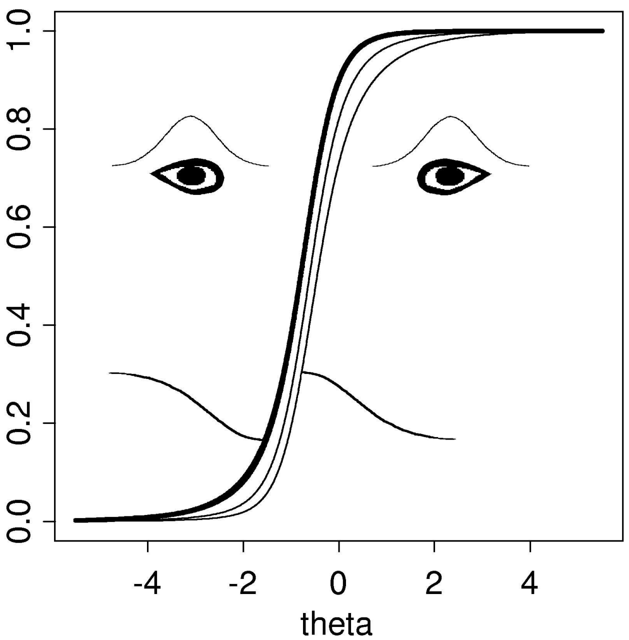
KATHOLIEKE UNIVERSITEIT LEUVEN
Faculteit Psychologie en Pedagogische Wetenschappen
Onderzoeksgroep Hogere Cognitie en Individuele Verschillen

Item response models for self-report data on emotional responses



Proefschrift aangeboden tot
het verkrijgen van de graad van
Doctor in de Psychologische Wetenschappen
door **Dirk Smits**
o.l.v. Prof. Dr. Paul De Boeck (promotor)

2003

Dirk Smits, **Item response models for self-report data on emotional reponses**. Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologische Wetenschappen, juni 2003. Promotor: Prof. dr. P. De Boeck

---

In dit proefschrift worden vier benaderingen besproken waarmee men de structuur van concepten zoals emoties kan onderzoeken in termen van componenten (appraisals en actietendensen). Deze vier benaderingen worden geïllustreerd met zelfrapporteringsdata over schuldgevoelens. Daarenboven bespreken we ook één benadering om de band tussen de gedragsmatige expressie van een emotie en de neiging daartoe te bestuderen. Deze laatste wordt geïllustreerd met zelf-rapportage data over verbaal agressief gedrag. Alle data werden verzameld aan de hand van situatie-response vragenlijsten, zodat de invloed die een situatie uitoefent op een component, een emotie, een actietendens of een gedrag steeds mee in rekening kan worden gebracht. De vijf benaderingen zijn allen gestoeld op item response theorie en bevatten parameters die specifiek zijn voor het individu en parameters die specifiek zijn voor het item (combinatie van een situatie met een component, een emotie of een gedrag). Een functie van beide soorten parameters bepaalt de kans dat men een component als aanwezig beschouwt in een situatie, een emotie ervaart in een situatie of een gedrag wil stellen of stelt in de situatie.

In Hoofdstuk 1 wordt de componentiële structuur van het schuldgevoel onderzocht. Deze structuur bestaat uit 3 componenten: de appraisal norm overtreding en de meer actie-gerichte componenten piekeren en de neiging om het goed te maken. Dit werd onderzocht aan de hand van het MIRID waarbij de itemparameter van een schulditem een gewogen som is van de itemparameters van de component-items (items die elk een specifieke component in een situatie representeren). De schatting van het MIRID en de robuustheid van de parameters t.o.v. schendingen van de normaliteitsassumptie voor de persoonsparameter vormt het onderwerp van Hoofdstuk 6. In het tweede hoofdstuk wordt het MIRID uitgebreid tot het RW-MIRID, zodat het gewicht van een component kan verschillen van persoon tot persoon. Op basis van dit model kunnen we afleiden dat er geen interindividuele verschillen zijn in het belang of gewicht van de componenten voor het schuldgevoel. In Hoofdstuk 3 onderzoeken we de relationele structuur van emoties aan de hand van modellen voor Locale Item Afhankelijkheden. De resultaten vervolledigen het beeld van het schuldgevoel als een gevoel waarvan de betekenis deels gekleurd wordt door de situatie waarin de emotie ervaren wordt.

In een vierde hoofdstuk maken we abstractie van interindividuele verschillen door de introductie van het Marginaal LLTM. Dit is een model waarbij de effecten van de item covariaten effecten zijn op het niveau van de populatie. Een belangrijk voordeel van deze benadering is dat de associaties tussen de items erg soepel kunnen gemodelleerd worden en dat de juistheid van het model voor de associaties de parameters van de gemiddelden-structuur niet noodzakelijk beïnvloedt.

In Hoofdstuk 2 wordt onderzocht of men verbaal agressief gedrag kan voorspellen op basis van de neiging om zich verbaal agressief te gedragen. Het effect van deze neiging op het gedrag verschilt van persoon tot persoon. De band tussen beide wordt in Hoofdstuk 5 verder bestudeerd aan de hand van het leermodel van Embretson (1991). Hierbij vonden we dat de neiging om zich verbaal agressief te gedragen beïnvloed wordt door factoren eigen aan de persoon en door factoren eigen aan de combinatie van een situatie en een gedrag, terwijl het inhiberen van deze actietendens voornamelijk bepaald wordt door factoren eigen aan de persoon.

We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance.

John A. Wheeler

Acknowledgments

First and foremost, I acknowledge the valuable role played by Paul De Boeck, my supervisor, whose encouragement and willing support kept me going to the end. His promptness in giving me feedback on my rough drafts facilitated the progress of the research. Also his considerations and comments about the studies were very inspiring and often provided interesting pathways to continue the research.

Second, I would like to mention the numerous co-authors of the manuscripts inside this dissertation for their willingness to share their knowledge and thoughts with me.

Third, the faculty, staff, and my colleagues in the research group Higher Cognition and Individual Differences, and Quantitative and in the research group Personality Psychology here at Department of Psychology, K.U. Leuven have together created a stimulating and exciting environment in which to work over the last four years. It has been a privilege to spend this time here. Special thanks are due to Laurence Claes, my 'room-mate', for her enjoyable and supportive companionship.

Fourth, I would like to thank the schools and the psychology students who kindly participated in our studies.

Fifth, this work was financially supported by the GOA 2000/2-grant from the K.U. Leuven: 'Psychometric models for the study of personality' and by the IAP P5/24-grant from the federal OSTC, Belgium. My sincere gratitude is due to all the taxpayers in Belgium who provide this financial support.

Finally, I owe most to my wife, Natalie, my parents, and my friends for their confidence in me and because they were always there when needed. It is to them, with much love and gratitude, that I dedicate this dissertation.

Leuven,
May, 2003

Contents

General introduction

Emotions are an important part of life, and as such they received a lot of attention in literature. Different aspects of emotions are investigated: etiology, biological basis and origin, neurological and hormonal structures involved, structure, behavioral aspects, expressions, etc. Several methods are developed to investigate these different aspects of emotions: for example, introspection, behavioral observation, self-report measures, neuro-imaging, etc. We will focus on self-report measures, and demonstrate the use of some data-analytic models and techniques, which stem from cognitive research, to investigate two aspects: the structure of emotions in terms of components and the expression of emotions in behavior. Our research is restricted to two negative emotions: guilt and anger. Four different modeling approaches will be presented to explore the componential and relational structure of emotions. All four will be illustrated with an analysis of data on guilt feelings. One approach will be presented to model the behavioral expression of emotions, illustrated with an analysis of data on verbal aggression.

In the first part of this introduction, First, the substantive results on emotions are described, whereas in the second part the modeling perspective will be discussed.

1. Research on emotions

1.1 Componential theories of emotions

According to theories about the componential structure of emotions, emotions can be characterized and differentiated on the basis of their association with components. The componential approach to emotion is perhaps most clearly articulated within appraisal theories of emotion. This theory states that emotions are based on a process of appraisal or evaluation of the circumstances in relation to the organism's own goals and needs (e.g., Ellsworth & Smith, 1988; Kuppens, Van Mechelen, Smits, & De Boeck, in press; Ortony, Clore, & Collins, 1988; Omdahl, 1995; Reisenzein & Hofmann, 1993; Roseman, 1984; Scherer, 1993, 1997, 1999; Smith & Ellsworth, 1985). The main idea is that each emotion is associated with a distinct pattern of appraisals. Following this line of thought, it has been argued that action tendencies should be considered as central emotion components as well, in that emotions can also be characterized in terms of their association with specific action tendencies (e.g., Fischer, 1991; Frijda, 1986; Frijda, Kuipers, & Schure, 1989; Lazarus, 1991; Oatley & Jenkins, 1996; Skiffington, Fernandez, & McFarland, 1998). In sum, every emotion is assumed to be characterized by a distinct pattern of appraisals and action tendencies. Both are summarized by the term *components* of emotions. Other aspects, like for example, physiological changes or experiences, bodily feedback, and many others can be important as well (e.g., Berkowitz, 1990; Izard, 1993), but are not our primary point of interest, as they are less accessible to self-report.

Appraisal theories link emotions explicitly to the situations in which they are experienced. For example, a situation may be experienced as threatening, and could therefore lead to anxiety, and a situation in which one believes to have violated a norm may lead to feelings of guilt. Using a situational approach with appraisals allows for explaining emotional responses to the situations and depending on the specific approach that is followed can shed light on the interaction of persons with situations.

1.1.1 Componential structure of emotions: the case of guilt

In the Chapter 1, an approach is presented to investigate the componential structure of concepts like emotions. A necessary condition that there are situational differences in how much the situations induce the components and the emotions (differences other than main effect differences, i.e. there may not be a perfect correlation over situations between the components). Further, it is assumed that there are individual differences in the proneness to the components and the emotions but this is not a necessary condition for the approach to apply. The approach is applied to a data set on situational guilt feelings. These data stem from my masters' thesis and are used further in this thesis to illustrate and investigate various modeling approaches.

We will first describe the work from the masters' thesis. Based on a survey of literature a componential theory about situational guilt feelings was constructed. Guilt was assumed to be based on three appraisals and two more action-oriented components. The three appraisals are responsibility, norm violation, negative self-evaluation and the action-oriented components are worrying (covert action) and the tendency to rectify (overt action) what one did or failed to do. In a first study, situations were collected in which persons reported guilt feelings. Ten situations were selected using criteria like understandability and variation in content. In a first exploratory study, the component of responsibility was found to be rated fairly high in the selected set of situations without much individual differences. Therefore, responsibility was considered a rather objective appraisal, primarily based on the situation, and not dependent on the person. Because we expected individual differences in proneness to guilt, we gave priority to subjective appraisals. Therefore, responsibility was not included as a component in the main study. Note that we do not argue that responsibility is not a component of guilt feelings. It can be a component without individual differences, or the ten situations may not be optimal to reveal such differences. Second, the correlation over situations between the ratings of norm violation and negative self-evaluation was very high (.98), so that they cannot be used both to be related with guilt feelings. It would be impossible to differentiate between the situational contributions of these two components. Therefore, norm violation, which was more often mentioned in literature as an important appraisal, was retained, whereas negative self-evaluation was omitted from the main study. The remaining three components are norm violation (an appraisal), worrying, and tendency to rectify (action tendencies).

In the main study based on a reanalysis for this thesis (the same is true for all what follows), the following results were found: (1) The individual differences in the components and guilt feelings can be described with one underlying dimension: the sensitivity of a person to guilt components and to guilt itself. (2) The three components were elicited to a different degree by the different situations: Some situations favor certain components whereas other situations favor other components. (3) The three components were sufficient to predict situational guilt feelings. (4) Finally, worrying was found to be the most important component, followed by the tendency to rectify, and norm violation. worrying and the tendency to rectify are more action-oriented (covert and overt, respectively), whereas norm violation is a pure (actionless) appraisal. This result suggests that the feeling of guilt is more than just an appraisal, which is in line with the view of Frijda et al. (1989).

In the previous analysis, it was assumed that the importance of a component for the emotion is fixed, the same for all persons. All individual differences are assumed to stem from the position of a person on the latent trait (one general latent trait was sufficient, component-specific latent traits were not needed). However, as explained in Chapter 2, the importance of certain components may differ from person to person, which leads to person-by-situation interaction. Because situations differ as to the components they elicit and people differ in how important these components are for the feeling, the interaction follows. For example, for some people norm violation may more important, whereas for other people the tendency to rectify is more important. This means that the basis for guilt (or call it the meaning of guilt) is different depending on the person, so that people will also differ with respect to the kind of situation they feel guilty about (given that situations differ on the components). As we had no a priori hypothesis about the specific components that would be person dependent, we tested them all three. The results indicate that the hypothesis of no such individual differences could not be rejected. Since no substantive individual differences are found in the components beyond the general latent trait (not in the component levels and not in their weights), it was tentatively concluded that guilt feelings can be explained by one general underlying guilt proneness. However, this conclusion will be modified in Chapter 3 on the relational structure of emotions.

1.1.2 Relational structure of emotions

In Chapter 3, an approach is presented to investigate which relations exist between the components and between the components and the emotion. The approach is again exemplified with the data on situational guilt feelings. Different relational structures are proposed and tested against each other. The same components as in the previous study are considered: norm violation, worrying, and the tendency to rectify.

The following structures are studied: a linear-sequence structure, a star structure, a cluster structure, and an item-family structure. Although we will use causal language for the relations in the structure, no causal evidence will be available. The relations can be associations based on other than causal relations. A linear-sequence structure implies that the components of guilt can be linearly ordered in how they affect each other and the resulting guilt feeling. A star structure implies that all components affect the guilt feelings, but not each other. A cluster structure implies that there exist clusters of components that affect each other and the emotion, without any relations between components belonging to different clusters. For the data set on guilt, the clusters are defined based on the kind of component: one cluster is formed by the appraisal of norm violation and the emotion guilt, and a second cluster is formed by the two action-oriented components, worrying and tendency to rectify, and the emotion guilt. Guilt forms the overlap between the two clusters. Finally, an item-family structure implies that all components and the emotion are related to the same degree within a situation, but the degree may differ over situations. For the dataset on guilt feelings, this item-family structure showed the best fit, better than a structure with relations between the components or between the components and the emotion (beyond the general guilt sensitivity). The implication is that apart from the general underlying guilt sensitivity also some situation-specific individual differences exist. This completes the earlier made tentative conclusion of one underlying trait.

1.1.3 Making abstraction of individual differences: a marginal approach

In Chapter 4, we make abstraction of individual differences, in order to apply what is called a marginal approach. The reason for employing this approach is that the dependencies we investigated using the relational approach may distort the results when not taken into account, or obscure the main effects if they are taken into account indeed. When abstraction is made of individual differences, using a marginal approach, the dependencies no longer have these effects. The price one pays is that the effects are population-level effects and not effects at the level of the person. This approach is exemplified with a subset of the data on guilt feelings (only three situations). The responses on the items measuring guilt and its components were predicted based on an effect of the situation and an effect of the item type (norm violation, brooding, tendency to rectify, and guilt feelings). Guilt was not decomposed into its components as in Chapter 1 and 2, but the guilt feelings were treated as the components are: as just another type of item. All effects were significant. In addition to the modeling of the item difficulty structure, also the relations between the items were modeled in correspondence with the item family structure, which was supported by the data.

1.2 Emotional behavior: the case of verbal aggression

Another substantive topic of the dissertation is the link between the action tendency and the emotionally motivated behavior. One kind of behavior associated with anger is investigated: verbal aggression. The choice for an action tendency related to anger can be motivated by two reasons: First, anger is a more frequent emotion than guilt (Zelenski & Larsen, 2000) and as such the opportunity to study its action tendencies and associated behaviors is much broader. Second, anger is often conceived as an action-oriented emotion, leading for example to aggression (Averill, 1983; Cornell, Peterson, & Richards, 1999; Kassinove, Sukhodolsky, Tsytsarev, & Solovyova, 1997; Kinney, Smith, & Donzella, 2001), whereas the behaviors associated with guilt are not that clearly delineated and are often more covert, and less open to observation. Verbal aggression is chosen instead of more severe forms of aggression, because it is more common and less socially undesirable to report. Although the effect it has may be smaller than the effect of physical aggression, because of its more common character it is an important phenomenon as well. I can hurt all kinds of relationships and be the source of many conflicts.

Three verbally aggressive (VA) behaviors were selected: cursing, scolding, and shouting. For each of these VA behaviors, the participants were asked two questions, one on the action tendency (wanting to the display the VA behavior in the situation), called a want-item, and one on the actual behavior (displaying the VA behavior in the situation), called a do-item. It was investigated how much the behavior depends on the action tendency, whether inhibition plays a role, and whether there are situational and behavioral and/or individual differences in inhibition.

The results concerning these questions are reported in two different chapters. The results reported in Chapter 2 are: (1) The VA action tendency has a clear predictive power for the VA behavior. (2) The do-items are more difficult than want-items, meaning that inhibition occurred. (3) There are individual differences in the weight of wanting to display verbal aggression (action tendency) for the prediction of actually displaying verbal aggression.

In Chapter 5, the phenomenon of inhibition is investigated more in depth to find an explanation for the inhibition. The following theory about verbal aggression and its inhibition is investigated: The tendency to be verbally aggressive and the inhibition of VA behavior can be influenced by behavior specific factors, situation specific factors, general individual differences (a latent trait), or a combination of any of the previous. The person specific factors can be thought of as traits (an action tendency trait and an inhibition trait), whereas the behavior and situation specific factors can be thought of as features like visibility of the behavior, or the degree in which the situation is frustrating, etc.

It was found that the VA action tendency is based on a latent trait (a verbal aggression trait), and that features of the combination of VA behaviors with situations also play a role, whereas inhibition is primarily determined by individual differences (an inhibition trait), and not so much by the situation or the

VA behavior in question. The approach was validated by correlating the verbal aggression trait and the inhibition trait with the following related measures: the scores of the participants on the Trait Anger scale of Spielberger (1980), the scales Anger In, Anger Out, Anger In Control, and Anger Out Control of the Self Expression and Control Scale of Van Elderen, Maes, Komproe, and Kamp (1997) (an adaptation of the Anger Expression Scale of Spielberger, Johnson, and Jacobs (1982)), and the Direct Aggression scale and the Indirect Aggression scale of the Buss-Durkee Hostility Inventory-Dutch (Lange, Hoogendoorn, Wiederspahn, & Beurs, 1995). The VA behavior-situation specific parameters were correlated with several situational properties like presence of others, degree of frustration elicited by the situation, etc. The verbal aggression trait turned out to be primarily related to Trait Anger, Direct and Indirect Aggression. The inhibition trait was primarily related to coping with anger: it was negatively correlated with Anger Out, and positively with Anger In and Control Anger Out. There was a slight positive correlation between the verbal aggression trait and the inhibition trait. As for the situational effect, the action tendency was positively correlated with how frustrating the situation is experienced and with instrumentality and the expressiveness of VA behavior in the situation. Negative correlations were found with expected dislike from others and negative self-evaluation.

It seems that our approach was successful in modeling and understanding the data from a situation-response questionnaire, including its external validation. The two basic concepts, the VA action tendency and the inhibition of VA behavior each have interesting correlations with external variables.

2. The modeling perspective

As formal basis for the just described substantive research into the structure of emotions and emotional behavior, a modeling approach was chosen based on Item Response Theory (IRT). An important aim of our studies was to show the suitability of the approach for data stemming from situation-response questionnaires on emotions and emotion-related behavior, while the approach was developed primarily for cognitive tests. For practical reasons, we have limited our modeling to binary data. It would require an additional extension, beyond the extensions we will describe, to handle also multi-categorical data. The extension is straightforward, but beyond the scope of this dissertation.

Most IRT models assume that the probability of giving a 1-response on an item is a function of two kinds of parameters: person specific parameters (commonly random effects) and item specific parameters (commonly fixed effects). The most simple model is the Rasch model (Rasch, 1960), which has one parameter per person, often called the ability or latent trait, and one parameter per item, often called the item difficulty. The person parameter is conceived of as a random effect in this dissertation, but also a fixed effect is possible as in a JML (Joint Maximum Likelihood) formulation. However, this approach is not to be recommended because of consistency problems of the estimation.

The response a person gives to an item is explained by an effect specific to the person and an effect specific to the item. The model equation of the Rasch model is the following:

$$P\left(Y_{ij}=1|\theta_i\right)=\frac{\exp\left(\theta_i+\beta_j\right)}{1+\exp\left(\theta_i+\beta_j\right)} \tag{1.1}$$

with $i = 1, \ldots, I$ the index for the person,
$j = 1, \ldots, J$ the index for the item,
θ_i the person parameter,
β_j the item parameter,
and Y_{ij} the response of person i to item j.

Studying emotions with a situation-response questionnaire, the person parameter can be conceived of as an emotion specific personal threshold, as the sensitivity of a person to the emotion in question, whereas the item parameter can be conceived of as the emotion inducing power of the situation. Note that this interpretation holds after multiplying the person parameter by -1, resulting in the parameterization $\beta_j - \theta_i$ instead of $\theta_i + \beta_j$. According to this reparameterization, if the emotion inducing power is larger than the emotion specific personal threshold, then the probability is higher than .5 for the person to experience the emotion in question. Note that the more common parameterization is $\theta_i - \beta_j$, in line with the interpretation of θ_i as the ability and β_j as the difficulty.

All models used in this dissertation are based on the Rasch model. They are further developments or modifications of this model. Four modeling approaches are investigated on their potentialities for the situational study of emotions and individual differences in these.

2.1 MIRID: a model for investigating the decomposition of concepts

In the section on the componential structure of emotions, a componential theory for situational guilt feelings was presented. This theory was tested with data from a situation-response questionnaire. An IRT model that is commonly used to test componential theories is the Linear Logistic Test Model (LLTM, Fischer, 1973, 1977). The LLTM requires that one knows the value of each component in each situation, which is not always the case. Therefore, an alternative model was developed: the Model with Internal Restrictions on Item Difficulties (MIRID, Butter, De Boeck, & Verhelst, 1998). MIRID assumes a relationship between items, not in the correlational sense, but in the sense that the effect one item has on the response probabilities is a function of the effect of other items. For the MIRID it is required that some items are *composite items* in that it is hypothesized that they are based on other more elementary items. The set of more elementary items are *component items*. For the guilt data for example, the question 'Do you worry in this situation?' is a component item, whereas the item 'Do you feel guilty in this situation?' is a composite item. In the MIRID, the item parameter of the composite item is modeled as a linear combination of the item parameters of component items. It is assumed that the *guilt-inducing power of a situation* is a weighted sum of contributions from different components, which can be expressed in a linear function:

$$\beta_{s0} = \sum_{k=1}^{K} \sigma_k \beta_{sk} + \tau \tag{1.2}$$

with $s = 1, \ldots, S$ the index for the situation,
$k = 1, \ldots, K$ the index for the type of component, with $k = 0$ for the composite items,
σ_k the weight or the contribution of the component of type k, to be interpreted as the importance,
β_{sk} the contribution of situation s to component k,
and τ an additive scaling parameter.

The principle as formulated in Equation 1.2 is built into the Rasch model. As a result, the β_j are no longer the basic parameters as in Rasch model (Equation 1.1), but instead the β_{sk} and the σ_k are. For the case there is one underlying trait, the *probability of a componential response* is considered a function of the person contribution θ_i and of the component-specific situational guilt-inducing power β_{sk}:

$$P\left(Y_{isk}=1|\theta_i\right)=\frac{\exp\left(\theta_i+\beta_{sk}\right)}{1+\exp\left(\theta_i+\beta_{sk}\right)} \tag{1.3}$$

The *probability of a composite response* is considered a function of the same person contribution θ_i and of a weighted sum of the component-specific situational guilt-inducing powers, represented in the parameter β_{s0}:

$$P\left(Y_{is0}=1|\theta_i\right)=\frac{\exp\left(\theta_i+\beta_{s0}\right)}{1+\exp\left(\theta_i+\beta_{s0}\right)} \tag{1.4}$$

with β_{s0} defined as in Equation 1.2

In Chapter 1, an extension of the MIRID is used that is developed by Butter (1994). It allows for unequal but fixed discrimination values, and it is called the OPLM-MIRID. The restriction to one underlying latent trait is not necessary, but it turned out empirically that MIRIDs with only one such latent trait have a reasonable good fit to the data. In Chapter 6, two estimation methods for the MIRID and the OPLM-MIRID are compared: a conditional maximum likelihood estimation (CML) and a marginal maximum likelihood estimation (MML). For the former, we wrote a stand-alone program in Delphi 5, whereas the latter estimation can be performed with PROC NLMIXED from SAS V8. In addition, the robustness of the parameter estimates to violations of the normality assumption for the person parameter is investigated. Only small differences were found concerning individual estimates for the person parameter. For the item parameters, no differences were found. Also, a method is presented to test the structure MIRID imposes to the data regarding the link between component items and composite items.

As the assumption that no individual differences occur in the weights of the components (σ_k) can be too strong, in Chapter 2, we have adapted the MIRID to allow for individual differences in the weights of the components. This extension is called the Random Weights MIRID (RW-MIRID). The RW-MIRID assumes that the weights of some of the components are random effects, meaning that they follow a normal distribution over persons, instead of being fixed effects. This model can be estimated with PROC NLMIXED from SAS V8.

2.2 The relational structure of concepts: using Local Item Dependencies

In Chapter 3, a methodology to investigate and test the relational structure of emotions is presented. Based on psychological knowledge, different plausible relational structures were constructed for the dataset on guilt feelings. These structures were translated into IRT models and tested against each other to find the best fitting model, and by consequence, the best fitting theory.

The methodology is based on extant IRT models as described amongst others by Kelderman (1984) and Hoskens and De Boeck (1997). The models are called Local Item Dependency (LID) models. To explain the notion of LID, we start with a basic assumption of most IRT models: the assumption of local stochastic independence. This means that the dependence between the responses of an individual is solely attributed to the underlying latent traits, without the responses on the other items containing any additional information for the probability of responses to the item in question, so that Equation 1.5 holds:

$$P\left(\mathbf{Y}_i = y_{i1},\ \ldots,\ y_{iJ}|\boldsymbol{\theta}_i\right) = \prod_{j=1}^{J} P\left(Y_{ij} = y_{ij}|\boldsymbol{\theta}_i\right) \tag{1.5}$$

with $\mathbf{Y}_i$ the vector containing all responses of person i, and $\boldsymbol{\theta}_i$ the vector of latent traits.

If Equation 1.5 does not hold, it is said that there is LID, because after partialling out the latent trait(s), covariances between the items remain. Often, LIDs are considered something to be avoided. However, we will demonstrate that they can be informative about the relational structure of emotions.

LIDs can be incorporated into, for example the Rasch model, by adding fixed effect parameters to the model that capture the interactions (another term for dependency) between the items. Table 1.1 shows the basic model formulation for the case there is a fixed interaction between a pair of items j and h.

TABLE 1.1. Model for fixed pairwise interaction.

Response pattern $(y_{ij},\ y_{ih})$	Formula
(0,0)	$1/v(\theta)$
(0,1)	$exp\left(\theta_i + \beta_j\right)/v(\theta)$
(1,0)	$exp\left(\theta_i + \beta_h\right)/v(\theta)$
(1,1)	$exp\left[\left(\theta_i + \beta_j\right) + \left(\theta_i + \beta_h\right) + \beta_{int}\right]/v\left(\theta\right)$

Note: β_{int} is the interaction parameter for the item pair, and v(θ) $=$ $1 + exp\left(\theta_i + \beta_j\right) + exp\left(\theta_i + \beta_h\right) + exp(2\theta_i + \beta_j + \beta_h + \beta_{int})$.

It is easy to see that when β_{int} is positive, the probability of observing the response pattern (1,1) increases and that when β_{int} is negative, the probability decreases, in comparison to the probability of the same event under the Rasch model. The implication of this interaction model is that the item parameters β_j and β_h are difficult to interpret, because they are no pure item effects anymore, but instead they are also dependent on the interaction.

Different patterns of interactions –i.e. different relational structures– can be defined by adding the corresponding fixed-effect interaction parameters to the model. If the responses are in agreement with one of these theory-based patterns of LIDs, two types of conclusions can be drawn. First, the corresponding theory is supported by the data, so that we gain insight in the relational structure of an emotion. Second, evidence for the internal validity of the questionnaire is found, as the responses are in agreement with the theory in question.

Note that the componential and the relational approach can be combined as well. This suggestion is mentioned at the end of Chapter 3. Although no such analysis is reported here, we performed such an analysis for the data on guilt. The results confirmed those from Chapter 1 and Chapter 3, but as explained, the dependency as implied by the relational structure hampers the interpretation of the parameters.

2.3 A marginal approach to model the effect of item covariates

All models described in the previous paragraphs are random-effect models, although LID models contain elements from what is called the conditional model approach (Diggle, Heagerty, Liang, & Zeger, 2002; Fahrmeir & Tutz, 2001). Three characteristics of these models lead us to look for a different approach: First, the item and the person parameters of the previously mentioned models cannot be separated from the dependency structure. A consequence is that they cannot be interpreted separately from the dependency. Second, because the parameters are affected by the dependency, one cannot study how the item parameters depend on item covariates independent of the dependency structure. Third, if the dependency structure is misspecified, this can have serious consequences for the other parameters (Thissen, Steinberg, & Mooney, 1989; Tuerlinckx & Boeck, 2001; Yen, 1993). Therefore, we have explored marginal model approaches which yields estimates that are not affected by the dependency structure. The price to pay is that these models are less appropriate to study individual differences and that the estimated item property effects are population-level effects instead of effects that apply to individual persons.

In Chapter 4, a marginal variant of the LLTM was formulated. This marginal approach is rather new in the context of psychometric modeling. We chose for the LLTM because it is a natural first step to the MIRID (see Section 2.1). It is a future perspective to extend the marginal approach to the MIRID. The marginal LLTM (M-LLTM) has two parts: (1) a model for the mean structure in which the marginal probabilities are related to the item covariates by the logit link function, and (2) a model for the associations between the observations, called the association structure.

The mean structure of this marginal LLTM can be defined as in Equation 1.6.

$$\text{Logit}\left[P\left(Y_{ij}=1\right)\right]=\sum_{k=1}^{K} q_{jk}\eta_k^* \tag{1.6}$$

with $k = 1, \ldots, K$ the index for the item covariate,
q_{jk} the value of item j on item covariate k,
and η_k^* the effect of item covariate k on the marginal probabilities.

The association between two items j and h is denoted with the parameter γ_{ijh}. The person subscript is added, because the model in principle allows for person-dependent association covariates. However, this extension is beyond the scope of this dissertation. The formula for the association structure can be written as follows:

$$f\left(\gamma_{ijh}\right)=\sum_{m=1}^{M} z_{jhm}\alpha_m \tag{1.7}$$

with $m = 1, \ldots, M$ the index for the association covariates,
z_{jhm} the value of association covariate m for the association between the responses to the items j and h,
α_m the effect of association covariate m,
and $f(.)$ a link function to link the association parameter γ_{ijh} to the association covariates.
Higher-order generalizations of the association parameter to more than two items can be denoted with subscripts added to z for all items involved in the association in question. Note that the models to be discussed also allow for person-specific association covariates. However, this extension is is beyond the scope of this dissertation.

Three different options for expressing the associations are discussed: marginal correlations, marginal log odds ratios and conditional log odds ratios (the log odds ratio conditional on zero responses for all other items). For each model, two or more estimation techniques are discussed: a full-likelihood based approach and an approach based on Generalized Estimation Equations (GEE, Hardin & Hilbe, 2003; Liang & Zeger, 1986; Zeger & Liang, 1986). Advantages and disadvantages of all approaches are described in Chapter 4.

The marginal approach has two major advantages: First, some marginal approaches yield consistent estimates for the parameters of the mean structure regardless of the correct specification of the association structure. By consequence, an incorrect assumption about the association structure will not bias the estimates of the effects of the item covariates like is the case in random-effect models. Second, these marginal models allow for a very flexible modeling of the associations between response in terms of item specific and person specific covariates. These more complex patterns are a serious complication for the random-effect models, since they would require the inclusion either of dependency parameters or of multiple random effects. These marginal models can be used to investigate the effect of covariates at the level of the population, independent of the associations between item responses, while it is still possible to either explore or even model these associations. Unfortunately, the effects can no longer be interpreted as effects at the level of the individuals.

2.4 Embretson's learning model as a framework

To model responses to action tendency questions and behavior questions, in Chapter 5, a model was used that was originally formulated for learning (Embretson, 1991). The action tendency is formally equivalent with the stage before learning, and the behavior with the stage after learning. Individual differences play a role in the first stage and in the change from the first stage to the second. The equivalence with the learning model is only formal in that learning has a positive effect, whereas the effect of inhibition (between the action tendency and the behavior) is negative on average.

A specific feature of the model is that for items from the first stage only one latent trait plays, whereas for items in the second stage, also a second latent trait plays (learning ability, in our case inhibition). We have applied the model in the following formulation:

$$\text{Logit}\left[P\left(Y_{ijk}=1 \mid \alpha_i,\kappa_i\right)\right]=\alpha_i+\beta_{sk}^{(want)}-d\left(\kappa_i+\beta_{sk}^{(do)}\right) \qquad (1.8)$$

with $d = 1$ for a do-item , and $d = 0$ for a want-item,
$\alpha_i \sim N(0, \sigma_\alpha^2)$ the personal VA action tendency or verbal aggression trait,
$\kappa_i \sim N(\mu_\kappa, \sigma_\kappa^2)$ the personal VA inhibition parameter, or inhibition trait. The mean of κ_i $(= \mu_\kappa)$ is the overall inhibition effect,
$\beta_{sk}^{(want)}$ the effect of the combination of a situation s and a kind of VA behavior k on the tendency to behave in a verbally aggressive way,
and $\beta_{sk}^{(do)}$ the inhibition effect of the combination of a situation s and a kind of VA behavior k on the VA behavior.

The $\beta_{jk}^{(do)}$ is an extension of Embretson's learning model, as in her model, there is only one item parameter for the two stages. Based on this model, we have developed a framework to test whether inhibition is either person dependent or item (situation and/or VA behavior) dependent or both.

3. Conclusions

We believe that the study of emotions through situational questionnaires can benefit from item-response modeling. For several research questions, there will be a close match between the question and a particular IRT model, so that one can test this hypothesis using the model, and for other questions, an appropriate model can be formulated. The availability of general model estimation tools like PROC NLMIXED from SAS V8, associated with the framing of IRT models as generalized (non)linear mixed models (McCulloch & Searle, 2001), contributes to the flexibility of item response modeling. Several of our findings are made possible thanks to the models we have used. Examples concern the role of guilt components, the lack of individual differences in the role they play, and the primarily individual (versus situational) nature of inhibition.

Finally, the exploration of marginal modeling for its potentialities has lead us to the belief that it deserves further attention as a flexible and promising approach, one that is underexplored and hardly used thus far to analyze questionnaire data.

On this dissertation

The thesis is a collection of manuscripts that are either submitted or already accepted for publication.

Chapters

In Chapter 1 a manuscript on the MIRID and the OPLM-MIRID in which the componential structure of situational guilt feelings is investigated. The model is estimated with a CML method that is also implemented in the MIRID CML Program discussed in Chapter 6. The reference of the manuscript is:
Smits, D. J. M., & De Boeck, P. (2003). A componential model for guilt. *Multivariate Behavioral Research, 38*, 161-188.

Chapter 2 is manuscript that is accepted for a book. The reference is given below. It describes the MIRID and the Random weights MIRID, which are both illustrated with the data on situational guilt feelings and with the data on verbal aggression.
Smits, D. J. M., & Moore, S. (accepted). MIRID: Latent item covariates with fixed effects. In P. D. Boeck & M. Wilson (Eds.)(in preparation), *Psychometrics using logistic mixed models.* New York: Springer-Verlag. (contract with Springer)

Chapter 3 is a manuscript on the relational structure of emotions. The reference is:
Smits, D. J. M., De Boeck, P., & Hoskens, M. (2003). *Examining the structure of concepts: using interactions between items.* Manuscript submitted for publication.

Chapter 4 is a manuscript on the marginal approach for the Linear Logistic Test Model. The reference is:
Smits, D. J. M., De Boeck, P., & Molenberghs, G. (2003). *Marginal approaches to the linear logistic test model.* Manuscript submitted for publication.

Chapter 5 describes an approach to model the behavioral expression of emotions based on the learning model of Embretson (1991). The verbal aggression

data were collected by Kristof Vansteelandt. The reference of the manuscript is:
Smits, D. J. M., De Boeck, P., & Vansteelandt, K. (2003). *The inhibition of verbally aggressive behavior*. Manuscript submitted for publication.

Finally, in Chapter 6 two estimation methods for the MIRID and the OPLM-MIRID are compared and the robustness of the parameters to violations of the normality assumption of the person parameter is investigated. For the CML estimation of the MIRID a stand-alone program was written, which can be found on the compact disc that accompanies the dissertation. The MML estimation can be performed with PROC NLMIXED of SAS V8. The reference is:
Smits, D. J. M., De Boeck, P., & Verhelst, N. D. (in press). *Estimation of the MIRID: A program and a SAS based approach. Behavior Research Methods, Instruments, and Computers.*

The co-authors of the various manuscripts are the following:
Paul De Boeck, supervisor of the thesis
Machteld Hoskens, researcher at CTB, Monterey, CA
Geert Molenberghs, professor at the Center of Statistics of the Limburg University Center
Stephen Moore, Ph.D. student at UC Berkeley, CA
Kristof Vansteelandt, postdoctoral researcher at the K.U. Leuven
Norman Verhelst, Methodologist at CITO, The Netherlands.

Preliminary notes

Note that the notation, and the formulation of the models will differ between the chapters. The reason is that the chapters are stand-alone manuscripts for various journals and a book. We have followed notational requests for the book and we have chosen the notation for the journal manuscripts to be as appropriate as possible for the manuscript in question, taking into account also the notation for similar models in the literature.

Also the estimation methods differ depending on the chapter. Two different estimation methods are used throughout the dissertation: CML and MML. The CML method is used in Chapter 1 and Chapter 6, whereas the MML method is used in Chapter 2, 3, 4, 5, and 6. However, re-estimating the models of Chapter 1 with an MML method resulted in similar results as the ones obtained from the CML estimation. For Chapter 6, the correspondence between the two is discussed more in detail, as the comparison of the two estimation methods is an important part of that chapter. All models used in this dissertation can be estimated with an MML method, but this does not hold for the CML method.

1.1 References

Averill, J. R. (1983). Studies on anger and aggression: implications for theories of emotion. *American Psychologist*, *38*, 1145-1160.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: a cognitive-neoassociationistic analysis. *American Psychologist*, *45*, 494-503.

Butter, R. (1994). *Item response models with internal restrictions on item difficulty.* Unpublished doctoral dissertation, Catholic University of Leuven, Leuven.

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

Cornell, D. G., Peterson, C. S., & Richards, H. (1999). Anger as a predictor of aggression among incarcerated adolescents. *Journal of Consulting and Clinical Psychology*, *67*, 108-115.

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

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

Embretson, S. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495-515.

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

Fischer, A. H. (1991). *Emotion scripts: A study of the social and cognitive facets of emotions.* Leiden, The Netherlands: DSWO Press.

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.* Bern, Switzerland: Huber.

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

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

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimation equations.* London: Chapman & Hall/CRC.

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

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

Kassinove, H., Sukhodolsky, D. G., Tsytsarev, S. V., & Solovyova, S. (1997). Self-reported anger episodes in russia and america. *Journal of Social Behavior and Personality*, *12*, 301-324.

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

Kinney, T. A., Smith, B. A., & Donzella, B. (2001). The influence of sex, gender, self-discrepancies, and self-awareness on anger and verbal aggressiveness among u.s. college students. *Journal of Social Psychology*, *141*, 245-275.

Kuppens, P., Van Mechelen, I., Smits, D. J. M., & De Boeck, P. (in press). The appraisal basis of anger: Specificity, necessity and sufficiency of components. *Emotion*.

Lange, A., Hoogendoorn, M., Wiederspahn, A., & Beurs, E. de. (1995). *Buss Durkee hostility and direction of hostility inventory-dutch (BDHI-D).* Lisse, The Netherlands: Swets & Zeitlinger.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46*, 352-367.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Oatley, K., & Jenkins, J. M. (1996). *Understanding emotions.* Oxford: Blackwell Publishers.

Omdahl, B. L. (1995). Cognitive appraisal theories of emotions. In B. L. Omdahl (Ed.), *Cognitive appraisal, emotion, and empathy* (p. 37-90). New Jersey: Lawrence Erlbaum.

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

Rasch, G. (Ed.). (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: 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.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Vol. 5. emotions, relationships, and health* (p. 11-36). Beverly Hills, CA: Sage.

Sas System V8 for windows [Computer program]. (1999). Carey, NC.

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: testing 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.

Skiffington, S., Fernandez, E., & McFarland, K. (1998). Towards a validation of multiple features in the assessment of emotions. *European Journal of Psychological Assessment*, *14*, 202-210.

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

Spielberger, C. D. (1980). *Preliminary manual for the state-trait anger scale (STAS).* Tampa, Florida: University of south Florida.

Spielberger, C. D., Johnson, E. H., & Jacobs, G. A. (1982). *The anger expression (AX) scale.* Tampa, Florida: University of south Florida.

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

Van Elderen, T., Maes, S., Komproe, I., & Kamp, L. van der. (1997). The development of an anger expression and control scale. *British Journal of Health Psychology*, *2*, 269-281.

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

Zeger, S. L., & Liang. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121-130.

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: a state and trait perspective from experience sampling data. *Journal of Research in Personality*, *34*, 178-197.

Chapter 1

A Componential IRT Model for Guilt

Dirk J. M. Smits
Paul De Boeck

ABSTRACT 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. [1]

[1]Correspondence concerning this article should be addressed to: Dirk J. M. Smits, K.U. Leuven, Department 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
The research is financially supported by a GOA 2000/2-grant from the K. U. Leuven: 'Psychometric models for the study of personality'.
I thank René Butter for his help with MIRID. We also thank the schools and the students for kindly participating to this study

1.1 Introduction

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: 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. Embreston (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 the correct response being 'the feminine of ...;' the corresponding second subtask would be 'the feminine of uncle is (a) sister, (b) aunt, (c) niece, (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 Embreston (1981) 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.

1.2 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 Schwartz (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, & Schure, 1989; Izard, 1993; Roseman et al., 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, De Riviera, & Mascolo, 1995; Smith & Lazarus, 1993; Wicker et al., 1983).
2. Guilt implies an *appraisal in terms of norm violation.* For example,

Lindsay-Hartz (1984) and 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, etc. (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 correlation 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.

1.3 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:
$$\text{Logit}\left[P\left(X_{ij}=1|\theta_i\right)\right] = a_j\left(\beta_j - \theta_i\right) \tag{1.1}$$
with $P\left(X_{ij}=1|\theta_i\right)$ indicating the probability of person i ($i = 1, \ldots, I$) feeling guilty in situation j ($j = 1, \ldots, J$), and $\text{logit}[P\left(X_{ij}=1|\theta_i\right)] = \ln\left\{P\left(X_{ij}=1|\theta_i\right)/\left[1-P\left(X_{ij}=1|\theta_i\right)\right]\right\}$; with θ_i 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:
$$\beta_j = \sum_{k=1}^{K} \sigma_k \beta_{jk} + \tau \tag{1.2}$$
with σ_k denoting the weight of the contribution of the component of type k ($k = 1, \ldots, K$); with β_{jk} denoting the k^{th} 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 1.1 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} :

$$\text{Logit}\left[P\left(X_{ijk}=1|\theta_i\right)\right]=a_{jk}\left(\beta_{jk}-\theta_i\right) \tag{1.3}$$

with $P\left(X_{ijk}=1|\theta_i\right)$ 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.

1.3.1 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

$$\beta_j^* = \sum_{k=1}^{K} \sigma_k \beta_{jk} + \tau + c = \sum_{k=1}^{K} \sigma_k \left(\beta_{jk}^*\right) + \tau^* \tag{1.4}$$

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

$$\tau^* = \tau + c \left(1 - \sum_{k=1}^{K} \sigma_k\right) \tag{1.5}$$

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 1.5 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.

1.3.2 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.

1.3.3 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 1.6, is identical to the structure of Equation 1.2, 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 (2). These complexity factor values are *a priori given*, and hence are constants, whereas the β_{jk} are to be estimated parameters. The matrix $\mathbf{Q}$ 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 1.6.

$$\beta_j = \sum_{k=1}^{K} \sigma_k q_{jk} + \tau \tag{1.6}$$

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, Equation 1.6 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.

1.3.4 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, 2003; 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 (2003). 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 R1c-statistic and the DIMTEST procedure (Stout, 1987). Like the Martin-Löf test, the R1c-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 R1c-statistic fits the framework of the generalized Pearson statistics. Because the R1c-statistic is not very sensitive for violations of unidimensionality (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 R1c-statistic did not reject the OPLM with the chosen discrimination values, did we conclude that the OPLM showed a reasonable fit.

1.3.5 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.

1.4 Collecting the situations

1.4.1 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, the feeling disappeared, 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.

1.4.2 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.

1.4.3 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.

1.5 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).

1.6 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 et al., 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.

1.6.1 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 having 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.

1.6.2 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).

1.6.3 Descriptive statistics

In Table 1.1, 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.

1.6.4 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). 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

TABLE 1.1. Means and standard deviations before dichotomization

Situation	Component 2	Component 4	Component 5	Guilt item
Break-up	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)
Movie	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)
Secret	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)
Pen pal	1.366 (.983)	1.336 (.891)	1.470 (1.047)	1.437 (.940)
Jacket	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)

apply as for the OPLM, but the composite item parameters are restricted to be a linear combination of the componential item parameters. As mentioned 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 1.2. It can be seen in Table 1.2 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 R1c-statistic was 117.105 (p = .48, df = 39). The assumption of unidimensionality was tested independently using the DIMTEST procedure (Stout, 1987; Stout et al., 1992, 1993). 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 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 1.3. 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

TABLE 1.2. 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
Break-up	1	2	2	2
Trumpet	2	1	1	1
Shoes	1	2	2	2
Movie	3	3	2	2
Discussion	2	2	2	2
Secret	2	2	2	2
Youth movement	3	3	3	3
Pen pal	3	2	2	3
Jacket	1	2	2	1
Homework	3	3	3	3

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.

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 1.3 and 1.4). For each situation, they are the weighted sum of the values for the three components (Table 1.3), using the weights of the components and the additive scaling parameter (Table 1.4). 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 1.4. 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-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 signific-

TABLE 1.3. Item parameters of the component items (and their standard errors) and reconstructed item parameters of the composite items as estimated with the OPLM-MIRID

Situation	Component 2	Component 4	Component 5	Composite item
Break-up	-.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
Movie	-.122 (.048)	.009 (.047)	.053 (.066)	-.091
Discussion	.563 (.078)	.775 (.081)	1.094 (.104)	.842
Secret	1.272 (.116)	1.006 (.095)	.830 (.089)	1.073
Youth movement	.352 (.055)	.604 (.061)	.170 (.051)	.412
Pen pal	-.193 (.049)	-.279 (.058)	-.108 (.065)	-.327
Jacket	-.077 (.123)	.814 (.087)	1.461 (.133)	.819
Homework	-.623 (.055)	-.718 (.053)	-.541 (.053)	-.821

TABLE 1.4. Weights of the components and their Standard Errors, estimated with the OPLM-MIRID

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

antly 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.

1.7 Discussion

Given the results in Table 1.4, 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 dif-

ferentiating 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, Wagner, & Gramzow, 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 1.3, the contribution of the components varies from situation to situation: for example, the *discussion* situation (hurting a friend in a discussion) and the *jacket* situation (a jacket 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 *jacket* 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 *jacket* situation, whereas feeling guilty in the *jacket* 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

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* (Embreston, 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.

1.8 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. 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/she was in love with you (and was taking the relationship very seriously). The break-up hurts him/her considerably. (Break-up)

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 shoestore. One day, a mother with four children enters the store. One of the kids wants Samson-shoes (Samson is a popular doll figuring in a Belgian TV-series for children). The mother leaves the child with you while she goes on 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 afterward. 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/her to go to the movies. You tell him/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/her, but you pretend not to see it. (Discussion)

6. A friend tells you something in confidence, and adds that he/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/her. Halfway he falls from the rope, and he passes out. (Youth movement)

8. You have a pen pal. You get bored of writing with him/her, and suddenly, you stop corresponding with him/her. After one and a half year, he/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 though a lot of trouble finishing it. You get a good grade for your assignment, the same grade as your friend. (Homework)

1.9 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 K. W. Fischer (Ed.), *Self conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (p. 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* (p. 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, Catholic University of Leuven, Leuven.

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., & 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. (1997). Conceptions and correlates of openness to experience. In *Handbook of personality psychology* (p. 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.

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

Embreston, S. E. (1981). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*, 179-197.

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

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

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.* Bern, Switzerland: Huber.

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

Fischer, G. H. (1995). Linear logistic latent trait models: Theory and applications. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (p. 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* (p. 225-244). New York: Springer-Verlag.

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

Frijda, N. H., Kuipers, P., & Schure, E. ter. (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 No. 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* (p. 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* (p. 368-392). New York: Guilford Press.

Izard, C. E. (1978). Guilt, conscience and morality. In C. E. Izard (Ed.), *Human emotions* (p. 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* (p. 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* (p. 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* (p. 274-300). New York: Guilford Press.

SAS System V8 for Windows (Computer program). (1999). Carey NC, USA: SAS Institute Inc.

Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (p. 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. In B. L. Omdahl (Ed.), *Cognitive appraisal, emotion, and empathy* (p. 37-90). New Jersey: Lawrence 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* (p. 3-19). Oxford: University Press.

Roseman, I. J., Wiest, C., & Schwartz, T. S. (1994). Phenomenology, behaviors, 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. (2003). A componential model for guilt. *Multivariate Behavioral Research*, *38*, 161-188.

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

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* (p. 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.

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* (p. 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). The Netherlands, Arnhem: CITO.

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

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

Chapter 2

Latent item predictors with fixed effects

Dirk J. M. Smits
Stephen Moore

ABSTRACT A nonlinear model, named the 'Model with Internal Restrictions on Item Difficulties' (MIRID) is presented. It is a generalization of the LLTM, as the item predictors are parameters to be estimated instead. Therefore, they will be called 'latent item predictors' or 'components'. Second, an extension of the MIRID, named the Random Weights MIRID, is presented, in which the weights of some components are assumed to be normally distributed over persons. Both models are illustrated with an example on situational guilt feelings and with an example on verbal aggression data.[1]

One of the basic models in item response modeling is the Linear Logistic Test Model (LLTM, Fischer, 1973, 1977). In comparison with the Rasch model (Rasch, 1960), the LLTM is based on a reduced number of item predictors. In the Rasch model, item indicators are used as predictors, and therefore each predictor weight is associated with only one item. These predictor weights are often called difficulty parameters, or locations. In the LLTM item difficulties are explained on the basis of item properties. The effect of an item is defined as the weighted sum of item properties. These properties function as item predictors, as predictors of the item effects. The weights of the properties express how important the properties are in explaining the probabilities of a response. The LLTM may be considered an item explanatory model.

---

[1]This manuscript is slightly adapted to fit in this thesis. For example, references to chapter numbers in the book by De Boeck and Wilson are omitted.
Correspondence concerning this manuscript should be addressed to: Dirk J. M. Smits, K.U. Leuven, Department 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
The research is financially supported by the IAP P5/24 from the federal OSTC, Belgium, and by a GOA 2000/2-grant from the K. U. Leuven: 'Psychometric models for the study of personality'

The requirement that the values of all item properties be known is both, a strength and a limitation. The strength is that a fully specified hypothesis can be tested, the weakness is that the hypothesis has to be fully specified. In the current chapter, a model will be introduced with latent item predictors (properties): The values of the item predictors do not have to be known a priori, but they may have unknown values instead. These item predictors with latent values will be called *latent item predictors* or *components*. Both terms will be used interchangeable. The current model is called the Model with Internal Restriction on Item Difficulties (MIRID, Butter, De Boeck, & Verhelst, 1998). The MIRID can be considered a model containing latent item predictors with fixed effects. It is considered an item explanatory model, as the LLTM but the item properties are latent.

MIRID in its standard form will be explained in the first part of the chapter. In the second part, an extension of MIRID will be explained in which the weights of some of the components are randomly distributed over persons instead of being fixed. This model is called the 'Random Weights MIRID' (RW-MIRID). The RW-MIRID is an extension of MIRID that parallels the RW-LLTM as an extension of the LLTM (Rijmen & De Boeck, 2002).

MIRID assumes a relationship between items. Some of the items are considered to be *composite items* in that it is hypothesized that they are based on one or more elementary items. The more elementary items are *component items*. The relations between all items are expressed in a *relationship matrix* of items by component items (see Figure 2.1). A cell contains a '1' if the item denoted by the row, depends on the component item denoted by the column, and a zero otherwise. The matrix relates items, and may not be considered a property matrix.

$$\begin{array}{l c} & \begin{array}{cccc} \text{Component 1} & \text{Component 2} & \text{Component 1} & \text{Component 2} \\ \text{item 1} & \text{item 2} & \text{item 4} & \text{item 5} \end{array} \\ \begin{array}{l} \text{item 1} \\ \text{item 2} \\ \text{item 3} \\ \text{item 4} \\ \text{item 5} \\ \text{item 6} \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

FIGURE 2.1. Example of a relationship matrix

The items are grouped in two ways: item families and components. A *compon-*

ent groups all items that rely solely on that component. Such items are called component items. For example in Figure 2.1. Component 1 groups the items 1 and 4 and Component 2 groups the items 2 and 5. An *item family* groups a composite item with the component items it is related with. An item family has only one composite item, which is related to component items from all components. Each component is represented in an item family with one and only one component item, a different one depending on the item family. For example, in Figure 2.1 there are two item families: (1) the items 1, 2, and 3, and (2) the items 4, 5, and 6. This pattern of relations between items, and between items and components is called here an *item family structure*. All model formulations in this chapter will be based on this structure. However, MIRID models can be formulated also for other kinds of relations between items, for example for a hierarchical family structure in which lower-order composite items function as component items in turn for higher-order composite items.

MIRID estimates the values of the latent item predictors (α), the latent counterpart of the manifest values of the item predictors in the LLTM, and it also estimates the fixed weights (β) of these predictors in determining the composite item's location. It will be explained that the values of the latent item predictors are the item parameters of the component items. Although MIRID does not directly estimate item parameters for the composite items, their locations or difficulties are modeled indirectly as the linear combination of the the latent item predictors.

The first application comes from a study on situational guilt feelings (Smits & De Boeck, 2003). A brief description of the situational guilt study will illustrate the item family structure described above. Subsequently, a formal representation of the model will be given, followed by the results of the MIRID analysis of the guilt data. As a second example, the chapter also illustrates MIRID using data on verbal aggression.

In their study, Smits and De Boeck (2003) tested a componential theory of situational guilt feelings by means of a questionnaire. The data are from 268 persons, 130 males and 138 females between the ages of 17 and 19. Situational guilt feelings are assumed to rely on three components: (1) whether one feels like having violated a moral, ethical, religious, or personal code in the situation, (2) whether one worries about what one did or failed to do in the situation, and (3) whether one wants to rectify what one did or failed to do in the situation. The questionnaire contains 10 hypothetical situations (see Appendix 1), and the participants were asked to respond to four questions per situation: one question for every component and one about guilt feelings.

This creates an item family structure: three component items and one composite item per item family, all four about the same situation. The item about feeling guilty in a situation is seen as a composite of the three component items associated with the same situation. Each item family corresponds with one situation.

2.1 The model

2.1.1 The systematic component

Like most item response models, including the LLTM, the MIRID has a fixed-effect part and a random-effect part in its systematic component (the one that determines the response probabilities). To explain the formula for the fixed-effect part, two new indexes are needed for the items: The index r (1, ..., R) denotes the components or the latent item predictors. The index s (1, ..., S) denotes the item family of which a component item or composite item is a part. For the composite items, the index r will be set equal to $R+1$. The index i (1, ..., I) denotes the item number, meaning that each value of i corresponds with a particular combination of the indices r and s. The index p (1, ..., P) denotes the person.

In addition, two matrices will be introduced: a latent item predictor matrix $\mathbf{A}$ and an componential weight matrix $\mathbf{\Psi}$. The *latent item predictor matrix* $\mathbf{A}$ is a matrix of items by latent item predictors. It contains the values (parameters) of the latent item predictors: the α_{rs}. See Figure 2.2 for a latent item predictor matrix for the example of six items, as presented in Figure 2.1. In addition, a constant predictor is added for the composite items. Each row of the matrix $\mathbf{A}$ can be conceived of as a row-vector, and will be denoted by $\mathbf{A}_s$.

$$
\begin{array}{ll}
 & \begin{array}{ccc} \text{Predictor 1} & \text{Predictor 2} & \\ \text{Component 1} & \text{Component 2} & \end{array} \\
\begin{array}{l} \text{Item family 1} \\ \text{Item family 2} \end{array} & \begin{pmatrix} \alpha_{11} & \alpha_{21} & 1 \\ \alpha_{12} & \alpha_{22} & 1 \end{pmatrix}
\end{array}
$$

FIGURE 2.2. Example of a latent item predictor matrix $\mathbf{A}$

In order to see that Figure 2.2 presents the latent item predictor matrix, we have expanded this matrix in Figure 2.3. The symbol $\mathbf{A}$ will further be used for the restricted matrix as in Figure 2.2.

$$
\begin{array}{lll}
 & & \begin{array}{ccc} \text{Predictor 1} & \text{Predictor 2} & \\ \text{Component 1} & \text{Component 2} & \end{array} \\
\text{Item family 1} & \text{item 1} & \multirow{6}{*}{$\begin{pmatrix} \alpha_{11} & \alpha_{21} & 0 \\ \alpha_{11} & \alpha_{21} & 0 \\ \alpha_{11} & \alpha_{21} & 1 \\ \alpha_{12} & \alpha_{22} & 0 \\ \alpha_{12} & \alpha_{22} & 0 \\ \alpha_{12} & \alpha_{22} & 1 \end{pmatrix}$} \\
 & \text{item 2} & \\
\text{(composite)} & \text{item 3} & \\
\text{Item family 2} & \text{item 4} & \\
 & \text{item 5} & \\
\text{(composite)} & \text{item 6} &
\end{array}
$$

FIGURE 2.3. Expanded latent item predictor matrix

The second matrix is the *componential weight matrix* $\mathbf{\Psi}$, which is a matrix of item type by latent item predictors. The item types are component item type 1 (referring to Component 1), component item type 2 (referring to Component 2), $\ldots$, component item type R (referring to Component R), composite item, see Figure 2.4 for the example of six items as presented in Figure 2.1 and 2.2. The matrix gives the weights of the predictors for each of the item types. Component items (row 1 and 2) have a weight of one for the corresponding component, and zero otherwise. This is reflected in an identity matrix in the upper left part. Composite items have a weight for each of the components (β_1, β_2) and for the constant (β_0). A row of $\mathbf{\Psi}$ is denoted with $\mathbf{\Psi}_r$.

$$
\begin{array}{ll}
 & \begin{array}{ccc} \text{Component 1} & \text{Component 2} & \end{array} \\
\begin{array}{l} \text{Component item type 1} \\ \text{Component item type 2} \\ \text{Composite item} \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_1 & \beta_2 & \beta_0 \end{pmatrix}
\end{array}
$$

FIGURE 2.4. Example of a componential weight matrix $\mathbf{\Psi}$

The product of $\mathbf{A}$ and $\mathbf{\Psi}'$ results in an item parameter matrix of item families by item types. The item parameter matrix that corresponds to Figure 2.2 and 2.4 is shown in Figure 2.5. In this matrix, the item parameters for all items can be found, organized per item family.

$$
\begin{array}{ll}
 & \begin{array}{ccc} \text{Component} & \text{Component} & \text{Composite} \\ \text{item type 1} & \text{item type 2} & \text{item} \end{array} \\
\begin{array}{l} \text{Item family 1} \\ \text{Item family 2} \end{array} & \begin{pmatrix} \alpha_{11} & \alpha_{21} & \alpha_{11}\beta_1 + \alpha_{21}\beta_2 + \beta_0 \\ \alpha_{12} & \alpha_{22} & \alpha_{12}\beta_1 + \alpha_{22}\beta_2 + \beta_0 \end{pmatrix}
\end{array}
$$

FIGURE 2.5. Example of an item parameter matrix

The fixed-effect part of MIRID can now be formulated as:

$$(\text{fixed-effect part})_{pi} = \beta'_i = \mathbf{A}_s \mathbf{\Psi}'_r \tag{2.1}$$

$\mathbf{A}_s$ corresponds with the s-th row of the latent item predictor matrix $\mathbf{A}$, and $\mathbf{\Psi}_r$ corresponds with the r-th row of the componential weight matrix $\mathbf{\Psi}$. The products $\mathbf{A}_s \mathbf{\Psi}'_r$ correspond to the cells in the item parameter matrix. For example, $\mathbf{A}_2 \mathbf{\Psi}'_1$ is α_{12}.

The fixed-effect part for the component item parameters is similar to the fixed-effect part of a Rasch model, as for the component items $\beta'_i = \alpha_{rs}$, see Equation 2.1. This implies that the values of the latent item predictors are also the item parameters of the corresponding component items. The fixed-effect part for the composite items is a linear combination of the component item parameters α_{rs} with weights β_r: $\beta'_i = \sum_{r=1}^{R} \beta_r \alpha_{rs} + \beta_0$. The composite item parameter is decomposed into the item parameters of the component items. In other words the effect of a composite item is explained in terms of latent item predictors and their weights.

As can be seen in Equation 2.1, the equation for the fixed-effect part is not linear in its parameters anymore, since a product of two parameters is involved. By consequence, the MIRID is not part of the family of generalized linear models, but it is a nonlinear model instead (McCulloch & Searle, 2001). Maris and Bechger (2003) showed that the MIRID is a curved exponential family model, which implies that, among other things, the conditional maximum likelihood method can be used to estimate β_r and α_{rs}.

The *random-effect part* of the MIRID is the same as for the Rasch model. It consists of θ_{p0}, called the random intercept or person parameter. We assume $\theta_{p0} \sim N\left(0, \sigma_\theta^2\right)$. The subscript 0 is used to differentiate the random intercept from the other types of random effects to be presented later.

2.1.2 Conditional formulation of the full model

The conditional formulation of the MIRID for the odds of a response of 1 is analogous to the corresponding formula for the Rasch model or the LLTM:

$$\text{logit}\left[P\left(Y_{pi} = 1 | \theta_{p0}, \beta'_i\right)\right] = \theta_{p0} - \beta'_i \tag{2.2}$$

with $\theta_{p0} \sim N(0, \sigma_\theta^2)$,
for the component items $\beta'_i = \alpha_{rs}$,

and for the composite items, $\beta'_i = \sum_{r=1}^{R} \beta_r \alpha_{rs} + \beta_0$, with i as an index for the pairs (s, r).

Equation 2.2 can be rewritten in terms of the previously used matrix notation as follows:

$$\text{logit}\left[P\left(Y_{pi} = 1 | \theta_{p0}, \mathbf{A}_s, \mathbf{\Psi}_r\right)\right] = \theta_{p0} - \mathbf{A}_s \mathbf{\Psi}'_r \tag{2.3}$$

2.1.3 Identifiability of the MIRID

The well-known indeterminacy of the Rasch model has implications for the item parameters of the MIRID (Butter et al., 1998). If we rescale so that $\beta^*_i = \beta'_i + c$, then it follows that for the component items $\alpha^*_{rs} = \alpha_{rs} + c$ and for the composite items that $\alpha^*_{R+1,s} = \sum_{r=1}^{R} \beta_r \alpha_{rs} + \beta_0 + c$ and also that $\alpha^*_{R+1,s} = \sum_{r=1}^{R} \beta_r \alpha^*_{rs} + \beta^*_0$, so that

$$\beta^*_0 = \beta_0 + c\left(1 - \sum_{r=1}^{R} \beta_r\right) \tag{2.4}$$

The weights are invariant under translations of the scale, but the constant is not. As for the Rasch model, a restriction is needed to render the model identifiable. If $\sum_{r=1}^{R} \beta_r = 1$, then $\beta^*_0 = \beta_0$, so that in this particular case, fixing the constant will not solve the indeterminacy. The constant may not be invariant under translations of the scale. Fixing the constant will render the model identifiable only if $\sum_{r=1}^{R} \beta_r \neq 1$. For the applications presented in this chapter, we will fix the mean of the distribution of the person parameter to zero. Another possibility is to fix one of the α_{rs} or the mean α_{rs} to a known value.

For the identification of MIRIDs with other relations between the items then the ones described by the item family structure, more restrictions may be needed to render the model identifiable. Bechger, Verhelst, and Verstralen (2001) have studied the identification of the Non-Linear Logistic Test Model, a larger family of models of which the MIRID with an item family structure is a particular case.

A problem related to the identification is the existence of equivalent MIRIDs. Bechger, Verstralen, and Verhelst (2002) described this problem for the LLTM, and Maris and Bechger (2003) extended it to the MIRID. The MIRIDs for different componential theories about an item set may be formally equivalent, so that they cannot be differentiated. For example, if we modify the latent item predictor matrix $\mathbf{A}$ (see Figure 2.2) into $\mathbf{A}^*$ and the componential weight matrix $\mathbf{\Psi}$ (see Figure 2.4) into $\mathbf{\Psi}^*$ as shown below, an equivalent MIRID is obtained.

$$\mathbf{A}^* = \begin{array}{l} \\ \text{Item family 1} \\ \text{Item family 2} \end{array} \begin{array}{c} \begin{array}{cc} \text{Component 1} & \text{Component 2} \end{array} \\ \begin{pmatrix} \alpha_{11} - \alpha_{21} & \alpha_{21} & 1 \\ \alpha_{12} - \alpha_{22} & \alpha_{22} & 1 \end{pmatrix} \end{array}$$

$$\mathbf{\Psi}^* = \begin{array}{l} \text{Component item type 1} \\ \text{Component item type 2} \\ \text{Composite item} \end{array} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ \beta_1 & (\beta_1 + \beta_2) & \beta_0 \end{pmatrix}$$

The resulting item parameter matrix is shown in Figure 2.5. The problem of equivalent MIRIDs is not surprising, since MIRID is a model with bilinear terms, so that rotational invariance may play. Therefore, it is preferred to use MIRID in a confirmatory way, which is implied when the item family structure is imposed.

2.2 Applications of the MIRID

2.2.1 Results for the example on situational guilt feelings

We will reparameterize the model using the difference $(\beta_i' - \theta_{p0})$ instead of $(\theta_{p0} - \beta_i')$, so that the probability of feeling guilty for a certain person in a specific situation is a function of the difference between a weighted sum of the componential contributions and the personal guilt threshold. For the reverse parameterization $\beta_i' - \theta_{p0}$, it holds that if the situational inductive power β_i' exceeds the personal threshold θ_{p0}, the probability of experiencing the compontential or composite emotion becomes higher than .5.

The MIRID was fitted with the SAS V8 NLMIXED procedure (see, e.g., Wolfinger, 1999), using an adaptive Gaussian quadrature method as described in Pinheiro and Bates (1995) with 15 quadrature points and Newton-Raphson as optimization technique (see, e.g., Bunday, 1984; Gill, Murray, & Wright, 1981). The program code is given in the Appendix 2. Note that based on the study of Smits and De Boeck (2003), we knew that the Rasch model, and by consequence also the MIRID, did not fit the data in absolute terms. However, for illustrative purposes, we will use the MIRID here as if it fits. In the second part of the chapter, a variant of the MIRID that allows for unequal but fixed discrimination values will be used, called the OPLM-MIRID (Butter, 1994; Smits & De Boeck, 2003).

The goodness-of-fit values of the original MIRID are 10549 (deviance), 10619 (AIC), and 10745 (BIC). These values are similar to the goodness-of-fit values for the Rasch model: 10546 (deviance), 10628 (AIC), 10775 (BIC), meaning that the MIRID fits the data about as well as the Rasch model. The values for the item parameters of the component items are given in Table 2.1. High values mean a high situational guilt inductive power. The values of the other

TABLE 2.1. Estimates and standard errors (S.E.) for the component item parameters or latent item predictor values (situational guilt example)

Situation	Norm Violation (S.E.)	Worrying (S.E.)	Rectify (S.E.)
1	-.089 (.144)	1.376 (.165)	.519 (.153)
2	-2.701 (.222)	-1.989 (.183)	-2.140 (.199)
3	-.728 (.147)	.125 (.141)	-.394 (.151)
4	.101 (.141)	.419 (.142)	.356 (.151)
5	1.483 (.166)	1.971 (.181)	2.549 (.225)
6	2.805 (.232)	2.262 (.200)	1.942 (.191)
7	1.319 (.163)	1.976 (.182)	.854 (.157)
8	-.136 (.144)	-.283 (.142)	.078 (.150)
9	.055 (.149)	1.901 (.186)	3.163 (.275)
10	-1.305 (.162)	-1.530 (.162)	-1.023 (.160)

parameters are given in Table 2.2. The effect of the composite items can be

TABLE 2.2. Estimates and standard errors (S.E.) for the componential weight parameters and variance of the person parameter (situational guilt example)

Parameter	Value (S.E.)
β_1(weight of norm violation)	.497 (.103)
β_2(weight of worrying)	.549 (.131)
β_3(weight of tendency to rectify)	.025 (.094)
β_0(constant)	.203 (.082)
σ_θ^2	1.124 (.120)

reconstructed based on the α from Table 2.1 and the β from Table 2.2. For example, the reconstruction for the composite item of the third item family is equal to (-.728 * .497) + (.125 * .549) + (-.394 * .025) + (.203) = -.100. As could be expected from the goodness of fit, there is a good correspondence between the item parameters of the component items and the composite items as estimated under the Rasch model and as estimated (component items) and reconstructed (composite items) under the MIRID: the correlation between both is .999. The correlation between the item parameters of the composite items as estimated under a Rasch model for all 40 items and the composite item parameters as reconstructed from the parameters of the just estimated MIRID is also .999.

Two of the componential weights are significant: the weights of norm violation and worrying are significant.[2] The weight of the second component is the largest, meaning that for our set of situations 'worrying about what one did' is the most important component of situational guilt feelings. The weight of composite item constant (.203) is the extra effect of the composite item, but its interpretation depends on centering issues. The interpretation is difficult in our case, since the centering was based on θ_{p0} by fixing its mean to zero and not on the latent item predictors. Finally, it is clear that guilt sensitivity as an underlying latent variable shows substantial individual differences (σ^2=1.124). The variance is statistically significant using the conservative Wald test for variances (Snijders & Bosker, 1999; Verbeke & Molenberghs, in press).

2.2.2 Example on verbally aggressive behavior

In a data set on verbal aggression, we have four situations (4 from the 15 studied in Chapter 5) and three different kinds of verbally aggressive reactions (cursing, scolding, and shouting). The four situations are given in Appendix 1. Each type of aggressive reaction is measured in two different ways, which were called response modes: (1) whether one wants to display the corresponding reaction in that situation (want-item), and (2) whether one actually would display the reaction (do-item). Hence, the total number of items is $4 \times 3 \times 2 = 24$. As we are interested in whether actually displaying an aggressive reaction can be explained by wanting to display that aggressive reaction, the items measuring the want-response mode will be considered the component items, whereas the items measuring the do-response mode are the composite items. In this example, there is only one component: wanting as a component of doing as a composite. We will call this component also the action tendency.

The combination of a situation and a kind of aggressive reaction defines an item family that contains two items: a component item or want-item (e.g., 'do you want to curse in this situation?') and a composite item or do-item (e.g.,

[2]Using the OPLM-MIRID also the weight of the third component was significant.

'do you curse in this situation?'). Each situation is associated with three item families: one for each kind of verbally aggressive reaction. As mentioned in the introduction, we want to know to which degree actually displaying an aggressive reaction can be explained by wanting to display that aggressive reaction. Therefore, we fitted a MIRID with wanting as a component. Note that we again used the parameterization $\beta'_i - \theta_{p0}$, to be in line with the interpretation of β'_i as the inductive power from the situation for a certain behavior and θ_{p0} as the personal threshold.

The goodness-of-fit values for this MIRID are 8116.3 (deviance), 8146.3 (AIC), and 8202.6 (BIC). These values approach the goodness-of-fit values for the Rasch model (8073.8 (deviance), 8123.8 (AIC), and 8217.7 (BIC)), meaning that the MIRID has a relatively good fit (based on the AIC and BIC). The values for the item parameters of the component items are given in Table 2.3, and the values for the other parameters are given in Table 2.4. There is a good correspondence between the item parameters of the component items and the composite items as estimated under the Rasch model and as estimated (component items) and reconstructed (composite items) under the MIRID: the correlation between the item parameters (estimated or reconstructed) of both models is equal to .987. The correlation between the item parameters of the composite items as estimated by a Rasch model for all items and the item parameters of the composite items as reconstructed by the MIRID is also high: .991.

TABLE 2.3. Estimates and standard errors (S.E.) for the component item parameters (verbal aggression example)

Situation	Reaction	Item Parameter	Value (S.E.)
1	Curse	$\alpha_{1,1}$	1.396 (.128)
1	Scold	$\alpha_{1,2}$	.762 (.118)
1	Shout	$\alpha_{1,3}$	-.015 (.116)
2	Curse	$\alpha_{1,4}$	1.396 (.132)
2	Scold	$\alpha_{1,5}$	.595 (.117)
2	Shout	$\alpha_{1,6}$	-.308 (.119)
3	Curse	$\alpha_{1,7}$	.459 (.116)
3	Scold	$\alpha_{1,8}$	-.605 (.125)
3	Shout	$\alpha_{1,9}$	-1.583 (.147)
4	Curse	$\alpha_{1,10}$	1.098 (.123)
4	Scold	$\alpha_{1,11}$	.055 (.117)
4	Shout	$\alpha_{1,12}$	-.973 (.133)

The weight of the want-response mode or action tendency is quite large and

TABLE 2.4. Estimates and standard errors (S.E.) for the componential weight parameters and variance of the person parameter (verbal aggression example)

Parameter	Value (S.E.)
β_1 (weight of want-response mode)	1.332 (.083)
β_0 (constant)	-.771 (.076)
σ_θ^2	1.890 (.193)

highly significant, meaning that it has a serious predictive power for the do-response mode ($\beta_1 = 1.332$). It follows from the value of β_1 that the do-items are better differentiated with respect to their inductive power than the want-items are. The effect of being a composite item (constant item predictor) is negative ($\beta_0 = -.771$).

Given these results, it must be further concluded that the inductive power is lower for the actual behavior (do-items) than for the action tendency (want-items): If a want-item has a negative α, the fact that β_1 is larger than 1, and that β_0 is smaller than zero, necessarily leads to a lower inductive power for the corresponding do-items. If a want-item has a positive α, the value of β_0 compensates for a β_1 of 1.332 up to values for α as high as 2.322. Finally, the variance of the general underlying trait is quite large ($\sigma^2 = 1.890$), and highly significant when relying on the conservative Wald test for variances. In sum, the verbally aggressive behavior (doing) seems to require a lower threshold than its action tendency, and the pairs of the behaviors and situations (items) are better differentiated in the actual expression (do-items) than in the action tendency (want-items).

2.3 Extension to Random weights MIRID (RW-MIRID)

The MIRID assumes that the weights of the latent item predictors are the same for all persons. However, as it makes sense plausible that people would differ as to the weight of the components, it would be interesting to allow for individual differences in the weights. For example, for some people worrying may be more important, whereas for other people the tendency to rectify may be more important, perhaps because they are more action-oriented. An extension of the MIRID, which is called the Random Weights MIRID (RW-MIRID), allows for the weights to be random variables. Except for its specific componential structure, the RW-MIRID is very similar to a multidimensional 2PL model, as in both the RW-MIRID and the 2PL person-specific parameters are the weights of latent item predictors. In the multidimensional 2PL these are the latent traits, and the item loadings (trait specific discriminations), respectively.

2.3.1 The systematic component

Each random weight, denoted by β_{pr}, can be split into a mean (the fixed-effect part β_r) and a deviation from that mean (the random-effect part θ_{pr}). In the remainder, the deviation from the mean will be considered the random weight. To construct the formula for the RW-MIRID, the componential weight matrix is now a person specific matrix, denoted with $\mathbf{\Psi}_p$. In the example of Figure 2.6, only the weight of the first latent item predictor is a random effect, all other weights are fixed effects.

$$\begin{array}{ll} & \begin{array}{ccc} \text{Component 1} & \text{Component 2} & \end{array} \\ \begin{array}{l} \text{Component item type 1} \\ \text{Component item type 2} \\ \text{Composite item} \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \beta_{p1} & \beta_2 & \beta_0 \end{pmatrix} \end{array}$$

FIGURE 2.6. Example of an componential weight matrix $\mathbf{\Psi}_p$ for the RW-MIRID

The formula for the RW-MIRID can be written as in Equation 2.5.

$$\text{logit}\left[P\left(Y_{pi}=1|\theta_{p0},\mathbf{A}_s,\mathbf{\Psi}_{pr}\right)\right]=\theta_{p0}-\mathbf{A}_s\mathbf{\Psi}'_{pr} \tag{2.5}$$

where for the component items: $\mathbf{A}_s\mathbf{\Psi}'_{pr}=\alpha_{rs}$

and for the composite items: $\mathbf{A}_s\mathbf{\Psi}'_{p,R+1}=\sum_r^R\alpha_{rs}\beta_{pr}+\beta_0$.

with i as an index for the pairs (s,r), and $\beta_{pr}=\theta_{pr}-\beta_r$.

Given that more than one random effect is included in the model, a multivariate normal distribution is assumed for $\boldsymbol{\theta}_p$, the vector of random effects. Another way to extend the MIRID into a multidimensional model is the following: Until now it is assumed that the same random intercept (θ_{p0}) applies to the component items and the composite items. This is not necessary. There are cases where dependent on the component a different random intercept (a different dimension) applies. Such a model is called the Multidimensional-MIRID (MULTI-MIRID; Butter, 1994), but this extension will not be discussed in this chapter. When the intercept of the composite items (β_{p0}) and the overall random intercept (θ_{p0}) are the only random effects, the model is equivalent with the learning model of Embretson (1991) for two stages: before and after learning.

2.4 Applications of the RW-MIRID

2.4.1 Example on situational guilt feelings

In the example on situational guilt feelings three components were present: norm violation, a tendency to worry, and a tendency to rectify. Because we have no a priori hypotheses about which component should have a random weight, three different models were estimated and compared. In each model a different component was assumed to have a random weight. As mentioned earlier, the OPLM-MIRID will be used here, instead of the previously used original MIRID. In contrast to the MIRID, the OPLM-MIRID allows for unequal but fixed discrimination values. The same discrimination values as in Chapter 1 are used, and not those used earlier in this chapter. To determine the fit of this OPLM-MIRID, it has to be compared with the One Parameter Logistic Model (OPLM: Verhelst & Glas, 1995; Verhelst, Glas, & Verstralen, 1994), an adaptation of the Rasch model that allows for fixed but unequal discriminations.

A nonadaptive Gaussian quadrature method with 15 quadrature points was chosen for the estimation of the OPLM-MIRID. The goodness-of-fit values for the OPLM-MIRID, and the three RW-OPLM-MIRIDs with one random weight are given in Table 2.5.

TABLE 2.5. Goodness-of-fit values for OPLM-MIRID and the RW-OPLM-MIRID on the situational guilt example

Model	Deviance	AIC	BIC
OPLM-MIRID	10451	10521	10647
RW-OPLM-MIRID			
Random weight for			
Norm Violation	10449	10523	10656
Worrying	10451	10525	10658
Tendency to Rectify	10451	10525	10658

All of the goodness-of-fit values mentioned in Table 2.5 are similar to the ones of the OPLM-MIRID, meaning that adding a random weight to the model did really not enhance the fit. By consequence, for this sample of persons and situations, it is not necessary to assume differences between persons for the weights of any of the three components.

2.4.2 Example on verbally aggressive behavior

In the example on verbally aggressive behavior, the model had only one component. We assume that for some people what they want has a larger effect on what they do than for other people. In contrast to the example on situational guilt, the RW-MIRID was estimated using an adaptive Gaussian quadrature method with 15 quadrature points. The adaptive method was used because with a low number of items, the (slow) adaptive method is still doable for the time it takes. The goodness-of-fit values for the RW-MIRID are 8027.5 (deviance), 8061.5 (AIC), 8125.4 (BIC), which are clearly better than those for the original MIRID: 8116.3 (deviance), 8146.3.1 (AIC), and 8202.6 (BIC), meaning that for some people what they want weights heavier in what they do than for other people.

The parameter estimates of the component items (values of the latent item predictor) are given in Table 2.6, and the values for the other parameters are given in Table 2.7. The variance of the weight is much smaller than the variance of the overall random intercept (1.018 vs. 2.031). The correlation between both is .098.

As in the MIRID with fixed component weights, again the action tendency has a serious effect. The inductive power for the actual behavior is lower (for the average person) than that for the action tendency, up to α-values of 1.782. This is because the negative β_0 (-1.032) compensates for a β_1 larger than 1 in all combinations of situations and behaviors. A person with a weight of 1.5 standard deviations below the mean β_1 does more than wanted below α-values of -1.104. The reactions of a person who is situated 1.5 standard deviations above the mean β_1 does more than wanted for α-values higher than .493, so that in six of the twelve situation-behavior combinations this person will do more than wanted to do, and in six of the twelve situation-behavior combinations, the action tendency will be inhibited in some way.

In sum, the results of the RW-MIRID confirm those of the fixed weight MIRID, except for the fact that clear individual differences appear in the effect the action tendency has on the behavior.

2.5 Concluding remarks

The main advantage of the MIRID and its variants, is that in cases where no exact knowledge is available about the values of the components (latent item

TABLE 2.6. Estimates and standard errors (S.E.) for the component item parameters (example on verbal aggression)

Situation	Reaction	Item parameter	Value (S.E.)
1	Curse	$\alpha_{1,1}$	1.715 (.126)
1	Scold	$\alpha_{1,2}$	.963 (.103)
1	Shout	$\alpha_{1,3}$	.230 (.094)
2	Curse	$\alpha_{1,4}$	1.736 (.135)
2	Scold	$\alpha_{1,5}$	.811 (.099)
2	Shout	$\alpha_{1,6}$	.016 (.096)
3	Curse	$\alpha_{1,7}$	.719 (.096)
3	Scold	$\alpha_{1,8}$	-.149 (.103)
3	Shout	$\alpha_{1,9}$	-1.236 (.152)
4	Curse	$\alpha_{1,10}$	1.340 (.115)
4	Scold	$\alpha_{1,11}$	.440 (.093)
4	Shout	$\alpha_{1,12}$	-.576 (.121)

TABLE 2.7. Estimates and standard errors (S.E.) for the componential weight parameters and variance/covariance of the person-dependent parameters (example on verbal aggression)

Parameter	Value (S.E.)
β_1 (mean weight of Component)	1.579 (.118)
$\sigma^2_{\theta_{p1}}$ (variance of component weight)	1.018 (.216)
β_0 (constant)	-1.032 (.092)
$\sigma^2_{\theta_{p0}}$ (variance of overall intercept)	2.031 (.217)
$\mathrm{cov}(\theta_{p0}, \theta_{p1})$	.141 (.147)

predictors), these values can be estimated.

An advantage specific to the RW-MIRID is that one can test whether the assumption of fixed weights is reasonable. A better fit of the RW-MIRID with random weights for one or more components would imply that there are differences in how important these components are.

Finally, the principle behind MIRID can easily be generalized to other basic models, like for example the 2PL (Birnbaum, 1968) or the multidimensional Rasch model. Because extending the MIRID by incorporating additional random effects is straightforward, the MIRID is a flexible tool for the decomposition of general concepts into more elementary aspects.

2.6 Further reading

The MIRID was originally published by Butter et al. (1998), based on Butter (1994). In Butter et al. (1998), a conditional maximum likelihood formulation and estimation method was explained, complemented with a simulation study. An application of the MIRID and an extension of the MIRID to the OPLM-MIRID, originally described by Butter (1994), and the 2PL-MIRID can be found in Smits and De Boeck (2003).

A comparison between two estimation methods for the MIRID and the OPLM-MIRID –a conditional maximum likelihood estimation (Smits, De Boeck, Verhelst, & Butter, 2001) and a marginal maximum likelihood estimation, implemented within PROC NLMIXED– can be found in Smits, De Boeck, and Verhelst (in press) (which is Chapter 6 in this dissertation).

Bechger et al. (2001) embedded the MIRID in a more general model called the non-linear logistic test model (NLTM). They derived the conditions the NLTM has to fulfill in order for the model to be identified. Maris and Bechger (2003) provide more specific conditions for the identifiability of MIRIDs with various kinds of relations, other than the item family structure, as no additional conditions are needed for MIRIDs with an item family structure, as explained earlier. They also discuss that different componential theories about an item set can lead to equivalent MIRIDs, which is true also for the LLTM (Bechger et al., 2002).

Finally, Maris and Bechger (2003) mention that as the MIRID is a restriction of the Rasch model, it is part of the curved exponential family, which implies, among other things, that conditional maximum likelihood estimation is possible for α and β. Also in the MIRID, the sum scores are sufficient statistics for the person parameters (θ_{p0}).

2.7 Appendix 1: Situations

2.7.1 Situations of example on guilt feelings

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/she was in love with you (and was taking the relationship very seriously). The break-up hurts him/her considerably. (Break-up)

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 shoestore. One day, a mother with four children enters the store. One of the kids wants Samson-shoes (Samson is a popular doll figuring in a Belgian TV-series for children). The mother leaves the child with you while she goes on 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/her to go to the movies. You tell him/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/her, but you pretend not to see it. (Discussion)

6. A friend tells you something in confidence, and adds that he/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/her. Halfway he falls from the rope, and he passes out. (Youth movement)

8. You have a pen pal. You get bored of writing with him/her, and suddenly, you stop corresponding with him/her. After one and a half year, he/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 though a lot of trouble finishing it. You get a good grade for your assignment, the same grade as your friend. (Homework)

2.7.2 Situations of example on verbal aggression

1. You are waiting at the bus stop and the bus fails to stop for you.

2. You miss your train because the clerk has given you faulty information.

3. The grocery store closes just as you are about to enter.

4. You use your last 10 cents to call a friend and the operator disconnects you.

2.8 Appendix 2: SAS programs

The SAS program for the MIRID is exemplified with the source code for the guilt example. The α are renumbered using one index to simplify the SAS code. They can easily be matched to the α_{rs} as given in the formulas. The code used is the following:

```
title 'MIRID guilt example';

Proc NlMixed data=MIRID method=gauss
technique=NewRap Qpoints=15 Optcheck;
Parms alpha1-alpha30=1 Beta1-Beta3=1 Beta0=1 VarTheta=1;

ex=exp(-theta
/*Co: dummy variable denoting the type of item: 0 for component items, 1 for
composite items*/
/*COMPONENT ITEMS*/
+(1-Co)*
/*Component 1 = Norm Violation*/
(I1*alpha1+I2*alpha2+I3*alpha3+I4*alpha4+I5*alpha5
+I6*alpha6+I7*alpha7+I8*alpha8+I9*alpha9+I10*alpha10

/*Component 2 = Worrying*/
+I11*alpha11+I12*alpha12+I13*alpha13+I14*alpha14
+I15*alpha15+I16*alpha16+I17*alpha17+I18*alpha18
+I19*alpha19+I20*alpha20

/*Component 3 = Tendency to Rectify*/
+I21*alpha21+I22*alpha22+I23*alpha23+I24*alpha24
+I25*alpha25+I26*alpha26+I27*alpha27+I28*alpha28
+I29*alpha29+I30*alpha30)

/*COMPOSITE ITEMS
+Co*
/*Beta1 = weight for Norm Violation*/
(I1*alpha1*Beta1+I2*alpha2*Beta1+I3*alpha3*Beta1
```

```
+I4*alpha4*Beta1+I5*alpha5*Beta1+I6*alpha6*Beta1
+I7*alpha7*Beta1+I8*alpha8*Beta1+I9*alpha9*Beta1
+I10*alpha10*Beta1

/*Beta2 = weight for Worrying*/
+I11*alpha11*Beta2+I12*alpha12*Beta2+I13*alpha13*Beta2
+I14*alpha14*Beta2+I15*alpha15*Beta2+I16*alpha16*Beta2
+I17*alpha17*Beta2+I18*alpha18*Beta2+I19*alpha19*Beta2
+I20*alpha20*Beta2

/*Beta3 = weight for Tendency to Rectify*/
+I21*alpha21*Beta3+I22*alpha22*Beta3+I23*alpha23*Beta3
+I24*alpha24*Beta3+I25*alpha25*Beta3+I26*alpha26*Beta3
+I27*alpha27*Beta3+I28*alpha28*Beta3+I29*alpha29*Beta3
+I30*alpha30*Beta3

/*constant*/
+Beta0));

/*INVERSE LOGIT TRANSFORMATION*/
p=ex/(1+ex);
model y~binary(p);
Random Theta~Normal(0, VarTheta) Subject=Person;
run;
```

Remember that in this application, we had 10 item families, 3 latent item predictors, and one item per component per item family. The dummy variables I1-I30 are used to select the correct α, and the dummy variable 'Co' is used to select the correct part of the formula: the part for the component items or the part for the composite items.

Also the SAS program for the RW-MIRID is exemplified with the source code for the guilt example. In the code it is assumed that the Component 'Worrying' has a random weight. The code used is the following:

```
Title 'RW-MIRID, the weight of Worrying is assumed to be random ';
Proc NlMixed data=RWMIRID method=gauss noad
```

```
technique=NewRap Qpoints=15 Optcheck;
Parms alpha1-alpha30=1 Beta1-Beta3=1 Beta0=1 VarTheta=1
CovThetaRWBeta2=1 VarRWBeta2=1;

ex=exp(-theta
/*Co: dummy variable denoting the type of item: 0 for component items, 1 for
composite items*/
/*COMPONENT ITEMS*/
+(1-Co)*
/*Component 1 = Norm Violation*/
(I1*alpha1+I2*alpha2+I3*alpha3+I4*alpha4+I5*alpha5
+I6*alpha6+I7*alpha7+I8*alpha8+I9*alpha9+I10*alpha10

/*Component 2 = Worrying*/
+I11*alpha11+I12*alpha12+I13*alpha13+I14*alpha14
+I15*alpha15+I16*alpha16+I17*alpha17+I18*alpha18
+I19*alpha19+I20*alpha20

/*Component 3 = Tendency to Rectify*/
+I21*alpha21+I22*alpha22+I23*alpha23+I24*alpha24
+I25*alpha25+I26*alpha26+I27*alpha27+I28*alpha28
+I29*alpha29+I30*alpha30)

/*COMPOSITE ITEMS
+Co*
/*Beta1 = weight for Norm Violation*/
(I1*alpha1*Beta1+I2*alpha2*Beta1+I3*alpha3*Beta1
+I4*alpha4*Beta1+I5*alpha5*Beta1+I6*alpha6*Beta1
+I7*alpha7*Beta1+I8*alpha8*Beta1+I9*alpha9*Beta1
+I10*alpha10*Beta1

/*Beta2 = Mean weight for Worrying*/
+I11*alpha11*Beta2+I12*alpha12*Beta2+I13*alpha13*Beta2
+I14*alpha14*Beta2+I15*alpha15*Beta2+I16*alpha16*Beta2
+I17*alpha17*Beta2+I18*alpha18*Beta2+I19*alpha19*Beta2
+I20*alpha20*Beta2
```

```
/*Random weight part: as the mean is already modeled above, the random weight of Worrying, as modeled here, is the deviation per person from this mean weight (the deviation θp2). It has a mean of zero.*/
+I11*alpha11*RWBeta2+I12*alpha12*RWBeta2
+I13*alpha13*RWBeta2+I14*alpha14*RWBeta2
+I15*alpha15*RWBeta2+I16*alpha16*RWBeta2
+I17*alpha17*RWBeta2+I18*alpha18*RWBeta2
+I19*alpha19*RWBeta2+I20*alpha20*RWBeta2


/*Beta3 = weight for Tendency to Rectify*/
+I21*alpha21*Beta3+I22*alpha22*Beta3+I23*alpha23*Beta3
+I24*alpha24*Beta3+I25*alpha25*Beta3+I26*alpha26*Beta3
+I27*alpha27*Beta3+I28*alpha28*Beta3+I29*alpha29*Beta3
+B30*alpha30*Beta3

/*constant*/
+Beta0));

/*INVERSE LOGIT TRANSFORMATION*/
p=ex/(1+ex);
model y~binary(p);
Random Theta RWBeta2 ~Normal([0, 0],
[VarTheta, CovThetaRWBeta2, VarRWBeta2]) Subject=Person;
run;
```

To fit a OPLM-MIRID or a RW-OPLM-MIRID, the term ‘theta’ in the code has to be replaced with the term ‘(a*theta)’, with ‘a’ corresponding to the known discrimination value of the current item.

2.9 References

Bechger, T. M., Verhelst, N. D., & Verstralen, H. H. F. M. (2001). Identifiability of nonlinear logistic test models. *Psychometrika*, *66*, 357-371.

Bechger, T. M., Verstralen, H. H. F. M., & Verhelst, N. D. (2002). Equivalent linear logistic test models. *Psychometrika*, *67*, 123-136.

Birnbaum, A. (1968). Some latent trait models. In F. M. Ford & M. R. Novick (Eds.), *Statistical theories of mental test scores* (p. 397-424). Reading: Addison-Wesley.

Bunday, B. D. (1984). *Basic optimisation methods.* London: Arnold.

Butter, R. (1994). *Item response models with internal restrictions on item difficulty.* Unpublished doctoral dissertation, Catholic University of Leuven, Leuven.

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

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495-515.

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.* Bern, Switzerland: Huber.

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

Maris, G., & Bechger, T. (2003). *Equivalent MIRID models* (Measurement and research department reports 2003-2). Arnhem, The Netherlands: CITO.

SAS System V8 for Windows (Computer program). (1999). Carey NC, USA: SAS Institute Inc.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Pinheiro, J. C., & Bates, D. M. (1995). Approximation to the log-likelihood function in the nonlinear mixed-effects model. *Jounal of Computational and Graphical Statistics*, *4*, 12-35.

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

Rijmen, F., & De Boeck, P. (2002). The random weights linear logistic test model. *Applied Psychological Measurement*, *26*, 271-285.

Smits, D. J. M., & De Boeck, P. (2003). A componential model for guilt. *Multivariate Behavioral Research*, *38*, 161-188.

Smits, D. J. M., De Boeck, P., & Verhelst, N. D. (in press). *Estimation of the MIRID: A program and a SAS based approach.* Behavior research methods, instruments, and computers.

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

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage Publishers.

Verbeke, G., & Molenberghs, G. (in press). The use of the score test for inference on variance components. *Biometrics*, *59.*

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* (p. 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). The Netherlands, Arnhem: CITO.

Wolfinger, R. (1999). *Fitting nonlinear models with the new NLMIXED procedure.* SUGI 24 Conference proceedings, Paper 287. Cary, NC.

Chapter 3

Examining the structure of concepts: using interactions between items

Dirk J. M. Smits
Paul De Boeck
Machteld Hoskens

ABSTRACT A framework is presented for modeling the relational structure of concepts using IRT models with interactions between the items, so-called models with local item dependency (LID). The proposed approach works for unidimensional as well as for multidimensional concepts. In order 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 of the components with 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. [1]

3.1 Introduction

Psychological concepts often contain different components. For example, Mischel and Shoda (1995) conceive of personality as "a stable system that mediates how the individual selects, construes and processes social information and gen-

[1]Correspondence concerning this manuscript should be addressed to: Dirk J. M. Smits, K.U. Leuven, Department 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
The research is financially supported by a GOA 2000/2-grant from the K. U. Leuven: 'Psychometric models for the study of personality'

erates 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" (quoted from Mischel & Shoda, 1995, p. 246). So, the concept of personality can be decomposed in subprocesses or components. Another way of looking at the concept of personality is less process-oriented and mainly dimensional instead, like 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 analyzed into what is called facets. For example, for Conscientiousness the facets are Competence, Order, Dutifulness, Achievement Striving, Self-Discipline, and Deliberation (Costa, McCrae, & Dye, 1991; Costa & McCrae, 1995). Many other psychological concepts can be decomposed into more basic aspects in a similar way, including emotion concepts. In the appraisal theory, for example, different emotions are supposed to be built upon 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, & Schure, 1989; Frijda & Zeelenberg, 2001; Izard, 1993; Omdahl, 1995; Reisenzein & Hofmann, 1993; Roseman, Antoniou, & Jose, 1996; Roseman & Smith, 2001; Scherer, 1993, 1997; Smith & Lazarus, 1993; etc.). 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 each 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 to decompose concepts with. Embreston (1980, 1984), for example, developed the MLTM (Multicomponent Latent

Trait Model). 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, like 3*(4+5), requires two different operations (4+5; 3*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, like for example trait anger. This means that a concept, which contains different components, can be unidimensional and is not necessarily multidimensional. We are not proposing unidimensionality as the most plausible structure, but it is a possibility, one may want to consider. In this article, we focus on an approach based on IRT (Item Response Theory), one that is especially appropriate for relational concepts. By relational concepts, we mean concepts with several components and with a possibly complicated pattern of relations between the components and the global concept. Here we will conceive of these relations as dependencies between the components, and 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) and Hoskens and De Boeck (2001) for the unidimensional and the multidimensional case respectively. These dependencies are called local dependencies, because they are not explained by the global underlying dimensions, for example one general underlying trait. Most often, these local dependencies are treated as problematic, because they complicate the parsimony of a simpler model. We will argue that local dependencies can tell us 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 specify-ing different theories about the relations between components and the concept and translating these relations into IRT models with local dependencies, so that the theories can be tested through the corresponding models.

Local 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 parameters. It is possible to capture such dependencies by fixed effect parameters (constant over all persons) so that there is no need to add person parameters to the model (see section on the models). We make a distinction between the multidimensionality as defined by the number of person parameters and multidimensionality as captured by fixed dependency parameters. We see two clear advantages to a local dependency approach. The first advantage is theoretical and concerns the flexibility and fine-grained nature of dependency models. The patterns of inter-item dependency that are dictated by a theory can be quite complex. Including local dependency parameters is a flexible way to translate a theory into a model without augmenting the number of parameters too much. With local dependency models it is possible to specify in a direct way all kinds of networks of inter-item relations, also networks that can hardly be specified by including more person parameters. The second advantage is practical. It is often cumbersome to estimate models with a high number of person parameters, whereas it is rather easy to estimate local item dependency models.

First, the approach will be explained, and second, an application is described with data collected about guilt feelings 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.

3.2 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 our aim 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, and 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 (Embreston, 1981, 1984). Component items are items for a single component that is assumed to underlie the concept, while composite items are items for the total concept. The test of 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). Our application is based on a questionnaire for situational guilt feelings, each item family is associated with one situation (the common stimulus) and it comprises four items: three component items 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, brooding, and a tendency to restitute. Therefore, the component items for each situation are:

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

and the composite item is:

- 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 our aim 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, since the number of person parameters does not increase.

3.3 The model

As a starting point, we take the Rasch model (Rasch, 1960) for binary data. In this model, the probability of a response xvi to an item i $(i = 1, \ldots, I)$ by person v $(v = 1, \ldots, V)$ can be written as in Equation 3.1:

$$P(X_{vi} = x_{vi} \mid \theta_v, \ \beta_i) = \frac{e^{[x_{vi}(\theta_v - \beta_i)]}}{1 + e^{(\theta_v - \beta_i)}} \tag{3.1}$$

In Equation 3.1 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 we will discuss; see Hoskens and De Boeck (1997) for models with heterogeneous item discrimination. For the interpretation we want to use, 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 on the other items containing any additional information for the probability of responses to the item in question, so that Equation 3.2 holds:

$$P(X_{v1} = x_{v1}, \ \ldots, \ x_{vI} \mid \theta_v) = \prod_{j=1}^{I} P(X_{vi} = x_{vi} \mid \theta_v) \tag{3.2}$$

If Equation 3.2 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 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). We will follow this latter approach, because we want to keep the explicit dimensionality restricted and the random-effect approach needs an extra dimension per group of dependent items. However, for other applications, the random-effect approach may be the one to be preferred.

Following Hoskens and De Boeck (1997) and when following this fixed-effect LID approach, the interaction (another term for dependency) between items can be constant or dimension dependent. Constant interaction is interaction 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, we will concentrate here on constant interaction. We have actually tested also dimension-dependent models, but without success, since they did not explain our data any better. This means that there are no individual differences in the degree of LIDs.

Table 3.1 shows the basic model formulation for the case there is constant interaction between a pair of items i and j.

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

TABLE 3.1. Model for constant pairwise interaction.

Response pattern $(x_{vi},\ x_{vj})$	Adjusted formula
(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)	$exp\,[(\theta_v - \beta_i) + (\theta_v - \beta_j) - \beta_{int}]\,/v\,(\theta)$

Note: β_{int} is the interaction parameter for each item pair, and $\mathrm{v}(\theta) = 1 + exp\,(\theta_v - \beta_i) + exp\,(\theta_v - \beta_j) + exp(2\theta_v - \beta_i - \beta_j - \beta_{int})$.

The implication of this interaction model is that the item parameters i and j are difficult to interpret, because they are no pure reflections of the difficulty anymore, 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, we will pursue the fixed-effect LID approach.

3.3.1 Model estimation

All the models presented by Hoskens and De Boeck (1997) and the ones we will present below can be estimated with existing IRT-programs like CONQUEST (Wu, Adams, & M, 1997), LOGIMO (Kelderman & Steen, 1993), or MULTILOG (Thissen, 1988). The Appendix of Hoskens and De Boeck (1997) 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, in press).

3.3.2 Testing the fit of the model

If two models are nested, a likelihood-ratio test can be used. 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 test the absolute goodness-of-fit, a bootstrap method (Efron & Tibshirani, 1993) will be used, as explained in the section on Estimation.

3.3.3 Specific interaction structures

Consider a questionnaire with the following structure: J item families (index $j = 1, \ldots, J$), each with K component 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 component item k from family j, while 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. We will describe four structures: a linear-sequence structure, a star structure, a cluster structure, and an item-family structure. The dependencies will be represented with arrows, 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 endpoint in the process, and that it interacts only with the 'last' component. The result is a linear-sequence structure as represented in Figure 3.1.

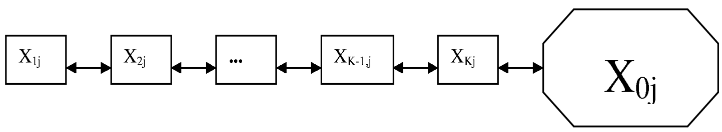


FIGURE 3.1. A linear-sequence structure: The size of the interactions between consecutive components and the concept can be different within and over item families (item families are denoted with index j)

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 3.2.

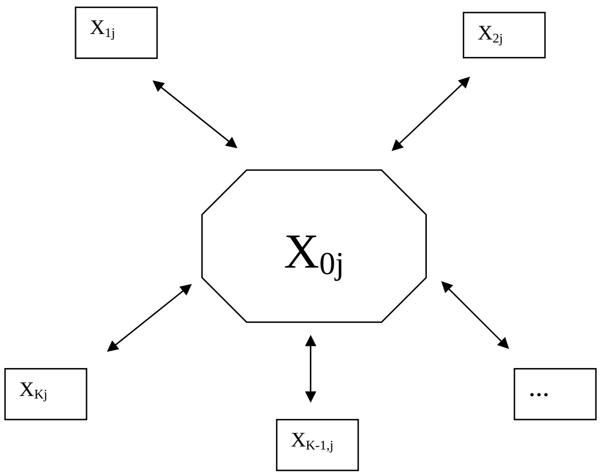


FIGURE 3.2. A star structure: The size of the interactions of the components with the concept can be different within and over item families (item families are denoted with index j)

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, while 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 3.3. An extension of this structure which we will not consider 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 3.4, 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 our application, the phenomenon is feeling guilty. A linear-sequence structure suggests that feeling guilty is an end product of the

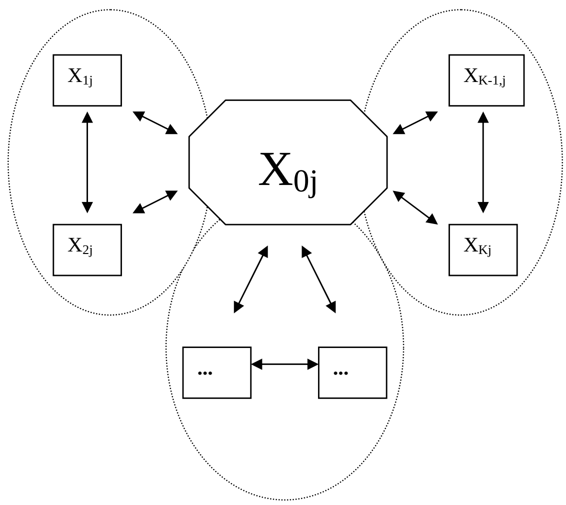


FIGURE 3.3. Cluster structure: The size of the interactions within each cluster between components and the concept can be different within and over item families (item families are denoted with index j)

components norm violation, brooding, and tendency to restitute, and that these components are ordered in a linear way in how they affect feeling guilty. For example, authors like 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 brooding (a covert act), whereas they both precede the tendency to restitute (an overt act), and guilt is the end product, which correspond with a linear-sequence structure. Note that not only a psychological order can lead to a linear-sequence structure, but that also other orders like, for example, 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 (brooding and the tendency to restitute), each complemented with feeling

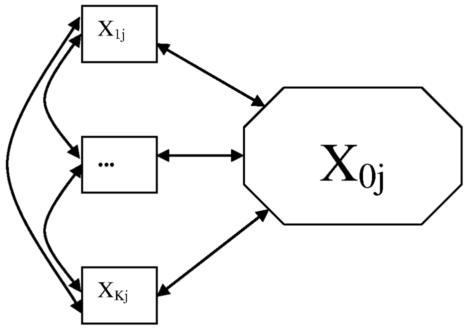


FIGURE 3.4. Item-family structure: The size of the interactions between the components and the concept is equal within an item family, but can be different over item families (item families are denoted with index j)

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 an action tendencies can be separate 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 dimension, as many dimensions are defined (and estimated) as there are item families. In our application, there are 10 item families, but in other cases there may be more.

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

3.4 Application: Modeling of guilt feelings

3.4.1 Data and preliminary analysis

We have selected 10 situations 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 were asked to describe three situations they felt guilty about, each stemming from a different domain of life: (1) work or study situation, (2) personal relationships, and (3) leisure time. 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. Subsequently, ten stories were selected using the following six criteria: understandability, equal representation of the three domains of life, variation in content and assumed guilt inductive power, conformity with the environment of 18-years old people, equal representation of stories stemming from males of females. The selected situations are listed in the Appendix. In the article, we will use 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 were omitted. The first component, self-responsibility, was omitted as it was found that the ratings were fairly constant over all the judges. Apparently, for our 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 we use. 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. Since norm violation seems more important from the literature on guilt, we decided to omit the negative self-evaluation appraisal. In sum, only one appraisal component will be included, next to two action-tendency components: brooding, and tendency to restitute.

The data we will analyze 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, 18-year old high-school students (130 males and 138 females) who answered on a four-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. In order to find out whether a model with one general latent trait complemented with LIDs would make a chance, we did a principal component analysis. For the non-dichotomized and the dichotomized data, the eigenvalues for the first 12 principal components are mentioned in

Table 3.2. From both PCAs, it may be concluded that there is a dominant first component. Therefore, we will start out with a model that has only one person parameter, but is supplemented with LIDs.

TABLE 3.2. Eigenvalues of the first 12 Principal Components of the non-dichotomized and the dichotomized data

Principal Component	Non-dichotomized 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

In the following, we will formulate various instantiations of the types of dependency models we presented earlier. 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, we will further present the PCA results, 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 we 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, it is required that the LIDs are strong enough in terms of explained variance and that clear item structures can be delineated from the PCA. Since we aim at illustrating a theory-based approach, we will continue with LIDs that are defined a-priori, in order to compare their goodness of fit.

3.4.2 Modeling

We will investigate five different relational structures. 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 order we expect is that the appraisal (norm violation) comes before the action tendencies and that the action tendency for a covert act (brooding) precedes the action tendency for an overt act (restitution), and that the feeling of guilt follows. This order is indicated as N-B-R-Guilt in Table 4. However, we also tried all possible orders with guilt in the last position. 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 do not only interact with guilt feelings but also with components of a similar kind. To define similarity of components, we group them 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: brooding 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 upon the hypothesis that guilt has a partially situation-specific meaning, possibly varying in degree depending on the situation. This model has equal interaction 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 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 3.4.2, 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.

We have described six dependency models, 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, we will explain the parameterization of the four basic types of models, 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. If we take the example from Figure 3.4.2, then we obtain the parameterization for one item family as shown in the upper part of Table 3.3. 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, B, R, and G (N for norm violation, B for brooding, R for tendency to restitute, and G for guilt); β_{NBj}, β_{BRj}, β_{RGj} represent the interaction parameters. As mentioned, in Table 3.3 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.

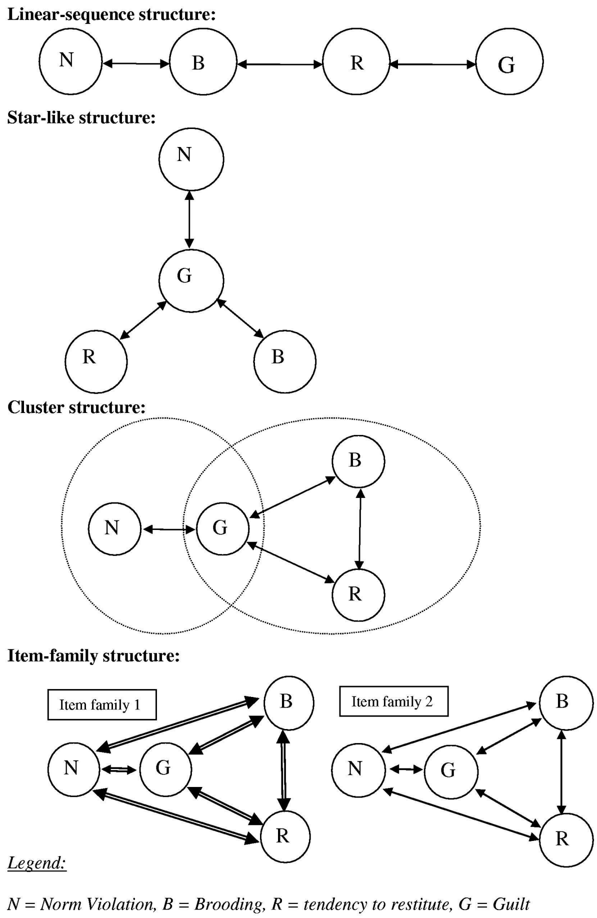


FIGURE 3.5. The interaction models for guilt

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 3.3 for three response patterns. The principle for construct-

TABLE 3.3. Parameterization of dependency models

Model	Norm violation	Brooding	Restitution	Guilt	Parameterization
Linear-sequence structure	1	1	0	0	$\exp(2\theta_v - \beta_{Nj} - \beta_{Bj} - \beta_{NBj})/v(\theta)$
the N-B-R-G sequence	0	1	1	0	$\exp(2\theta_v - \beta_{Bj} - \beta_{Rj} - \beta_{BRj})/v(\theta)$
	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	$\exp(2\theta_v - \beta_{Bj} - \beta_{Gj} - \beta_{BGj})/v(\theta)$
	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	$\exp(2\theta_v - \beta_{Bj} - \beta_{Gj} - \beta_{BGj})/v(\theta)$
	0	0	1	1	$\exp(2\theta_v - \beta_{Rj} - \beta_{Gj} - \beta_{RGj})/v(\theta)$
	0	1	1	0	$\exp(2\theta_v - \beta_{Bj} - \beta_{Rj} - \beta_{BRj})/v(\theta)$
Item-family structure	1	1	0	0	$\exp(2\theta_v - \beta_{Nj} - \beta_{Bj} - \beta_{INTj})/v(\theta)$
	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	$\exp(2\theta_v - \beta_{Bj} - \beta_{Rj} - \beta_{INTj})/v(\theta)$
	0	1	0	1	$\exp(2\theta_v - \beta_{Bj} - \beta_{Gj} - \beta_{INTj})/v(\theta)$
	0	0	1	1	$\exp(2\theta_v - \beta_{Rj} - \beta_{Gj} - \beta_{INTj})/v(\theta)$

ing 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 brooding, tendency to restitute, and guilt feelings. The parameterization is exemplified for four response patterns in the third part of Table 3.3.

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 3.3 for six response patterns.

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), in order 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.

3.4.2.1 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 Aitken (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 inter-item correlations of our data will be compared to the inter-item correlations as expected from those models. Therefore, a bootstrap methodology will be used (Efron & Tibshirani, 1993): Based on the parameter estimates obtained under our models, we will generate 500 new datasets for each model. These replicated data are used to derive the 95% confidence interval for each inter-item correlation. The empirical pairwise inter-item correlations will be compared with these 95% confidence intervals in order to see which model leads to a similar pattern of inter-item correlations as our data. The proportion of empirical inter-item 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.

3.4.3 Results

In Table 3.4 the results for the different LID models are summarized. The three best fitting models are indicated with an asterisk.

TABLE 3.4. Fit of the various models: main-effects model, linear-sequence structure, star structure, cluster structure, item-family structure, and two control models

Model	-2LogL	Number of par.	AIC
No interactions	10545.7	41	10627.7
Linear-sequence structures			
N-B-R-Guilt	9805.8	71	9947.8
N-R-B-Guilt	9609.8	71	9751.8
B-N-R-Guilt	9860.4	71	10002.4
B-R-N-Guilt	9756.4	71	9898.4
R-N-B-Guilt	9611.0	71	9753.0
R-B-N-Guilt	9702.0	71	9844.0
Star structure	9481.1	71	9623.1
Cluster structure	9365.5	81	9527.5
Cluster structure			
with int. constant over situations*	9432.8	45	9522.8
Item-family structure*	9329.3	51	9431.3
Control models			
Order of presentation model		See N-B-R-Guilt above	
Situational-consistency model*	9366.5	42	9450.5

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-B-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-B order (norm violation, tendency to restitute, and brooding) seems the best. The star structure does slightly better than the linear-sequence structure, but the cluster structure outperforms both. Therefore, also the restricted variant with all interaction parameters constant over situations was tested. The AIC-value of this restricted variant is lower than the one of the non-restricted 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 non-restricted version mentioned above, but the difference is only minor. From these results the cluster structure, as well as the item-family structure seem reasonable structures for the concept of guilt.

In order to further investigate the goodness of fit of these three models, the previously presented bootstrap methodology was followed. For the cluster structure, the 95% confidence intervals cover 87% of the empirical correlations in total, but only 65% of the empirical correlations within the same situation. The corresponding percentages for the situational-consistency structure are 87% and 70%, which is not much better than for the previous model. Finally, for the item-family structure, the corresponding percentages are 89% and 92%. 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 3.5.

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 3.5 shows the parameter estimates of the component items (norm violation, brooding, 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 is. 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 5) are the situations 2 (trumpet) and 10 (homework), whereas the highest guilt inducing situations are the situations 1 (break-up) and 6 (secret), which are both about close relationships whereas the situations 2 and 10 are not. As explained earlier, the item parameters cannot safely be interpreted as difficulties, since they depend on the interaction. In case the item parameters correspond with the item difficulties, a perfect neg-

TABLE 3.5. Estimated values for the item and interaction parameters and their standard errors (for the item-family structure)

Item family	Norm violation		Brooding		Tendency to restitute		Guilt		Interactions	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
1.	1.82	.13	-.06	.15	1.10	.13	.70	.14	-.85	.03
2.	3.73	.23	2.59	.18	2.75	.18	2.80	.19	-1.33	.05
3.	2.41	.14	1.12	.13	1.86	.13	1.34	.13	-1.05	.03
4.	2.06	.13	1.54	.14	1.54	.14	1.09	.14	-1.16	.03
5.	1.93	.16	1.20	.18	.25	.21	.89	.19	-1.32	.03
6.	0.60	.23	1.48	.19	1.96	.17	.42	.24	-1.46	.03
7.	1.95	.15	.66	.18	2.82	.14	.85	.18	-1.49	.03
8.	1.92	.14	2.18	.14	1.55	.14	1.78	.14	-1.23	.03
9.	2.85	.13	.43	.17	-1.17	.26	1.18	.15	-1.13	.03
10.	2.44	.15	2.84	.16	2.05	.14	2.84	.16	-1.35	.03

ative 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.

In order to interpret the interactions in terms of Item Characteristic Curves (ICCs), we have compared the ICCs of the main-effects model (Rasch model) with these of the item-family structure. 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 which contain a 1-response for item 1, so $P(Yv11 = 1|\theta) = P(1000|\theta) + P(1100|\theta) + P(1010|\theta) + P(1001|\theta) + P(1110|\theta) + P(1011|\theta) + P(1101|\theta) + P(1111|\theta)$, where $P(1000|\theta)$ is $P(Y_{vN1} = 1,\ Y_{vB1} = 0,\ Y_{vR1} = 0,\ Y_{vG1} = 0|\theta)$, etc.

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. This is because all interactions are positive. 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 2PL model (although the marginal ICCs deviate slightly from the logistic form). This may explain why in an analysis of our data, 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 inter-item dependencies we want to study. 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 3.9, 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

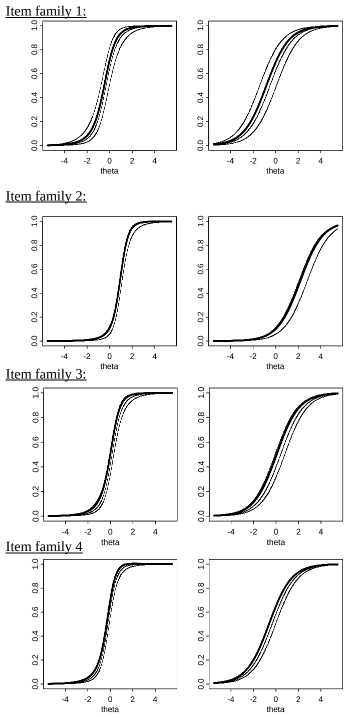


FIGURE 3.6. ICCs for the items per item family: for the item-family structure (left column) and the main-effects models (the Rasch model) (right column), item family 1 to 4

estimated with the Rasch model and the person parameters as estimated with the item-family structure, is .99.

As mentioned earlier, from a PCA we obtained one dominant principal component. The loadings of the items on this component were all positive and varied from .63 to .27 for the non-dichotomized data, and from .58 to .15 for the dichotomized data. Based on a scree test, one could choose either for a one-component solution or for 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 non-dichotomized and the dichotomized data, respectively, whereas the corresponding loadings of the items on their situational component

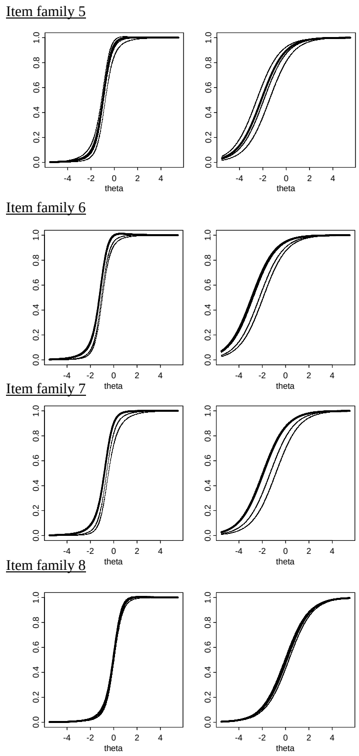


FIGURE 3.7. ICCs for the items per item family: for the item-family structure (left column) and the main-effects models (the Rasch model) (right column), item family 5 to 8

varied from .86 to .62 and from .84 to .50, respectively. A large majority of the cross-loadings was smaller than .1.

Like in the item-family structure, the dominant unrotated component can be interpreted 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 in to a PCA structure, this is not necessarily the case, for example because the number of PCA components would be higher or because there would be

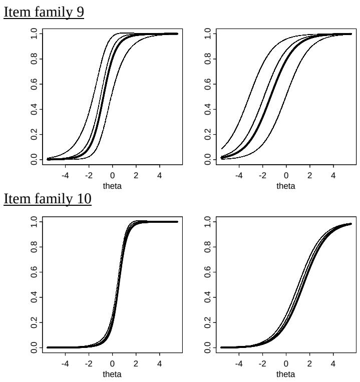


FIGURE 3.8. ICCs for the items per item family: for the item-family structure (left column) and the main-effects models (the Rasch model) (right column), item family 9 and 10

too much overlap between the PCA components. It is an empirical result that the PCA could have told us what the kind of structure was, but this is not by definition so.

3.5 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 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 inter-item correlations overall, but it 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 also this structure 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

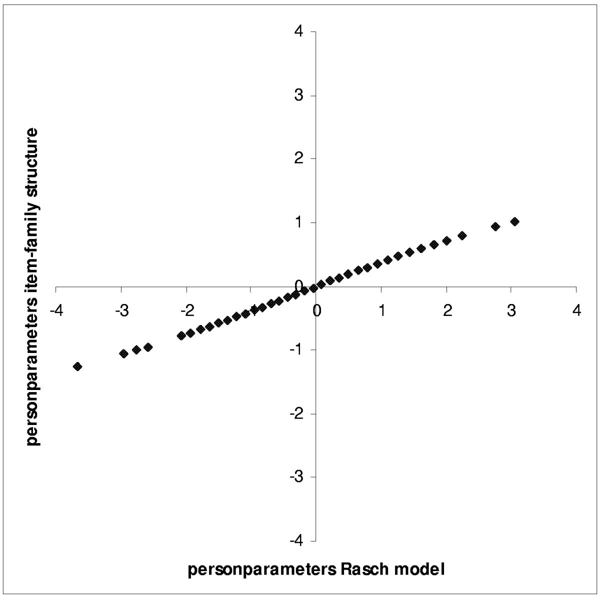


FIGURE 3.9. Person parameters of the Rasch model versus person parameters of the item-family structure (empirical Bayes estimates)

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 gives us the opportunity to model situational specificity, while still 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 in addition as many as there are situations (for the item-family structure), and even more for the other dependency patterns.

Testing LID models, as we did, 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 we have gathered. 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 we gain insight 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 our case, the conclusion is that the structure of guilt is situation specific to some extent.

However, it can not be excluded that the structure we found stems from response consistency within situations, but also in that case, LID modeling is 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 we followed is similar to a structural equation modeling (SEM, Everitt & Dunn, 1991; Du Toit, Du Toit, Jöreskog, & Sörbom, 1999) approach with correlated error terms for what we called the LIDs. 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 (Muthen & Muthen, 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 our results. However, when the dependency pattern is more complicated, or when a rather simple structure has to be compared with more complicated ones, as in our study, it may be appealing to choose an approach that does not increase the number of random effects. The approach we have followed corresponds to what is called the conditional approach in the statistical literature as an alternative for the random-effect approach and the marginal approach (Diggle, Heagerty, Liang, & Zeger, 2002; Fahrmeir & Tutz, 2001).

In the kind of LID models we have used any pattern of inter-item dependencies can be specified, also 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 for our data that 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 where a person has a .5 probability of answering the item correctly. However, due to the LIDs, this interpretation is not valid any more, since 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 3.4.2. 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 family 2 and 10) whereas for some other item families, one can see a shift to the right (item family 5, 6 and 7). Situations 2 and 10 are the weakest when it comes to inducing guilt (column 8, Table 3.5), 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.

3.6 Appendix

The ten descriptions we selected in the first study are listed below (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/she was in love with you (and was taking the relationship very seriously). The break-up hurts

him/her considerably. (Break-up)

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 shoestore. One day, a mother with four children enters the store. One of the kids wants Samson-shoes (Samson is a popular doll figuring in a Belgian TV-series for children). The mother leaves the child with you while she goes on 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/her to go to the movies. You tell him/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/her, but you pretend not to see it. (Discussion)

6. A friend tells you something in confidence, and adds that he/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/her. Halfway he falls from the rope, and he passes out. (Youth movement)

8. You have a pen pal. You get bored of writing with him/her, and suddenly, you stop corresponding with him/her. After one and a half year, he/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 though a lot of trouble finishing it. You get a good grade for your assignment, the same grade as your friend. (Homework)

3.7 References

Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Proceedings of symposium on applications of statistics* (p. 27-41). Amsterdam, The Netherlands: 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* (p. 245-275). New York: Academic Press.

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

Barrett, K. C. (1995). A functionalist approach to shame and guilt. In K. W. Fischer (Ed.), *Self conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (p. 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* (p. 255-273). New York: Guilford Press.

Bock, R. D., & Aitken, 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, Cermak, C., & 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: Oxford University Press.

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

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

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

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

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

Embreston, 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* (p. 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: University Press.

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

Frijda, N. H., Kuipers, P., & Schure, E. ter. (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* (p. 141-155). 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 item response models for polytomous items. *Applied Psychological Measurement*, *2*, 19-37.

Izard, C. E. (1978). Guilt, conscience and morality. In C. E. Izard (Ed.), *Human emotions* (p. 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.

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.

Kempf, W. F. (Ed.). (1977). *Structural models of thinking and learning.* Bern, Switzerland: 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* (p. 274-300). New York: Guilford Press.

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., p. 139-153). New York: Guilford Press.

Mischel, 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.

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

Omdahl, B. L. (1995). Cognitive appraisal theories of emotions. In B. L. Omdahl (Ed.), *Cognitive appraisal, emotion, and empathy* (p. 37-90). New Jersey: Lawrence Erlbaum.

Ortony, A., Clore, G. L., & Collins, A. (Eds.). (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. (Ed.). (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: 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. (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* (p. 3-19). Oxford: University Press.

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: testing theoretical predictions across cultures. *Cognition and Emotion*, *11*, 113-150.

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

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* (p. 114-139). New York: Guilford Press.

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. A. (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.

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

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

Chapter 4

Marginal approaches to the Linear Logistic Test Model

Dirk J. M. Smits
Paul De Boeck
Geert Molenberghs

ABSTRACT Binary data stemming from questionnaires are often analyzed with Item Response Theory (IRT) models, like the LLTM (Linear Logistic Test Model, Fischer, 1977; Fischer & Molenaar, 1995). The parameters of the LLTM have an interpretation conditional on the value of the person parameter. When interested in the effect of certain variables at the level of the population, or when only interested in the item difficulty structure of a questionnaire, a marginal version of the LLTM can be used instead. With marginal models, the effects of the item covariates can be investigated separately from the correlations between the responses. In contrast to the more common random-intercept LLTM, the correlations between the responses can be modeled in a very flexible way, and violations of the assumed association structure do not necessarily influence the estimated effects of the item covariates. Three different ways of approaching the associations between the responses are discussed: marginal correlations, conditional log odds ratios, and marginal log odds ratios. Finally, the approach is illustrated with an example on guilt feelings.[1]

[1]Correspondence concerning this manuscript should be addressed to: Dirk J. M. Smits, K.U. Leuven, Department 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
The research is financially supported by the IAP P5/24 from the federal OSTC, Belgium, and by a GOA 2000/2-grant from the K. U. Leuven: 'Psychometric models for the study of personality'

4.1 Introduction

Social scientists often make use of data stemming from questionnaires to investigate their research questions. There are various possibilities for the response format of a questionnaire, but many are categorical. We will discuss modeling techniques for categorical response data, in particular for binary data (Yes/No, Correct/Incorrect, etc.), although the technique can be easily extended to polytomous data. An approach that is quite popular for modeling binary data in such a context is Item Response Theory (IRT). In most IRT models it is assumed that the probability for a person to give a 1-response to an item is a function of a person-specific parameter and an item-specific parameter. The simplest model of IRT is the Rasch model (Rasch, 1960), which comprises one parameter per person, often called the ability, and one parameter per item, often called the item difficulty. The response a person gives to a certain item is explained based on an effect specific to the person and an effect specific to the item. In some IRT models, for example in the Linear Logistic Test Model (LLTM, Fischer, 1977; Fischer & Molenaar, 1995), the item effects are explained based on some known item properties or item covariates. The responses are explained in terms of a person effect and the effects of some known item covariates. In fact, the Rasch model can be seen as a special case of the LLTM in which the item covariates are dummy variables for each item.

When formulated within a marginal maximum likelihood framework (Baker, 1992), the LLTM can be conceived as a generalized linear mixed model or a random-effect model (McCulloch & Searle, 2001): The person parameters are assumed to be normally distributed with a certain mean and a certain variance, whereas the effects of the item covariates are fixed. The person parameter is a random effect and, more in particular, a random intercept. Therefore, we will call this model the *random-intercept LLTM* (RI-LLTM). An explicit expression for the RI-LLTM is:

$$\operatorname{logit}\left[P\left(Y_{ij}=1 \mid \theta_i\right)\right]=\theta_i+\sum_{k=1}^{K} q_{jk}\eta_k \tag{4.1}$$

with $i = 1, \ldots, I$ the index for the person,
$j = 1, \ldots, J$ the index for the item,
$k = 1, \ldots, K$ he index for the item covariate,
θ_i the person effect or random intercept: $\theta_i \sim N\left(\mu_\theta, \sigma_\theta^2\right)$,
q_{jk} the value of item j on item covariate k,

and η_k the effect of item covariate k.

The effects of the item covariates denote the influence of the covariates conditional on the value of the random intercept. In other words, the effects of the item covariates are to be interpreted as an effect at the level of the individual. The effects of the item covariates are *individual-level effects*, and not population-level effects (Diggle, Heagerty, Liang, & Zeger, 2002; Fahrmeir & Tutz, 2001; Hardin & Hilbe, 2003; Neuhaus, Kalbfleisch, & Hauck, 1991; Zeger, Liang, & Albert, 1988). The same holds for other IRT models with random effects. Also for all general linear mixed models using conditional maximum likelihood (CML) (Baker, 1992) a similar interpretation of the effects of the item covariates holds. In the CML formulation, the likelihood is conditional on the sufficient statistic for the person parameter, the sum score of the person. The result is a conditional model, but without any assumption about the form of the distribution of the person parameter. Therefore, the effects of the item covariates correspond to the effect of an item covariate conditional on the sum score of the person. Since there is a one-to-one mapping between the person parameters and the values of the sufficient statistic, the effects of the covariates have the same meaning as in the RI-LLTM with an MML formulation.

4.2 Marginal Models

For models with a nonlinear link function and with random affects, such as the logistic model, the effect of covariates on the mean over individuals (the population-level effect) is not equal to the mean effect of the covariates for each of the individuals (the individual-level effect) (Diggle et al., 2002; Fahrmeir & Tutz, 2001; McCulloch & Searle, 2001; Neuhaus et al., 1991; Snijders & Bosker, 1999). Models explicitly designed for population-level effects are *marginal models*. Marginal models are useful when inferences about the population(s) are the focus of interest. Therefore they are sometimes also called *population-averaged models*. The term 'marginal' emphasizes that the mean response is the marginal probability of giving a 1-response to a certain item. This can be linked to the common practice in classical test theory to use the *percentage correct* for an item as an estimator of the probability of a correct response. This percentage correct is a marginal statistic. In contrast, the random-effect models have parameters for the effects of covariates on the conditional probabilities conditional on the value(s) of the random effect(s).

Marginal models ought not to be confused with a so-called marginal maximum likelihood (MML) formulation of a model (Baker, 1992; Bock & Aitken, 1981; Bock & Lieberman, 1970). The MML formulation is an integration over terms that contain conditional covariate effects and random effect(s). The covariate effects correspond to the mean over the effects and not to the effect on the mean.

As will be explained later, some marginal models have the very attractive feature that the estimates of the covariates effects are consistent regardless of the correct specification of the association structure, whereas others do not. Marginal models that result in consistent estimates regardless of the correct specification of the association structure, provide a very flexible way to deal with this association structure. The structure of the item difficulties can be investigated independent of the structure of the individual differences. Our primary motivation to look at marginal models is to estimate the item difficulty structure when one may expect associations between items beyond those that can be explained by an underlying latent trait (the random intercept), as will be illustrated in the application. It is not surprising that such associations would occur in an LLTM concept. In the LLTM the item effects are explained in terms of item covariates. These covariates often define groupings of items on the basis of common properties. These properties may be a source of extra correlation.

However, as is already mentioned, the estimates of a marginal model are different from those of a random-effect model. As will be explained later, the marginal effects have a reduced value in comparison with the conditional effects. However, it is not true that marginal model parameters are 'biased' relative to random-effect model parameters. In fact, the true population parameters are different between marginal models and random-effect models.

4.2.1 Model

The marginal modeling approach is described, among others, by Aerts, Geys, Molenberghs, and Ryan (2002), Diggle et al. (2002), and Fahrmeir and Tutz (2001). The marginal version of the LLTM (M-LLTM) consist of two parts: (1) a generalized linear model in which the marginal probabilities are related to the item covariates by the logit link function, called the *mean structure* (See, Equation 4.2), and (2) a model for the associations between the observations, which is often also a generalized linear model, called the *association structure*

(See, Equation 4.3). For example, one can assume equal correlations, a different correlation for each item pair, etc.

The formula for the mean structure of the M-LLTM can be written as follows:

$$\operatorname{logit}\left[P\left(Y_{ij}=1\right)\right]=\sum_{k=1}^{K} q_{jk}\eta_k^* \tag{4.2}$$

with η_k^* the marginal effect, to be distinguished from the conditional effect η_k in Equation 4.1. Another difference is the omission of the random intercept θ_i. The fact that $\eta_k \neq \eta_k^*$ results from the omission of θ_i. Note that the models that will be discussed, also allow for person covariates. However, as this possibility is beyond the scope of the current manuscript, it will not be discussed here any further.

As for the associations structure, one has to decide first on how the association structure is formalized: in terms of marginal correlations, conditional log odds ratios (log odds ratios conditional upon zero-responses on all other items), or marginal log odds ratios (not conditional upon responses on other items). These are the more common parameterizations. Other options exist, but will not be discussed here. In all three just mentioned approaches, the association parameters (correlations or log odds ratios) can be considered a function of some association covariates. Denoting the association parameter for two items j and h with γ_{ijh}, the formula for the association structure can be written as follows:

$$f\left(\gamma_{ijh}\right)=\sum_{m=1}^{M} z_{jhm}\alpha_m \tag{4.3}$$

with $m = 1,\ ...,\ M$ the index for the association covariates,
z_{jhm} the value of association covariate m for the association between the responses to the items j and h,
α_m the effect of association covariate m,
and $f\left(.\right)$ a link function to link the association parameter γ_{ijh} to the association covariates.

Higher-order generalizations of the association parameter to more than two items can be denoted with subscripts added to z for all items involved in the association in question. Note that the models to be discussed also allow for person-specific association covariates. This feature will not be discussed here any further as it is beyond the scope of the current manuscript.

In the literature, three association structures are commonly used: (a) an *independence structure*, which means that no association is assumed between the responses, so that $\alpha_m = 0$, for all m; (b) an *exchangeable association structure*, which means that all pairwise associations are equal, so that $\gamma_{ijh} = \alpha$ for all pairs $j \neq h$, whereas all higher order associations are equal to zero, and (c) an *unspecified association structure*, which means that all pairwise associations can have a different value, so that $\gamma_{ijh} = \alpha_{jh}$, whereas all higher order associations equal zero. Other association structures can be constructed as well, depending on a specific hypothesis one has about the association structure.

4.2.2 Estimation

To estimate the M-LLTM, two choices are to be made: (1) which kind of association parameter one wants to use (correlations or log odds ratios), and (2) which kind of estimation approach one prefers. For the latter, two options will be discussed: a full-likelihood based approach and a Generalized Estimation Equations based approach (GEE). The former can be used for marginal correlations and conditional log odds ratios, and the latter for marginal correlations and marginal log odds ratios.

Various software tools are available for the estimation. Different association structures can be easily estimated with one and the same software tool. For example, PROC GENMOD in SAS V8; XtGee in Stata, Oswald in S-Plus; ALR, Geepack, or Yags in R, Mareg (Fieger, Heumann, & Kastner, 1996; Kastner, Fieger, & Heumann, 1997; Kastner, Heumann, & Fieger, 1999) can be used for the estimation of the M-LLTM. See Horton and Lipsitz (1999) and Ziegler and Grömping (1998) for more information about programs that can be used for the estimation of M-LLTM.

To summarize, the central idea is to model the marginal expectation of each binary variable in terms of item covariates (mean structure, Equation 4.2) and the associations between responses in terms of association covariates (association structure, Equation 4.3). Both sets of covariates can show some overlap, but they do not have to.

Before discussing the different options to model and to estimate the M-LLTM, we will first discuss the relation between conditional and marginal parameter estimates, and second, an approach with a hybrid parameterization that combines a random-intercept model with a separate association structure specified in terms of conditional log odds ratios. This hybrid model is in spirit close to the 'mixed model' of Fitzmaurice and Laird (1993), see later.

4.3 Comparison with other models

4.3.1 Marginal and conditional parameters

The effects of the item covariates as estimated with the RI-LLTM are *conditional parameters* (conditional on the random intercept), denoted with η_k as in Equation 4.1. They can be related to the effects of the same item covariates as estimated with the M-LLTM, called *marginal parameters* and denoted with η_k^* as in Equation 4.2. Neuhaus et al. (1991) showed that the marginal parameters (η_k^*) are always closer to zero than or equal to the corresponding conditional parameters (η_k) (equality holds only if $\eta_k = 0$) and that the discrepancy between η_k and η_k^* increases with the variance of θ. Zeger et al. (1988) derived an approximate relation between the two:

$$\eta_k^* \approx \left(1 + c^2\sigma_\theta^2\right)^{-1/2} \eta_k \tag{4.4}$$

where the variance of θ is equal to σ_θ^2, and $c = 16\sqrt{3}/\left(15\pi\right)$, so that $c^2 \approx .346$ (See also, Diggle et al., 2002; McCulloch & Searle, 2001). For normal-ogive models, $\eta_k^* = \left(1 + \sigma_\theta^2\right)^{-1/2} \eta_k$ (Snijders & Bosker, 1999) meaning that $c = 1$ for these models, whereas for logistic models $1/c$ is close to the well-known multiplicative factor 1.7 to approach a normal-ogive model with a logistic model.

4.3.2 Random-effect models

In random-effect models such as the RI-LLTM, one conditional equation suffices to model both the effects of the item covariates and the association structure of the data (Equation 4.1). Extensions of the RI-LLTM into random-slope versions (Rijmen & De Boeck, 2002) are more flexible with respect to the association structure, whereas they still explain the effects of the item covariates and the association structure with one equation. Also the random effects testlet model of Bradlow, Wainer, and Wang (1999) is of this type. In all random-effect models, it is assumed that given the person parameter, there are no remaining associations between the responses. This assumption is called 'local stochastic independence' (LSI, Lord & Novick, 1968). This assumption is violated if more complicated associations between the responses exist that cannot be explained with random effects. Neglecting the residual associations leads to a distortion of the parameter estimates (Thissen, Steinberg, & Mooney, 1989; Tuerlinckx & Boeck, 2001; Yen, 1993).

4.3.3 Conditional models

Also *conditional models* are an alternative that can take into account these remaining associations (Diggle et al., 2002; Fahrmeir & Tutz, 2001; Hoskens & De Boeck, 1997, 2001; Verhelst & Glas, 1993, 1995; Wilson & Adams, 1995). Conditional models are models in which observed variables are modeled conditional upon one or more other observed random variables (Fahrmeir & Tutz, 2001). The term 'conditional' refers here to conditioning on the observed values of other random variables. This may not be confused with the term 'conditional' when referring to covariate effects in a random-effect model. In the latter case conditional means conditional on the random effect(s). Unfortunately, these models suffer from two problems: First, they are not upward compatible in that the item parameters do not have the same interpretation as there LSI counterparts (Ip, 2002; McCullagh & Nelder, 1989). The condition of reproducibility, which is that the joint distribution of a subset of responses depends only on the parameters specific for the selected subset, is not fulfilled since the effect parameters cannot be reproduced independent of the association structure. For example, the item parameter is not equal any more to the value on the person parameter scale that corresponds with the .5 probability of giving a 1-response, since the probability of giving a 1-response depends also on the dependency parameters. Second, these models for local item dependencies inherit the problem mentioned for the random-effect models, that the estimates of covariate effects are distorted when actually the item responses are dependent beyond what can be expected based on the random effects and the local item dependency parameters (misspecification of the association structure).

4.3.4 Hybrid models

A solution for both problems can be found in a *hybrid models* as developed by Ip (2002), starting from a log-linear representation of the multinomial distribution of the responses (see also, Fitzmaurice & Laird, 1993; Fitzmaurice, Laird, & Rotnitzky, 1993; Molenberghs & Ritter, 1996). For two binary items, the probability distribution function is shown in Equation 4.5.

$$P\left(Y_{i1}=y_{i1},Y_{i2}=y_{i2}|\theta_i\right)= \tag{4.5}$$
$$P\left(00|\theta_i\right)^{(1-y_{i1})(1-y_{i2})}\,P\left(10|\theta_i\right)^{y_{i1}(1-y_{i2})}\,P\left(01|\theta_i\right)^{(1-y_{i1})y_{i2}}\,P\left(11|\theta_i\right)^{y_{i1}y_{i2}}$$

where $P\left(00|\theta_i\right)$ is a short notation for $P\left(y_{i1}=0,\ y_{i2}=0|\theta_i\right)$, and $P\left(00|\theta_i\right)+P\left(10|\theta_i\right)+P\left(01|\theta_i\right)+P\left(11|\theta_i\right)=1$. Taking the log on both sides, and regrouping terms leads to the following parameterization (Ip, 2002):

$$\log\left[P\left(Y_{i1}=y_{i1},Y_{i2}=y_{i2}|\theta_i\right)\right]=y_{i1}\omega_{i1}+y_{i2}\omega_{i2}+y_{i1}y_{i2}\omega_{i12}-\kappa(\theta_i) \tag{4.6}$$

where $\omega_{i1}=\log\left[P\left(10|\theta_i\right)/P\left(00|\theta_i\right)\right]$, $\omega_{i2}=\log\left[P\left(01|\theta_i\right)/P\left(00|\theta_i\right)\right]$, $\kappa\left(\theta_i\right)=-\log\left[P\left(00|\theta_i\right)\right]$, and $\omega_{i12}=\log\left[P\left(00|\theta_i\right)P\left(11|\theta_i\right)/P\left(10|\theta_i\right)P\left(01|\theta_i\right)\right]$. This model is called the generalized log-linear model (GLLM, Holland, 1990; Laird, 1991). For a given θ_i, the GLLM belongs to the exponential family of distributions with the canonical parameters $\boldsymbol{\omega}_i=\left(\omega_{i1},\ \omega_{i2},\ \omega_{i12}\right)$ (Andersen, 1980; Lehmann, 1983).

The GLLM can be used to specify many different IRT models: models that allow for local item dependencies, and models that do not allow for local item dependencies. For example, setting ω_{i12} equal to zero, and modeling ω_{ij} as $\theta_i+\sum_{k=1}^{K}q_{jk}\eta_k$, the RI-LLTM model is obtained. Setting ω_{i12} equal to γ, the constant interaction model of Kelderman (1984) and Hoskens and De Boeck (1997) is obtained. A generalization of Equation 4.6 to more than two items can be found in Ip (2002). It is clear that the effects of the item covariates are influenced by the specified association structure. For example, if the model-based probability of giving the same response to both items (00 or 11) increases due to a local item dependency, also the interaction parameter ω_{i12} will increase, resulting in lower probabilities for the response patterns with different responses (10 and 01). Since ω_{i1} and ω_{i2} equal the log odds of 10 versus 00 and of 01 versus 00 respectively, it is clear that they are both affected by the local dependency of the items 1 and 2.

Ip (2002) solved this problem with a mixed parametrization, in terms of mean parameters and canonical parameters. ω_{i1} and ω_{i2} are transformed into their respective mean parameters, $\mu_{i1}=P\left(Y_{i1}=1|\theta_i\right)$ and $\mu_{i2}=P\left(Y_{i2}=1|\theta_i\right)$ and ω_{i12} is retained, resulting in a hybrid parameterization in terms of $\left(\mu_{i1},\mu_{i2},\omega_{i12}\right)$ (see also, Fitzmaurice & Laird, 1993; Fitzmaurice et al., 1993; Molenberghs & Ritter, 1996). The model is called hybrid, because of its mixed parameterization in terms of mean parameters and canonical parameters. This model is a partly but not a fully marginal model. The mean parameters are not marginal over the random effect, but they are marginal over the responses to other items, and therefore they are invariant under the association structure. The association structure is isolated in a separate aspect of the model. The effect parameters are therefore both conditional parameters (conditional on the random effect) and free of the problems associated with the conditional local item dependency models.

The hybrid model for the LLTM reads as follows: logit $(\mu_{ij}) = \theta_i + \sum_{k=1}^{K} q_{jk}\eta_k$, logit $(\mu_{ih}) = \theta_i + \sum_{k=1}^{K} q_{hk}\eta_k$, and $\omega_{ijh} = \sum_{m=1}^{M} z_{jhm}\alpha_m$.

The advantage of this parameterization is that the effects of the item covariates are not influenced by the specified association structure (Ip, 2002). It offers the possibility of modeling the item difficulty structure independent of the association structure as reflected in the local item dependencies. As such, it is an attractive approach that meets our concern of studying the item difficulty structure in a way that is not dependent on the specific association structure. However, a serious disadvantage is that no closed form solution exists for the joint distribution in terms the marginal and canonical parameters (Fitzmaurice et al., 1993; Ip, 2002). A consequence is that the estimation becomes computationally cumbersome if the number of items, or the number of association parameters is large.

4.4 Association structure

Seven characteristics will be used to compare the different kinds of M-LLTM that will be discussed below.

1. The availability of fit characteristics. For the full-likelihood based approaches, one can use likelihood based measures of goodness of fit, for example the deviance, the AIC (Akaike's Information Criterion, Akaike, 1977), the BIC (Bayesian Information Criterion, Schwartz, 1978). However, for the GEE approaches, these statistics cannot be used.

2. The kind of association parameter used (marginal correlations, conditional log odds ratios, or marginal log odds ratios).

3. The interpretability of the association parameters.

4. The constraints on the association parameters. For some models, the parameter space of the association parameters is constrained by the marginal probabilities. This is the case for all models using marginal association parameters (marginal correlation and marginal log odds ratios), but the degree of the constraints differs.

5. Consistency of the effects estimates. Some approaches yield consistent estimates regardless of the correct specification of the association structure, whereas others do not.

6. Feasibility depending on the number of items. Some approaches are computationally feasible only for a small number of items, whereas other approaches can be used for moderate to high number of items.

7. Reproducibility. The issue is whether the joint distribution of a subset of responses depends on the whole parameter set (not reproducible) or only on the parameters specific for the selected subset (reproducible). Models that are not reproducible require that all subjects receive the same number of items, at least by design.

For all models to be discussed as instantiations of these two approaches, the seven characteristics are summarized into Table 4.1. We will first discuss full-likelihood based approaches, and next the Generalized Estimation Equations (GEE) based approaches.

A feature related to the fifth characteristic (the consistency of the estimates of the effects of the item covariates) is the estimation of the error variance of the effects of the covariates. Even for marginal models that result in consistent estimates for the effects of the item covariates, regardless of the correct specification of the association structure, it holds that the model-based estimates of the error variances will be biased if the association structure is not correctly specified. To solve this problem so-called sandwich estimators (e.g., Hardin & Hilbe, 2003; Liang & Zeger, 1986; Pan, 2001a; Royall, 1986; White, 1982) or jackknife estimators (e.g., Kastner & Ziegler, 1999; Lipsitz, Dear, & Zhao, 1994) are developed, which do not suffer from this misspecification problem. The sandwich estimator is sometimes called the 'empirical variance estimate,' since it combines the variance estimate from the model with a variance matrix constructed from the data. For marginal models in which a misspecification of the association structure affects the parameters of the mean structure, the sandwich estimator does not solve the just mentioned misspecification problem.

TABLE 4.1. Characteristics of M-LLTM

Characteristic	M-LLTM variants						
	Quadrat. exp.	Mixed par.	GEE1 correl.	GEE2 correl.	GEE1 log odds	GEE2 log odds	ALR
1. Fit statistics	Likelihood based: Deviance, AIC, BIC	Likelihood based: Deviance, AIC, BIC	Barnhart and Williamson (1998), Horton et al. (1999), Pan (2001a, 2001b, 2002), Rotnitzky and Jewell (1990), Williamson et al. (2003), Zheng (2000)	none	none	none	none
2. Association parameters	marginal correlations	conditional log odds ratios	marginal correlations	marginal correlations	marginal log odds ratios	marginal log odds ratios	marginal log odds ratios
3. Interpretability	easy	difficult	easy[1]	easy	easy[1]	easy	easy
4. Constrained by marginal probabilities	strong	no	strong	strong	moderately	moderately	moderately
5. Consistency if associations misspecified	no	yes	yes	no	yes	no	yes
6. Maximum number of items	very small	very small	large	small	large	small	large
7. Equal number of items/subject required	yes	yes	no	no	no	no	no

Note[1]: In principle, the association parameters of the GEE1 approaches yield an easy interpretation, but as mentioned for GEE1, the working assumptions about the associations between the responses may not be interpreted as the real association structure.

4.4.1 Full-likelihood based approaches

4.4.1.1 The Bahadur model

A first attempt to model the mean structure and the association structure separately was undertaken by Bahadur (1961). In his model, the joint distribution of a response pattern is expressed in terms of marginal probabilities, pairwise correlations, and higher-order correlations. The joint distribution for a response pattern $\mathbf{Y} = \mathbf{y}$ can be written as:

$$\prod_{j=1}^{J} \mu_j^{y_j} \left(1 - \mu_j\right)^{(1-y_j)} \tag{4.7}$$
$$\left(1 + \sum_{j<h} \rho_{jh} e_j e_h + \sum_{j<h<l} \rho_{jhl} e_j e_h e_l + \ldots + \rho_{12\ldots J} e_1 e_2 \ldots e_J\right)$$

where $\mu_j = E\left(y_j\right) = P\left(y_j = 1\right)$; $e_j = \left(y_j - \mu_j\right) / \left[\mu_j \left(1 - \mu_j\right)\right]^{\frac{1}{2}}$; and $\rho_{jh} = E\left(e_j e_h\right) = r\left(y_j, y_h\right), \ \ldots, \ \rho_{12\ldots J} = E\left(e_1 e_2 \ldots e_J\right)$.
Thus, the joint distribution can be evaluated in closed form in terms of the $2^J - J - 1$ marginal correlations and higher order moments. One can put restrictions on the different patterns of correlations by setting some of the pairwise and higher-order correlations to zero. A drawback of this approach is that the marginal correlations among binary responses are constrained in complicated ways by the marginal probabilities. By consequence, modeling the marginal probabilities in terms of some item covariates, will not lead to a model in which the marginal correlations are independent of the item covariate effects as would be convenient (Diggle et al., 2002; Fahrmeir & Tutz, 2001; Fitzmaurice et al., 1993). A general study of this phenomenon is given in Declerck, Aerts, and Molenberghs (1998). Note that this model is not mentioned in Table 4.1, as to our knowledge, it is of interest primarily for theoretical reasons.

4.4.1.2 Models with a log-linear form

An alternative to Bahadur's representation is the log-linear specification, which assumes that the joint distribution of a response pattern $\mathbf{Y}_i$ is specified as follows (following the parameterization of Fitzmaurice & Laird, 1993): Suppose that all individuals receive the same number of items (J) with a binary response format. The resulting response pattern of a person is $\mathbf{Y}_i = \left(y_{i1}, \ \ldots, \ y_{iJ}\right)^T$. Second, assume that for every item we have the values of K item covariates

(q_{jk}). Assume that the marginal distribution of a binary response y_{ij} is:

$$\begin{aligned} f\left(y_{ij}|q_{jk}\right) &= \mu_{ij}^{y_{ij}}\left(1-\mu_{ij}\right)^{1-y_{ij}} \\ &= \exp\left\{y_{ij} - \log\left[1+\exp\left(\nu_{ij}\right)\right]\right\} \end{aligned} \tag{4.8}$$

where $\nu_{ij} = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \sum_{k=1}^{K} q_{jk}\eta_k^*$ and $\mu_{ij} = E\left(y_{ij}\right) = P\left(y_{ij}=1|q_{jk},\ \eta_k^*\right)$

Now, the joint distribution of a response pattern $\mathbf{Y}_i$ is of log-linear form:

$$f\left(\mathbf{Y}_i,\ \Psi_i,\ \Omega_i\right) = \exp\left[\Psi_i^T\mathbf{y}_i + \Omega_i^T\mathbf{w}_i - A\left(\Psi_i,\ \Omega_i\right)\right] \tag{4.9}$$

where Ψ_i and Ω_i are vectors of canonical parameters:
$\Psi_i^T = \left(\psi_{i1},\ldots\psi_{iJ}\right)^T$ with $\psi_{ij} = \text{logit}\left[P\left(y_{ij}=1 \mid y_{il}=0,\ l\neq j\right)\right]$: the conditional probability given that the remaining responses $y_{il}, l\neq j$ are all zero and $\Omega_i^T = \left(\omega_{i12},\ldots,\omega_{iJ-1\ J},\ldots,\omega_{i12\ldots J}\right)^T$ can be interpreted in terms of conditional log odds ratios as

$\omega_{ijh} = \log\left[\frac{P(y_{ij}=1,\ y_{ih}=1 \mid y_{il}=0,\ l\neq j,h)P(y_{ij}=0,\ y_{ih}=0 \mid y_{il}=0,\ l\neq j,h)}{P(y_{ij}=1,\ y_{ih}=0 \mid y_{il}=0,\ l\neq j,h)P(y_{ij}=0,\ y_{ih}=1 \mid y_{il}=0,\ l\neq j,h)}\right]$

is a sum of log odds ratios given that the remaining responses $y_{il}, l\neq j,h$ are all zero, $\mathbf{w}_i = \left(y_{i1}y_{i2},\ldots,y_{iJ-1}y_{iJ},\ldots,y_{i1}y_{i2}\ldots y_{iJ}\right)^T$ is a R x 1 vector of all second and higher-order cross products of responses of a certain person i, and $A\left(\Psi_i,\ \Omega_i\right)$ is a normalizing constant: $\exp\left[A\left(\Psi_i,\ \Omega_i\right)\right] = \sum\exp\left(\Psi_i^T y_i + \Omega_i^T w_i\right)$, where the summation is over all 2^J possible values of $\mathbf{Y}_i$.

This representation will be used for the following two models. The general strategy for both is to transform Ψ_i to marginal probabilities for modeling the effects of the item covariates, whereas Ω_i is used to account for the pairwise and higher-order associations between the responses. Note that the subscript i for the person is retained as the models based on this representation allow for person-specific covariates for the mean structure, as well as for the association structure.

The mixed parameter model of Fitzmaurice and Laird (1993)

In this model, the associations are formalized in terms of conditional log odds ratios. Fitzmaurice and Laird (1993) developed a full likelihood-based method for analyzing correlated binary data, in which the associations between responses are modeled in terms of conditional log odds ratios (see definition of Ω_i). The joint distribution of a response pattern $\mathbf{Y}_i$ is assumed to follow Equation 4.9. The Ψ_i are transformed into marginal probabilities, but to specify the associations between the responses, conditional log odds ratios are used. Unlike the approach of Zhao and Prentice (1990) (next paragraph), the third and higher-order associations are not restricted to be zero. This model is called the *mixed parameter model* as for the specification of the model both marginal and canonical parameters are used.

The mixed parameter model allows for varying degrees of dependence among the responses and also for setting certain association parameters to zero or explaining the association parameters in terms of association covariates. For a discussion of the likelihood equations and an algorithm for estimating the parameters of this model, we refer to Fitzmaurice and Laird (1993).

A first advantage of this approach (see Table 4.1), is that it is a likelihood based approach and therefore likelihood-based fit statistics can be used. Second, unlike the correlations in the Bahadur model and in the quadratic exponential model (see next paragraph), the conditional log odds ratios are not constrained by the marginal probabilities (Fitzmaurice et al., 1993). Third, the estimates for the effects of the item covariates are consistent if the mean structure is correctly specified, regardless of whether the association structure (Ω_i) is correctly specified. This has to be contrasted with the random-effect approach and with the subsequent approach based on marginal correlations, in which the item difficulty structure (mean structure) parameters are affected if the associations among the responses are not correctly specified by the model.

There are also some limitations related to this model: First, the association parameters do not have an attractive or meaningful interpretation. They represent the log odds ratio given that all other responses are zero. However, if the primarily point of interest is the effects of the item covariates, the interpretability issue of the association parameters is of no importance (Fitzmaurice & Laird, 1993). In such a case a simple model for the association structure suffices. When the investigator is also interested in the association structure, using marginal log odds ratios is to be preferred. Second, computations grow exponentially with the number of items, and quickly become impractical (Fahrmeir & Tutz, 2001). Third, the mixed parameter model is not 'reproducible,' meaning that the joint distribution for a subset of items depends on the whole parameter set, and not only on the parameters specific to this subset. If $\mathbf{Y}_i^*$ is a $J^* \times 1$ subset of the response vector $\mathbf{Y}_i$ of subject i, where $J^* < J$, then

$$f\left(\mathbf{Y}_i^*\right) \neq \exp\left[\Psi_i^{*T}\mathbf{y}_i^* + \Omega_i^{*T}\mathbf{w}_i^* - A\left(\Psi_i^*,\ \Omega_i^*\right)\right] \tag{4.10}$$

where Ψ_i^{*T}, Ω_i^{*T}, and $\mathbf{w}_i^*$ are the corresponding subsets of Ψ_i^T, Ω_i^T, and $\mathbf{w}_i$

respectively. Hence, omitting an item will change the interpretation of all conditional log odds ratio parameters. By consequence, the model is only applicable when every subject received the same number of items, at least by design (Fitzmaurice et al., 1993).

The quadratic exponential model of Zhao and Prentice (1990)

Zhao and Prentice (1990) proposed a model called the *quadratic exponential model* in which the third and higher-order associations are set to zero (see also, Gourieroux, Monfort, & Trognon, 1984). It connects as a special case to the model of Fitzmaurice and Laird (1993). In the quadratic exponential model, the association structure is reduced to pairwise associations and modeled in terms of marginal correlations. The canonical parameters Ω_i and Ψ_i are mapped through a one-to-one transformation to the moment parameters $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, i.e. the vector of marginal probabilities and the covariance matrix. The moment parameters $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ are modeled in terms of item covariates and association covariates, which both yield a marginal interpretation.

Zhao and Prentice (1990) derived pseudo maximum likelihood estimates for the effects of the item covariates (η_k^*) and for the effects of association covariates (α_m), which are consistent and asymptotically normal under regularity conditions (Gourieroux et al., 1984).

This approach has two major advantages: First, it is a likelihood based approach, so that one can use likelihood-based fit statistics like the deviance, the AIC, and the BIC. Nested models can be compared with likelihood-ratio tests. Second, the association parameters are modeled in terms of marginal correlations, which are easy to interpret. However, the approach has also serious drawbacks: As for the Bahadur model, also for the quadratic exponential model it holds that the marginal probabilities constrain the marginal correlations in complicated ways if the number of items is larger than 2 ($J \geq 2$) (Diggle et al., 2002; Fahrmeir & Tutz, 2001; Fitzmaurice et al., 1993; Ziegler, Kastner, & Blettner, 1998). Second, the consistency properties for the effects of the item and association covariates depend on the correct specification of both the mean and the association structure (Fitzmaurice et al., 1993; Zhao & Prentice, 1990). Thus, a misspecification of the marginal association structure can lead to incorrect estimates of the effects of the item covariates. Therefore, this model does not meet our concerns. Third, like for the mixed parameter model, this procedure is computationally feasible only for small numbers of items. Fourth, since the canonical parameters depend on the number of responses, the subjects need to receive the same number of items (Fitzmaurice et al., 1993; Zhao & Prentice, 1990) (See Table 4.1).

4.4.1.3 Other full-likelihood based approaches

Note that it is not our aim to give an exhaustive enumeration of all existing marginal models for correlated binary data that can be estimated with a full-likelihood based method. In this paragraph two additional, interesting models will be mentioned. The first model is developed by Ekholm, Smith, and McDonald (1995), Ekholm, McDonald, and Smith (2000), and Ekholm, Jokinen, McDonald, and Smith (in press). In this model, associations are modeled in terms of dependency ratios, which can be interpreted as the probability of observing 1-responses on all items involved in the association compared to the probability of observing a 1-response on each item as can be expected based on an independence model, for example the dependency ratio between two items 1 and 2 equals $P\left(y_1 = 1,\ y_2 = 1\right) / P\left(y_1 = 1\right) P\left(y_2 = 1\right)$. However, as this model does not allow for a flexible modeling of the association structure in terms of item-specific association covariates, it will not be discussed here any further. A second model to be mentioned is the Dale model described in Dale (1986) for bivariate data and extended by Molenberghs and Lesaffre (1994, 1999) to multivariate data. Similar to one of the GEE-approaches to be discussed below, the Dale model models associations in terms of marginal log odds ratios. As the model specification is, except for the estimation method, close to the GEE approach with marginal log odds ratios, it will not be discussed here any further.

4.4.2 Generalized estimation equations based approaches

An alternative to the full-likelihood based methods was given by Liang and Zeger (1986), Zeger and Liang (1986). [2] They developed a multivariate analogue of quasi-likelihood methods as described in Wedderburn (1974) which can lead to computational more simple procedures to estimate the effects of the item covariates and the association covariates. These quasi-likelihood methods are called *Generalized Estimation Equations* (GEE). They require only the

[2] Another non-likelihood method, called empirical generalized least squares, was developed by Koch, Landis, Freeman, Freeman, and Lehnen (1977). However, this approach requires that each covariate pattern is non-sparse, which is not feasible for covariates with many levels or continuous covariates. As such, it will not be discussed here.

specification of the form of the mean function and the way in which the variance of the responses depend upon the mean, but a complete specification of the likelihood function is not required. Therefore, these methods use only part of the information, in contrast to full-likelihood based approaches. In all commonly used GEE approaches the associations are modeled in terms of marginal association parameters: marginal correlations or marginal log odds ratios.

There exist different kinds of GEE. Two kinds will be discussed: first-order GEE (GEE1) and second-order GEE (GEE2). Almost all GEE1 approaches focus on the estimation of the effects of the item covariates (the first moments). The effects of the item covariates on the marginal probabilities are modeled while using working assumptions about the association structure (the second moments). The working assumptions about the association structure specified in terms of association covariates do not need to be correct, nor can they be tested. They are only used to reach more efficient estimates for the effects of the item covariates. A critical feature of GEE1 is that the effects of the item covariates and the effects of the association covariates are treated as orthogonal to one another, even when they are not (Liang, Zeger, & Qaqish, 1992), resulting in consistent estimates for the effects of the item covariates regardless of the correct specification of the association structure.

The GEE2 approaches, on the other hand focus on both, the effects of the item covariates and the effects of the association covariates. The effects of the item covariates and the effects of the association covariates are jointly estimated, and as such they can both be tested. The joint estimation results in more efficient and consistent estimates for both kinds of effects provided that the model for the mean structure and the association structure are correctly specified. The reason is that GEE2 approaches do not treat the two kinds of effects as orthogonal. For GEE2, working assumptions about the third and fourth moments are necessary, but the correctness of these working assumptions does not affect the consistency of the estimates for the effects of item covariates and association covariates. However, the correctness of the working assumption does affect the efficiency of the estimates (Hall & Severini, 1998; Prentice & Zhao, 1991).

The GEE approach in general has three major disadvantages: First, not all information available in the data is used for the estimation of the model: For GEE1, only information about the first moments (the means and the variances) is used, and for GEE2, only information about first and the second moments (the pairwise associations) is used. Information about higher order associations is neglected. The implication is that the efficiency of the estimates is reduced. A second and more serious drawback is that one leaves the framework of the full-likelihood methods, and therefore the likelihood-based fit statistics (deviance, AIC, BIC) no longer apply. Instead, the fit of different models is commonly investigated by looking at the significance of the effects of the covariates, based on Wald type arguments. Recently, some fit statistics specific for GEE are described in literature (e.g., Barnhart & Williamson, 1998; Horton et al., 1999; Pan, 2001a, 2001b, 2002; Rotnitzky & Jewell, 1990; Williamson et al., 2003; Zheng, 2000). Most of them are developed for the GEE1 approach with the working assumptions on marginal correlations as association parameters, and they are not developed yet for models with working assumptions on marginal log odds ratios, such as the model of Carey, Zeger, and Diggle (1993). A generalization toward these models is desirable. Third, as already mentioned, all GEE approaches use marginal association parameters. For binary data, they are constrained to a certain degree by the marginal probabilities (Fitzmaurice et al., 1993; Liang et al., 1992; Ziegler et al., 1998). Note that these constraints are not specific for GEE approaches, but hold for all models using marginal association parameters. Fourth, most GEE1 approaches provide no information about the association structure. In contrast, GEE2 approaches do provide this information, but the price to pay is that they do not yield consistent estimates of the item covariates effects regardless of the correct specification of the association structure. Furthermore, for GEE2 it holds that, due to the non-orthogonality of the effects of the item covariates and the effects of the association covariates, the sandwich estimator for the error variance of the effects of the item covariates and the effects of the association covariates is not always robust to misspecification of the association structure (Hardin & Hilbe, 2003), although for GEE1 it is. Finally, for moderate to high number of items, GEE2 becomes computationally infeasible, as large matrices have to be calculated and inverted (due to the inclusion of third-order and fourth-order moments). However, GEE2 is still computationally less cumbersome than the full-likelihood based approaches (Fitzmaurice et al., 1993; Liang et al., 1992) (see Table 4.1).

In the next paragraphs, we will discuss four approaches: GEE1 and GEE2, each combined with either marginal correlations or marginal log odds ratios as association parameters. As a fifth approach, an extension of the GEE1 approach combined with logistic regressions will be discussed that does allows for the estimation of the association parameters, *and* is computationally feasible for large numbers of items.

4.4.2.1 GEE1 and GEE2 with marginal correlations as association parameters

Liang and Zeger (1986), Zeger and Liang (1986) focus primarily on the effects of the item covariates and follow a GEE1 approach. The associations between the responses are considered nuisance elements in the model. A quasi-score equation is derived for the estimation of the effects. For binary data the estimating equation for the effects, denoted with $U(\boldsymbol{\eta}^*)$, is the following:

$$U(\boldsymbol{\eta}^*) = \sum_{i=1}^{N} Q_i^T \Delta_i V_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0 \tag{4.11}$$

where Q_i is a $J \times K$ matrix with the values of the J items to the K item covariates, $\mathbf{y}_i$ is the response pattern of person i, $\boldsymbol{\mu}_i = E(\mathbf{y}_i) = (\mu_{i1},\ \mu_{i2},\ \ldots,\ \mu_{iJ})^T$, Δ_i is the diagonal of the variance-covariance matrix: $diag\left(\sigma^2_{y_{i1}},\ \ldots,\ \sigma^2_{y_{iJ}}\right)$, and V_i is a working assumption about the covariance matrix of the responses, to be chosen by the investigator. One can express this working covariance matrix as follows:

$$V_i = \Delta_i^{1/2} \Re_i(\boldsymbol{\alpha}) \Delta_i^{1/2} \tag{4.12}$$

where $\Re_i$ is a $J \times J$ working correlation matrix, and $\boldsymbol{\alpha}$ represents a vector of parameters associated with the specified model for $\Re_i$. Different correlation structures can be expressed as $h(\Re_i) = \mathbf{Z}_i \boldsymbol{\alpha} = \sum_{m=1}^{M} z_{ijhm} \alpha_m$, with j and h as item indices, $\mathbf{Z}_i$ is a matrix containing the values of the *association covariates* (to be distinguished from the item covariates), $\boldsymbol{\alpha}$ is the vector containing the effects of the association covariates, and $h(.)$ is a link function to link the correlation matrix to the association covariates, for example the Fischer-Z transformation. Note that the subscript i for the person is retained in the Equations 4.11 and 4.12 as the GEE based models allow for person-specific covariates for the mean structure, as well as for the association structure.

The association structure as modeled with the association covariates, is *only a working assumption*, used to reach more efficient estimates for the effects of the item covariates. They are not the primary interest of this approach. The parameters of the association structure ($\boldsymbol{\alpha}$) are replaced by a $J^{\frac{1}{2}}$-consistent estimator, based on the Pearson residuals and the working assumption, under the assumption that $\boldsymbol{\eta}^*$ is known (Liang & Zeger, 1986; Prentice, 1988; Zeger & Liang, 1986; Ziegler et al., 1998). Iterating between the estimation equation for $\boldsymbol{\eta}^*$ and the estimator for $\boldsymbol{\alpha}$ leads to estimates for the effects of the item covariates and the association covariates. However, as the estimates for $\boldsymbol{\alpha}$ are not jointly estimated with the $\boldsymbol{\eta}^*$ and depend strongly on the assumed working assumptions that do not need to be correct, nor can be tested, the estimates of the effects of the association covariates cannot be interpreted as reflecting the real association structure. As a consequence, most software packages do not estimate the error variances for the effects of the association covariates.

Two attractive features of this GEE1 approach are that it yields consistent estimates for the effects of the item covariates regardless the association structure is correctly specified or not, and that it is not required that every respondent receives the same number of items.

The just described GEE1 framework has been extended by Zhao and Prentice (1990) and Prentice and Zhao (1991) toward an approach in which effects of the item covariates and effects of the association covariates are jointly estimated (GEE2). It leads to estimation equations of the form:

$$\sum_{i=1}^{N} \begin{pmatrix} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}^*} & 0 \\ \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\eta}^*} & \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\alpha}} \end{pmatrix}^T \begin{pmatrix} V_i & C_i \\ C_i^T & B_i \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{s}_i - \boldsymbol{\sigma}_i \end{pmatrix} = 0 \qquad (4.13)$$

where $\mathbf{s}_i$ is the vector containing the elements $s_{ijh} = (y_{ij} - \mu_{ij})(y_{ih} - \mu_{ih})$, $\boldsymbol{\sigma}_i$ is the vector containing the elements $\sigma_{ijh} = E(s_{ijh})$, V_i is the working variance covariance matrix defined as in Equation 4.12, C_i is the working covariance matrix of $\mathbf{y}_i$ and $\mathbf{s}_i$, and B_i is the working covariance matrix of $\mathbf{s}_i$ (Fitzmaurice et al., 1993; Prentice & Zhao, 1991).

The main advantages of this GEE2 approach is that the associations are modeled in terms of marginal correlations, which yield an easy interpretation. A serious drawback is that the estimates of the item covariate effects may fail to be consistent when the association structure is incorrect, even if the model for the mean is correctly specified. Other (dis)advantages are described in the general paragraph on GEE and in Table 4.1. Note that although these GEEs with marginal correlations are actually a moment-based version of the Bahadur model, the constraints on the marginal correlation are less and weaker than for the Bahadur model.

Recently, an alternative to GEE2, called Extended Generalized Estimation Equations (EGEE) is developed that does not suffer from the just described consistency problem (Hall, 1999, 2001; Hall & Severini, 1998). In fact, this method is a special case of GEE1 (Hall, 2001). Since the convergence of this method is not as reliable as other GEE1 approaches (Hall, 1998) and as this method is not currently implemented in available software packages (Hardin & Hilbe, 2003), it will not further be discussed here. Instead, another GEE1-based approach with the desired consistency feature, one with marginal log odds ratios as association parameters will be described, as it too allows for the estimation of the effects of the association covariates (and its error variances), converges more reliably, and is implemented indeed in available software packages. This approach will be presented as a fifth one, after the common GEE1 and GEE2 for marginal log odds ratios are described.

4.4.2.2 GEE1 and GEE2 with marginal log odds ratios as association parameters

An alternative measure for the association between two binary responses are the marginal log odds ratios defined as,

$$\gamma_{ijh} = \log\left[\frac{P\left(y_{ij}=1,\ y_{ih}=1\right)P\left(y_{ij}=0,\ y_{ih}=0\right)}{P\left(y_{ij}=1,\ y_{ih}=0\right)P\left(y_{ij}=0,\ y_{ih}=1\right)}\right] \tag{4.14}$$

This log odds ratio is called a *marginal log odds ratio* because the odds ratio is not conditional on the other items. Its use was suggested by Lipsitz, Laird, and Harrington (1991). Marginal log odds ratios can be seen as a compromise between the conditional log odds ratios, which are unconstrained but which have interpretations that depend on the number of items, and marginal correlations, which are severely constrained by the marginal probabilities. The constraints on marginal log odds ratios are weaker and their interpretation is independent of the number of items (Diggle et al., 2002; Hardin & Hilbe, 2003).

Apart from the likelihood-based estimation method of Molenberghs and Lesaffre (1994) (Dale model), several GEE approaches are developed to model and to estimate the association structure in terms of marginal log odds ratios (See, e.g., Lipsitz et al., 1991; Liang et al., 1992; Carey et al., 1993). If only information is requested about the mean structure, again GEE1 can be used to estimate the effects of the item covariates, see the previous paragraph. The characteristics of this approach are indicated in Table 4.1. When also interested in the association structure of the data, again, one cannot interpret the working assumption about the associations between the responses as a real association structure. Instead, GEE2 can be used for the estimation of the association structure (e.g., Zhao & Prentice, 1990; Liang et al., 1992). However, GEE2 has some serious drawbacks as mentioned in the general discussion of GEE and as indicated also in Table 4.1 for GEE2 with marginal correlations.

4.4.2.3 GEE1 combined with logistic regressions

Following the suggestions by Firth (1992) and Diggle (1992) mentioned in the discussion of the paper of Liang et al. (1992), Carey et al. (1993) have developed an approach that is closely related to GEE1, that can handle large numbers of items and that yields consistent parameters estimates of the mean structure even when the association structure is misspecified, while still the parameters of the association structure (second moments) can be estimated. At first sight, this is in contradiction with the GEE1 approach, but as orthogonality is assumed between the effects of the item covariates and the effects of the person covariates and the effects of the item covariates are estimated with a GEE1 whereas for the estimation of the effects of the association covariates a logistic regression is used, the method is sometimes classified as a GEE1 approach.

For the simplest case, in which there is a constant marginal log odds ratio for all item pairs (γ), the approach alternates between two steps:

1. For a given γ, the effects of the item covariates (η_k^*) are estimated as in a marginal logistic regression model using a GEE1.

2. For a given set of η_k^*, the log odds ratio parameter γ is estimated using a logistic regression of y_{ij} on each y_{ih} $(h > j)$, while using the term $\log\left[\frac{P(y_{ij}=1,\ y_{ih}=0)}{P(y_{ij}=0,\ y_{ih}=0)}\right]$ as an offset (regression coefficient = 1), see Carey et al. (1993) for more details.

The algorithm is referred to as *alternating logistic regressions* (ALR) and is implemented in PROC GENMOD of SAS V8, in the Oswald package for S-Plus (Smith, Robertson, & Diggle, 1996), and in the ALR package for R (Carey, 2002). In the first step of the ALR algorithm, a GEE1 is used for the estimation of the effects of the item covariates (η_k^*). The resulting estimates are consistent even if the association structure is misspecified. Furthermore, these estimates are reasonably efficient if the association structure is well approximated. In a second step of the ALR algorithm, an offset logistic regression is used to estimate the effects of the association covariates. In practice, the algorithm developed by Carey et al. (1993) converges quickly when ordinary logistic regression estimates are used as starting values for the effects of the item covariates and when zero is used as a starting value for the effects of the association covariates (Carey et al., 1993).

The main advantages of this approach are the following (See Table 4.1): First, because of the use of marginal log odds ratios, all parameters have a straightforward interpretation. Second, the ALR approach results in consistent estimates for the effects of the item covariates regardless of the correct specification of the association structure. Third, unlike the GEE2 approach, the ALR approach is computationally feasible for large numbers of items.

The approach has also some limitations: Beside the limitations stemming from the fact that the method is a GEE method, one other limitation is that for binary data with three or more items, also this method suffers from the fact that the marginal log odds ratios are constrained by the marginal probabilities (Liang et al., 1992; Ziegler et al., 1998), although the constraints are only moderate in comparison with those when marginal correlations are used (Diggle et al., 2002; Hardin & Hilbe, 2003).

4.5 Application

We will demonstrate the use of the M-LLTM estimated with two methods: the approach of Fitzmaurice and Laird (1993) (full-likelihood based, mixed parameter M-LLTM) and the ALR approach of Carey et al. (1993) (GEE1, ALR M-LLTM). The dataset to illustrate these methods consist of responses from 268 persons to 40 items of a situation-response questionnaire. The questionnaire comprises 10 situations. For each of the ten situations, questions were asked regarding four different guilt-related reactions: (1) whether one would feel as if a norm is violated in the situation, (2) whether one would brood about what happened in the situation, (3) whether one would feel the tendency to rectify what happened in the situation, and (4) whether one would feel guilty in the situation. These four reactions will be called norm violation, brooding, tendency to rectify and guilt, respectively. As we wanted to fit also a full-likelihood based model, we reduced the number of items by a random selection of three situations resulting in 12 items: four questions for each of the three situations. For more than 12 items the mixed parameter M-LLTM turned out computationally too demanding for a Pentium IV computer with 512 Mb memory. For the GEE1 based ALR M-LLTM and the RI-LLTM, we did not encounter the same problem so that the full dataset was used to estimated the model. Because the analyses based on all data led to similar conclusions as the one for the 12 items, we will report only on the data for three of the situations, so that we can compare the three methods. The three situations were:

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/she was in love with you (and was taking the relationship very seriously). The break-up hurts him/her considerably.

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.

3. During the holidays, you are working as a salesperson in a clothing and shoestore. One day, a mother with four children enters the store. One of the kids wants Samson-shoes (Samson is a popular doll figuring in a Belgian TV-series for children). The mother leaves the child with you while she goes on to look for shoes 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 afterward without having checked whether they fit. 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 does not matter, and that everyone is capable of mistaking the size of shoes.

To begin with, we fitted a random-intercept Rasch model on the 12 items: Fitting a Rasch model, the person parameter (random intercept) can be interpreted as the guilt sensitivity of the corresponding person, and the difficulty parameter as the guilt inducing power of the situation (with a parameterization so that the item parameter is added instead of subtracted in the logit). This model has been compared with a RI-LLTM, with the situations and the type of question as item covariates, assuming there are no interactions between both kinds of item covariates. Three dummy covariates were defined for the four reactions and also two dummy variables for the three situations. The combination of the fourth reaction (guilt) with the first situation functions as the reference level. Both models were fitted with PROC NLMIXED from SAS V8, using an adaptive Gaussian quadrature method with 20 quadrature points as described in Pinheiro and Bates (1995). Table 4.2 shows the deviance, the AIC, and the BIC.

From Table 4.2, one can conclude that the RI-LLTM has a comparatively good fit (lower AIC and BIC). In Table 4.3, one can see that all item covariates

TABLE 4.2. Goodness-of-fit statistics for the random-intercept Rasch model, the RI-LLTM, the mixed parameter marginal Rasch model, and the mixed parameter M-LLTM

Model	Deviance	AIC	BIC
random-intercept Rasch model	3420	3447	3493
RI-LLTM	3430	3444	3469
Mixed parameter Marginal Rasch model, exchangeable associations	967	993	1039
Mixed parameter M-LLTM, exchangeable associations	975	989	1015
Mixed parameter M-LLTM, situational associations	690	710	746

have a significant effect. Note that because the guilt question and situation 1 are used as reference levels, we have no estimates for their effects and their value is fixed to zero. The results indicate that one is less inclined to make the appraisal of norm violation (-.77), and to feel a tendency to rectify (-.26), than to feel guilty. On the other hand, brooding (.29) seems to be more likely than feeling guilty.

As both the random-intercept Rasch model and the RI-LLTM assume that the associations between the items can be explained by one person parameter with the same weight for all items, the closest corresponding marginal Rasch model and M-LLTM assume a constant association between all item pairs. This is called an exchangeable association structure. Furthermore, Neuhaus (1993) showed that a RI-LLTM and a M-LLTM with an exchangeable association structure have about the same estimation efficiency and power to detect effects of item covariates.

Next, we have estimated the marginal Rasch model and the M-LLTM, both, with a likelihood-based approach: the mixed parameter M-LLTM (Fitzmaurice & Laird, 1993), and with a GEE1 approach: the ALR M-LLTM (Carey et al., 1993). The mixed parameter M-LLTM is estimated with WinMareg (Kastner et al., 1997), whereas the ALR M-LLTM is estimated with PROC GENMOD from SAS V8. Standard errors of all effects are estimated with the sandwich estimator (Carey et al., 1993; Fitzmaurice & Laird, 1993; Hardin & Hilbe, 2003; Royall, 1986; White, 1982). Note that we did not include any third-order or higher-order associations, as they can only be estimated with the approach of Fitzmaurice and Laird (1993), and not with the GEE approach.

The deviance, AIC and BIC for the mixed parameter marginal Rasch model and the mixed parameter M-LLTM are mentioned in Table 4.2. They may not be compared with those of the random intercept models. According to these criteria, the marginal Rasch model has a similar fit as the M-LLTM, both with exchangeable associations, which confirms the results obtained with the corresponding random-intercept models. For the ALR M-LLTM or the ALR Rasch model, no fit indices are mentioned, as this approach is not likelihood based, and measures like the deviance, AIC, BIC are therefore not available. As the estimates for the item covariates are very similar to those obtained with a third marginal approach, a discussion of the effects is postponed until this third approach and its goodness of fit is described.

When interested in the association structure, a structure with a constant log odds ratio for all item pairs can seem too restrictive for this application.

TABLE 4.3. Effects of the item covariates for the RI-LLTM and the M-LLTM with an exchangeable and a situation specific association structure (SE = standard error)[1]

Covariate	RI-LLTM		mixed par. M-LLTM				ALR M-LLTM			
			exch.		sit.		exch.		sit.	
	effect	SE	effect	SE	effect	SE	effect	SE	effect	SE
Intercept	.87	.12	.72	.10	.71	.11	.72	.10	.71	.10
Norm violation	-.77	.13	-.65	.09	-.64	.09	-.65	.09	-.64	.09
Brooding	.29	.12	.24	.08	.24	.08	.24	.08	.23	.08
Tendency to rectify	-.26	.12	-.22	.10	-.22	.10	-.22	.10	.22	.10
Guilt	.00	-	.00	-	.00	-	.00	-	.00	-
Situation 1	.00	-	.00	-	.00	-	.00	-	.00	-
Situation 2	-3.00	.13	-2.51	.16	-2.52	.16	-2.51	.16	-2.52	.16
Situation 3	-.95	.10	-0.78	.12	-.78	.12	-.78	.12	-.78	.12

note[1]: The standard error is estimated with the sandwich estimator as described in Carey et al. (1993), Fitzmaurice and Laird (1993), Hardin and Hilbe (2003), which is consistent regardless the correct specification of the association structure.

Alternatively, an association structure with equal between-situation associations and situation dependent but homogeneous within-situation associations is more realistic, as a shared situation can introduce an additional dependency between the responses. This association model can be formalized as follows:

$$\gamma_{jh} = \alpha_{overall} + I_{jh} (\alpha_s) \tag{4.15}$$

I_{jh} is an association covariate which is equal to one if the items j and h are about the same situation, and zero otherwise so that the α_s ($s = 1, \ \ldots, \ S$) are the parameters to model the association induced by situation s. The $\alpha_{overall}$ parameter is meant to reflect the overall association independent of the situational structure. For three situations, this model allows for four different values for the pairwise associations γ_{jh}: three values for the within-situation associations (one for each situation: $\alpha_{overall} + \alpha_1$, $\alpha_{overall} + \alpha_2$ and $\alpha_{overall} + \alpha_3$), and one value for all pairwise associations between items from different situations ($\alpha_{overall}$). We will call this model the M-LLTM with a situational association structure. This model has a clearly better goodness-of-fit than the one with an exchangeable structure, as can be seen in Table 4.2.

As the results for the effects of the item covariates are similar for all three marginal approaches, they will be discussed together. We will use the term 'log odds ratios' for both the conditional log odds ratios (mixed parameter M-LLTM) and the marginal log odds ratios (ALR M-LLTM). In Table 4.3, one can see that all M-LLTM covariates have a significant effect. Note that the estimates from all three M-LLTM approaches are smaller than those from the RI-LLTM, as expected. Applying Equation 4.4, the RI-LLTM effects can be transformed into their M-LLTM counterparts. This leads to results that differ not more than .02 from the effect estimates that are directly obtained with the marginal approaches.

The better approximation of the association structure has not really an effect on the estimated effects of the item covariates (Table 4.3), which is expected because consistent estimates for the parameters of the mean structure are obtained even if the association structure is not correctly specified (Carey et al., 1993; Fitzmaurice & Laird, 1993).

The association parameters for the M-LLTM with an exchangeable association structure and the situational M-LLTM are given in Table 4.4. Note that the association parameter estimates for the mixed parameter M-LLTM are different from the association parameters estimated for the ALR M-LLTM, because in the mixed parameter M-LLTM the associations are modeled with conditional log odds ratios, whereas in the ALR M-LLTM the associations are modeled with marginal log odds ratios. Three conclusion can be drawn from Table 4.4: (1) If the items belong to different situations, their association (log odds ratio) is still positive (.06 and .09), which is in agreement with a random-intercept model. (2) The associations are higher within the same situation. For example, the log odds ratio for situation 1, as derived from the mixed parameter M-LLTM, is equal to .06 + .93 = .99 (or .09 + 1.22 = 1.31 for the ALR M-LLTM). This finding is an indication that the assumption of local stochastic independence of the RI-LLTM is violated. This is precisely the kind of situation where the M-LLTM is useful, since it is far more flexible than a random-effect model when it comes to the association structure. With the M-LLTM these additional associations can be modeled rather easily, whereas for the RI-LLTM, this is more difficult as one either has to include local item dependency parameters into the model or one has to include as many random effects as there are situations plus one for the intercept. (3) In Table 4.4, one can see that the situation seems to influence the dependencies between responses and differently so depending on the situation.

TABLE 4.4. Association parameters estimates from the mixed parameter M-LLTM (conditional log odds ratios) and from the ALR M-LLTM (marginal log odds ratio) with exchangeable association structure, and with situational association structure (SE = standard error)

Log odds ratio	mixed par. M-LLTM				ALR M-LLTM			
	exch.		sit.		exch.		sit.	
parameter	α	SE	α	SE	α	SE	α	SE
$\alpha_{overall}$	.35	.02	.06	.02	.78	.09	.30	.11
α_1			.93	.08			1.22	.19
α_2			1.40	.12			2.19	.28
α_3			1.13	.08			1.69	.19

A serious drawback of the marginal models is that they do not provide a basis for the measurement of persons on one or more underlying latent traits. However, there are certainly circumstances where one wants to test a general theory about how item responses depend on features of the items. This is when a marginal modeling approach is useful, especially if local item dependencies are possible. Additionally, the approach can be also informational about the association structure, also when the dimensionality is rather high, so that also inferences can be made about the structure of individual differences, although without measuring these differences.

4.6 Conclusions

Although binary data are often modeled with IRT models, for research questions concerning population effects or for research questions regarding the item difficulty structure without person measurement purposes, the marginal models with GEE as an estimation approach are a valuable and flexible alternative. An interesting kind of flexibility is that one can allow for complex patterns of associations between responses. These more complex patterns are a serious complication for the random-effect models, since they would require either the inclusion of dependency parameters or multiple random effects. For a small number of items even a full likelihood approach as that one by Fitzmaurice and Laird (1993) can be recommended, whereas for larger number of items, the GEE approach, and especially the ALR approach, provides a valuable alternative.

4.7 References

Aerts, M., Geys, H., Molenberghs, G., & Ryan, L. M. (2002). *Topics in modelling of clustered binary data.* London: Chapman & Hall.

Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Proceedings of symposium on applications of statistics* (p. 27-41). Amsterdam, The Netherlands: North Holland.

Andersen, E. B. (1980). *Discrete statistical models with social science applications.* Amsterdam, The Netherlands: North Holland.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dicotomous items. In H. Solomon (Ed.), *Studies on item analysis and prediction* (p. 158-168). Stanford, California: Stanford University Press.

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

Barnhart, H. X., & Williamson, J. M. (1998). Goodness-of-fit tests for GEE modeling with binary responses. *Biometrics*, *54*, 326-335.

Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of the EM algorithm. *Psychometrika*, *46*, 443459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, *35*, 179-197.

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

Carey, V. (2002). *Alternating logistic regression for multivariate binary data: ALR* [Computer program]. Retrieved Januari 20, 2003 from http://www.biostat.harvard.edu/∼carey/.

Carey, V., Zeger, S. L., & Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, *80*, 517-526.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, *42*, 909-917.

Declerck, L., Aerts, M., & Molenberghs, G. (1998). Behaviour of the likelihood ratio test statistic under a bahadur model for exchangeable binary data. *Journal of Statistical Computations and Simulations*, *61*, 15-38.

Diggle, P. J. (1992). Discussion of 'Multivariate regression analysis for categorical data' by Liang, Zeger & Qaqish. *Journal of the Royal Statistical Society B*, *54*, 28-29.

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

Ekholm, A., Jokinen, J., McDonald, J. W., & Smith, P. W. F. (in press). Joint regression and association modelling for longitudinal ordinal data. *Biometrics.*

Ekholm, A., McDonald, J. W., & Smith, P. W. F. (2000). Association models for a multivariate binary response. *Biometrics*, *56*, 712-718.

Ekholm, A., Smith, P. W. F., & McDonald, J. W. (1995). Marginal regression analysis of a multivariate binary response. *Biometrika*, *82*, 847-854.

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

Fieger, A., Heumann, C., & Kastner, C. (1996). *MAREG and WinMAREG.* Retrieved Januari 30, 2003 from Ludwig-Maximilians-Universitat Munchen, Institut fur Statistik web site: http://www.stat.uni-muenchen.de/∼andreas/mareg/winmareg.html.

Firth, D. (1992). Discussion of 'Multivariate regression analysis for categorical data' by Liang, Zeger & Qaqish. *Journal of the Royal Statistical Society, Series B*, *54*, 24-26.

Fischer, G. H. (1977). *Linear logistic latent trait models: Theory and applications.* Bern, Switzerland: Huber.

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments and applications.* New York: Springer-Verlag.

Fitzmaurice, G. M., & Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, *80*, 141-151.

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

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudomaximum likelihood methods: theory. *Econometrika*, *52*, 681-700.

Hall, D. B. (1998). *Estimation equations for clustered data based on extended quasilikelihood.* Paper presented at the XIXth International Biometric Conference. Cape Town, South Africa.

Hall, D. B. (1999). On GEE-based regression estimators under first moment misspecification. *Communications in Statistics: Theory and Methods*, *28*, 1021-1042.

Hall, D. B. (2001). On the application of extended quasi-likelihood to the clustered data case. *The Canadian Journal of Statistics*, *29*, 77-97.

Hall, D. B., & Severini, T. A. (1998). Extended generalized estimation equations for clustered data. *Journal of the American Statistical Association*, *93*, 1365-1375.

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimation equations.* London: Chapman & Hall/CRC.

Holland, P. (1990). The dutch identity: a new tool for the study of item response models. *Pychometrika*, *55*, 5-18.

Horton, N. J., Bebchuk, J. D., Jones, C. L., Lipsitz, S. R., Catalano, P. J., Zahner, G. E. P., & Fitzmaurice, G. M. (1999). Goodness-of-fit for GEE: An example with mental health service utilization. *Statistics in Medicine*, *18*, 213-222.

Horton, N. J., & Lipsitz, S. R. (1999). Review of software for generalized estimating equation regression models. *The American Statistician*, *53*, 160-169.

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

Hoskens, M., & De Boeck, P. (2001). Multidimensional componential item response models for polytomous items. *Applied Psychological Measurement*, *2*, 19-37.

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

Kastner, C., Fieger, A., & Heumann, C. (1997). MAREG and WINMAREG, a tool for marginal regression models. *Computational Statistics and Data Analysis*, *24*, 235-241.

Kastner, C., Heumann, C., & Fieger, A. (1999). *New features in MAREG 0.2.0.* Retrieved januari 30, 2003 from Ludwig-Maximilians-Universitat Munchen, Institut fur Statistik web site: http://www.stat.uni-muenchen.de/∼andreas/mareg/winmareg.html.

Kastner, C., & Ziegler, A. (1999). *A comparison of jackknife estimators of variance for GEE2.* Sfb 386, nr. 167. retrieved januari 30, 2003 from http://www.imbs.mu-luebeck.de/staff/ziegler/report.html.

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

Koch, G. G., Landis, J. R., Freeman, J. L., Freeman, D. H., & Lehnen, R. G. (1977). A general methodology for the analysis of experiments with repeated measurements of categorical data. *Biometrics*, *33*, 133-158.

Laird, N. M. (1991). Topics in likelihood-based methods for longitudinal data analysis. *Statistica Sinica*, *1*, 33-50.

Lehmann, E. L. (1983). *Theory of point estimation.* California: Wadsworth.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

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

Lipsitz, S. R., Dear, K. B., & Zhao, L. P. (1994). Jackknife estimators of variance for parameter estimates from estimation equations with applications to clustered data. *Biometrics*, *50*, 842-846.

Lipsitz, S. R., Laird, N. M., & Harrington, D. P. (1991). Generalized estimation equations for correlated binary data: using the odds ratio as a measure of association in unbalanced mixed models with nested random effects. *Biometrika*, *78*, 153-160.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Molenberghs, G., & Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using a multivariate plackett distribution. *Journal of the American Statistical Association*, *89*, 633-644.

Molenberghs, G., & Lesaffre, E. (1999). Marginal modelling of multivariate categorical data. *Statistics in Medicine*, *18*, 2237-2255.

Molenberghs, G., & Ritter, L. (1996). Likelihood and quasi-likelihood based methods for analysing multivariate categorical data, with the association between outcomes of interest. *Biometrics*, *52*, 1121-1133.

Neuhaus, J. M. (1993). Estimation efficiency and tests of covariate effects with clustered binary data. *Biometrics*, *49*, 989-996.

Neuhaus, J. M., Kalbfleisch, J. D., & Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, *59*, 25-35.

Pan, W. (2001a). Akaike's information criterion in generalized estimation equations. *Biometrics*, *57*, 120-125.

Pan, W. (2001b). Model selection in estimating equations. *Biometrics*, *57*, 529-534.

Pan, W. (2002). Goodness-of-fit tests for GEE with correlated binary data. *Scandinavian Journal of Statistics*, *29*, 101-110.

Pinheiro, J. C., & Bates, D. M. (1995). Approximation to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*, 12-35.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, *44*, 1033-1048.

Prentice, R. L., & Zhao, L. P. (1991). Estimating equations for parameters in the mean and covariances of multivariate discrete and multivariate responses. *Biometrics*, *47*, 825-839.

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

Rijmen, F., & De Boeck, P. (2002). The random weights linear logistic test model. *Applied Psychological Measurement*, *26*, 271-285.

Rotnitzky, A., & Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, *77*, 485-497.

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review*, *54*, 221-226.

Sas OnlineDoc^{TM} Version 8. (1999). Carey, NC.

Sas System V8 for windows [Computer program]. (1999). Carey, NC.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Smith, D. M., Robertson, W. H., & Diggle, P. J. (1996). *Oswald: Objectoriented software for the analysis of longitudinal data in S* (Technical Report MA

96/192). Department of Mathematics and Statistics, University of Lancaster, LA1 4YF, United Kingdom.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage Publications Inc.

Stata statistical software: Release 5.0 [Computer program]. (1997). College Station, TX.

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

Verhelst, N. D., & Glas, C. A. W. (1993). A dynamic generalization of the rasch model. *Psychometrika*, *58*, 395-415.

Verhelst, N. D., & Glas, C. A. W. (1995). Dynamic generalizations of the rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (p. 181-201). New York: Springer-Verlag.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*, *61*, 439-447.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrika*, *50*, 1-25.

Williamson, J., Lin, H.-M., & Barnhart, H. X. (2003). A classification statistic for GEE categorical response models. *Journal of Data Science*, *1*, 149-165.

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

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

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121-130.

Zeger, S. L., Liang, K. Y., & Albert, P. (1988). Models for longitudinal data: a generalized estimation equation approach. *Biometrics*, *44*, 1049-1060.

Zhao, L. P., & Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika*, *77*, 642-648.

Zheng, B. (2000). Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine*, *19*, 1265-1275.

Ziegler, A., Kastner, C., & Blettner, M. (1998). The generalized estimation equations: an annotated bibliography. *Biometrical Journal*, *40*, 115-139.

Ziegler, A., & Grömping, U. (1998). Generalized estimation equations in commercial statistical software packages. *Biometrical Journal*, *40*, 247-262.

Chapter 5

The inhibition of verbally aggressive behavior

Dirk J. M. Smits
Paul De Boeck
Kristof Vansteelandt

ABSTRACT We studied the inhibition of verbal aggression defined as not displaying verbal aggression, while one would want to. The approach we used was based on a situation-response questionnaire containing 15 anger provoking situations and three verbal aggressive reactions. Two questions were asked for each combination of a situation and a reaction: one about wanting to react in a verbally aggressive way and one about actually displaying the reaction. This questionnaire was administered to 316 participants. The data were analyzed with inhibition conceptualized as a trait. Trait inhibition was negatively correlated with external measures of Anger Out and positively with Control of Anger Out.[1]

5.1 Introduction

Verbal aggression is a rather common but problematic behavior (Infante & Rancer, 1996). It is a common behavior because it is a rather easy and not very dangerous expression of anger, as only words or sounds are involved. As for how problematic verbal aggression is one should differentiate between verbal aggression to oneself or to others. Cursing at oneself, for example, is a possible reaction to one's own behavior when this behavior is considered negative and attributed to oneself. For an outsider this verbal aggression may still be interpreted as unfriendly and as an indication that the verbally aggressive person

[1]Correspondence concerning this manuscript should be addressed to: Dirk J. M. Smits, K.U. Leuven, Department 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 The research is financially supported by a GOA 2000/2-grant from the K. U. Leuven: 'Psychometric models for the study of personality'

is easily irritated or bad-tempered. When directed at others, it can seriously hurt other people, and it may arouse a mutual aggressive reaction and further escalation (DeTurck, 1987; Fitness & Fletcher, 1993; Infante, 1987; Infante & Rancer, 1996; Roloff, 1996).

Not all aggressive inclinations a person experiences are expressed (Averill, 1982, 1983; Kassinove, Sukhodolsky, Tsytsarev, & Solovyova, 1997). One may withhold an aggressive reaction for several reasons, often even without any conscious reflection, for example, because the other person has a higher social rank (Allan & Gilbert, 2001). Here, we are interested in withholding one's verbal aggression defined as wanting to be verbally aggressive (curse, scold, shout), while not behaving accordingly, as 'wanting' without 'doing'. According to King, Emmons, and Woodley (1992), this wanting without doing can be due to two kinds of inhibition: (1) behavioral inhibition, which may be considered as an inhibition of the overt expression of emotional experiences, and (2) emotional inhibition, which may be considered as control over naturally occurring emotional reactions. Inhibition has the advantage that it prevents the problems that may follow from verbal aggression, but it may create new problems. For example, inhibiting verbal aggression by turning one's anger (emotional inhibition) inwards instead of expressing it may have consequences for one's health (Begley, 1994; Culbertson & Spielberger, 1996; Engebretson, Matthews, & Scheier, 1989; Greenglass, 1996; Julkunen, 1996; Martin et al., 1999; Venable, Carlson, & Wilson, 2001). One can assume that avoiding negative consequences is a primary cause for inhibition (see, e.g., Averill, 1982, 1983; Beatty & McCroskey, 1997). For example, one can inhibit the tendency to be verbally aggressive to avoid being disliked by others, or to avoid an aggressive counterreaction of others (Deffenbacher, Oetting, Lynch, & Morris, 1996; Fitness & Fletcher, 1993; Infante & Rancer, 1996).

To situate and delineate the logic of our study, we first make a conceptual analysis of verbal aggression in terms of its constituents. Verbal aggression will be conceptualized here as based on three constituents: the anger feelings, the verbal aggressive action tendency and the verbal aggressive act (see Figure 5.1). The anger feelings are assumed to feed the tendency to be verbally aggressive (Averill, 1983; Cornell, Peterson, & Richards, 1999; Fitness & Fletcher, 1993; Frijda, 1986; Kassinove et al., 1997; Kinney, Smith, & Donzella, 2001), and the action tendency in turn is at the basis of the act (Frijda, 1986). This view implies two links: between the feeling and the action tendency, and between the action tendency and the act.

First, the link between the feeling and the action tendency is not necessarily a sequence, nor is it necessarily causal (Kuppens, Van Mechelen, Smits, & De Boeck, in press). An alternative view is that the tendency is a part of the feeling (Frijda, 1986, 1993; Frijda, Kuipers, & Schure, 1989; Rubin, 1986). The reason for distinguishing between the feelings and the tendency is that different factors may affect the feelings and the tendency. Some factors in the person or in the situation may affect the tendency without affecting (the other part of) the feeling, and vice versa. For example, one way to understand *coping with anger* is that it is of importance for the kind of tendency to which the anger feelings lead. We do not further specify the link between the feelings and the action tendency, other than that the tendency is based on or is part of the feelings, without the feeling necessarily imply the action tendency, since factors may play that counteract the action tendency, but not the (other part of the) feeling. These factors inhibit the action tendency in or following the feelings. They constitute the action tendency inhibition (see Figure 5.1).

Second, the link between the tendency and the act is one with the action tendency at the basis of the act, but again without the tendency leads necessarily to the act. Also here other factors may play. For example, one may fear the reactions of others. These factors can inhibit the act given the action tendency. They constitute the behavioral inhibition (Figure 5.1).

In fact an earlier kind of inhibition may play. One that inhibits the anger feelings, so that the two links we discussed become irrelevant, since the anger feelings do not even arise. This kind of inhibition is called emotional inhibition (Figure 5.1). This emotional inhibition and the behavioral inhibition are the two kinds of inhibition King et al. (1992) discuss.

We will concentrate on behavioral inhibition and therefore on the link between the action tendency and the behavior. Various situations will be presented, and the participants in the study will be asked whether they would want to be verbally aggressive (want to curse, want to scold, and want to shout) and whether they would actually display the corresponding behavior (curse, scold, shout).

Following the previous conceptual analysis as depicted in Figure 5.1, the basis for either displaying verbal aggression or not is twofold: the action tendency and the behavioral inhibition. We will now link the action tendency and behavioral inhibition to factors in the person and in the situation.

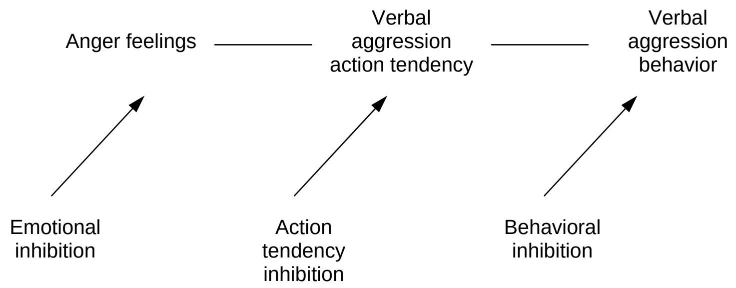


FIGURE 5.1. Conceptual analysis of the verbal aggression process

5.1.1 Person factors

As for the verbal aggression (VA) action tendency, an evident person factor is *trait anger*. Trait anger can be defined as the disposition to experience anger, or in other words, the relative stable individual differences in the tendency to experience anger (Forgays, Forgays, & Spielberger, 1997; Spielberger, Jacobs, Russell, & Crane, 1983). Therefore, we hypothesize that trait anger is positively correlated with the VA action tendency, as indicated in Table 5.1. We have no basis to expect also a correlation with behavioral inhibition.

A second type of factors, one that may play a role in both the action tendency and the behavioral inhibition, refers to *coping with anger*. In the literature on anger, a taxonomy of coping style variables has been discussed and studied (Averill, 1983; Boddeker & Stemmler, 2000; Deffenbacher et al., 1996; Forgays et al., 1997; Spielberger, Johnson, & Jacobs, 1982; Spielberger et al., 1983, 1985; Spielberger, Krasner, & Solomon, 1988). A first coping with anger variable is *Anger Out*, or the tendency to express one's anger outwards. The VA action tendency is an important step toward the outward expression, so that a positive correlation is expected between Anger Out and the VA action tendency (Kinney et al., 2001) (see Table 5.1). Furthermore, we expect a negative correlation between Anger Out and behavioral inhibition, since the behavioral inhibition counteracts the the outward expression.

A second coping with anger variable is *Anger In*, the contrast of Anger Out in that it refers to the tendency to turn one's anger inwards (e.g., bottling anger up, boiling inside). In literature this concept is often denoted as the regulation of the anger by suppressing it (Begley, 1994; Greenglass, 1996; Julkunen, 1996; Schwenkmezger & Hank, 1996). This implies that the VA action tendency is counteracted and therefore we expect a negative correlation with this action tendency (see Table 5.1). Given that anger is turned inwards so that the outward action tendency is prevented, there seems no reason to inhibit the expression of this action tendency. Therefore, we have no basis to expect a correlation between Anger In and behavioral inhibition (see Table 5.1).

A third variable is Anger Control. It refers to the control of one's anger. Van Elderen, Maes, Komproe, and Kamp (1997) have distinguished between Anger Out Control, and Anger In Control. *Anger Out Control* is the extent to which one controls the externalization of anger. Therefore, it may be expected to correlate negatively with the VA action tendency, and positively with behavioral inhibition (the opposite pattern as for Anger Out). *Anger In Control* is the extent to which one controls the internalization of anger. This is a new concept in the literature on coping with anger. Because Anger In Control refers to control of the internalization, we have no basis to expect correlations with the VA action tendency and the behavioral inhibition (see Table 5.1).

A third type of person factors refers to the aggressive style. The aggressive style can be direct or indirect (Averill, 1983; Dehghani & Lange, 1993; Lange, Dehghani, & De Beurs, 1995; Ramanaiah, Conn, & Schill, 1987). Direct Aggression can be defined as the physical or verbal expression of aggression, which can be recognized as such by others. An example from the Buss Durkee Hostility Inventory (Buss & Durkee, 1957; Lange, Hoogendoorn, Wiederspahn, & Beurs, 1995) is 'When angry, I say mean things.' Therefore, we expect that Direct Aggression is positively correlated to the VA action tendency and that it is negatively correlated to the behavioral inhibition (see Table 5.1). Indirect Aggression is covert and would therefore correlate positively with the VA action tendency, which is also covert (Averill, 1982; Frijda, 1986). An example item of the Buss Durkee Hostility Inventory (Buss & Durkee, 1957; Lange et al., 1995) is 'I'm more often irritated then people know.' We do not expect a correlation with behavioral inhibition (see Table 5.1), as the indirect style is not necessarily a consequence of inhibition.

5.1.2 Situation variables

From the situational perspective, differences in the VA action tendency may be seen as related to how frustrating a situation is, and to the value of the aggressive act. The value can be of an instrumental or of an expressive kind. In sum, three situational factors will be linked with the situational activation of a VA action tendency: how frustrating the situation is, how instrumental the aggressive action in question is, and how well it serves as an expression of feelings within the situation.

As far as inhibition is considered situational, it may be linked to problems verbal aggression may lead to. First, one of the most common problems is dislike from the part of others or loss of being liked, as a consequence of the fact that being verbally aggressive in the situation would hurt other people (Infante, 1987). Second, verbal aggression may also be inhibited because it is a transgression of norms one has, with as a consequence a negative self-evaluation (Infante, 1987; Campbell & Muncer, 1987; Roloff, 1996). Therefore, we expect that situational effects on inhibition are related to expected dislike from others and to expected negative self-evaluation in the situation in question. To support the interpretation of these relations, correlations will be derived of expected dislike and expected negative self-evaluation with ratings of how much someone in the situation would feel hurt by a verbally aggressive act, and with ratings of how much a norm would be violated in the situation, respectively. Since all these ratings are situation specific, their means over judges may be considered situational properties.

It may be expected as well that more objective properties, such as presence of the person to whom the VA behavior is directed and the presence of witnesses may play a role, for example because they are necessary conditions for the VA behavior to have an effect on dislike from the part of others.

Studying the inhibition of verbal aggression is of interest in several respects: First, since verbal aggression is a rather light form of aggression it is more open to observation and to self-report than more serious types of aggression. It is less socially undesirable to report and it is also more common, so that the opportunity to study its inhibition is much broader than for the more severe types of aggression. Second, verbal aggression pertains to daily life and may determine a broad range of relationships one has with colleagues, family members, and friends. It may have an effect on these relationships more often than more serious forms of aggression. The effect it has may be smaller, and therefore it may be less important in total, but it is more widespread and it pertains more to daily life than do stronger forms of aggression.

5.1.3 An indirect self-report approach

We will study behavioral inhibition without asking direct self-report questions on behavioral inhibition. A direct approach would be an inventory with items as: 'How much do you feel inhibited to curse in this situation?'. Instead we ask two questions, one whether one would want to react with verbal aggression, and one on whether one would actually display verbal aggression. The easiest indirect measure of inhibition is the difference between the two responses, but several problems are associated with difference scores (Bereiter, 1963; Embretson, 1991, 2000; Lord, 1963). Therefore, we chose a modeling approach inspired by Embretson (1991), to be explained in the section on modeling.

An inventory will be presented with a description of 15 possible frustrating situations. For each of these three verbally aggressive behaviors will be presented: cursing, scolding, and shouting. Including more than three would seriously expand the number of items. For each combination of a situation and a behavior, two questions were asked: a 'want' question ('Would you want to curse / scold / shout in this situation?'), and a 'do' question ('Would you curse / scold / shout in this situation?'). In total 90 items will be presented: 45 want-items and 45 do-items. The construction of the inventory is in line with the facet methodology (Canter, 1985; Guttman, 1981; Guttman & GreenBaum, 1998; Shye, Elizur, & Hoffman, 1994). A further explanation is given in the method section.

5.2 Theories, formal models, and validation approach

5.2.1 Theories

The theories will be focused on the discrete events of experiencing a VA tendency (wanting to curse, scold, shout) and displaying a VA behavior (cursing, scolding, shouting). The theories we will present are theories for the probability of these events. In the inventory, the participants will be asked whether they agree these events would occur in the corresponding situations. The basic assumptions of all theories we will formulate to explain the probabilities are that (1) people differ in the strength of their VA action tendency, called the *personal VA action tendency*, which will be considered a VA trait (Beatty & McCroskey, 1997; Infante, 1987; Infante & Rancer, 1996; Kinney et al., 2001), and that (2) situations differ in how much they activate the VA action tendency, which will be considered the situational VA activation. We assume throughout that the situational VA activation may depend on the behavior type (curse, scold, shout), and therefore, the term *situational VA behavior activation* will be used. The probability of a person's VA action tendency for a given behavior and situation is considered a function of the sum of the personal VA action tendency and the situational VA behavior activation:

probability of VA action tendency for a given VA behavior in a given situation =
f(personal VA action tendency + situational VA behavior activation)

Note that three simpler formulations of the same theory are possible: with the situational VA behavior activation being reduced to situation main effects, to behavior main effects, or to the combination of both. Because these three formulations will turn out to be inferior from the data analysis, we will not further consider them.

The first behavioral inhibition theory we formulate is a *null theory*, in which it is assumed that no inhibition at all occurs. It will be used as a base-level theory, as a reference point for other more complicated theories.The probability of displaying the behavior is considered identical to the probability of having the corresponding action tendency:

probability of a given VA behavior in a given situation =
f(personal VA action tendency + situational VA behavior activation)

The second theory is the *constant inhibition theory.* It differs from the null theory in that a constant inhibition holds for all persons, and for all combinations of situations and VA behavior types. The probability of displaying a VA behavior is lowered in comparison with the probability of having the corresponding VA action tendency. A constant effect (constant VA inhibition) is subtracted, when not the tendency but the act is considered:

probability of a given VA behavior in a given situation =
f(personal VA action tendency + situational VA behavior activation - constant VA inhibition)

The third theory is the *situational inhibition theory*, implying that each situation may have its own inhibitory effect for each VA behavior, without any individual differences. The inhibition is assumed to depend on the combination of the situation and the behavior, although the latter is not expressed in the label for the theory:

probability of a given VA behavior in a given situation =
f(personal VA action tendency + situational VA behavior activation - situational VA behavior inhibition)

The effect is hypothesized to be negative, which is why a term 'situational VA behavior inhibition' is subtracted. In principle also three subset theories can be formulated: one with situational main effects on inhibition, one with behavioral main effects on inhibition, and one with both kinds of main effects. However, all of these are empirically inferior to the more global situational inhibition theory, so that they will not be considered any further.

The fourth theory is the complement of the former, and is called the *personal inhibition theory.* All inhibition is assumed to stem from the person, independent of the situation and the VA behavior:

probability of a given VA behavior in a given situation =
f(personal VA action tendency + situational VA behavior activation - personal VA inhibition)

Also the personal effect is hypothesized to be negative, or more correctly, its mean is hypothesized to be negative, so that we subtract a term called 'personal VA inhibition.' However, as this theory implies individual differences, some people may actually turn out to display more verbally aggressive behavior then they want: The variance around the mean may be so large that some show an effect opposite to the mean effect.

The fifth and final theory is the *combined theory*, stating that the inhibition depends both on the person and the combination of a situation and a VA behavior. This theory is a combination of the two former theories:

probability of a given VA behavior in a given situation =
f (personal VA action tendency + situational VA behavior activation - situational VA behavior inhibition - personal VA inhibition)

In principle a sixth theory is possible, one that allows for person-by-situation-by-behavior interactions. Such a theory would be a theory with so much flexibility that it could explain everything concerning inhibition. It is a so-called saturated theory and it cannot be rejected by the data. As such it will not further be considered.

We will relate the external person variables and the external situation variables to the corresponding theoretical concepts from the theories: the external person variables and the VA action tendency and the personal VA inhibition, the situational properties and the situational VA behavior activation and the situational VA behavior inhibition.

5.2.2 Formal models

In this section, it will be described how we model the data based on the theories we just described. All models we will use are models for the expected value of a binary dependent variable, or in other words, the probability of a response. This probability is first transformed as follows:

$$\text{logit}\,[P\,(Y=1)] \;=\; \ln\{P\,(Y=1)\,/\,[1-P\,(Y=1)]\},$$

which is a way to obtain an unbounded, real valued, dependent variable (Menard, 2001), as in the linear regression model. The logistic dependent variable, $\text{logit}\,[P\,(Y=1)]$, is modeled then as a linear function of a number of predictors. It is possible to allow regression weights to vary over persons, according to some distribution, which is mostly the normal distribution. These models are logistic regression models because of the logistic transformation, and they are mixed models because some of the regression weights may vary over persons. In the literature on Generalized Linear Mixed Models (McCulloch & Searle, 2001), these varying effects are called random effects, while the other are called fixed effects. Many item response models are logistic mixed models (Rijmen, Tuerlinckx, De Boeck, & Kuppens, in press). For example, in the Rasch model in its marginal maximum likelihood formulation (Baker, 1992) the predictors are item indicators (equal to 1 for the item in question, and equal to 0 for the other items), while the weights of the indicators are the item parameters, and the intercept is a normally distributed latent trait (an intercept that varies over persons).

5.2.2.1 Basic model: for the want-items

We will first describe the model for the want-items. This model will be part of the model for the do-items as well.

In all theories we have explained, it is assumed that persons can have different values for the personal VA action tendency. We will denote the value of this tendency as α_i, with i as an index for the person ($i = 1, \ldots, I$). It is further assumed that α is normally distributed (a normal random variable). Further, also the combination of a situation and a type of behavior plays a role, with an effect denoted as $\beta_{jk}^{(want)}$. Index j is used for situations ($j = 1, \ldots, J$) and index k is used for the behavior type ($k = 1, \ldots, K$). The $\beta_{jk}^{(want)}$ are the regression weights or the effects that each refer to one pair of a situation and a behavior:

$$\text{Logit}\left[P\left(Y_{ijk}^{(want)} = 1 \mid \alpha_i\right)\right] = \alpha_i + \beta_{jk}^{(want)} \tag{5.1}$$

with $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$ as the personal VA action tendency.

Equation 5.1 has an identification problem, since the mean of the $\beta_{jk}^{(want)}$ can compensate for the mean of the α_i, since adding a constant to every $\beta_{jk}^{(want)}$ and subtracting the same constant from every α_i results in exactly the same values for the logits. Therefore, the mean of α_i (μ_α) is fixed to zero.

The model as described in Equation 5.1, is better know as the Marginal Maximum-Likelihood formulation of the Rasch model (Baker, 1992), a basic model of item response theory (Baker, 1992; Fischer & Molenaar, 1995). The α_i corresponds with the ability, and if we reparameterize the β_{jk} as $\beta_{jk}^* = -\beta_{jk}$, the β_{jk}^* correspond with the item difficulties.

5.2.2.2 Model for the do-items

Given that we wanted to investigate the inhibition of VA behavior, not only want-items are needed, but also do-items. Remember that each do-item corresponds to a want-item, and that it shares the same situation and behavior with that item. The general idea is that the same factors that play in the want-item also play in the do-item, plus more. The additional factors are related to behavioral inhibition. In all cases the want-items and the do-items will be analyzed simultaneously, using the two types of models (for the want-items and for the do-items) as two submodels of one overall model for the whole dataset.

In the *null theory* there is no inhibition, so that the model for the do-items is exactly the same as for the want-items:

$$\text{Logit}\left[P\left(Y_{ijk}^{(do)} = 1 \mid \alpha_i\right)\right] = \alpha_i + \beta_{jk}^{(want)} \tag{5.2}$$

In the *constant inhibition theory*, inhibition is assumed to be constant value, independent of the person and the combinations of a situation and a behavior. Therefore, the model reads as:

$$\text{Logit}\left[P\left(Y_{ijk}^{(do)} = 1 \mid \alpha_i\right)\right] = \alpha_i + \beta_{jk}^{(want)} - \tau \tag{5.3}$$

with τ as the inhibition constant.

In the *situational inhibition theory*, inhibition is assumed to be situation and behavior dependent. This can be modeled by subtracting a parameter that is specific for the combination of a situation and a behavior:

$$\text{Logit}\left[P\left(Y_{ijk}^{(do)} = 1 \mid \alpha_i\right)\right] = \alpha_i + \beta_{jk}^{(want)} - \delta_{jk} \tag{5.4}$$

with δ_{jk} as the situation and VA behavior specific inhibition parameter. The mean δ_{jk} is the overall inhibition effect.

In the *personal inhibition theory*, it is assumed that the inhibition of VA behavior is a source of individual differences. Therefore, for the do-items a person-dependent parameter κ_i will be subtracted. Like for the personal VA action tendency, this parameter is assumed to be normally distributed over persons:

$$\text{Logit}\left[P\left(Y_{ijk}^{(do)} = 1 \mid \alpha_i, \kappa_i\right)\right] = \alpha_i + \beta_{jk}^{(want)} - \kappa_i \tag{5.5}$$

with $\kappa_i \sim N(\mu_\kappa, \sigma_\kappa^2)$ as the personal VA inhibition parameter. The mean of κ_i is the overall inhibition effect. It is expected that $\mu_\kappa > 0$. The larger the value for κ_i, the stronger is the personal VA inhibition.

Finally, the *combined theory* is a combination of the two previously mentioned theories. In the combined theory, inhibition is both person dependent, and situation and behavior specific. This model is presented in Equation 5.6:

$$\text{Logit}\left[P\left(Y_{ijk}^{(do)} = 1 \mid \alpha_i, \kappa_i\right)\right] = \alpha_i + \beta_{jk}^{(want)} - \delta_{jk} - \kappa_i \tag{5.6}$$

with $\kappa_i \sim N(\mu_\kappa, \sigma_\kappa^2)$.

In a similar way as the model in (5.1), the model in (5.6) has an identification problem. The mean of the δ_{jk} can compensate for μ_κ. Therefore, μ_κ is fixed to zero. By consequence, the mean of the δ_{jk} indicates the overall inhibition effect.

We expect that the model for the combined theory is the best model. Therefore, we will describe the validation in terms of external variables as if this model applies.

5.2.3 Validation approach

To validate the previously mentioned theories, the parameters of the formalized theories can be correlated with the external variables that are discussed earlier. Here, we will link these external variables explicitly to the various parameters. We do not expect very high correlations with the external person variables for the following reasons: First, only three behaviors are included. Second, the personal effects refer only to verbal aggression while the external variables are more general. As such, correlations with an absolute value in the range of .10 to .35, like the trait-behavior correlations (Mischel, 1968; Mischel & Peake, 1982; Kenrick & Funder, 1988) are expected. Finally, as for the correlation with the situation properties, the situations are rather similar in that they all are of the kind that verbal aggression can be expected, so that a restriction of range may attenuate the correlations.

In terms of the parameters, we expect the following correlations. First, the personal VA action tendency is estimated through α_i, and it is expected to be correlated positively with the following trait measures: Trait Anger, Anger Out, Direct Aggression, and Indirect Aggression, and negatively with Anger In and Anger Out Control. Second, the personal behavioral inhibition is estimated through κ_i, and it is expected to be positively correlated with Anger Out Control, and negatively with Anger Out and Direct Aggression. Third, the situational VA behavior activation is estimated through β_{jk}, and it is expected to be positively correlated with how frustrating the situation is and with the expressiveness and instrumentality of the VA behavior in the situation. Fourth, as far as inhibition is considered situational, it is estimated through δ_{jk}, and it is expected to be correlated positively with expected dislike from others and with expected negative self-evaluation in the situation in question. In addition, we will include two measures for the presence of other persons: one to measure the presence of witnesses, and one to measure the presence or absence of the person to whom the VA behavior is directed. Both are expected to correlate positively with situational VA behavior inhibition.

5.3 Method

5.3.1 Subjects

The participants were 316 first-year psychology students at the home university of the authors. Participation in the study was a partial fulfillment of a requirement to participate in research. The sample consists of 73 males and 243 females, which reflects the proportion of the two sexes among psychology students. The average age was 18.4 (sd = 1.2).

5.3.2 Measures and procedure

The data collection was part of a larger study by the third author. Only the characteristics relevant for the present study are mentioned here.

5.3.2.1 Situation-response questionnaire

Fifteen situations were taken from two situation-response inventories of Endler and Hunt (1968). The situations were constructed for a student population from which the sample in the present study is drawn. Endler and Hunt (1968) selected situations students may be familiar with through direct or indirect experience. The 15 situations are listed in the Appendix.

Six reactions were selected: want to curse, want to scold, want to shout, curse, scold, and shout.

The 15 situations were crossed with the six reactions yielding 90 items in total. The items were presented in random order to the students. For each item three response categories were provided (0 = no, 1 = to some extent, and 2 = to a strong extent). For the data analysis, the response categories 1 and 2 are grouped because we were interested in whether or not the VA behavior occurs and not so much in its intensity.

5.3.2.2 External person variables

Three questionnaires on anger, anger expression, and aggression were used as measures of external subject characteristics: the Zelf-Analyse vragenlijst (ZAV, Van Der Ploeg, Defares, & Spielberger, 1982), which is the Dutch adaptation of the State-Trait Anger scale (STAS, Spielberger, 1980), to have a measure of Trait Anger; the self-expression and control scale (SECS, Van Elderen et al., 1997), an adaptation of the Anger Expression Scale (AX) of Spielberger et al. (1982) containing the subscales Anger In, Anger Out, Control Anger In, and

Control Anger Out; and the Buss-Durkee Hostility Inventory-Dutch (BDHI-D, Lange et al., 1995). The latter inventory contains three subscales labeled Direct Aggression, Indirect Aggression, and Social Desirability. As the Social Desirability scale is very short (5 items) and has a rather low internal consistency (Cronbach's alpha = .50; Lange et al., 1995), it will be excluded from the analyzes.

5.3.2.3 VA behavior - situations characteristics

Ten judges evaluated all situations on several properties: (1) the amount of frustration elicited by the situation, (2) how well the VA behavior in general expresses the feelings elicited by the situation, (3) how instrumental the VA behavior is in each of the situations, (4) expected dislike from others, (5) expected negative self-evaluation, (6) the degree in which VA behavior hurts others, (7) norm violation, (8) the presence of witnesses, and (9) the presence of the person to whom the VA behavior is directed. Note that for the judgments of the situations, no distinction was made depending on the type of VA behavior. All judgments were made on a 4-point scale varying from 0 to 3. For each property, the mean of the ten judges was used as a measure. The reliability of the judgments for the nine features are the following .87 (frustration), .73 (expressiveness), .92 (instrumentality), .80 (dislike form others), .83 (negative self-evaluation), .85 (hurt others), .85 (norm violation), .92 (witness), and .97 (presence of person to whom the VA behavior is directed). These coefficients are the Cronbach's Alphas for internal consistency.

5.3.2.4 Analysis

The data of the situation-response questionnaire were analyzed with PROC NLMIXED (Wolfinger, 1999) from SAS V8 (1999). A quasi Newton-Raphson optimization technique was used, together with a nonadaptive Gauss-Hermite quadrature approximation as described in Pinheiro and Bates (1995) with 15 quadrature points to estimate the parameters of the different models. The fit of the different models can be compared using two information criteria: the AIC statistic (Akaike's information criterion; Akaike, 1977) and the BIC statistic (Bayesian information criterion; Schwartz, 1978). Given that both information criteria contain a penalty for the number of parameters, an optimal balance between model fit and model parsimoniousness is obtained. The lower the value of these statistics, the better the model fits the data. The specific values of each person on both person dependent parameters (α_i and κ_i) were be obtained by requesting Empirical Bayes Estimates.

Subsequently, the values for the external trait measures –i.e. the sum scores on subscales of the ZAV, the SECS, and the BDHI-D– were correlated with the estimated values of the personal VA action tendency (α_i), and the personal VA inhibition (κ_i). The nine situational properties were correlated with the estimated values of the situational VA behavior activation (β_{jk}) and with the situational VA behavior inhibition (δ_{jk}).

TABLE 5.1. Expected correlations between α_i, κ_i and external person variables

	TA	AO	AI	CAO	CAI	DA	IDA
α_i	+	+	-	-	.	+	+
κ_i	.	-	.	+	.	-	.

'+' indicates a positive correlation, '-' indicates a negative correlation, '.' means that no specific hypothesis is formulated about the correlation.
α_i: person parameter for the action tendency;
κ_i: person parameter for the behavioral inhibition;
TA = Trait Anger (ZAV); AO = Anger Out (SECS); AI = Anger In (SECS); CAO = Control Anger Out (SECS); CAI = Control Anger In (SECS); DA = Direct Aggression (BDHI-D); IDA = Indirect Aggression (BDHI-D).

5.4 Results

5.4.1 Modeling the situation-response questionnaire

The models that are presented earlier will be used to analyze the data. The values of the fit statistics for all models are listed in Table 5.2.

TABLE 5.2. Fit statistics of the models

Model	AIC	BIC
Null model	30354	30527
Constant inhibition	29762	29938
Situational inhibition	29647	29989
Personal inhibition	29077	29261
Combined	28968	29317

As can be seen in Table 5.2, all the models with inhibition included fit the data clearly better than the null model. Within the group of inhibition models, the models containing a person-dependent inhibition parameter fit clearly better than the models without personal VA inhibition. The effect of adding situational inhibition is much smaller (AIC), or even opposite (BIC). According to the AIC values, the combined model has the best fit. However, according to the BIC value, the personal inhibition model has a somewhat better fit. This difference is due to the stronger penalty the BIC gives to the number of parameters. Hence, depending on how important one considers it for a model to be parsimonious, either the combined model (less parsimonious) or the personal inhibition model (more parsimonious) should be preferred. We opt for the personal inhibition model, since this model is more parsimonious, and since the two types of situation-by-behavior parameter estimates (the estimates of β_{jk} and δ_{jk}) of the combined model were correlated rather strongly (-.56) and showed an analogous, but opposite, pattern of correlations with the situational properties. Therefore, we don't have a strong basis to differentiate empirically between the situational VA activation and the situational VA inhibition. The values of the model parameters of the personal inhibition model are given in the Tables 5.3, and 5.4.

TABLE 5.3. Situational VA behavior activation parameter estimates for the personal inhibition model

Par.	Est.	SE	Par.	Est.	SE	Par.	Est.	SE
$\beta_{1,1}$	.65	.11	$\beta_{6,1}$	1.51	.12	$\beta_{11,1}$	.57	.11
$\beta_{1,2}$	.06	.11	$\beta_{6,2}$	.75	.11	$\beta_{11,2}$	.34	.11
$\beta_{1,3}$	-1.04	.12	$\beta_{6,3}$	-.13	.11	$\beta_{11,3}$	-.31	.11
$\beta_{2,1}$	.89	.11	$\beta_{7,1}$	-1.05	.12	$\beta_{12,1}$	.34	.11
$\beta_{2,2}$	.70	.11	$\beta_{7,2}$	-1.89	.13	$\beta_{12,2}$	-.96	.12
$\beta_{2,3}$	-.19	.11	$\beta_{7,3}$	-2.62	.16	$\beta_{12,3}$	-1.36	.12
$\beta_{3,1}$	1.54	.12	$\beta_{8,1}$	.13	.11	$\beta_{13,1}$	1.58	.12
$\beta_{3,2}$	.14	.11	$\beta_{8,2}$	-.74	.11	$\beta_{13,2}$	.59	.11
$\beta_{3,3}$	-.14	.11	$\beta_{8,3}$	-1.56	.13	$\beta_{13,3}$	-.44	.11
$\beta_{4,1}$	.56	.11	$\beta_{9,1}$	.42	.11	$\beta_{14,1}$	1.48	.12
$\beta_{4,2}$	-.49	.11	$\beta_{9,2}$	-.84	.12	$\beta_{14,2}$	.11	.11
$\beta_{4,3}$	-1.20	.12	$\beta_{9,3}$	-1.92	.14	$\beta_{14,3}$	-.19	.11
$\beta_{5,1}$	1.27	.11	$\beta_{10,1}$	.63	.11	$\beta_{15,1}$	-.81	.12
$\beta_{5,2}$	.54	.11	$\beta_{10,2}$	-.08	.11	$\beta_{15,2}$	-1.14	.12
$\beta_{5,3}$	.27	.11	$\beta_{10,3}$	-.98	.12	$\beta_{15,3}$	-2.26	.15

The first index refers to the situation (see Appendix), the second to the type of VA behavior (curse=1; scold=2; shout=3).

For the interpretation of the estimates given in Table 5.3, one has to keep in mind that the higher β_{jk}, the higher the situational VA behavior activation. The β_{jk} vary from -2.62 to 1.58. Their mean value is equal to -.16, meaning that on the average (over all situations and behaviors) a person with an average

VA action tendency has a probability of .54 of wanting to display VA behavior. The mean β_{jk} per kind of VA behavior are .65 for wanting to curse, -.19 for wanting to scold, and -.94 for wanting to shout. These estimates correspond with mean probabilities of .66, .45 and .28 for respectively wanting to curse, wanting to scold, or wanting to shout for a person with a mean action tendency. In Figure 5.2, the probabilities of the VA behavior activation for the person with an average VA action tendency are plotted per kind of VA behavior, for the 15 situations. It is clear from the figure that there is a lot of variability in the action tendencies, primarily depending on the situation.

FIGURE 5.2. Probability of wanting to curse, wanting to scold, wanting to shout for the average person in all 15 situations

The estimates for the personal VA action tendency (α_i) and the personal VA inhibition (κ_i) are summarized in Table 5.4. To interpret the parameters of Table 5.4, one has to keep in mind that the higher κ_i, the higher the personal VA inhibition, and the higher α_i, the higher is the personal VA action tendency of the subject. Both variables have a variance significantly different from zero ($p < .0001$). Note that the test of the variance estimate by using its standard error is a conservative test, since the variance is bounded by zero (Snijders & Bosker, 1999; Verbeke & Molenberghs, 2000, in press). The variance of the personal VA action tendency is larger than the variance for the personal VA inhibition, meaning that people differ more in action tendency than they differ with respect to inhibition. The correlation between the action tendency and inhibition is .42 ($p < .001$), meaning that the inhibition is stronger for a stronger action tendency. Also the mean of the personal VA inhibition (.73) is highly significant ($p < .001$). As a result, in the average situation the person with the mean action tendency and the mean personal inhibition, has a probability of .29 of displaying VA behavior (averaged over the three types of VA behavior), which is about .2 less than the probability wanting to display the VA behavior.

TABLE 5.4. Estimates of distributional parameters for the personal inhibition model

Distributional parameter	Estimate	SE
Var(α)	1.95	.14
μ_κ	.73	.06
Var(κ)	1.03	.10
Cov(α,κ)	.59	.10

We also looked for gender differences in the means of the person dependent parameters. For example, replacing α_i by $\alpha_i^* + \beta^{(male)}Gender$, with 'Gender' coded as male=1; female=0, leads to a model which can be used for detecting gender differences in the mean of the personal verbal aggression activation. We fitted two different models: one in which we allowed for gender differences in the mean of the personal verbal aggression activation and one in which we allowed for gender differences in the mean personal VA inhibition, as a model which allows for both kinds of gender differences is not identified. The first model did not have a better fit than the original personal inhibition model without gender differences (AIC=29079, BIC=29267), and the latter only had a slightly better fit (AIC=29057, BIC=29245), meaning that there are no gender differences in the mean of the personal verbal aggression activation, and some small gender differences in the mean of the personal VA inhibition. The difference between the means is .69. The inclusion of this gender difference has no further effects on the results, so that we have reported only the results of the joint analysis.

5.4.2 Correlations with external person variables

The correlations between the parameter estimates and the external variables are given in Table 5.5. Comparing Table 5.5 with Table 5.1, it can be concluded that except for Anger In, all expected correlations are significant, and if no correlation is expected, the empirical correlation is not significant. For Anger In, the opposite pattern was found, as it does correlate significantly with inhibition but not with the action tendency. It seems that Anger In suppresses the act given the action tendency, rather than to prevent an outward action tendency. Anger In seems to be a way of coping with aggressive action tendencies (while preventing their expression) rather than with anger itself (by turning it inwards).

TABLE 5.5. Correlations between person-dependent parameters and external person variables

	TA	AO	AI	CAO	CAI	DA	IDA
α_i	0.17**	0.14*	0.08	-0.12*	-0.09	0.21**	0.21**
κ_i	0.08	-0.20**	0.18**	0.14*	0.08	-0.16**	0.04

* $p < .05$ (two-tailed)
** $p < .01$ (two-tailed)

5.4.3 Correlations with external situational variables

The situational VA behavior activation as estimated through β_{jk} was positively correlated with frustration (.38), with instrumentality (.43), and with expressiveness (.44), all $p < .01$. From a multiple regression analysis, it can be concluded that together they explain 36% of the variance in the β_{jk} (adjusted R^2=.31).

As we opt for a model without situational VA behavior inhibition, we cannot relate dislike from others and negative self-evaluation to a situational inhibition parameter. Correlating the situational VA behavior activation with dislike from other and with negative self-evaluation to, a correlation of -.50 was found for dislike from others and a correlation of -.40 for negative self-evaluation, both $p < .01$. The correlation of situational VA behavior activation with the presence of a witness is -.19 (n.s.), and with the presence of the person to whom the VA behavior is directed is -.41 ($p < .01$). From a multiple regression analysis, it can be concluded that together, all situational properties explain 51% of the variance in the β_{jk} (adjusted R^2=.42).

Finally, the correlation between expected dislike from others and expected hurting others is .78, and the correlation between expected negative self-evaluation and norm violation to .74; both $p < .01$.

5.5 Discussion

The model we selected assumes that the VA action tendency in a certain situation is a function of the person and of the combination of a situation and a behavior. When it comes to actually displaying the VA behavior, also inhibition plays a role, but this inhibition depends only on the person and not on the combination of a situation and a type of VA behavior. Although the combined model is an intuitively appealing model and has a better fit when the penalty for number of parameters is not high, the personal inhibition model is preferred for two reasons: First, it is a more parsimonious model with about the same fit. Second, for the combined model the correlations of situational VA inhibition with the situational properties expected dislike from others and expected negative self-evaluation are about the opposite of the correlations of situational VA activation with same situational properties. Moreover, the estimates of both parameters have a high negative correlation. It turned out difficult to differentiate between the situational action tendency and the situational inhibition, which detracts for the attractiveness of the combined model.

A similar problem did not occur for the two types of individual differences. The variances of the personal VA action tendency and the personal VA inhibition are significant, meaning that VA behavior depends on two person-dependent processes or traits: the personal VA action tendency, supplemented with a trait for the inhibition of verbal aggressive inclinations (personal VA inhibition). The two have a moderate positive correlation, so that the action tendency is compensated somewhat by the inhibition. People with a higher aggressive tendency tend to have a somewhat stronger inhibitory tendency. Comparing the correlations of these two person-dependent parameter estimates with the external person variables (see Table 5.5), one can see that they have a differentiated pattern of correlations. The personal VA action tendency (α_i) is primarily correlated with Direct and Indirect Aggression, and with Trait Anger, whereas the personal VA inhibition is primarily correlated with Anger Out (negatively) and Anger In (positively). It seems that the inhibition is primarily related with the direction of the anger (in or out), whereas the action tendency is related to aggressiveness, whether of a direct or indirect kind.

The results also shed light on the external variables, especially on the coping with anger variables and the aggressive style variables. From the correlations with the action tendency and with the behavioral inhibition, interpretations can be made about the activating and the inhibitory nature of the variables.

A positive correlation with the action tendency suggests an activating role, and when the correlation with behavioral inhibition is negative, this activating role is of a general kind as it includes the behavioral expression. A negative correlation with the action tendency suggests an inhibitory role, but one that is important earlier in the scheme of Figure 5.1 than behavioral inhibition. When the correlation with behavioral inhibition is positive at the same time, the inhibitory role is a rather general one, as it includes inhibition in an earlier and a later stage.

First, Trait Anger seems to have an activating role in the earlier part of the VA scheme, since it is not negatively correlated with behavioral inhibition. Second, Anger Out and Anger In seem primarily related to behavioral inhibition. To have an Anger Out coping style means to follow one's verbal aggression tendency and not to inhibit that tendency. To have an Anger In coping style means to inhibit one's verbally aggressive inclinations, so that they are not expressed. As far as verbal aggression is concerned, both styles seem to concern what happens with the action tendency: expression (Anger Out) or inhibition (Anger In), and not so much with the action tendency itself. One might have expected that an aggressive action tendency as such fits an Anger-Out style and contradicts an Anger-In style, but this is much less the case than that these styles are associated with the expression of the action tendency. As for Anger Out Control, the inhibitory role is rather general, as expected, but the correlations are low. The additional control variable (Control Anger IN) does not show significant correlations. Third, the aggression style variables are of an activating kind. The Direct, as well as the Indirect aggression style are positively correlated with the action tendency. The difference is that a direct aggression style also activates the expression of the action tendency, while the indirect style does not.

It is remarkable that the VA action tendency and the VA behavioral inhibition are correlated positively, although not very highly. This finding can be related to the frequency of occurrence: the more people feel the urge to act verbally aggressive, the more they will find themselves in situations in which this tendency or the resulting behavior will be inhibited. The fact that the correlation is not very high, means that the action tendency and inhibition are two related, but also separate processes that can be influenced by different factors.

Gender differences were only found for personal VA inhibition, resulting in a model in which females in general show a somewhat stronger inhibitory tendency than males do. This difference in strength of inhibitory tendency between males and females, can be explained based on the finding of Crane-Ross, Tisak, and

Tisak (1998) that females, more than males, indicated that aggressive behavior is less acceptable, and a more of a cause for, amongst other things, concerns with respect to negative self-evaluation and negative effects on others.

The situational part of the personal inhibition theory comprises only a VA activation component. No situational inhibition component is included in the personal inhibition theory. Nevertheless, the situational VA action tendency is not only positively related to how frustration the situation is, and to the instrumental and expressive value of the behavior (as expected), but also a negative correlation with expected dislike form others and expected negative self-evaluation was found. This suggest that the situation may have an inhibitory effect earlier in the scheme of Figure 5.1. Expected dislike from others and expected negative self-evaluation, are exactly the kind of situational properties that may stimulate inhibition. Consequently, one can argue that these properties had their influence already on action tendency inhibition. As our approach cannot distinguish between the situationally induced action tendency and situationally induced action tendency inhibition, both are summarized in one parameter, the situational VA action tendency. Overall, it seems that situationally induced inhibition mainly occurs earlier, at the level of the action tendency, whereas there is still a substantial type of person-induced inhibition that plays a role later, in preventing the expression of an action tendency.

Finally, note that the estimation of both the action tendency and the behavioral inhibition are based on a limited number of behaviors in a restricted set of situations. As already mentioned in the introduction, this may limit the magnitude of the correlations with the more general external trait measures. On the other hand, working with a limited number of behaviors and situations allowed us to make a fine-grained analysis, one that looks at individual situations and behaviors, and one that can elegantly grasp the conceptual difference between action tendencies and the resulting behaviors, as well as the differences between effects induced by the person versus effects induced by the particular situation-behavior combination.

One should also realize that not only our sample of situations and behaviors is small, but that also the age range in our study is limited (about 16 to 20 years). These limitations as to the situations, the types of VA behavior, and the age of the participants prevent us from making strong claims on verbal aggression and its inhibition in general. Nevertheless, the approach we followed seemed successful in capturing the personal and situational aspects of verbal aggression and its inhibition.

5.6 Conclusion

Although the research has limitations (self report data, restricted population, only 15 situations, 3 types of verbally aggressive behaviors), it seems that our approach was successful in modeling and understanding the data from a situation-response questionnaire, including its external validation. The two basic concepts, the verbal aggressive action tendency and the inhibition of verbal aggressive behavior seem slightly correlated and each have interesting correlations with external variables. Inhibition seems to be related to coping with anger and the action tendency to be verbally aggressive on the other hand is related to trait anger, direct and indirect aggression style, and to several situational properties.

5.7 Appendix

1. Someone has lost an important book of yours.
2. You have just found out that someone has told lies about you.
3. You are driving to a party and suddenly your car has a flat tire.
4. You arrange to meet someone and he/she does not show up.
5. You are trying to study and there is incessant noise.
6. You are waiting at the bus stop and the bus fails to stop for you.
7. You are in a restaurant and have been waiting a long time to be served.
8. You are very tired and just asleep when awakened by some friends passing by.
9. The grocery store closes just as you are about to enter.
10. Someone has splashed mud over your new clothes.
11. Someone makes an error and blames it on you.
12. You are reading a mystery novel and find that the last page of the book is missing.
13. You miss your train because the clerk has given you faulty information.

14. You are typing a term paper and your typewriter breaks.

15. Someone pushes ahead of you in a theater ticket line.

5.8 References

Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Proceedings of symposium on applications of statistics* (p. 27-41). Amsterdam, The Netherlands: North Holland.

Allan, S., & Gilbert, P. (2001). Anger and anger expression in relation to perceptions of social rank, entrapment and depressive symptoms. *Personality and Individual Differences*, *32*, 551-565.

Averill, J. R. (1982). *Anger and aggression: An essay on emotion.* New York: Springer-Verlag.

Averill, J. R. (1983). Studies on anger and aggression: implications for theories of emotion. *American Psychologist*, *38*, 1145-1160.

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

Beatty, M. J., & McCroskey, J. C. (1997). It's in our nature: verbal aggressiveness as temperamental expression. *Communication Quarterly*, *45*, 446-460.

Begley, T. M. (1994). Expressed and suppressed anger as predictors of health complaints. *Journal of Organizational Behavior*, *15*, 503-516.

Bereiter, C. (1963). Some persistent dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change* (p. 203-212). Madison: University Wisconsin Press.

Boddeker, I., & Stemmler, G. (2000). Who responds how and when to anger? the assessment of actual anger response styles and their relation to personality. *Cognition and Emotion*, *14*, 737-762.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology*, *21*, 343-349.

Campbell, A., & Muncer, S. (1987). Models of anger and aggression in the social talk of women and men. *Journal for the Theory of Social Behaviour*, *17*, 489-511.

Canter, D. (Ed.). (1985). *Facet theory: Approaches to social research.* New York: Springer.

Cornell, D. G., Peterson, C. S., & Richards, H. (1999). Anger as a predictor of aggression among incarcerated adolescents. *Journal of Consulting and Clinical Psychology*, *67*, 108-115.

Crane-Ross, D., Tisak, M. S., & Tisak, J. (1998). Aggression and conventional rule violations among adolescents: social-reasoning predictors of social behavior. *Aggressive Behavior*, *24*, 347-365.

Culbertson, F. M., & Spielberger, C. D. (1996). Relations of anger expression to depression and blood pressure in high school students. *Stress and emotion: Anger, anxiety, and curiosity*, *16*, 193-202.

Deffenbacher, J. L., Oetting, E. R., Lynch, R. S., & Morris, C. D. (1996). The expression of anger and its consequences. *Behavioral Research and Therapy*, *34*, 575-590.

Dehghani, B., & Lange, A. (1993). Het meten van aggresiviteit: validatie van de nederlandse versie van de Buss-Durkee Hositility Inventory [Measuring aggression: validation of the dutch adaptation if the Buss-Durkee Hositility Inventory]. *Gedrag & Gezondheid*, *21*, 298-304.

DeTurck, M. A. (1987). When communication fails: Physical aggression as a compliance-gaining strategy. *Communication Monographs*, *54*, 106-112.

Embretson, S. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495-515.

Embretson, S. (2000). Multidimensional measurement from dynamic tests: abstract reasoning under stress. *Multivariate Behavioral Research*, *35*, 505-542.

Endler, N. S., & Hunt, J. M. (1968). S-r inventories of hostility and comparisons of the proportions of variance from persons, behaviors, and situations for hostility and anxiousness. *Journal of Personality and Social Psychology*, *9*, 309-315.

Engebretson, T. O., Matthews, K., & Scheier, M. F. (1989). Relations between anger expression and cardiovascular reactivity: reconciling inconsistent findings through a matching hypothesis. *Journal of Personality and Social Psychology*, *57*, 513-521.

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments and applications.* New York: Springer-Verlag.

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

Forgays, D. G., Forgays, D. K., & Spielberger, C. D. (1997). Factor structure of the state-trait anger expression inventory. *Journal of Personality Assessment*, *69*, 497-507.

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

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

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

Greenglass, E. R. (1996). Anger suppression, cynical distrust and hostility: implications for coronary heart disease. *Stress and emotion: Anger, anxiety, and curiosity*, *16*, 205-225.

Guttman, L. (1981). What is not what in theory construction. In I. Borg (Ed.), *Multidimensional data representations, when and why.* Ann Arbor: Mathesis Press.

Guttman, R., & GreenBaum, C. W. (1998). Facet theory: Its development and current status. *European Psychologist*, *3*, 13-36.

Infante, D. A. (1987). Aggressiveness. In J. C. McCroskey & J. A. Daly (Eds.), *Personality and interpersonal communication* (p. 157-194). Newbury Park, CA: Sage.

Infante, D. A., & Rancer, A. S. (1996). Argumentativeness and verbal aggressiveness: a review of recent theory and research. In B. R. Burleson & A. W. Kunkel (Eds.), *Communication yearbook* (p. 319-352). Thousand Oaks, CA, US: Sage.

Julkunen, J. (1996). Suppressing your anger: good manners, bad health? *Stress and emotion: Anger, anxiety, and curiosity*, *16*, 227-240.

Kassinove, H., Sukhodolsky, D. G., Tsytsarev, S. V., & Solovyova, S. (1997). Self-reported anger episodes in russia and america. *Journal of Social Behavior and Personality*, *12*, 301-324.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from the controversy: lessons from the person-situation debate. *American Psychologist*, *43*, 23-34.

King, L. A., Emmons, R. A., & Woodley, S. (1992). The structure of inhibition. *Journal of Research in Personality*, *26*, 85-102.

Kinney, T. A., Smith, B. A., & Donzella, B. (2001). The influence of sex, gender, self-discrepancies, and self-awareness on anger and verbal aggressiveness among u.s. college students. *Journal of Social Psychology*, *141*, 245-275.

Kuppens, P., Van Mechelen, I., Smits, D. J. M., & De Boeck, P. (in press). The appraisal basis of anger: specificity, necessity and sufficiency of components. *Emotion.*

Lange, A., Dehghani, B., & De Beurs, E. (1995). Validation of the dutch adaptation of the buss-durkee hostility inventory. *Behaviour Research and Therapy*, *33*, 229-233.

Lange, A., Hoogendoorn, M., Wiederspahn, A., & Beurs, E. de. (1995). *Buss Durkee hostility and direction of hostility inventory-dutch (BDHI-D).* Lisse, The Netherlands: Swets & Zeitlinger.

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (p. 21-38). Madison: University Wisconsin Press.

Martin, R., Wan, K., David, J. P., Wegner, E. L., Olson, B. D., & Watson, D. (1999). Style of anger expression: relation to expressivity, personality, and health. *Personality and Social Psychology Bulletin*, *25*, 1196-1207.

McCulloch, C., & Searle, S. (Eds.). (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Menard, S. (2001). *Applied logistic regression analysis.* University papers series on quantitative applications in the social sciences, 07-106. Thousand Oaks, CA: Sage.

Mischel, W. (Ed.). (1968). *Personality and assessment.* New York: Wiley.

Mischel, W., & Peake, P. K. (1982). Beyond the déjà vu in the search for cross-situational consistency. *Psychological Review*, *89*, 730-755.

Pinheiro, J. C., & Bates, D. M. (1995). Approximation to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, *4*, 12-35.

Ramanaiah, N. V., Conn, S. R., & Schill, T. (1987). On the content saturation of the Buss-Durkee hostility inventory scales. *Psychological Reports*, *61*, 591-594.

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

Roloff, M. E. (1996). The catalyst hypothesis: conditions under which coercive communication leads to physical aggression. In D. D. Cahn & S. A. Lloyds (Eds.), *Family violence from a communication perspective* (p. 20-36). Thousand Oaks, CA: Sage.

Rubin, J. (1986). The emotion of anger: some conceptual and theoretical issues. *Professional Psychology: Research and Practice*, *17*, 115-124.

Sas OnlineDoc^{TM} Version 8. (1999). Carey NC, USA: SAS Institute Inc.

Sas System V8 for windows (Computer program). (1999). Carey NC, USA: SAS Institute Inc.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Schwenkmezger, P., & Hank, P. (1996). Anger expression and blood pressure. *Stress and emotion: Anger, anxiety, and curiosity*, *16*, 241-259.

Shye, S., Elizur, D., & Hoffman, M. (1994). *Introduction to facet theory: content design and intrinsic data analysis in behavioral research.* Applied social research methods, Vol. 35. Thousand Oaks, CA, US: Sage Publications, Inc.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage Publishers.

Spielberger, C. D. (1980). *Preliminary manual for the state-trait anger scale (STAS).* Tampa, Florida: University of south Florida.

Spielberger, C. D., Jacobs, G. H., Russell, S. F., & Crane, R. S. (1983). Assessment of anger: the state-trait anger scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 2, p. 20-36). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

Spielberger, C. D., Johnson, E. G., Russell, S. F., Crane, R. S., Jacobs, G. H., & Worden, T. J. (1985). The experience and expression of anger: Construction and validation of an anger expression scale. In M. Chesney & R. H. Roseman (Eds.), *Anger and hostility in cardiovascular and behavioral disorders* (p. 5-29). New York: McGraw-Hill.

Spielberger, C. D., Johnson, E. H., & Jacobs, G. A. (1982). *The anger expression (AX) scale.* Tampa, Florida: University of south Florida.

Spielberger, C. D., Krasner, S. S., & Solomon, E. P. (1988). The experience, expression and control of anger. In M. P. Janisse (Ed.), *Health psychology: individual differences and stress* (p. 89-108). New York: Springer-Verlag.

Van Der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (1982). *Zelf analyse vragenlijst (ZAV)* [Dutch version of the state-trait anger scale]. Lisse, The Netherlands: Swets & Zeitlinger.

Van Elderen, T., Maes, S., Komproe, I., & Kamp, L. van der. (1997). The development of an anger expression and control scale. *British Journal of Health Psychology*, *2*, 269-281.

Venable, V. L., Carlson, C. R., & Wilson, J. (2001). The role of anger and depression in recurrent headache. *Headache*, *41*, 21-30.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer-Verlag.

Verbeke, G., & Molenberghs, G. (in press). The use of the score test for inference on variance components. *Biometrics.*

Wolfinger, R. (1999). *Fitting non linear models with the new nlmixed procedure.* SUGI 24 Conference proceedings, Paper 287. Cary, NC.

Chapter 6

Estimation of the MIRID: A program and a SAS based approach

Dirk J. M. Smits
Paul De Boeck
Norman D. Verhelst

ABSTRACT The MIRID CML program is a program for the estimation of the parameter values of two different componential IRT models: the Rasch-MIRID and the OPLM-MIRID (Butter, De Boeck, & Verhelst, 1998; Butter, 1994). To estimate the parameters of both models, the program uses a CML approach. The model parameters can also be estimated with a marginal maximum likelihood approach which can be implemented in the PROC NLMIXED procedure of SAS V8.
Both, the MIRID CML program and the MML SAS approach are explained and compared in a simulation study. The results showed that they did about equally well in estimating the values of the item parameters, but that there are some differences in the estimation of the person parameters, as could be expected from the differential assumptions regarding the distribution of the persons. The SAS MML approach is much slower than the MIRID CML program, but it is more flexible on the other hand. [1]

[1] Correspondence concerning this article should be addressed to: Dirk J. M. Smits, K.U. Leuven, Department 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
The research is financially supported by a GOA 2000/2-grant from the K. U. Leuven: 'Psychometric models for the study of personality'

6.1 Introduction

Butter et al. (1998) described the Model with Internal Restrictions on Item Difficulties (MIRID) as a componential model for binary data. In the MIRID, parameters of some items are defined to be a linear combination of the parameters of other items. The model requires that two sets of items are defined: component items and composite items. A *composite item* is an item that measures a concept that can be decomposed into components. A *component item* is an item that measures one of these components. The item parameters of the composite items are decomposed into parts attributed to the component items (the item parameters of the component items). For example, '10*(5+3)' as a composite item has two component items: '5+3' and '10*8'. The first component is of the addition type, the second of the multiplication type. One can formulate many different items along the same line using different basic numbers, for example the composite item '7*(6+8)', and its component items '6+8', '7*14'. The componential approach applies to the affective domain as well. For example, feeling guilty in a given situation may stem from feeling that a norm is violated, from a tendency to brood about what one did, and from a tendency to restitute what one did wrong, each related to the same given situation (Smits & De Boeck, 2003). The question "Do you feel like having violated a moral, an ethic, a religious and/or a personal code in situation A?" is a component item of the norm violation type and it is associated with the composite item "Do you feel guilty in situation A?". For the approach to work, for each composite item associated with a particular cognitive task or with an affective situation, a number of component items has to be formulated with respect to the same cognitive task or situation.

In general, J item families (j=1, …, J) are defined, so that within each family there is one composite item, to be conceived of as a dependent variable, and K component items, one for each of the K component types (k=1, …, K), to be conceived of as the independent variables. For the composite items, k is set to zero. The total number of items is equal to J*(K+1).

Suppose we have a questionnaire with five item families and three types of components, then the total number of items equals 5 * (3 [number of components] + 1 [composite item]) = 20. The structure of such a questionnaire is given in Table 6.1.

The crucial assumption of the MIRID is that the item parameter of a composite item is a linear function of the item parameters of the associated component

TABLE 6.1. Structure of a questionnaire with component and composite items

Situation (Item family)	Component type 1	Component type 2	Component type 3	Composite item
Situation 1 (Item family 1)	Item 1	Item 2	Item 3	Item 4
Situation 2 (Item family 2)	Item 5	Item 6	Item 7	Item 8
Situation 3 (Item family 3)	Item 9	Item 10	Item 11	Item 12
Situation 4 (Item family 4)	Item 13	Item 14	Item 15	Item 16
Situation 5 (Item family 5)	Item 17	Item 18	Item 19	Item 20

items (Butter et al., 1998): the MIRID models the composite item parameters as a linear function of the component item parameters:

$$\beta_{j0} = \sum_{k=1}^{K} \sigma_k \beta_{jk} + \tau \tag{6.1}$$

β_{j0} is the item parameter of the composite item from item family j,
β_{jk} is the item parameter of the component item of type k from item family j,
σ_k is the weight of the component item parameters of type k in determining the composite item parameters,
τ is a normalization constant.

Equation 6.1 is a building block for IRT models with an item threshold (difficulty) parameter. This MIRID principle can be built into various types of IRT models. It imposes in all cases a restriction on the model of which it is a part. We restrict the discussion here to the Rasch model (Rasch, 1960), yielding a Rasch-MIRID, and to the OPLM (One Parameter Logistic Model, Verhelst & Glas, 1995; Verhelst, Glas, & Verstralen, 1994), yielding the OPLM-MIRID (Butter, 1994; Smits & De Boeck, 2003). Like the Rasch model, the OPLM is a model with fixed item discrimination values, but unlike the Rasch model, these fixed values can differ over items.

In the Rasch model, each item (composite item and component item) has its own item parameter, so that the Rasch model models the probability for person i of giving a correct answer to item jk, as in Equation 6.2.

$$P(Y_{ijk} = 1|\theta_i, \beta_{jk}) = \frac{\exp(\theta_i - \beta_{jk})}{1 + \exp(\theta_i - \beta_{jk})} \tag{6.2}$$

θ_i is the person parameter of person i, often called the ability,
β_{jk} is the item parameter associated with item jk with k now varying from 0 to K.

We can group the item parameters into a column vector $\boldsymbol{\beta}$ with (variable) length $R = J(K + 1)$. We can construct an indicator vector $\mathbf{x}_{jk}$ per item jk with a length R equal to that of the item parameter vector $\boldsymbol{\beta}$. The cells contain a '1' for the item parameter of the current item and a '0' otherwise, such that a multiplication of transposed $\mathbf{x}_{jk}$, denoted by $\mathbf{x}'_{jk}$, with the item parameter vector $\boldsymbol{\beta}$ results in the item parameter of item jk. All item indicator vectors can be grouped into one design matrix $\mathbf{X}$, with the rows of this matrix equal to the $\mathbf{x}'_{jk}$. As a result, Equation 6.2 can be reformulated into Equation 6.3. This formula will be used when explaining how to estimate models with SAS. In the same section an example of the item design matrix for the Rasch model and for the Rasch-MIRID will be given (see Figure 6.1).

$$P(Y_{ijk} = 1|\theta_i, \boldsymbol{\beta}) = \frac{\exp\left(\theta_i - \mathbf{x}'_{jk}\boldsymbol{\beta}\right)}{1 + \exp\left(\theta_i - \mathbf{x}'_{jk}\boldsymbol{\beta}\right)} \tag{6.3}$$

Building the MIRID principle into the Rasch model, the item parameters of the composite items are restricted to be a linear combination of the item parameters of the component items. The formula for the Rasch-MIRID is given in Equation 6.4. Remember that $k = 0$ for composite items.

$$P(Y_{ijk} = 1|\theta_i, \beta_{jk}, \sigma_k, \tau) = \frac{\exp\left(\theta_i - \beta_{jk}\right)}{1 + \exp\left(\theta_i - \beta_{jk}\right)} \quad \text{with } k = 0, \ldots, K \tag{6.4}$$

$$\text{with } \beta_{j0} = \sum_{k=1}^{K} \sigma_k \beta_{jk} + \tau \; ; \qquad \text{(composite items)}$$

$$\beta_{jk} = \beta_{jk} \qquad \text{(component items)}$$

For the Rasch-MIRID, a new item parameter vector $\boldsymbol{\beta}$ needs to be constructed, which contains the item parameters of the component items and in the last position the normalization constant, so that $R = JK + 1$. The item indicator vector $\mathbf{x}'_{jk}$ differs according to the kind of item. For the component items, it is similar to the item indicator vector of the Rasch model: '0' in all positions except for a '1' in the position that corresponds to the item parameter of item jk. For the composite items, the item indicator vector contains the weight of the components at the positions of the component item parameters of the same item family as the composite item, and a '1' in the last position. As a consequence, the multiplication of $\mathbf{x}'_{jk}$ with $\boldsymbol{\beta}$ results in β_{jk} for the component items and in $\sum_{k=1}^{K} \sigma_k \beta_{jk} + \tau$ for the composite items. The last column of $\mathbf{X}$ indicates the kind of item: for component items it contains a '0' and for composite items a '1'. An example of such an item design matrix is given in the section about the estimation of the Rasch-MIRID with SAS (see Figure 6.2). The formula for the Rasch-MIRID corresponds to Equation 6.3, but with a modified item design matrix $\mathbf{X}$ and a modified item parameter vector $\boldsymbol{\beta}$.

The OPLM differs from the Rasch model in that a priori and fixed degrees of discrimination are included, that may differ depending on the item. These a priori values are the elements a_{jk} of the item discrimination vector $\mathbf{a}$. Similarly, the OPLM-MIRID differs from the Rasch-MIRID, again only in that the a priori and fixed degrees of discrimination may differ depending on the item. The model equation for the OPLM and the OPLM-MIRID is given in Equation 6.5:

$$P(Y_{ijk} = 1|\theta_i, \boldsymbol{\beta}) = \frac{\exp\left[a_{jk}\left(\theta_i - \mathbf{x}'_{jk}\boldsymbol{\beta}\right)\right]}{1 + \exp\left[a_{jk}\left(\theta_i - \mathbf{x}'_{jk}\boldsymbol{\beta}\right)\right]} \tag{6.5}$$

In the next paragraphs, a program for estimating the model parameters of Rasch-MIRID and the OPLM-MIRID will be presented. The program is based on a CML formulation (Conditional Maximum Likelihood, see, e.g., Baker, 1992; Fischer & Molenaar, 1995; Verhelst, 1993) of the Rasch-MIRID and the OPLM-MIRID. Apart from this, also SAS V8 can be used for the estimation of the model parameters of these models, assuming a normal distribution for the person parameter (Wolfinger, 1999) and following a MML approach (Marginal Maximum Likelihood, see, e.g., Baker, 1992; Fischer & Molenaar, 1995; Verhelst, 1993). Both the MIRID CML program and the SAS procedure for MML will be explained next. In a final section, both approaches will be compared in a small simulation study.

6.2 The MIRID CML program

The MIRID CML program (Smits, De Boeck, Verhelst, & Butter, 2001) is a Windows-based program. It is written in Borland Delphi 5.0 and tested under Windows 95, 98, 2000 and NT 4.0. About 8 MB of free disk space is needed to install the program.

6.2.1 Model estimation

Four models can be estimated with the MIRID CML program: the Rasch model, the Rasch-MIRID, the OPLM and the OPLM-MIRID. The item parameter values and their standard errors are estimated, using a CML approach, and the Davidon-Fletcher-Powell (a quasi Newton-Raphson optimization technique, Bunday, 1984) or the Newton-Raphson optimization technique (Gill, Murray,

& Wright, 1981; Bunday, 1984). Using a CML approach, the item parameters are estimated by conditioning upon the examinees' sufficient statistics (sum of a priori degrees of discrimination for succeeded items). Once the item pool is calibrated, the person parameter estimate corresponding to each sufficient statistic can easily be obtained. A weighted maximum likelihood estimation procedure was implemented for the estimation of the person parameters and their standard errors (Warm, 1989).

6.2.2 Input

Before starting the estimation, one needs to specify the name of the data file, the number of persons, and the number of items in the data file, the discrimination values of the items, and the name of the output file. The data files need to be in plain text format. The requested structure of the data files is such that the rows are formed by the persons and the columns are formed by the items. All responses are typed one next to the other, without any spacing between them. The order of the items (columns) is: first, all component items of component type 1, ordered according to the item family they belong to, next all component items of component type 2 in the same order, and so on, and finally all composite items again in the same order.

As the data sets need to satisfy a rather rigid structure, a module is included in the program to rearrange data files with a different ordering.

6.2.3 Output

During the estimation process, a screen with the current value of the log-likelihood function of the model is shown, so that the state of convergence can be followed. After the estimation procedure has reached the convergence criterion, the output is automatically displayed in the built-in text editor. First, the estimated parameter values are shown: the item parameters of the component items (the β_{jk}), and the linear coefficients (the σ_k and τ), all with their standard errors. Person parameters estimates are optional, and, when requested, they are followed by their standard errors in a separate section after the item parameter estimates. As mentioned earlier, the program provides Warm estimates (Warm, 1989). Warm estimates of the person parameter can also be obtained, for example by using another CML IRT program, such as LPCM-Win (Fischer, Ponocny-Seliger, Ponocny, & Parzer, 1998), or OPLM (Verhelst et al., 1994).

Second, information is given about the fit of the estimated model. If the estimated model is a Rasch-MIRID or an OPLM-MIRID, the fit of this model is compared with the fit of the corresponding basic model (the Rasch model, and the OPLM, respectively) using a likelihood-ratio test, since because of the MIRID principle, the MIRID variants are nested within the corresponding original model. A more specific test will be presented in the section on the SAS MML approach.

6.2.4 Simulation module

The program also contains a module to simulate data. In addition, error can be added to an existing dataset, as explained in detail in the manual (Smits et al., 2001).

6.2.5 Availability

Two versions of the program are available: one for computers running Windows 95 or Windows NT 4.0, and one for computers running Windows 98 or Windows 2000. Except for some animations, the two versions are equivalent. The program can be obtained by e-mailing the author (Miridprogram@hotmail.com), or by sending two 3.5 inch high density diskettes and a self-addressed stamped diskette mailer to Dirk Smits, Department of Psychology (H.C.I.V.), Tiensestraat 102, B-3000 Leuven, Belgium. The MIRID CML program comes with a manual in a PDF file.

6.3 The SAS MML-approach

The parameters of the Rasch model, the Rasch-MIRID, the OPLM, and the OPLM-MIRID can also be estimated using SAS V8. For a discussion on how to use SAS for IRT models see Rijmen, Tuerlinckx, De Boeck, and Kuppens (in press). The SAS software package includes a procedure, called PROC NLMIXED, to fit nonlinear mixed models. Nonlinear mixed models are regression models that are non-linear in the predictors, for example because of a logit link, and with regression weights that are of a mixed nature depending on the predictor: fixed effects or random effects. When the non-linearity is due to the link function, as in the Rasch model and the OPLM, the models are generalized linear models. In the Rasch-MIRID and the OPLM-MIRID, there is a second type of non-linearity, because of the product of the parameters σ_k and β_{jk}, see Equations 6.1 and 6.4, so that they are not part of the family of generalized linear models. All item parameters can be considered fixed effects and the person parameter can be regarded as a random intercept, normally distributed over persons.

For all four models, PROC NLMIXED estimates the item parameters and the parameters of the person parameter distribution (and their standard errors) by using an approximation of the likelihood function based on a normally distributed random intercept. This means that PROC NLMIXED uses a Marginal Maximum Likelihood approach (MML, see e.g. Baker, 1992; Verhelst, 1993) to estimate all these parameters. The item parameters are estimated by integrating the likelihood function over a prespecified person parameter distribution, here the normal distribution. PROC NLMIXED estimates also the mean of the person distribution (if not fixed for identification reasons) and either the standard deviation or the variance, and their standard errors. Also individual person parameter estimates can be obtained by requesting empirical Bayes estimates. Various integral approximations, optimization techniques, and approximations for the first and second derivatives of the likelihood function are available in PROC NLMIXED, some of which will be discussed below.

Information about the fit of the estimated model is given by the maximized value of the log-likelihood function (transformed into a deviance), as well as by the information criteria of Akaike (AIC, Akaike, 1977) and Schwarz (BIC, Schwartz, 1978). These statistics can be used to compare the fit of different models (SAS OnlineDoc^{TM} Version 8). A more specific test that requires the estimation of several model variants, will be presented later.

In the remainder of this section, the structure of the data set and the SAS statements will be explained briefly.

6.3.1 Input

The structure of the input file needed for an analysis is the following:

1. There is a separate row for each observation (for each person by item combination).
2. The first column contains a label for the person.
3. The second column contains the observations for the person by item combination in question. Although not required by the program, we will use

a fixed order for the items within each person: the same order as for the MIRID CML program.

4. Finally, there is a column containing the discrimination value of the item that is involved in the observation (discrimination vector $\mathbf{a}$).

The remaining columns of the input file contains the design matrix $\mathbf{X}$. For the models considered here the design matrix is identical for all persons, and therefore it is repeated for each person. For the Rasch model and for the OPLM, the design matrix is defined as follows (See also section about the Rasch model):

1. There is one row for each item and as many columns as there are item parameters.

2. An element of a row equals '1' if the corresponding item parameter is needed for the item corresponding the row in question, and '0' otherwise (see item indicator vector $\mathbf{x}'_{jk}$).

Since in the Rasch model, there is one item parameter per item, for each person this results in an identity matrix with the same number of columns as the number of items. An example with 20 items and 284 persons is presented in Figure 6.1. The 20 items are organized in five item families and three types of components. The additional column with discrimination values is omitted since, for the Rasch model, these values are all equal.

For the Rasch-MIRID and the OPLM-MIRID the design matrix is defined as follows:

1. There is one row for each item and as many columns as there are component item parameters plus one.

2. An element of a row equals '1' if the corresponding component item parameter is needed for the item corresponding the row in question, and '0' otherwise (see item indicator vector $\mathbf{x}'_{jk}$). Again, this part of the design matrix is an identity matrix, but only for the component items (item indicator vector $\mathbf{x}'_{jk}$). Note that since we cannot include the weights σ_k in the SAS design matrix for the composite items, they are replaced with ones. In the SAS code representing the likelihood formula, the weights will be explicitly added, so that for the composite items, this modified item indicator vector containing only ones and zeros will be multiplied by σ_k times the component item parameters β_{jk}.

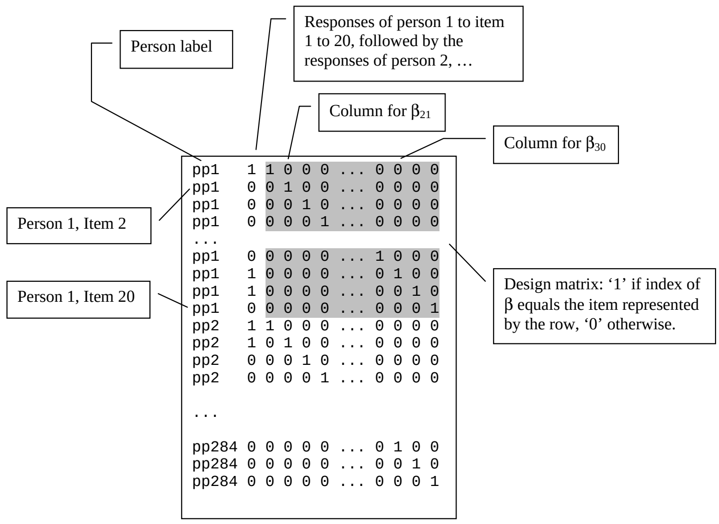


FIGURE 6.1. Example of a SAS data set for the Rasch model

3. An additional column is needed to denote the kind of item. The elements of this column equal '0' if the item is a component item and '1' if the item is a composite item (last element of $\mathbf{x}'_{jk}$, denoted as X0 in SAS code).

For a dataset containing 20 items and five item families and with three types of components, the input file for a Rasch-MIRID with the responses and the design matrix looks as in Figure 6.2. The additional column with discrimination values is omitted since, for the Rasch-MIRID, these values are all equal.

The structure of the input file for the OPLM and the OPLM-MIRID are the same as for the Rasch model and the Rasch-MIRID, respectively, except for the additional column with a priori discrimination values.

6.3.2 SAS statements (see SAS OnlineDoc^{TM}, 1999)

First, the *DATA* procedure has to be called to read the data file. In this procedure, the directory and name of the data file, and the names of variables (columns) it contains are to be specified. For the Rasch model, the SAS code can be found in Listing 1, and for the Rasch-MIRID, the code can be found in Listing 2. For the OPLM and for the OPLM-MIRID, the DATA procedure is

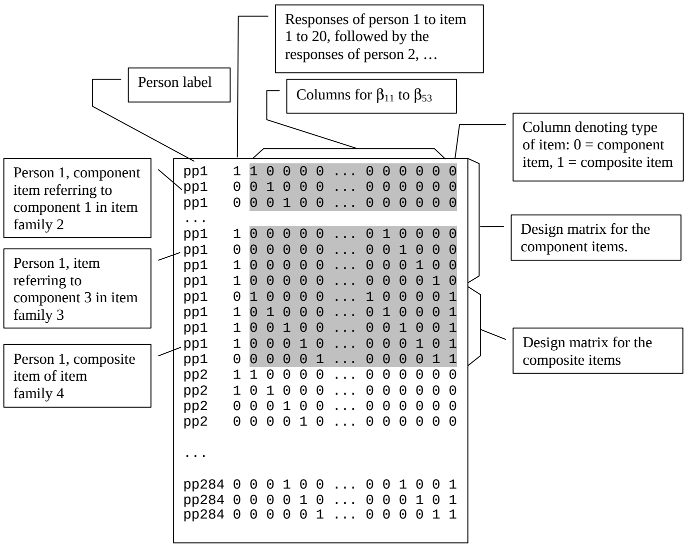


FIGURE 6.2. Example of a SAS data set for the RASCH-MIRID

identical to that for the Rasch model or the Rasch-MIRID respectively, except for one additional column: the additional column with discrimination values. The discrimination value must be mentioned in the INPUT statement. The code can be found in the Listings 3 and 4, respectively.

Subsequent to the DATA procedure, PROC NLMIXED should be called. To construct the SAS code for the Rasch model and the OPLM Equations 6.3 and 6.5 should be used. SAS needs the formula for the probability of giving a correct answer to an item. The dummy variables X correspond to the vectors $\mathbf{x}'_{jk}$ and are used to select the parameters needed for the current item. Since the SAS code does not use vector (or matrix) notations, and not a summation either, the vector multiplication of the item indicator vector $\mathbf{x}'_{jk}$ with the item parameter vector $\boldsymbol{\beta}$ needs to be spelled out completely.

The PROC NLMIXED will be explained now, and for the model-specific part, we will first use the Rasch model for the example with 20 items, three components and five item families. Later, also the model specifications for the other three models will be presented. As it would be to difficult to specify the β's with two indices in the SAS code, we will replace the indices j and k with one index r, $r = 1, \ldots, R$. If we order the component item parameters similar to the order of the component items in the dataset, it is easy to link the β_r to the original β_{jk}.

The statements for the Rasch model are mentioned in Listing 5. We will now go through the statements and explain them statement by statement.

Applying PROC NLMIXED, some choices are to be made about the estimation procedure. These choices can be discussed independent of the model to be estimated.

6.3.2.1 General options

The first general option is 'METHOD='. It is used to specify the method for integral approximation. We choose for the Gauss-Hermite quadrature approximation (GAUSS) as described in Pinheiro and Bates (1995) in combination with the 'NOAD' option, so that the quadrature points are centered at zero for each random effect and so that the current random-effects variance matrix is used as the scale matrix. This is also the default integration method.

With the general option 'QPOINTS=' the number of quadrature points used during the evaluation of the integral can be specified. In combination with the Gauss-Hermite quadrature approximation of the integral, this number equals the number of points used in each dimension of the random effects (we have only one, for the intercept). We choose to set this option equal to 20, to obtain a reasonable precision in describing the distribution of the random effects, and so that the estimation time is not increased too much.

The general option 'TECHNIQUE=' in combination with 'UPDATE=' can be used to determine the optimization technique. Eight different techniques are available: among which conjugate gradient (CONGRA), Newton-Raphson optimization (NEWRAP), Newton-Raphson optimization with ridging (NRRIDG) and quasi Newton-Raphson (QUANEW, which is the default option). For the quasi Newton-Raphson, in addition the 'UPDATE=' option is needed. Eight different possibilities are available, but not all update methods can be combined with all optimizers. Here and in the simulation study, a quasi Newton-Raphson approach, together with the Davidon-Fletcher-Powell update of the inverse Hessian matrix is used. This approach is also implemented in the MIRID CML program and, in contrast to the original Newton-Raphson optimization method that involves the calculation of the second derivatives of the log-likelihood function which is very time consuming, the quasi Newton-Raphson optimization techniques need only the first derivatives to be calculated. The default value for the 'UPDATE' option is the Double Broyden, Fletcher, Goldfarb and Shanno

(DBFGS) update of the inverse Hessian matrix. For the other available options and alternatives, see SAS OnlineDoc^{TM}(1999).

6.3.2.2 Model specific statements

To specify a model, three statements are needed: PARMS, MODEL, and RANDOM. The 'PARMS' statement identifies all model parameters and their starting values. Between the 'PARMS' and the 'MODEL' statement, the model equation is given (ex= ..., p=ex/(1+ex)). The 'MODEL' statement defines the dependent variable and how it depends on the result of the model equation. In our case the response variables are Bernoulli variables with the probability as described in Equation 6.2, which is the 'p' from the SAS code: y ∼ binary(p). In the 'RANDOM' statement, the distribution of the random effect is specified (theta ∼ normal(0,VarTheta)). Only a normal distribution is supported by SAS, and the variance can either be specified a priori, or it can be defined as a parameter to be estimated. In our case, we constrain the mean of the person distribution to be zero, to render the model identifiable, and we define the variance as a parameter to be estimated. Also the standard deviation (and its standard error) can be obtained as follows (theta ∼ normal(0, StdTheta**2)). The 'SUBJECT' option within the 'RANDOM' statement is needed to specify when the random effect obtains new realizations. As in the Rasch model each person has its own person parameter, the person variable defines the realizations of the random effect (SUBJECT=Person).

In order to fit a Rasch-MIRID to the data set mentioned above, only the statements referring to the specific model (PARMS and the model equation) differ. The SAS statements are given in Listing 6.

For the OPLM and OPLM-MIRID, the 'PARMS' statement is equal to the one of the Rasch model or the Rasch-MIRID, respectively. In the INPUT statement the degrees of discrimination need to be added. The SAS statements representing the model equation of the OPLM are given in Listing 7.

To demonstrate the flexibility of PROC NLMIXED, we will introduce a second way to test the MIRID structure besides the comparison with the basic model. One can test the MIRID structure with PROC NLMIXED by freeing one of the composite item parameters at a time and re-estimate the new model. These new models should not have a better fit than the original MIRID. The relaxation can be made for more than one item family at the time. A likelihood-ratio test can be used to test the difference in fit. If we free the first composite item parameter, the previous code for the Rasch-MIRID for example, should be modified into Listing 8.

As a result, the MIRID restrictions do not apply for the composite item of the first item family. This procedure of leaving one or more out is a way of testing whether the weights σ_k are equally valid for all composite items. That the weights would not be equal for all composite items is the most likely source of misspecification of the MIRID.

6.3.3 Output

The output of PROC NLMIXED contains the estimates of all parameters (the item parameters, the weights, the normalization constant and the variance of the person parameter distribution), the corresponding standard errors, a Wald test for testing the significance of the parameter estimates, and the value of the first derivative for the current parameter after the final iteration. In addition, four relative fit statistics are given: the deviance, defined as -2*log-likelihood value, the AIC value (Akaike's Information Criterion, Akaike, 1977), the AICC value (a finite-sample corrected version of AIC, Burnham & Anderson, 1998), and the BIC value (Schwartz' information criterion, Schwartz, 1978).

Note that the Wald test in PROC NLMIXED for the variance estimates does not give the correct p-value. The reference distribution that is used for the null hypothesis is a normal distribution, while a variance cannot be smaller than zero. An appropriate way of testing whether there are individual differences (random versus fixed intercept) is described by Verbeke and Molenberghs (2000) in terms of a likelihood ratio test. The reference distribution of this likelihood ratio test is a mixture of two χ^2-distributions, one with zero degrees of freedom and one with one degree of freedom, leading to p-values which are half the size of the p-values obtained under the classical χ^2_1 approximation to the null distribution (Verbeke & Molenberghs, in press). As the Wald test asymptotically equals the previously mentioned likelihood ratio test, a similar result applies for the PROC NLMIXED output. As a consequence, the correct p-value is half the size of the one mentioned in the PROC NLMIXED output.

As PROC NLMIXED provides the deviance for each model, and as the Rasch-MIRID is a restriction of the Rasch model, and the OPLM-MIRID of the OPLM, one can test the fit of the MIRID against the more general model using a likelihood-ratio test.

6.4 The MIRID CML program and the SAS MML approach compared

To compare the MIRID CML program with the SAS MML approach implemented in PROC NLMIXED, 140 datasets were simulated under the Rasch-MIRID, in which 2 features were varied: the number of persons, and the kind of distribution from which the person parameters were sampled. 80 data sets contained 200 persons and 60 contained 100 persons. The person parameters were sampled from three different distributions: a normal distribution with a mean of zero and a standard deviation of one (40 data sets with 200 persons, and 20 with 100 persons), a truncated normal distribution (20 data sets with 200 persons, and 20 with 100 persons), with the negative half omitted from the previous distribution, and a bimodal distribution, obtained by sampling half of the values for the person parameters from a normal distribution with mean 0, and the other half from a normal distribution with mean 4 (the standard deviations of both were equal to 1) (20 data sets with 200 persons, and 20 with 100 persons). All 140 datasets contained 40 items, 10 item families, and 3 types of components. The component item parameters (β), the weights (σ), and the normalization constant (τ) were sampled from a normal distribution with mean equal to zero and a standard deviation equal to one. Note that as we will use the group of data sets containing 200 persons and stemming from a normal distribution as reference condition, 40 datasets were included. In this small simulation study, we concentrated on the Rasch-MIRID, as the MIRID is our primarily point of interest and as the OPLM-MIRID is very similar to the Rasch-MIRID.

To differentiate amongst the different conditions, the following names will be used: (1) the "normal group", denoting the datasets containing 200 persons, with the person parameters sampled from a normal distribution (40 data sets), (2) the "truncated group", denoting the datasets containing 200 persons, with the person parameters sampled from a truncated normal distribution (20 data sets), (3) the "bimodal group", denoting the datasets containing 200 persons, with the person parameters sampled from a bimodal distribution (20 data sets). (4) The remaining three groups are named similarly, but the number of persons (100) is added as a suffix resulting in the "normal 100 group" (20 data sets), the "truncated 100 group" (20 data sets), and the "bimodal 100 group" (20 data sets).

All data sets were analyzed with the MIRID CML program and PROC NLMIXED using the previously mentioned options. In both, the MIRID CML program and PROC NLMIXED, we used a Quasi Newton-Raphson optimization technique together with the Davidon-Fletcher-Powell update method for the inverse Hessian matrix. In SAS, we choose the starting values for the component item parameters, the weights, and the normalization constant to be one, while in the MIRID CML program, first, a Rasch model was fitted and the values for item parameters obtained under the Rasch model are the basis for the starting values for all parameters of the Rasch-MIRID (based on a regression of the composite item parameters on the component item parameters). In this way, we obtained estimates for all item parameters and their standard error. An estimate for the variance of the person parameter distribution was only obtained by PROC NLMIXED (starting value = 1). Subsequently, estimates of the individual person parameters were calculated. The MIRID CML program uses a weighted likelihood approach (Warm, 1989) to estimate the values for the person parameters, often called Warm estimates. This results in one value for each possible sum score from the complete questionnaire. In PROC NLMIXED from SAS V8, empirical Bayes Estimates can be obtained for the individual realizations of a random effect -here the person parameter.

We expect the MIRID CML program to be superior with respect to the goodness-of-recovery for the data sets generated from a non-normal distribution: the data sets stemming from a bimodal distribution and the data sets stemming from a truncated normal distribution for the person parameter. The misspecification of the distribution should affect primarily the estimates of the person parameters. The MIRID CML and the Warm estimation method used in the MIRID CML program for the estimation of the person parameters make no assumptions about the distribution of the person parameters, whereas PROC NLMIXED in SAS imposes a normal prior distribution for these parameters, which does not correspond with the generating distribution. This effect should especially be visible in datasets with a smaller number of persons. More specifically, we expect PROC NLMIXED to result in an underestimation of the variance of the person parameter distribution for the bimodal distribution and the truncated normal distribution. For the data generated under a normal distribution an equal goodness-of-recovery is expected, but there could be an underestimation of variance of the random effect distribution as produced by PROC NLMIXED. As to the Warm estimates, we do not expect them to perform poorly in any condition, as Hoijtink and Boomsma (1996) found for example that Warm estimates perform reasonably well for sets of at least 15 items.

The fit of the models will be examined with five different statistics, related to the different kinds of parameters (the component item parameters, the weights, and the person parameters): First, as the component item parameters are defined up to an additive constant, the estimated values for the component item parameters will be correlated with the generating values. The higher these correlations, the better is the goodness-of-recovery for these parameters. Second, the ratio between the variances of the estimated versus the generating values for the component item parameters is calculated. The ratio of the variances is needed to detect differences in variance, which cannot be detected by a correlation. The closer this ratio is to one, the better is the goodness-of-recovery. Third, we calculated the mean squared differences between the original and the estimated values for the weights of the component item parameters only, because these parameters remain invariant under scale transformations (Butter et al., 1998). The higher the mean squared differences, the worse is the goodness-of-recovery of the model. Fourth, as the person parameters are defined up to an additive constant, the estimated values (Warm estimates or empirical Bayes estimates) for the person parameters are correlated with the generating values. The higher these correlations, the better is the goodness-of-recovery for these parameters. Fifth and finally, the variance of the generating person parameters is compared directly to the variance of the Warm estimates, the variance of the empirical Bayes estimates, and the variance of the random effect distribution as estimated by PROC NLMIXED. The latter estimate is direct, while the former to require an estimation of the model first.

A more extensive and more extensively documented simulation study about estimating the parameters of the Rasch-MIRID can be found in the article of Butter et al. (1998).

6.4.1 Results

In Table 6.2 the mean correlation between the estimated and the generating item parameters values over all datasets of the same kind are given, together with their standard deviations. Two-tailed Fischer Z transformations were made before testing the differences between the mean correlations. The standard deviations mentioned in Table 6.2 are the standard deviations of the correlations before the Fischer Z transformations.

To test the differences between the different conditions, we performed an analysis of variance for split plot designs on the Fischer Z transformed correlations, with the kind of generating distribution and the number of persons as between-

TABLE 6.2. Mean correlations between generating and estimated parameter values of the component-item parameters over all data sets of the same kind

Data Set	MIRID CML prog. Mean cor. (Std)	PROC NLMIXED Mean cor. (Std)	N
Normal	.986 (.005)	.986 (.005)	40
Truncated	.987 (.004)	.984 (.014)	20
Bimodal	.974 (.009)	.974 (.009)	20
Normal 100 pers	.974 (.008)	.974 (.007)	20
Truncated 100 pers	.972 (.008)	.972 (.008)	20
Bimodal 100 pers	.961 (.013)	.961 (.013)	20

subject factors and the estimation method (PROC NLMIXED vs MIRID CML program) as a within-subject factor. Only the main effects of the two between-subject factors turned out to be significant. From post-hoc t-tests, we can conclude that the difference between the normal and the truncated normal group is not significant, whereas the bimodal group does significantly worse. As to the number of persons, the goodness-of-recovery is significantly worse when the number of persons decreases from 200 to 100. The one within-subject factor does not yield a significant difference. Both approaches (PROC NLMIXED vs MIRID CML Program) do about equally well.

In Table 6.3, the mean values of the ratio of the variance of the component item parameters as estimated by both approaches, compared to the variance of the generating item parameters over all data sets of the same kind, are displayed. To test the differences between the different conditions, we performed an analysis of variance (split plot design) on the variance ratios, with the same design as for the previous ANOVA. Again, only the main effects of the two between-subject factors turned out to be significant. Both approaches perform somewhat less well if the person parameter distribution deviates from the normal distribution, and if the number of persons decreases. In addition, according to the F-tests per single data set, the ratios never differ significantly from 1 (all p-values are even larger than .24).

In Table 6.4, the mean values for the mean squared differences between the original and the estimated weights of the component item parameters over the different data sets of one kind are shown. Using PROC NLMIXED, the estimates for one dataset deviated strongly from the generating values. Excluding this one data set, the mean of the mean squared difference and its standard deviation drops to the values displayed in Table 6.4. An analysis of variance with the same design as the previous ANOVAs revealed no significant effects.

TABLE 6.3. Mean ratios of variances between generating and estimated parameter values of the component-item parameters over all data sets of the same kind

Data Set	MIRID CML prog. Mean ratio (Std)	PROC NLMIXED Mean ratio (Std)	N
Normal	1.034 (.075)	1.034 (.075)	40
Truncated	1.076 (.086)	1.073 (.081)	20
Bimodal	1.061 (.092)	1.055 (.093)	20
Normal 100 pers	1.048 (.164)	1.048 (.164)	20
Truncated 100 pers	1.139 (.189)	1.141 (.189)	20
Bimodal 100 pers	1.158 (.216)	1.158 (.230)	20

TABLE 6.4. Means of the mean squared differences between generating and estimated weights of the component item parameters over all data sets of the same kind

Data Set	MIRID CML prog. Mean sq. dif. (Stdev)	PROC NLMIXED Mean sq. dif. (Stdev)	N
Normal	.033 (.049)	.027 (.031)	40
Truncated	.031 (.029)	.031 (.028)	19
Bimodal	.055 (.060)	.053 (.060)	20
Normal 100 pers	.052 (.100)	.050 (.095)	20
Truncated 100 pers	.098 (.237)	.099 (.244)	20
Bimodal 100 pers	.126 (.314)	.115 (.269)	20

To summarize, we were not able to find differences in goodness-of-recovery between the MIRID CML program and the PROC NLMIXED MML approach for the component item parameters and the weights. Therefore, both approaches can be concluded to perform equally well for the estimation of the item parameters and the weights.

As for the person parameters, the estimated values were correlated with the generating values. The means of the correlations and the corresponding standard deviations are displayed in Table 6.5. An analysis of variance on the Fischer Z transformed correlations with the same design as for the previous ANOVAs revealed a significant main effect for the kind of distribution and a significant interaction effect between the kind of distribution and the kind of estimation method. Based on post-hoc tests, the correlations are lower for the truncated distributions. For this distribution also a small difference is found between the two estimation methods (.841 vs .832, and .852 vs .845, for 200 and 100 persons respectively), although it is not significant. Note that in one of the datasets of the bimodal group, no Warm estimates could be computed due to computational problems.

Finally, the variance of the originally simulated person parameters was com-

TABLE 6.5. Mean correlations between generating and estimated person parameters over all data sets of the same kind

Data Set	MIRID CML prog. W.E. (Std)	PROC NLMIXED E.B.E. (Std)	N
Normal	.931 (.009)	.933 (.009)	40
Truncated	.841 (.021)	.832 (.020)	20
Bimodal	.963 (.007)	.967 (.007)	19
Normal 100 pers	.929 (.015)	.931 (.015)	20
Truncated 100 pers	.852 (.025)	.845 (.025)	20
Bimodal 100 pers	.961 (.008)	.964 (.007)	20

W.E. = Warm estimates; E.B.E. = empirical Bayes estimates

pared with the variance of the Warm estimates, with the variance of the empirical Bayes estimates, and with the variance of the person parameter distribution as estimated by PROC NLMIXED. In Table 6.6 the mean difference between the estimated variance and the variance of the simulated person parameters is shown. Each mean difference is also tested against zero. A positive value reflects an overestimation of the variance and a negative value reflects an underestimation.

An analysis of variance with the same design as for the previous ANOVAs revealed two significant main effects (kind of distribution and estimation method) and three significant interaction effect (all interactions with the estimation method). A post hoc analysis revealed that all row-wise differences, except for one, in Table 6.6 are significantly different. The variance of the Warm estimates overestimates the variance of the generating values in the normal group and in the truncated normal group. In the bimodal group, there is a slight underestimation. The variance as estimated from the empirical Bayes estimates, is always underestimated, especially for the smaller number of persons. A similar underestimation, but much smaller, is found for the estimate of the random effect variance from PROC NLMIXED. However, in three of the six rows the difference with zero is not significant.

We also investigated the absolute deviations from the variance of the originally simulated parameters. An analysis of variance with the same design as for the previous ANOVAs revealed two significant main effects (kind of distribution and estimation method) and one significant interaction effect (between the kind of distribution and the estimation method). The PROC NLMIXED estimate of the random effect variance is always the closest to the expected variance, except for the bimodal 100 group, where the variance of the Warm estimates is the closest to the expected variance. The bias of the empirical Bayes estimates

TABLE 6.6. Means of the differences between variance of the estimated person parameters and the variance of generating person parameters over all data sets of the same kind

Data Set	MIRID CML prog.	PROC NLMIXED		N
	W.E. (Stdev)	E.B.E (Std)	E.V. (Std)	
Normal	.065* (.177)	-.059* (.141)	.020 (.140)	40
Truncated	.111** (.037)	-.129** (.043)	-.030** (.039)	20
Bimodal	-.065** (.101)	-.167** (.077)	-.071** (.095)	19
Normal 100 pers	.066** (.054)	-.090** (.052)	-.011 (.062)	20
Truncated 100 pers	.109** (.047)	-.133** (.059)	-.034* (.054)	20
Bimodal 100 pers	-.084** (.105)	-.187** (.118)	.078 (.216)	20

W.E. = Warm estimates; E.B.E. = empirical Bayes estimates; E.V. = random effect variance as estimated by PROC NLMIXED

* = mean difference is significantly different from 0 at .05 level

** = mean difference is significantly different from 0 at .01 level is always the largest, except in the normal group where it is equal to the bias in the Warm estimates.

6.4.2 Discussion

First, it was expected that the CML approach would do better in recovering the generating parameter values when the data were generated from a non-normal person parameter distribution, but there was actually no effect for the item parameters and the weights. The difference we found concerns the person parameters. It was very small in terms of correlations, and restricted to the datasets with a truncated distribution. The difference was larger for the estimated variance of the person parameters.

In general, the direct estimation of the variance of the person parameter distribution from PROC NLMIXED gave the best results. The variance of the empirical Bayes estimates underestimates the variance of the generating parameters in all kinds of datasets, also those generated from a normal distribution. As expected, the estimated variance is smaller for the datasets generated from the two non-normal distributions than for the datasets generated from the normal distribution, except for the variance as estimated by PROC NLMIXED in the bimodal 100 group.

In contrast, the Warm estimates overestimate the variance in the normal and in the truncated group, and underestimate the variance in the bimodal groups. Both were not predicted. The underestimation in the bimodal groups can be related to the fact that Warm estimates are negatively biased for large, positive θ-values (Warm, 1989). Since there are more such values expected for the data generated with the bimodal distribution, the effect of the negative bias is expected to be relatively large, which explains the underestimation of the variance. The overestimation of the variance in the normal sample and the truncated sample is similar to the results obtained by Hoijtink and Boomsma (1996) with a normal generating distribution. We did not find any explanation in literature for this overestimation.

Despite the differences found, we cannot conclude one approach to be better in general than the other. Nevertheless, we can conclude that for our data MIRID CML, supplemented with the Warm estimates, should be preferred if estimates for the person parameters are requested. Warm estimates showed equal (for the normal and the truncated group) or less bias (for the bimodal group) in terms of overestimation or underestimation of the variance of the person parameter distribution than the empirical Bayes estimates. If one does not need individual estimates, both approaches are inferior to the estimate obtained by from PROC NLMIXED for the variance of the random effect.

A remarkable difference between the MIRID CML program and PROC NLMIXED was the time needed for the estimation of the models: the MIRID CML program, which first fits the Rasch model and only then the Rasch-MIRID, takes two to three minutes for a single simulated data set. With PROC NLMIXED, we fitted only the Rasch-MIRID and this took 15 to 30 minutes for a single simulated data set, not including the empirical Bayes estimates for the random effect.

A major advantage of the SAS approach is that PROC NLMIXED is a very broad procedure that can be used for fitting many other generalized linear and non-linear models with fixed and random effects (see e.g. Rijmen et al., in press). One can for example test the MIRID structure with PROC NLMIXED by freeing one of the composite item parameters at a time and re-estimating the new models, as explained earlier. The price to pay for this generality is computing time.

6.5 Conclusions

Both approaches are useful for fitting MIRIDs, and do (about equally) well according to the goodness-of-recovery statistics for the item parameters and the weights. Taking the person parameters into account, small differences between both approaches are found: the CML approach supplemented with Warm estimates can be preferred when individual estimates of the person parameter are requested, and PROC NLMIXED can be preferred when an estimate of the variance of the person parameter distribution suffices.

A major advantage of PROC NLMIXED is that it is very flexible because of the many different options, and the many different model variants it can fit. On the other hand, PROC NLMIXED is rather time consuming. The MIRID CML program is less flexible as it has less options and can fit only MIRIDs, but it is faster.

6.6 Listings

6.6.1 Listing 1: SAS statements for reading data for the Rasch model

Comments are written between /*...*/:

```
DATA Rasch; /*name of the data set within the SAS
environment*/
INFILE 'c:\data\Rasch.dat'; /*name and location of
datafile*/
INPUT Person $ y X1-X20;
/*Variables: Person: person label (followed by $ because
person is a string (character)), y: responses, X1-X20 are
the dummy variables that form the columns of the design
matrix*/
RUN;
```

6.6.2 Listing 2: SAS statements for reading data for the Rasch-MIRID

Comments are written between /*...*/:

```
DATA RaschMirid; /*name for SAS data set*/
INFILE 'c:\data\raschmirid.dat'; /*name and location of
data file*/
INPUT Person $ y X1-X15 X0;
/*Variables: Person: person label, y: responses, X1-X15:
dummy variables for component item parameters, X0: dummy
variable denoting the composite item*/
RUN;
```

6.6.3 Listing 3: SAS statements for reading data for the OPLM

Comments are written between /*...*/:

```
DATA Oplm; /*name of the data set within the SAS
environment*/
INFILE 'c:\data\oplm.dat'; /*name and location of
datafile*/
INPUT Person $ y X1-X20 A;
/*Variables: Person: person label (followed by $ because
person is a string (character)), y: responses, X1-X20 are
the dummy variables that form the columns of the design
matrix, A: discrimination values*/
RUN;
```

6.6.4 Listing 4: SAS statements for reading data for the OPLM-MIRID

Comments are written between /*. . . */:

```
DATA OplmMirid; /*name for SAS data set*/
INFILE 'c:\data\oplmmirid.dat'; /*name and location of data
file*/
INPUT Person $ y X1-X15 X0 A;
/*Variables: Person: person label, y: responses, X1-X15:
dummy variables for component item parameters, X0: dummy
variable denoting the composite item, A: discrimination
values*/
RUN;
```

6.6.5 Listing 5: SAS statements for estimating the Rasch model

Comments are written between /*. . . */:

```
PROC NLMIXED DATA=Rasch METHOD=gauss NOAD QPOINTS=20
TECHNIQUE=QuaNew UPDATE=dfp;
/*Specification of data and estimation procedure*/
PARMS Beta1-Beta20=1 VarTheta=1; /*Parameters and their
starting values*/
ex=exp(theta-X1*Beta1-X2*Beta2-X3*Beta3-X4*Beta4-X5*Beta5
-X6*Beta6-X7*Beta7-...-X18*Beta18-X19*Beta19-X20*Beta20);
p=ex/(1+ex); /*Formula of Rasch model, see Equation 6.3*/
MODEL y ~ binary(p); /*the Rasch model is a model for
binary data*/
RANDOM theta ~ normal(0,VarTheta) SUBJECT=Person;
/*specification of distribution of the random intercept θ:
The persons are normally distributed with mean zero and
variance equal to VarTheta. The subject option specifies
over which variable the random effects are distributed. If
the option OUT=SAS-data set is specified, empirical Bayes
estimates for the realizations of the person parameter are
calculated and stored in the specified SAS data set*/
RUN;
```

6.6.6 Listing 6: SAS statements for estimating the Rasch-MIRID

Comments are written between /*. . . */:

```
PROC NLMIXED DATA=RaschMIRID METHOD=gauss NOAD QPOINTS=20
TECHNIQUE=QuaNew UPDATE=dfp;
/*Specification of data and estimation procedure*/
PARMS Beta1-Beta15=1 Sigma1-Sigma3=1 Tau=1 VarTheta=1;
/*Specification of the parameters of the Rasch-MIRID and
their starting values; Beta1-Beta15 are the component item
parameters, Sigma1-Sigma3 are the weights of the three
types of components and Tau is the normalization constant*/
ex=exp(theta+(1-X0)*(-X1*Beta1-X2*Beta2-X3*Beta3-X4*Beta4
-X5*Beta5-...-X14*Beta14-X15*Beta15)
/*part specific to component items*/
+X0*(-X1*Beta1*Sigma1-X2*Beta2*Sigma1-X3*Beta3*Sigma1
-X4*Beta4*Sigma1-X5*Beta5*Sigma1-X6*Beta6*Sigma2
-X7*Beta7*Sigma2-X8*Beta8*Sigma2-X9*Beta9*Sigma2
-X10*Beta10*Sigma2-X11*Beta11*Sigma3-X12*Beta12*Sigma3
-X13*Beta13*Sigma3-X14*Beta14*Sigma3-X15*Beta15*Sigma3
-Tau));
/*part specific to composite items*/
p=ex/(1+ex); /*inverse logit transformation*/
MODEL y ~ binary(p); /*the Rasch-MIRID is a model for
binary data*/
RANDOM theta ~ normal(0,VarTheta) SUBJECT=Person;
RUN;
```

6.6.7 Listing 7: SAS statements for model equation of the OPLM

Comments are written between /*. . . */:

```
ex=exp(A*(theta-X1*Beta1-X2*Beta2-X3*Beta3-X4*Beta4
-X5*Beta5-X6*Beta6-X7*Beta7-...-X18*Beta18-X19*Beta19
-X20*Beta20));
p=ex/(1+ex);
```

6.6.8 Listing 8: SAS statements for estimating a Rasch-MIRID in which the first composite item parameter is freed

Comments are written between /*...*/:

```
PROC NLMIXED DATA=RaschMIRID METHOD=gauss NOAD QPOINTS=20
TECHNIQUE=QuaNew UPDATE=dfp;
/*Specification of data and estimation procedure.*/
PARMS Beta1-Beta15=1 Sigma1-Sigma3=1 Tau=1 VarTheta=1
CompBeta1=1;
/*CompBeta1 is the freed composite item parameter of the
first item family*/
ex=exp(theta+(1-X0)*(-X1*Beta1-X2*Beta2-X3*Beta3-X4*Beta4
-X5*Beta5-...-X14*Beta14-X15*Beta15)
/*part specific to component items, nothing changes*/
+X0*(-X1*CompBeta1-X2*Beta2*Sigma1-X3*Beta3*Sigma1
-X4*Beta4*Sigma1-X5*Beta5*Sigma1-X7*Beta7*Sigma2
-X8*Beta8*Sigma2-X9*Beta9*Sigma2-X10*Beta10*Sigma2
-X12*Beta12*Sigma3-X13*Beta13*Sigma3-X14*Beta14*Sigma3
-X15*Beta15*Sigma3-Tau));
/*part specific to composite items: X1*Beta1*Sigma1-
X6*Beta6*Sigma2-X11*Beta11*Sigma3 is omitted and replaced
by X1*CompBeta1 */
p=ex/(1+ex); /*inverse logit transformation*/
MODEL y ~ binary(p);
RANDOM theta ~ normal(0,VarTheta) SUBJECT=Person;
RUN;
```

6.7 References

Akaike, H. (1977). On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Proceedings of symposium on applications of statistics* (p. 27-41). Amsterdam, The Netherlands: North Holland.

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

Bunday, B. D. (1984). *Basic optimisation methods.* London: Arnold.

Burnham, K., & Anderson, D. (1998). *Model selection and inference: a practical information-theoretic approach.* New York: Springer-Verlag.

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

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

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments and applications.* New York: Springer-Verlag.

Fischer, G. H., Ponocny-Seliger, E., Ponocny, I., & Parzer, P. (1998). *LPCM-WIN* (Computer program). Vienna, Austria.

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

Hoijtink, H., & Boomsma, A. (1996). Statistical inference based on latent ability estimates. *Psychometrika*, *61*, 313-330.

Pinheiro, J. C., & Bates, D. M. (1995). Approximation to the log-likelihood function in the nonlinear mixed-effects model. *Jounal of Computational and Graphical Statistics*, *4*, 12-35.

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

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

Sas onlinedoc^{TM} *version 8.* (1999). Carey NC, USA.

Sas system v8 for windows (Computer program). (1999). Carey NC, USA: SAS Institute Inc.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Smits, D. J. M., & De Boeck, P. (2003). A componential model for guilt. *Multivariate Behavioral Research*, *38*, 161-188.

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

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer-Verlag.

Verbeke, G., & Molenberghs, G. (in press). The use of the score test for inference on variance components. *Biometrics*, *59.*

Verhelst, N. D. (1993). Itemresponstheorie [item response theory]. In T. J. H. M. Eggen & P. F. Sanders (Eds.), *Psychometrie in de praktijk [psychometrics in practice]* (p. 83-176). Arnhem, The Netherlands: CITO.

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* (p. 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). The Netherlands, Arnhem: CITO.

Warm, T. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*, 427-450.

Wolfinger, R. (1999). *Fitting non linear models with the new NLMIXED procedure.* SUGI 24 Conference proceedings, Paper 287. Cary, NC.

Uitgebreide samenvatting

Emoties vormen een belangrijk deel van ons dagdagelijks bestaan. Diverse aspecten ervan werden reeds bestudeerd en staan beschreven in verscheidene artikels en boeken, waaronder hun ontstaan, biologische basis en herkomst, neurologisch en hormonaal substraat, structuur, gedragsmatige eigenschappen, uitdrukkingswijzen, en nog vele andere. Een veelheid aan methodes werd ontwikkeld om al deze aspecten te onderzoeken. In deze dissertatie focussen we ons op de zelfrapporteringsmethode. Ons doel is aan te tonen dat bepaalde data-analyse technieken die ontwikkeld zijn voor cognitief onderzoek, ook nuttig kunnen zijn om de structuur en de expressie van emoties te onderzoeken. We beperken ons tot twee negatieve emoties, met name het schuldgevoel en woede. Vier verschillende benaderingen om de componentiële en de relationele structuur van emoties te onderzoeken zullen worden voorgesteld. Deze worden elk geïllustreerd met een analyse van data over schuldgevoelens. Daarnaast beschrijven we één benadering om de gedragsmatige expressie van emoties te modelleren. Deze zal geïllustreerd worden een analyse van data over verbale agressie.

In het eerste deel van deze samenvatting beschrijven we de inhoudelijke resultaten. In het tweede deel beschrijven we de verschillende modellen die we gebruikten.

1. Onderzoek naar emoties

1.1 Componentiële theorieën over emoties

Verscheidene theorieën over de componentiële structuur van emoties stellen dat emoties gekenmerkt worden door en tevens van elkaar onderscheiden kunnen worden aan de hand van een specifiek patroon van componenten. Deze benadering is het meest uitgesproken in de zogenaamde appraisal-literatuur. Appraisaltheorieën stellen dat een emotie ontstaat doordat de omgeving door het individu beoordeeld of 'appraised' wordt in relatie tot de eigen doelen en noden (e.g., Ellsworth & Smith, 1988; Kuppens, Van Mechelen, Smits, & De Boeck, in press; Ortony, Clore, & Collins, 1988; Omdahl, 1995; Reisenzein & Hofmann, 1993; Roseman, 1984; Scherer, 1993, 1997, 1999; Smith & Ellsworth, 1985). De basisgedachte van deze theorieën is dat elke emotie gekenmerkt wordt door een specifiek patroon van beoordelingen of appraisals. Echter, sommige auteurs stellen dat niet alleen appraisals, maar ook actietendensen (de neiging om een bepaalde actie te ondernemen) belangrijke componenten zijn van emoties, daar men emoties ook kan karakteriseren aan de hand van de specifieke actietendens(en) waarmee ze geassocieerd zijn (e.g., Fischer, 1991; Frijda, 1986; Frijda, Kuipers, & Schure, 1989; Lazarus, 1991; Oatley & Jenkins, 1996; Skiffington, Fernandez, & McFarland, 1998). Dit alles kunnen we als volgt samenvatten: we veronderstellen dat elke emotie gekenmerkt wordt door een specifiek patroon van appraisals en actietendensen. Appraisals en actietendensen zullen allebei omschreven worden met de term *componenten* van emoties. In tegenstelling tot wat beweerd wordt in sommige appraisaltheorieën veronderstellen we niet dat appraisals en actietendensen de enige componenten van emoties zijn. Vele andere aspecten, zoals bijvoorbeeld lichamelijke gewaarwordingen en veranderingen, kunnen evenzeer een belangrijke rol vervullen (Berkowitz, 1990; Izard, 1993), maar daar deze zich moeilijk tot niet laten vatten met de zelfrapporteringsmethode gaan we daar niet verder op in.

In appraisaltheorieën worden emoties expliciet gelinkt aan de situatie waarin ze ervaren worden. Bijvoorbeeld, een situatie die als gevaarlijk kan ervaren worden, kan tot angst leiden, en een situatie waarin men gelooft een norm te hebben overtreden, kan gevolgd worden door schuldgevoelens. Door gebruik te maken van een situationele benadering gecombineerd met appraisals kunnen emotionele responsen op de situatiesverklaard worden. Afhankelijk van de gevolgde benadering kunnen we ook inzicht verwerven in de interactie tussen personen en situaties.

1.1.1 De componentiële structuur van emoties: het schuldgevoel

In Hoofdstuk 1 wordt een benadering voorgesteld om de componentiële structuur van emoties te onderzoeken. Voor deze benadering is het noodzakelijk dat situaties van elkaar verschillen in de mate waarin ze deze componenten en emoties induceren. Let wel, het gaat hier om andere verschillen dan hoofdeffecten, dit wil zeggen dat er geen perfecte correlatie mag bestaan tussen componenten over situaties. Verder veronderstellen we dat mensen onderling verschillen in de mate waarin ze gevoelig zijn voor bepaalde componenten en emoties. Dit is echter geen noodzakelijke veronderstelling voor de huidige benadering. De benadering werd toegepast op een dataset over situationele schuldgevoelens. De data werden verzameld in het kader van mijn licentiaatsthesis en worden in deze verhandeling gebruikt om verscheidene modelleringstechnieken te illustreren en de mogelijkheden ervan te onderzoeken.

Eerst beschrijven we kort het werk verricht voor mijn licentiaatsthesis. Op basis van een literatuuronderzoek werd een situationele theorie over schuldgevoelens opgesteld. Volgens deze theorie was het schuldgevoel gebaseerd op drie appraisals en twee meer actiegerichte componenten. De drie appraisals zijn verantwoordelijkheid, normovertreding en negatieve zelf-evaluatie. De twee meer actiegerichte componenten zijn piekeren (covert) en de neiging om goed te maken wat men verkeerd deed of naliet te doen (overt). In een eerste studie werden situaties verzameld waarin personen zich schuldig gevoeld hadden. Hieruit werden tien situaties geselecteerd met behulp van criteria zoals begrijpelijkheid en variatie in inhoud. In een eerste verkennende studie vonden we dat de component verantwoordelijkheid in alle situaties vrij hoog werd ingeschat, en dat er weinig verschillen tussen mensen waren met betrekking tot deze beoordelingen. Daarom beschouwden we verantwoordelijkheid als een meer objectieve appraisal, voornamelijk gebaseerd op de situaties en niet beïnvloed door persoonskenmerken. Omdat we interindividuele verschillen verwachtten in de schuldgevoeligheid, verkozen we subjectieve appraisals boven objectieve. Dit leidde ertoe dat verantwoordelijkheid niet als component in het hoofdonderzoek werd opgenomen. Let wel, we beweren niet dat verantwoordelijkheid geen component kan zijn van het schuldgevoel. Het kan een component zijn die geen interindividuele verschillen vertoont, of het kan zijn dat de tien situaties niet optimaal zijn om zulke verschillen aan het licht te brengen. Een tweede resultaat was dat de correlatie tussen de appraisals normovertreding en negatieve zelf-evaluatie zeer hoog was (.98), zodat het niet mogelijk was om beide componenten te relateren aan het schuldgevoel, daar het onmogelijk zou zijn om de bijdragen van beide componenten van elkaar te onderscheiden. Omdat de appraisal normovertreding in de literatuur als een belangrijkere appraisal beschouwd wordt, behielden we deze en werd de appraisal negatieve zelf-evaluatie weggelaten uit het hoofdonderzoek. De drie componenten die we behielden zijn normovertreding (appraisal), piekeren en de neiging om het goed te maken (actietendensen).

Het hoofdonderzoek, gebaseerd op heranalyse voor deze verhandeling (dit geldt voor al wat volgt) leidde tot de volgende bevindingen: (1) De interindividuele verschillen in het schuldgevoel en de componenten van dit schuldgevoel kunnen met één onderliggende dimensie beschreven worden, met name de gevoeligheid van een persoon voor de componenten van het schuldgevoel en voor het schuldgevoel zelf. (2) De drie componenten worden in verschillende mate uitgelokt door de verschillende situaties: Sommige situaties bevorderen de ene component, terwijl andere situaties andere componenten bevorderen. (3) Deze drie componenten volstonden om de situationele schuldgevoelens te voorspellen. (4) Piekeren was de belangrijkste component, gevolgd door de neiging om het goed te maken en tenslotte normovertreding. Piekeren en de neiging om het goed te maken zijn meer actiegericht (respectievelijk covert en overt), terwijl normovertreding een pure appraisal is. Dit resultaat suggereert dat het schuldgevoel meer is dan louter een appraisal of een beoordeling van een situatie. Deze bevinding stemt overeen met de argumentatie van Frijda et al. (1989).

In de vorige analyse veronderstelden we dat het belang van een component voor een emotie dezelfde is voor alle personen of met andere woorden een fixed effect is. We veronderstelden dat alle interindividuele verschillen veroorzaakt worden door de positie van de personen op de onderliggende dimensie of trek (één latente trek was voldoende, component-specifieke trekken waren niet nodig). Echter, zoals uitgelegd in Hoofdstuk 2, kan het belang van een component verschillen van persoon tot persoon. Dit leidt tot een interactie tussen personen en situaties, daar situaties verschillen van elkaar met betrekking tot de componenten die ze uitlokken, en personen van elkaar verschillen in het belang dat deze componenten hebben voor de resulterende emotie. Een voorbeeld: voor sommige mensen is normovertreding belangrijker, terwijl voor anderen de neiging om het goed te maken belangrijker is. Dit betekent dat de basis van het schuldgevoel (of de betekenis van het schuldgevoel) anders is naargelang de persoon in kwestie. Het gevolg is dat verschillende personen zich in andere situaties schuldig zullen voelen (gegeven dat de situaties verschillen met betrekking tot de componenten). Daar we geen a priori hypotheses hadden over voor welke componenten het belang persoonsafhankelijk is, toetsten we dit voor alledrie. De hypothese dat er geen individuele verschillen zijn in het belang van een component kon voor geen van de drie componenten verworpen worden. Daarom kunnen we de structuur van het schuldgevoel, zoals hier gemodelleerd, beschouwen als een algemene structuur die gelijk is voor alle personen. Daar er geen interindividuele verschillen gevonden werden, buiten diegene beschreven door de algemene onderliggende trek, kan men tentatief concluderen dat schuldgevoelens gebaseerd zijn op slechts één onderliggende gevoeligheid. Deze conclusie zal echter gewijzigd worden in hoofdstuk 3 over de relationele structuur van emoties.

1.1.2 Relationele structuur van emoties

In Hoofdstuk 3 wordt een benadering voorgesteld om de relaties tussen componenten en tussen de componenten en de emotie te onderzoeken. Deze wordt opnieuw geïllustreerd met de data over situationele schuldgevoelens. Verschillende relationele structuren werden opgesteld en ten opzichte van elkaar getoetst. We gebruiken dezelfde componenten als in de vorige studie, namelijk normovertreding, piekeren en de neiging om het goed te maken.

De volgende structuren werden bestudeerd: een lineaire-sequentie-structuur, een ster-structuur, een cluster-structuur en een item-familie-structuur. Hoewel de verschillende structuren in causale termen zullen beschreven worden, hebben we geen evidentie voor causale verbanden. De associaties kunnen dus ook gebaseerd zijn op andere dan causale relaties. Een lineaire-sequentie-structuur houdt in dat de componenten van het schuldgevoel en het resulterende schuldgevoel op een lineaire wijze geordend kunnen worden met betrekking tot hoe ze elkaar en het resulterende schuldgevoel beïnvloeden. Een ster-structuur impliceert dat alle componenten het schuldgevoel beïnvloeden, maar niet elkaar. Een cluster-structuur impliceert dat er clusters van componenten bestaan die elkaar en ook de emotie beïnvloeden, zonder dat er relaties bestaan tussen componenten die tot verschillende clusters behoren. Voor de dataset over schuldgevoelens definieerden we de clusters op basis van het type component: één cluster wordt gevormd door normovertreding (appraisal) en het schuldgevoel, en een tweede cluster door piekeren, de neiging om het goed te maken (actiegerichte componenten) en het schuldgevoel. Het schuldgevoel vormt dus de overlap tussen de twee clusters. Tenslotte, een item-familie-structuur houdt in dat alle componenten en de emotie in gelijke mate met elkaar gerelateerd zijn binnen een situatie. Deze mate kan echter verschillen van situatie tot situatie. Deze laatste structuur paste het beste bij de data, beter dan structuren met relaties tussen de componenten of tussen de componenten en de emotie (bovenop de algemene schuldgevoeligheid). Dit betekent dat er buiten de interindividuele verschillen gebaseerd op de onderliggende schuldgevoeligheid, ook situatie-specifieke interindividuele verschillen bestaan. Dit vervolledigt de eerder vermelde tentatieve conclusie.

1.1.3 Abstractie makend van interindividuele verschillen

In Hoofdstuk 4 maken we abstractie van interindividuele verschillen door gebruik te maken van een marginale benadering. De reden hiervoor is dat de afhankelijkheden die we gebruikten bij het bestuderen van de relationele structuur enerzijds de resultaten kunnen verstoren indien ze niet in het model opgenomen worden en anderzijds de interpretatie van hoofdeffecten bemoeilijken indien ze wel in het model opgenomen worden. Wanneer we abstractie maken van interindividuele verschillen door gebruik te maken van een marginale benadering hebben de afhankelijkheden niet langer deze effecten. De prijs die men hiervoor betaalt is dat de effecten effecten zijn op het niveau van de populatie en niet op het niveau van de persoon. De benadering wordt geïllustreerd met een deel van de schuld-data (slechts drie situaties). De antwoorden op de items die het schuldgevoel en de verschillende componenten meten, werden voorspeld op basis van een effect van de situatie en een effect van het itemtype (normovertreding, piekeren, de neiging om het goed te maken en het schuldgevoel). Het schuldgevoel wordt dus niet ontleed in componenten zoals in Hoofdstuk 1 en 2, maar wordt op gelijkaardige wijze behandeld als de componenten, namelijk als een itemtype. Alle effecten waren significant. Bovenop de gemiddeldenstructuur werden ook de relaties tussen de items gemodelleerd in overeenstemming met een item-familie-structuur. Deze structuur voor de associaties tussen de items werd ondersteund door de data.

1.2 Emotioneel gedrag: verbale agressie

Een andere, meer inhoudelijke topic van deze dissertatie is het verband tussen de actietendens en het emotioneel gemotiveerd gedrag. Een specifiek gedrag, nauw verbonden met woede, zal onderzocht worden, met name verbale agressie. De keuze voor een actietendens gerelateerd aan woede werd ingegeven door twee redenen: Ten eerste ervaart men vaker woede dan schuldgevoelens (Zelenski & Larsen, 2000). Bijgevolg is er meer gelegenheid om de actietendensen en de ermee gepaard gaande gedragingen te onderzoeken. Ten tweede wordt woede regelmatig beschreven als een actiegerichte emotie, die bijvoorbeeld tot verbale agressie kan leiden (Averill, 1983; Cornell, Peterson, & Richards, 1999; Kassinove, Sukhodolsky, Tsytsarev, & Solovyova, 1997; Kinney, Smith, & Donzella, 2001). Gedragingen geassocieerd met het schuldgevoel daarentegen zijn minder duidelijk afgebakend en meer covert. Bijgevolg zijn ze moeilijker te observeren. We verkozen verbale agressie boven de ernstigere vormen van agressie, omdat deze vorm gewoner en minder sociaal onwenselijk is. Hoewel het effect van verbale agressie kleiner kan zijn dan het effect van bijvoorbeeld fysieke agressie, kan het toch een belangrijk fenomeen zijn wegens zijn meer 'gewone' karakter.

Verbale agressie kan schade toebrengen aan allerhande relaties en een bron zijn van veel conflicten.

Drie verbaal agressieve (VA) gedragingen werden geselecteerd: vloeken, schelden en het uitschreeuwen. Voor elk VA gedrag werden twee vragen gesteld aan de proefpersonen: één over de actietendens (het willen stellen van VA gedrag in de situatie), genaamd een willen-item, en één over het effectief gestelde gedrag, genaamd een doen-item. We onderzochten in welke mate het gedrag afhangt van de actietendens, of inhibitie een rol speelt en of er situationele en gedragsafhankelijke en/of interindividuele verschillen zijn wat betreft inhibitie.

De resultaten worden vermeld in twee verschillende hoofdstukken. De resultaten vermeld in Hoofdstuk 2 zijn de volgende: (1) De VA actietendens heeft een duidelijk voorspellende kracht voor het VA gedrag. (2) Doen-items werden moeilijker bevonden dan willen-items, zodat men kan stellen dat er sprake was van inhibitie. (3) Er zijn interindividuele verschillen in het gewicht van zich verbaal agressief willen gedragen bij de predictie van VA gedrag.

In Hoofdstuk 5 wordt inhibitie meer in detail onderzocht. De volgende theorie over verbale agressie en de inhibitie ervan werd opgesteld: De neiging om zich verbaal agressief te gedragen en de inhibitie van dit gedrag kan beïnvloed worden door gedragspecifieke factoren, situatiespecifieke factoren, persoonspecifieke factoren of door factoren die eigen zijn aan een combinatie van twee van de voorgaande factoren. De persoonspecifieke factoren kunnen beschouwd worden als trekken (een actietendenstrek en een inhibitietrek). De situatiespecifieke en de gedragspecifieke factoren zijn kenmerken zoals de zichtbaarheid van het gedrag of de mate waarin een situatie frustrerend is, enz.

We vonden dat de VA actietendens gebaseerd is op een latente trek (verbale agressietrek) en dat kenmerken eigen aan de combinatie van VA gedragingen en situaties ook een rol spelen. Inhibitie daarentegen wordt vooral bepaald door interindividuele verschillen (inhibitietrek), en minder door de situatie of het gedrag in kwestie. De benadering werd gevalideerd door de verbale agressietrek en de inhibitietrek te correleren met de volgende gerelateerde metingen: de scores van de proefpersonen op de Trait Anger schaal van Spielberger (1980), op de schalen Anger In, Anger Out, Anger In Control en Anger Out Control van de Zelf-Expressie en Controle Schaal van Van Elderen, Maes, Komproe, and Kamp (1997) –een aanpassing van de Anger Expression Schaal van Spielberger, Johnson, and Jacobs (1982)–, en op de Directe Agressie schaal en de Indirecte Agressie schaal uit de Buss-Durkee Hostility Inventory-Dutch (Lange, Hoogendoorn, Wiederspahn, & Beurs, 1995). De parameters specifiek voor de combinatie van VA gedrag en situaties werden gecorreleerd met verscheidene situationele eigenschappen zoals de aanwezigheid van getuigen, de mate waarin de situatie frustrerend is, enz. De verbale-agressie-trek was voornamelijk gecorreleerd met Trait Anger, Directe Agressie en Indirecte Agressie. Inhibitie bleek voornamelijk gerelateerd aan coping met woede: het was negatief gecorreleerd met Anger Out en positief met Anger In en Anger Out Control. Tussen de verbale agressietrek en de inhibitietrek ervan was er een zwakke correlatie. De VA actietendens geïnduceerd door de combinatie van de situatie met het VA gedrag was positief gecorreleerd met de mate waarin de situatie als frustrerend ervaren werd en met de instrumentaliteit en expressiviteit van het gedrag in de situatie. Negatieve correlaties werden gevonden met verwachte antipathie van anderen en negatieve zelf-evaluatie.

Samengevat kunnen we stellen dat onze benadering vrij succesvol was in het modelleren en inzichtelijk maken van de data van deze situatie-response vragenlijst, alsmede in de externe validatie ervan. De twee basisconcepten, de VA actietendens en de inhibitie van VA gedrag, vertoonden beide interessante correlaties met externe variabelen.

2. Het modelleringsperspectief

Als formele basis voor het juist beschreven inhoudelijk onderzoek naar de structuur van emoties en emotioneel gedrag, kozen we voor een modelleringsbenadering gebaseerd op Item Response Theorie (IRT). Deze is voornamelijk ontwikkeld voor het modelleren van data van cognitieve testen. Een belangrijk doel van deze dissertatie is daarom aantonen dat deze benadering ook gepast is voor data van situatie-response vragenlijsten over emoties en aan emotie gerelateerd gedrag. Omwille van praktische redenen beperken we ons tot binaire data. Een bijkomende uitbreiding zou nodig zijn voor het modelleren van multi-categoriale data.

De meeste IRT modellen veronderstellen dat de kans om een 1-antwoord op een item te geven een functie is van twee soorten parameters: persoonspecifieke parameters (meestal random effecten) en itemspecifieke parameters (meestal fixed effecten). Het meest eenvoudige IRT model is het Rasch model (Rasch, 1960). Dit model bevat één parameter per item, die men vaak de moeilijkheid van dat item noemt. Daarenboven bevat het Rasch model ook één parameter per persoon, die men gewoonlijk de vaardigheid of de waarde van een persoon op de latente trek noemt. Hier wordt de persoonsparameter beschouwd als een random effect, maar een fixed effect is ook mogelijk, zoals in een joint maximum likelihood formulering. Deze is echter niet aan te raden wegens consistentieproblemen.

Het antwoord van een persoon op een item wordt bijgevolg verklaard door een effect specifiek voor het item en een effect specifiek voor de persoon. De modelvergelijking van het Rasch model is de volgende:

$$P\left(Y_{ij}=1|\theta_i\right)=\frac{\exp\left(\theta_i+\beta_j\right)}{1+\exp\left(\theta_i+\beta_j\right)} \tag{7.1}$$

met $i = 1, \ldots, I$ de index voor de persoon,
$j = 1, \ldots, J$ de index voor het item,
θ_i de persoonsparameter,
β_j de itemparameter,
en Y_{ij} het antwoord van persoon i op item j.

Wanneer men emoties bestudeert met situatie-response vragenlijsten, dan kan men de persoonsparameter interpreteren als een emotie-specifieke drempel van de persoon of als de gevoeligheid van de persoon voor de emotie. De itemparameter kan geïnterpreteerd worden als de emotie-inducerende kracht van een situatie. Deze interpretatie geldt pas nadat de persoonsparameters met -1 vermenigvuldigd zijn. Dit resulteert in de parameterisatie $\beta_j - \theta_i$ in de plaats van $\theta_i + \beta_j$. Uit deze herparameterisatie kan men afleiden dat indien de emotie-inducerende kracht van een situatie groter is dan de emotie-specifieke drempel van een persoon, de kans dat deze persoon de emotie zal ervaren groter is dan .5. De meest gebruikte parameterisatie is echter $\theta_i - \beta_j$, waarbij men θ_i kan interpreteren als de vaardigheid van persoon i en β_j als de moeilijkheid van item j.

Alle modellen die we in deze dissertatie gebruikten zijn gebaseerd op het Rasch model. Ze zijn verdere uitbreidingen of modificaties van dit model. Vier benaderingen tot modellering werden onderzocht op hun mogelijkheden voor het bestuderen van emoties in situaties, en van interindividuele verschillen in emoties.

2.1 MIRID: een model voor de decompositie van concepten

In de paragraaf over de componentiële structuur van emoties werd een componentiële theorie voor situationele schuldgevoelens voorgesteld. Deze theorie werd getoetst aan de hand van een dataset afkomstig van een situatie-response vragenlijst. Om zulke componentiële theorieën te toetsen gebruikt men vaak het Lineair Logistisch Test Model (LLTM, Fischer, 1973, 1977). Het LLTM veronderstelt echter dat men de waarde van elke component in elke situatie kent, iets wat niet altijd het geval is. Daarom werd een nieuw IRT model ontwikkeld: het Model met Interne Restricties op Item Moeilijkheden (MIRID, Butter, De Boeck, & Verhelst, 1998). Het MIRID veronderstelt een bepaalde relatie tussen items, niet in de correlationele betekenis, maar in de zin dat het effect dat een item heeft op de response-kansen een functie is van het effect dat andere items hebben. Het MIRID stelt dat sommige items *composiet-item*s zijn, dit wil zeggen dat ze gebaseerd zijn op meer elementaire items. De groep meer elementaire items noemen we *component-items*. Voor de schuld-data bijvoorbeeld is het item: 'Pieker je in deze situatie?' een component-item en het item 'Voel je je schuldig in deze situatie?' een composiet-item. De itemparameter van een composiet-item wordt gemodelleerd als een lineaire combinatie van itemparameters van component-items. We veronderstellen dus dat de *schuld-inducerende kracht van een situatie* een gewogen som is van de bijdragen van de verschillende componenten. Dit kan men zoals in de volgende lineaire functie uitdrukken:

$$\beta_{s0} = \sum_{k=1}^{K} \sigma_k \beta_{sk} + \tau \qquad (7.2)$$

met $s = 1, \ldots, S$ de index voor de situatie,
$k = 1, \ldots, K$ de index voor het type van de component, $k = 0$ voor composiet-items,
σ_k het gewicht of de bijdrage van component k. Dit kan men interpreteren als het belang van de component,
β_{sk} de bijdrage van situatie s aan component k,
en τ een schaalconstante.

Indien we het principe dat geformuleerd is in Vergelijking 2 in het Rasch model inbouwen, dan zijn de β_j niet langer de basisparameters van het model. De β_{sk} en de σ_k vervullen nu deze rol. In het geval er slechts één onderliggende dimensie is, is de kans op een bepaalde componentiële response een functie van de persoonspecifieke bijdrage θ_i en de component-specifieke situationele schuld-inducerende kracht β_{sk}:

$$P\left(Y_{isk}=1|\theta_i\right)=\frac{\exp\left(\theta_i+\beta_{sk}\right)}{1+\exp\left(\theta_i+\beta_{sk}\right)} \tag{7.3}$$

De kans op een bepaalde composiet-response is een functie van dezelfde persoonspecifieke bijdrage θ_i en een gewogen som van de component-specifieke situationele schuld-inducerende krachten, voorgesteld door de parameter β_{s0}:

$$P\left(Y_{is0}=1|\theta_i\right)=\frac{\exp\left(\theta_i+\beta_{s0}\right)}{1+\exp\left(\theta_i+\beta_{s0}\right)} \tag{7.4}$$

met β_{s0} gedefinieerd als in Vergelijking 2

In Hoofdstuk 1 wordt een uitbreiding van het MIRID, genaamd het OPLM-MIRID, gebruikt. Dit OPLM-MIRID is ontwikkeld door Butter (1994). Het verschil met het gewone MIRID is dat deze uitbreiding toestaat dat de discriminatiewaarden van elkaar verschillen, maar wel a priori gefixeerde constanten zijn. De beperking dat er slechts één onderliggende latente trek is, is geen noodzakelijke beperking, maar MIRIDs met één latente trek vertoonden een goede fit voor deze dataset. In Hoofdstuk 6 worden twee schattingsmethoden voor het MIRID en het OPLM-MIRID met elkaar vergeleken: een methode gebaseerd op een conditionele maximum likelihood formulering (CML) en een methode gebaseerd op een marginale maximum likelihood formulering (MML). Voor de eerste methode werd een programma ontwikkeld in Delphi 5. De tweede schattingsmethode kan geïmplementeerd worden in de PROC NLMIXED procedure van SAS V8. Daarenboven werd de robuustheid van de parameterschattingen voor schendingen van de normaliteitsassumptie van de persoonsparameter onderzocht. Tussen beide benaderingen waren slechts kleine verschillen wat betreft de schatting van de persoonsparameters, terwijl we voor de itemparameters geen verschillen vonden. Tenslotte stellen we in Hoofdstuk 6 een methode voor om de structuur die MIRID aan de data oplegt (component-items versus composiet-items) te toetsen.

De assumptie dat de gewichten van de componenten fixed effecten zijn, kan in bepaalde toepassingen te streng zijn. Daarom hebben we in Hoofdstuk 2 het MIRID aangepast zodat interindividuele verschillen in de gewichten in rekening kunnen worden gebracht. Deze uitbreiding noemen we het MIRID met Random Gewichten (RW-MIRID). Het RW-MIRID veronderstelt dus dat de gewichten van sommige componenten random effecten zijn. Dit wil zeggen dat ze een normale verdeling volgen over personen. Dit model kan geschat worden met PROC NLMIXED.

2.2 De relationele structuur van concepten: het gebruik van Locale Item Afhankelijkheden

In Hoofdstuk 3 stellen we een methodologie voor om de relationele structuur van emoties te onderzoeken en te toetsen. Op basis van psychologische kennis ontwikkelden we verscheidene plausibele relationele structuren voor de schulddata. Deze structuren werden vertaald naar IRT modellen en ten opzichte van elkaar getoetst om zo het best passende model te vinden en bijgevolg ook de best passende theorie.

De methode is gebaseerd op bestaande IRT modellen die onder andere beschreven werden door Kelderman (1984) en Hoskens en De Boeck (1997). De modellen worden modellen voor Locale Item Afhankelijkheden (LIA) genoemd. Om deze notie uit te leggen starten we met een basisveronderstelling van de meeste IRT modellen: de assumptie van locale stochastische onafhankelijkheid. Dit betekent dat het model veronderstelt dat alle afhankelijkheden tussen de antwoorden van een individu enkel en alleen kunnen toegeschreven worden aan de onderliggende latente trek(ken). De ant-woorden die de persoon geeft op andere items mogen geen bijkomende informatie bevatten voor de kansen van de verschillende mogelijke antwoorden op het huidige item. Dit wordt uitgedrukt in Vergelijking 5:

$$P\left(\mathbf{Y}_i = y_{i1},\ \ldots,\ y_{iJ}|\boldsymbol{\theta}_i\right) = \prod_{j=1}^{J} P\left(Y_{ij} = y_{ij}|\boldsymbol{\theta}_i\right) \tag{7.5}$$

met $\mathbf{Y}_i$ de vector die alle antwoorden van persoon i bevat,
en $\boldsymbol{\theta}_i$ de vector die de latente trekken of persoonsparameters bevat.

Indien Vergelijking 5 niet opgaat, dan zijn er LIA, want er blijven afhankelijkheden tussen de items bestaan nadat de latente trek(ken) in rekening zijn gebracht. Vaak beschouwt men zulke LIA als iets dat moet vermeden worden. We zullen echter aantonen dat LIA informatief kunnen zijn met betrekking tot de relationele structuur van emoties.

We kunnen LIA in, bijvoorbeeld, het Rasch model incorporeren door fixed effect parameters, die de interacties tussen items (een andere term voor afhankelijkheid) vatten, aan het model toe te voegen. In Tabel 1 staat de vergelijking voor een model met een vaste interactie tussen de items j en h.

TABLE 7.1. Model voor vaste paarsgewijze interactie.

Response patroon $(y_{ij},\ y_{ih})$	Vergelijking
(0,0)	$1/v(\theta)$
(0,1)	$exp\,(\theta_i + \beta_j)\,/v(\theta)$
(1,0)	$exp\,(\theta_i + \beta_h)\,/v(\theta)$
(1,1)	$exp\,[(\theta_i + \beta_j) + (\theta_i + \beta_h) + \beta_{int}]\,/v\,(\theta)$

Noot: β_{int} is de interactieparameter voor het itempaar, en v(θ) $= 1 + exp\,(\theta_i + \beta_j) + exp\,(\theta_i + \beta_h) + exp(2\theta_i + \beta_j + \beta_h + \beta_{int})$.

Men kan makkelijk zien dat indien β_{int} positief is, de kans op het antwoordpatroon (1,1) verhoogt en indien β_{int} negatief is, deze kans verlaagt, dit alles in vergelijking met de kansen onder het Rasch model. Dit interactiemodel leidt er wel toe dat de interpretatie van de itemparameters β_j en β_h moeilijk wordt: deze parameters zijn geen zuivere weerspiegeling meer van de moeilijkheidsgraad, daar ook de interactie hierin meespeelt.

Verschillende interactiepatronen –i.e. verschillende relationele structuren– kunnen gedefinieerd worden door de overeenstemmende fixed effect parameters aan het model toe te voegen. Indien de afhankelijkheden in de data overeenstemmen met één van zulke op basis van een theorie opgestelde LIA-patronen, dan kan men twee conclusies trekken: Ten eerste wordt de theorie ondersteund door de data en bijgevolg krijgen we inzicht in de relationele structuur van een emotie. Ten tweede vinden we zo evidentie voor de interne validiteit van de vragenlijst daar de antwoorden overeenstemmen met een psychologische theorie.

De componentiële en de relationele benadering kunnen ook gecombineerd worden. Dit wordt op het einde van Hoofdstuk 3 gesuggereerd. Hoewel zo'n analyse niet vermeld wordt in Hoofdstuk 3, voerden we deze uit op de schuld-data. De resultaten hiervan bevestigden deze van Hoofdstuk 1 en 3. Aan de andere kant bemoeilijken de afhankelijkheden, geïmpliceerd door de relationele structuur, de interpretatie van de andere parameters.

2.3 Een marginale benadering van het effect van item covariaten

Alle modellen die we in de vorige paragrafen beschreven zijn modellen met random effecten. Sommige LIA-modellen bevatten ook wel elementen van wat men conditionele modellen noemt (Diggle, Heagerty, Liang, & Zeger, 2002; Fahrmeir & Tutz, 2001). Drie eigenschappen van deze random-effect modellen zorgden ervoor dat we naar een andere benadering uitkeken. De eerste eigenschap is dat de itemparameters en de persoonsparameters niet kunnen gescheiden worden van de afhankelijkheidsstructuur. Bijgevolg kunnen ze ook niet los van de afhankelijkheden geïnterpreteerd worden. Een tweede eigenschap is dat aangezien de parameters aangetast worden door de LIA, men niet kan onderzoeken hoe ze zich verhouden tot item covariaten die onafhankelijk zijn van de LIA. Een derde eigenschap is dat indien de afhankelijkheidsstructuur in random-effect modellen niet correct gespecificeerd is, dit ernstige gevolgen kan hebben voor alle andere parameters (Thissen, Steinberg, & Mooney, 1989; Tuerlinckx & Boeck, 2001; Yen, 1993). Daarom verkenden we een marginale modelbenadering die tot parameterschattingen kan leiden die niet beïnvloed worden door de afhankelijkheidsstructuur. De prijs die we hiervoor moeten betalen is dat deze modellen minder geschikt zijn om interindividuele verschillen te onderzoeken en dat de effecten van de itemcovariaten effecten zijn op het niveau van de populatie in plaats van effecten die gelden voor specifieke personen.

In Hoofdstuk 4 wordt een marginale variant van het LLTM geformuleerd. Deze marginale benadering is nieuw in de context van psychometrische modellen. We kozen voor het LLTM omdat dit een natuurlijke eerste stap is naar het MIRID (zie sectie 2.1), maar in de toekomst willen we deze marginale benadering uitbreiden naar het MIRID. Het marginaal LLTM (M-LLTM) bestaat uit twee delen: (1) een model voor de gemiddelden, waarin de itemcovariaten gerelateerd worden aan de marginale kansen door middel van de logit-link functie, genaamd de gemiddeldenstructuur, en (2) een model voor de associaties tussen de observaties. Dit noemen we de associatiestructuur.

De gemiddeldenstructuur van het M-LLTM kan gedefinieerd worden als in Vergelijking 6.

$$\text{Logit}\left[P\left(Y_{ij}=1\right)\right]=\sum_{k=1}^{K} q_{jk}\eta_k^* \tag{7.6}$$

met $k = 1, \ldots, K$ de index voor de itemcovariaat,
q_{jk} de waarde van item j op itemcovariaat k,
en η_k^* het effect van itemcovariaat k op de marginale kansen.

De associatie tussen twee items j en h duiden we aan met de parameter γ_{ijh}. Het subscript i voor de persoon werd toegevoegd omdat het model in principe ook persoonspecifieke covariaten toestaat. Deze uitbreiding valt echter buiten het bereik van deze dissertatie. De vergelijking voor de associatiestructuur kunnen we als volgt opschrijven:

$$f\left(\gamma_{ijh}\right)=\sum_{m=1}^{M} z_{jhm}\alpha_m \tag{7.7}$$

met $m = 1, \ ..., \ M$ de index voor de associatiecovariaten,
z_{jhm} de waarde van associatiecovariaat m voor de associatie tussen de antwoorden op de items j en h,
α_m het effect van associatiecovariaat m,
en $f\,(.)$ een linkfunctie waarmee de associatieparameter γ_{ijh} gelinkt wordt aan de associatiecovariaten.

Hogere orde generalisaties van de associatieparameter naar meer dan twee items kunnen aangeduid worden door een index aan z toe te voegen voor elk item dat betrokken is in de desbetreffende associatie.

Drie verschillende mogelijkheden om de associaties te modelleren worden besproken: marginale correlaties, marginale log odds ratios en conditionele log odds ratios (de log odds ratio gegeven dat men op alle ander items nul antwoordt). Voor elk model bespreken we twee schattingsmethoden: een likelihood-methode en een methode gebaseerd op gegeneraliseerde schattingsvergelijkingen (GEE, Hardin & Hilbe, 2003; Liang & Zeger, 1986; Zeger & Liang, 1986). Voor- en nadelen van alle benaderingen worden beschreven in Hoofdstuk 6.

De marginale benadering heeft twee belangrijke voordelen. Ten eerste leiden sommige marginale benaderingen tot consistente schattingen van de parameters van de gemiddeldenstructuur, onafhankelijk van het feit of de associatiestructuur correct gespecificeerd is of niet. Een incorrecte specificatie van de associatiestructuur zal bijgevolg de schattingen van de itemcovariaten niet beïnvloeden, wat wel het geval is bij random-effect modellen. Ten tweede kan men in deze marginale modellen de associaties tussen de items op een erg soepele wijze modelleren aan de hand van item- en persooncovariaten. Deze complexe patronen vormen een ernstig probleem voor random-effect modellen, want om deze te modelleren moet men ofwel extra fixed effect parameters aan het model toevoegen, ofwel het model uitbreiden met meerdere random effecten. Men kan marginale modellen dus gebruiken om het effect van variabelen op het niveau van de populatie te schatten, onafhankelijk van de associaties tussen de antwoorden op de items, terwijl het tegelijkertijd mogelijk blijft om deze associaties te verkennen en zelfs te modelleren. De effecten kan men echter niet meer interpreteren als effecten op het niveau van het individu.

2.4 Een kader gebaseerd op het leermodel van Embretson

In Hoofdstuk 5 gebruikten we een model dat oorspronkelijk geformuleerd was om een leerproces te vatten (Embretson, 1991) om zelf-rapporteringsdata over actietendensen en gedrag te modelleren. De actietendens is in deze toepassing formeel equivalent met het stadium voor het leren optreedt, en het gedrag met het stadium na het leren. Interindividuele verschillen spelen een rol in het eerste stadium en in de overgang van het eerste naar het tweede stadium. Deze equivalentie met het leermodel is slechts formeel, want leren heeft gewoonlijk een positief effect, terwijl het effect van inhibitie (tussen de actietendens en het gedrag) gemiddeld gezien negatief is, tenminste voor zover het gaat om verbale agressie.

Een specifieke eigenschap van het model is dat voor items uit het eerste stadium slechts één latente trek meespeelt, terwijl voor items uit het tweede stadium de tweede latente trek (leervaardigheid of hier inhibitie) ook een rol speelt. We pasten dit model toe in de volgende formulering:

$$\text{Logit}\left[P\left(Y_{ijk}=1 \mid \alpha_i,\kappa_i\right)\right]=\alpha_i+\beta_{sk}^{(willen)}-d\left(\kappa_i+\beta_{sk}^{(doen)}\right) \tag{7.8}$$

met $d=1$ voor doen-items, en $d=0$ voor willen-items,
$\alpha_i \sim N(0,\sigma_\alpha^2)$ de persoonspecifieke VA actietendens of m.a.w. de verbale agressietrek,
$\kappa_i \sim N(\mu_\kappa,\sigma_\kappa^2)$ de persoonspecifieke VA inhibitie parameter of inhibitietrek. Het gemiddelde van κ_i $(=\mu_\kappa)$ is het algemene inhibitie-effect,
$\beta_{sk}^{(willen)}$ het effect van de combinatie van een situatie s en een type VA gedrag k op de neiging om zich verbaal agressief te gedragen,
en $\beta_{sk}^{(doen)}$ het effect van de combinatie van een situatie s en een type VA gedrag k op het VA gedrag.

De $\beta_{jk}^{(doen)}$ is een uitbreiding van het leermodel van Embretson, want in haar model is er slechts één itemparameter voor beide stadia. Gebaseerd op het model in Vergelijking 8, werkten we een kader uit om te kunnen toetsen of inhibitie een effect is dat afhangt van de persoon, van het item (combinatie van een situatie en een bepaald type VA gedrag) of van beide.

3. Conclusies

We geloven dat het bestuderen van emoties aan de hand van situationele vragenlijsten voordeel kan halen uit het gebruik van IRT modellen. Voor veel onderzoeksvragen zal er een sterke overeenkomst bestaan tussen de vraag en een specifiek IRT model, zodanig dat men de gestelde hypothese(s) kan toetsen. Voor andere onderzoeksvragen waarvoor geen overeenkomstig model voorhanden is, kan men een gepast model formuleren. De beschikbaarheid van erg algemene modelschattingssoftware zoals PROC NLMIXED uit SAS V8 –indien men IRT modellen als gegeneraliseerd (niet) lineaire modellen beschouwd (McCulloch & Searle, 2001)– draagt zeker bij tot de flexibiliteit van IRT. Verscheidene van onze bevindingen werden mogelijk gemaakt door de gebruikte modellen. Voorbeelden hiervan zijn de rol van de verschillende schuldcomponenten, het feit dat zij geen interindividuele verschillen vertonen en het voornamelijk persoonsgebonden (versus situatiegebonden) karakter van inhibitie.

Het verkennen van de mogelijkheden van de marginale modelbenadering, overtuigde ons ervan dat deze benadering, hoewel ze heden amper geëxploreerd en gebruikt is in de psychometrie, verder aandacht verdient voor het analyseren van data van vragenlijst, daar ze erg flexibel en veelbelovend is.

7.1 References

Averill, J. R. (1983). Studies on anger and aggression: implications for theories of emotion. *American Psychologist*, *38*, 1145-1160.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: a cognitive-neoassociationistic analysis. *American Psychologist*, *45*, 494-503.

Butter, R. (1994). *Item response models with internal restrictions on item difficulty.* Unpublished doctoral dissertation, Catholic University of Leuven, Leuven.

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

Cornell, D. G., Peterson, C. S., & Richards, H. (1999). Anger as a predictor of aggression among incarcerated adolescents. *Journal of Consulting and Clinical Psychology*, *67*, 108-115.

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

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

Embretson, S. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495-515.

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

Fischer, A. H. (1991). *Emotion scripts: A study of the social and cognitive facets of emotions.* Leiden, The Netherlands: DSWO Press.

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.* Bern, Switzerland: Huber.

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

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

Hardin, J. W., & Hilbe, J. M. (2003). *Generalized estimation equations.* London: Chapman & Hall/CRC.

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

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

Kassinove, H., Sukhodolsky, D. G., Tsytsarev, S. V., & Solovyova, S. (1997). Self-reported anger episodes in russia and america. *Journal of Social Behavior and Personality*, *12*, 301-324.

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

Kinney, T. A., Smith, B. A., & Donzella, B. (2001). The influence of sex, gender, self-discrepancies, and self-awareness on anger and verbal aggressiveness among u.s. college students. *Journal of Social Psychology*, *141*, 245-275.

Kuppens, P., Van Mechelen, I., Smits, D. J. M., & De Boeck, P. (in press). The appraisal basis of anger: Specificity, necessity and sufficiency of components. *Emotion.*

Lange, A., Hoogendoorn, M., Wiederspahn, A., & Beurs, E. de. (1995). *Buss Durkee hostility and direction of hostility inventory-dutch (BDHI-D).* Lisse, The Netherlands: Swets & Zeitlinger.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American Psychologist*, *46*, 352-367.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Oatley, K., & Jenkins, J. M. (1996). *Understanding emotions.* Oxford: Blackwell Publishers.

Omdahl, B. L. (1995). Cognitive appraisal theories of emotions. In B. L. Omdahl (Ed.), *Cognitive appraisal, emotion, and empathy* (p. 37-90). New Jersey: Lawrence Erlbaum.

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

Rasch, G. (Ed.). (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: 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.

Roseman, I. J. (1984). Cognitive determinants of emotion: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Vol. 5. emotions, relationships, and health* (p. 11-36). Beverly Hills, CA: Sage.

Sas System V8 for windows [Computer program]. (1999). Carey, NC.

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: testing 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.

Skiffington, S., Fernandez, E., & McFarland, K. (1998). Towards a validation of multiple features in the assessment of emotions. *European Journal of Psychological Assessment*, *14*, 202-210.

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

Spielberger, C. D. (1980). *Preliminary manual for the state-trait anger scale (STAS).* Tampa, Florida: University of south Florida.

Spielberger, C. D., Johnson, E. H., & Jacobs, G. A. (1982). *The anger expression (AX) scale.* Tampa, Florida: University of south Florida.

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

Van Elderen, T., Maes, S., Komproe, I., & Kamp, L. van der. (1997). The development of an anger expression and control scale. *British Journal of Health Psychology*, *2*, 269-281.

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

Zeger, S. L., & Liang. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, *42*, 121-130.

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: a state and trait perspective from experience sampling data. *Journal of Research in Personality*, *34*, 178-197.