as if it was not noticed. (*discussion*).

6. A friend tells you something in confidence. He/she asks you not to tell it to anyone. After he/she's gone you do tell it to some else. (*confidence*).

7. You're a member of a youth movement. One day the leaders hang up a rope between two trees, so you can glide from one tree to another. Some guys make the stop of the pulley unclear. You see them busy, but you don't help them. The following guy, who wants to glide to the other tree, hasn't seen that the stop was made unclear. You don't warn him/her. Halfway he falls from the rope, and is out of consciousness. (*youth movement*).

8. You're having a friend you correspond with. Suddenly, you stop writing letters, because you don't feel like corresponding anymore. One and a half a year later he/she writes you again, and again, you don't respond to his/her letter. (*correspondent*).

9. Before going to a party, you borrow a jacket from a friend. At the party, you put the jacket on a chair, and go dancing. Wanting to leave, you discover the jacket isn't on the chair anymore. In all probability, it is stolen. (*jack*).

10. One evening, you're not in the mood to do your homework. The following day, you copy the homework of a friend who has spent a lot of time doing it. You get a good mark for this homework, as good as the friend you copied the task from. (*homework*).

Copyright of Multivariate Behavioral Research is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.