Tilburg University

Micro-macro multilevel analysis for discrete data

Bennink, M.

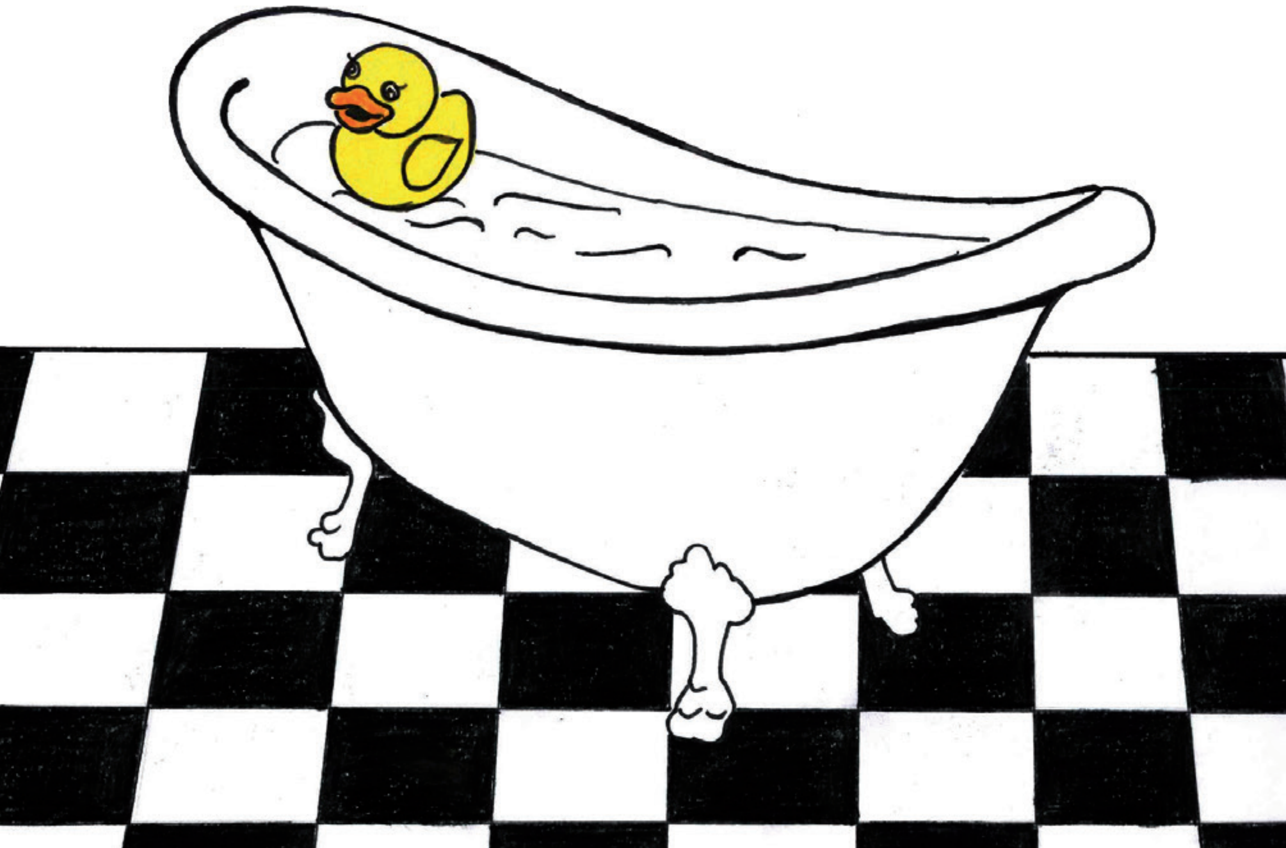
Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Bennink, M. (2014). *Micro-macro multilevel analysis for discrete data*. Ridderprint.

Micro-macro multilevel analysis for discrete data

Margot Bennink



MICRO-MACRO MULTILEVEL ANALYSIS FOR DISCRETE DATA

MICRO-MACRO MULTILEVEL ANALYSIS FOR DISCRETE DATA

This research is funded by The Netherlands Organization for Scientific Research (NWO 400-09-018).

Printing was financially supported by Tilburg University.

ISBN: 978-90-5335-893-1
Printed by: Ridderprint BV, Ridderkerk, the Netherlands
Cover design: Vivian Klaassen and Nikki Vermeulen

MICRO-MACRO MULTILEVEL ANALYSIS FOR DISCRETE DATA

MICRO-MACRO MULTILEVEL ANALYSE
VOOR DISCRETE DATA

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University
op gezag van rector magnificus, prof. dr. Ph. Eijlander,
in het openbaar te verdedigen ten overstaan van
een door het college voor promoties aangewezen commissie
in de aula van de Universiteit

op vrijdag 10 oktober 2014 om 14.15 uur

door

Margot Bennink

geboren op 28 juni 1985 te Rotterdam

Promotor:	prof.dr. J. K. Vermunt
Copromotor:	dr. M. A. Croon
Overige leden van de Promotiecommissie:	prof.dr. E. Ceulemans
	prof.dr.ir. J.-P. Fox
	dr. D.L. Oberski
	prof.dr. T.A.B. Snijders
	prof.dr. M.J.P.M. van Veldhoven

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

In many research situations in the social and behavioral sciences, data are collected within hierarchically ordered systems. Examples are data sets on children nested within schools, employees nested within teams or organizations, patients nested within therapists or hospitals, citizens nested within regions, clients nested within stores, but also repeated measurements nested within subjects. Multilevel analysis deals with these kind of nested observations (Goldstein, 2011).

The added value of multilevel modelling compared to standard statistical techniques is twofold. First, the dependencies among individuals within a group are taken into account. For example, two employees working in the same organization might be more similar with respect to work satisfaction than two employees working in different organizations. This violates the assumption of independent errors as made in standard regression analysis. Second, it allows investigating relationships among variables at different levels of such an hierarchical structure; that is, relationships between characteristics of schools and children, organizations and employees, or patients and therapists.

As far as the relationships between characteristics of individuals and groups are concerned, depending on the research question at hand, one of two rather different mechanisms may be of interest. The first option is that variables at the group level (or macro level) are assumed to affect one or more outcome variables at the individual level (or micro level). For example, the school's teaching system affects the pupils' learning rates, or the team's autonomy affects the work satisfaction of team members. Following Snijders and Bosker (2012), we refer to these types of situations as macro-micro relationships. The second possible mechanism concerns individual-level characteristics affecting group-level outcomes. For example, the motivation of children affects the teaching style of teachers, or the work load of employees affects the team's productivity. These are referred to as micro-macro relationships.

Although, both macro-micro and micro-macro relationships are of interest in social and behavioral science research, the overwhelming majority of models developed for the analysis of multilevel data sets concern the macro-micro situation. This dissertation will contribute to the methodology for investigating micro-macro relationships, with a special emphasis on discrete data. It contains data examples from a broad range of research fields within the social and behavioral sciences such as sociology, educational measurement, and organizational studies. This illustrates the need for, and the widely applicability of, these statistical methods.

1.1 Micro-Macro Analysis

Traditionally, a micro-macro relationship is analyzed by a single-level analysis in one of the following two ways. The first method involves aggregating the micro-level predictor to the macro level using the group mean or any other measure of central location. Subsequently, a group-level analysis is performed in which the group-level outcome is regressed on the aggregated individual-level predictor. A serious problem with this approach is that measurement error in the aggregated scores is not accounted for. This implies that the group members are assumed to provide perfect information about their group, while this assumption is unlikely to hold in practice (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011). Sampling fluctuation might be an issue as well when not all individuals within a group are investigated. In case of discrete data an additional problem arises, since it is not clear how to aggregate discrete variables. For example, for nominal variables with more than two categories, the group mean has no substantive interpretation. A group mode can be used instead, but measurement and sampling error is still not accounted for. Instead of aggregating, the second method for dealing with micro-macro situations is that the macro-level outcome is disaggregated to the micro level and an individual-level analysis is performed in which the disaggregated outcome is regressed on the individual-level predictor. Disaggregation violates one of the basic assumptions of regression analysis, namely that the units are independent (Keith, 2006). Consequently, Type-I errors are severely inflated leading to too liberal tests (Krull & MacKinnon, 1999; MacKinnon, 2008).

Croon and van Veldhoven (2007) proposed a two-level latent variable model for micro-macro analysis that appropriately handles the multilevel structure of the problem at hand. The scores of the group members on a micro-level predictor Z_{ij} are used as exchangeable indicators for a continuous group-level latent variable ζ_j. In this notation, the subscript j refers to the group level and subscript i to the individual level. Since a latent variable is used at the group level to represent the individual-level variable, measurement error and sampling error in the (latent) aggregated scores are taken into account. This part of the model is referred to as the within-group part of the model. In the between-group part of the model, the group-level latent scores are related to the group-level outcome Y_j, but it is also possible to include other (independent) group-level variables, represented by X_j. A graphical illustration of a model with a single group-level predictor X_j, a single individual-level predictor Z_{ij} and a single group-level outcome Y_j is shown in Figure 1.1.

Whereas this latent variable approach has been an import ant step forward, its main limitation is that it assumes that the endogenous variables in the model are continuous. Though in certain applications this may (at least approximately) be correct, there are also many situations in which an approach suited for discrete data is preferred. For example, when the micro-level predictor or the macro-level outcome is discrete (dichotomous, nominal, or ordinal), or when for theoretical or practical reasons it is warranted to classify the groups into (latent) categories instead of placing them on a continuous scale. The aim

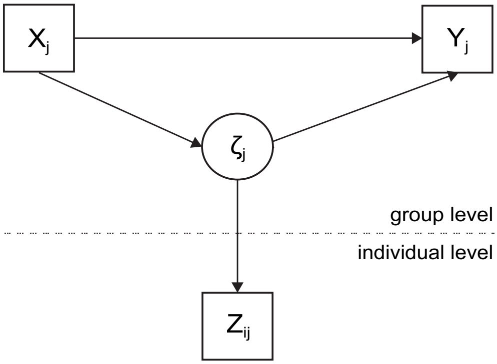


Figure 1.1: Micro-Macro Latent Variable Model

of the current dissertation is to generalize the approach from Croon and van Veldhoven (2007) to the situation in which the (latent) variables are not continuous, but discrete.

1.2 Discrete Data

To generalize the latent variable approach proposed by Croon and van Veldhoven (2007) to discrete data, a latent class model, instead of a factor-analytic model, is used to define a discrete latent variable at the group level. Aggregating a discrete individual-level predictor to the group level with a latent class model does not only make it possible to account for measurement and sampling error in the aggregates scores, it also overcomes the difficulties that arise when a manifest mean or mode is used for the aggregation.

When the model shown in Figure 1.1 is formulated for categorical data, the latent variable ζ_j is now a categorical variable which values define a set of C discrete latent classes at the group level, $c = 1, \cdots, C$. The individual scores Z_{ij} for a particular group are denoted by the vector $\mathbf{Z}_j$ and treated as 'unreliable' indicators of the group score ζ_j. A discrete group-level predictor X_j is added to the model. In most applications the relationships among the macro-level variables will be modeled by logit models or, eventually, by more complex log-linear models. For an arbitrary group j, the relevant conditional probability distribution for the manifest variables Y_j and $\mathbf{Z}_j$ given X_j is:

$$P(Y_j, \mathbf{Z}_j|X_j) = \sum_{c=1}^{C} P(Y_j, \zeta_j = c|X_j)P(\mathbf{Z}_j|\zeta_j = c). \tag{1.1}$$

The terms on the right hand side of the equation are the between and within part that can be further decomposed as

$$P(Y_j, \zeta_j = c|X_j) = P(\zeta_j = c|X_j)P(Y_j|X_j, \zeta_j = c), \tag{1.2}$$

and

$$P(\mathbf{Z}_j|\zeta_j = c) = \prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j = c). \tag{1.3}$$

This model is the baseline model throughout the dissertation.

1.3 Outline of the Dissertation

This dissertation consists of four journal articles that together give a coherent insight in micro-macro multilevel analysis for discrete data. Since the chapters are stand alone articles, they can be read independently. This creates some overlap in the text and some inconsistency in notation. A short overview of the chapters is given below.

In Chapter 2, the latent variable approach is presented and compared to the single-level aggregation and disaggregation methods in a simulation study. In a second simulation study, the latent variable approach is evaluated in the baseline model from Figure 1.1 by studying bias in the group-level parameter estimates, and the power and Type-I error rates of the statistical tests. In the end of the chapter, the latent variable approach is applied to personal network data. In the remaining chapters, the baseline model is extended to more complex situations that can be found in applied research.

In Chapter 3, two extensions are presented to handle multiple individual-level variables, so multiple Z_{ij}-variables. As in the baseline model, the individual-level data are summarized at the group level using a single discrete latent variable ζ_j at the group level. In the first extension, the multiple Z_{ij}-variables are directly used as indicators for ζ_j, such as done in Figure 1.1 when the model contained a single Z_{ij}. To capture the within-group (co)variation among the Z_{ij}-variables, either all two-way associations among the Z_{ij}-variables or an individual-level latent variable needs to be incorporated in the model. In the second extension, the Z_{ij}-variables are used indirectly at the group level by using them as indicators for an individual-level variable. This individual-level latent variable is aggregated to the group level by using it as a single indicator for ζ_j. Both extensions are applied to empirical data from either marketing research or research to human resource practices in small firms.

Thus far, the full models are estimated in one step and in Chapter 4 is explored how to estimate the proposed latent class models in a stepwise matter by estimating the within-group model before estimating the between-group model. This is in fact already done when an individual-level predictor is aggregated with a manifest variable, but the stepwise latent class approach also corrects the group-level estimates for error in the aggregated scores. This is discussed first in the context of the baseline model from Figure 1.1 and second in the context of a model with two group-level latent variables. The stepwise model with two latent group-level variables is applied to empirical data from organizational research.

In Chapter 5, an application from the field of educational measurement is presented in which the group-level latent classes are the main objective of the study. Schools are classified into (latent) classes (ζ_j) based on multiple student-level items (Z_{ij}-variables). Contrarily to the baseline model from Figure 1.1, ζ_j is not related to a group-level outcome Y_j. This can be seen as a micro-macro relationship in which the group-level outcome of interest is a latent, instead of a manifest, variable. Furthermore, the latent classification of schools is controlled for the ability of the students within the schools and can also be used to detect uniform and nonuniform school-level item bias. This shows that ζ_j can also be used for other purposes than aggregating the individual-level variables to the group level.

CHAPTER 2

A Latent Variable Approach to Micro-Macro Analysis

Abstract

A multilevel regression model is proposed in which discrete individual-level variables are used as predictors of discrete group-level outcomes. It generalizes the model proposed by Croon and van Veldhoven for analyzing micro-macro relations with continuous variables by making use of a specific type of latent class model. A first simulation study shows that this approach performs better than more traditional aggregation and disaggregation procedures. A second simulation study shows that the proposed latent variable approach still works well in a more complex model, but that a larger number of level-2 units is needed to retain sufficient power. The more complex model is illustrated with an empirical example in which data from a personal network are used to analyze the interaction effect of being religious and surrounding yourself with married people on the probability of being married.

This chapter is published as Bennink, M., Croon, M. A., & Vermunt, J. K. (2013). Micro-macro multilevel analysis for discrete data: A latent variable approach and an application on personal network data. *Sociological Methods & Research, 42*(4), 431-457.

2.1 Introduction

In many research situations in the social and behavioral sciences, data are collected within hierarchically ordered systems. For example, data may be collected on individuals nested within groups. Repeated measures carried out on the same individuals can also be treated as nested observations within these individuals. Data collected in a personal or egocentric network are hierarchical as well, since data are collected on individuals (egos) and on persons from the network of these individuals (alters) or on ties (ego-alter relations). This data collection procedure is an example of a multilevel design in which the observations on the alters or ties are nested within the egos (Hox & Roberts, 2011; Snijders, Spreen, & Zwaagstra, 1995). In the current article, data are considered hierarchical when both the level-2 units and the level-1 units are a (random) sample of the population of possible level-2 and level-1 units.

In these two-level settings, two basically different situations can be distinguished. In the first situation, independent variables defined at the higher (macro) level are assumed to affect dependent variables defined at the lower (micro) level. For example, whether firms have a salary bonus system or not may affect the individual productivity of the employees working in these firms (Snijders & Bosker, 2012). Snijders and Bosker (2012) refer to these relationships as macro-micro relations, but they are also referred to as 2-1 relations since a level-2 explanatory variable affects a level-1 outcome variable. In the last few decades many efforts have been made to develop multilevel models for this kind of hierarchical ordering of variables, and although the bulk of this work has emphasized multilevel linear regression models for continuous variables, multilevel regression models for discrete response variables have also been proposed (Goldstein, 2011; Snijders & Bosker, 2012). Standard multilevel software as implemented in, for instance, SPSS, MLwiN (Rasbash, Charlton, Browne, Healy, & Cameron, 2005), and Mplus (Muthén & Muthén, 1998-2012) is available to estimate these multilevel models.

In the second situation, referred to as a micro-macro situation by Snijders and Bosker (2012), independent variables defined at the lower level are assumed to affect dependent variables defined at the higher level. These relations, which can also be referred to as 1-2 relations, have received less attention in the statistical literature than the models for analyzing 2-1 relations. This is rather odd since this type of relation occurs frequently in the social and behavioral sciences. For instance, consider organizational research that tries to link team performance or team effectiveness to some attributes or characteristics of the individual team members (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; van Veldhoven, 2005; Waller, Conte, Gibson, & Carpenter, 2001). Also in educational psychology these micro-macro relations may be of interest when, for example, the global school effectiveness is studied in relation to the attributes of the individual students and teachers (Rutter & Maughan, 2002).

Two traditional approaches for analyzing micro-macro relationships are commonly in use: either, the individual-level predictors are aggregated to the group level, or the group-level outcome variables are disaggregated to the individual level and the analysis is concluded with a single-level regression analysis at the appropriate level. More recently, Croon and van Veldhoven (2007) presented an alternative latent variable approach for analyzing micro-macro relations with continuous outcomes. This approach has only been fully worked out yet for the case of linear relationships among continuous explanatory and outcome variables. The present article discusses how to extend this latent variable approach to the analysis of discrete data.

In the remainder of this article, the aggregation, disaggregation and latent variable approaches to deal with a micro-macro hypothesis are described, applied to discrete data and evaluated and compared in a simulation study. Subsequently, a discrete group-level predictor is added to the micro-macro model and this extended model is evaluated in a second simulation study and illustrated with an empirical example on personal network data.

2.2 Analyzing Micro-Macro Relations

2.2.1 Aggregation and Disaggregation

For the analysis of micro-macro relations, two traditional approaches are currently being applied: either the individual-level predictors are aggregated to the group level or the group-level outcome variables are disaggregated to the individual level, and the final analysis is concluded with a single-level regression analysis at the appropriate level.

The first approach to deal with micro-macro relations is to aggregate the individual-level predictors to the group level by assigning a mode, median or mean score to every group based on the scores of the individuals within the group. It is then assumed that the assigned scores perfectly reflect the construct at the group level. This assumption is not realistic in practice, since the group-level construct does not represent the heterogeneity within groups. Moreover, the group-level construct may be affected by measurement error and sampling fluctuation (Lüdtke et al., 2011). Additionally, the number of observations on which the final regression analysis is carried out decreases since the groups are treated as the units of analysis. Consequently, the power of the statistical tests involved may sharply decrease (Krull & MacKinnon, 1999). Aggregation also has the disadvantage that the information about the individual-level variation within the groups is completely lost.

When disaggregating the outcome variable, each individual in a group is assigned his group-level score, which in the further analysis is treated as if it was an independently observed individual score. Since the scores of all individuals within a particular group are the same, the assumption of independent errors among individuals (Keith, 2006), as made in regression analysis, is clearly violated. This violation leads to inefficient estimates, biased standard errors, and overly liberal inferences for the model parameters (Krull & MacKinnon, 1999; MacKinnon, 2008). Moreover, by analyzing the data at the individual level in this manner, the total sample size is not corrected for the dependency among the individual observations within a group, which causes the power of the analysis to be artificially high.

2.2.2 Latent Variable Approach

Recently, Croon and van Veldhoven (2007) presented an alternative approach for analyzing micro-macro relations with continuous outcomes which overcomes many of the problems associated with aggregation or disaggregation. Their latent variable approach is illustrated by the model shown graphically in Figure 2.1. This model covers the situation with a single explanatory variable at the individual level (Z_{ij}) affecting a single outcome variable at the group level (Y_j). In the notation used here, the subscript j refers to the groups, while the subscript i refers to individuals within a group.

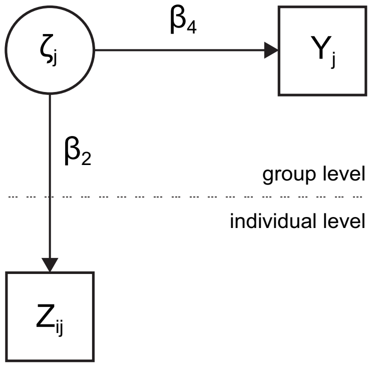


Figure 2.1: Graphical Representation 1-2 Model

To analyze the relationship between the individual-level independent variable and the group-level outcome, the scores on Z_{ij} are treated as exchangeable indicators for a latent group-level variable ζ_j. The exchangeability assumption implies that the relation between the individual-level observation and the group-level latent variable is assumed to be the same for all individuals within a group. In this way, all individuals are treated as equivalent sources of information about the group-level variable, and none of them is considered as providing more accurate judgments in this respect than his co-members. This assumption is warranted when all group members play similar or identical roles in the group and is probably less vindicated when the group members differ with respect to their functioning in the group. The latent group-level variable ζ_j is treated as a predictor or explanatory variable for the group-level outcome variable Y_j. In this way, the individual-level observations on Z_{ij} are not assumed to reflect the group-level construct ζ_j perfectly, but within group heterogeneity, and sampling variability are allowed to exist. This model actually consists of two parts: a measurement part which relates the individual-level scores on Z_{ij} to the latent variable ζ_j at the group level, and a structural part in which Y_j is regressed on ζ_j.

The latent variable approach can be generalized to situations in which the variables from the measurement or the structural part of the model are not necessarily continuous. With respect to the measurement model, the four different measurement models which are obtained by independently varying the scale type of the observed variable Z_{ij} and the latent variable ζ_j, are shown in Table 2.1. The basic idea is that groups can be classified or located on either a continuous or discrete latent scale at the group level and that the group members are acting as 'imperfect' informants or indicators of their group's position on this latent group-level scale. Furthermore, the information the group members provide about the group's position can also be considered as being measured on either a continuous or a discrete scale.

When both the observed variable Z_{ij} and the latent variable ζ_j are assumed to be continuous, as in Croon and van Veldhoven (2007), a linear factor model links the individual-level scores to the group-level score. Alternatively, one might assume that a discrete latent variable at the group level underlies a continuous observed variable at the individual level. In this situation the measurement part of the model is described by a latent profile model (Bartholomew & Knott, 1999). In situations in which the

Table 2.1: Measurement Model

	$\zeta_j = continuous$	$\zeta_j = discrete$
$Z_{ij} = continuous$	linear factor model	latent profile model
$Z_{ij} = discrete$	item response model	latent class model

observed explanatory variables at the individual level are discrete, either a latent class model (Hagenaars & MacCutcheon, 2002) or an item response model (Embretson & Reise, 2000) might be considered. A latent class model is appropriate when the underlying latent variable at the group level is discrete as well, whereas an item response model is appropriate when the underlying latent variable is assumed to be continuous.

With respect to the structural part of the model, the regression of Y_j on ζ_j at the group level can be conceived in different ways depending on the measurement level of the outcome variable Y_j. For a continuous outcome variable, Croon and van Veldhoven (2007) defined a linear regression model, but when the group-level outcome variable Y_j is discrete, (multinomial) logit or probit regression models are more appropriate to regress Y_j on ζ_j, irrespective of the scale type of ζ_j. All these models fit within the general framework of generalized latent variable models described by Skrondal and Rabe-Hesketh (2004).

2.3 Discrete Variables

The focus of the current paper will be on the application of the latent variable approach to discrete data by combining a latent class model for the measurement model with a (multinomial) logistic regression model at the group level. Readers interested in specifying a continuous latent variable underlying discrete observations are referred to Fox and Glas (2003) and Fox (2005). Our discussion of the model for discrete variables first considers the case in which all variables are dichotomous before discussing the more general case.

Consider again the model shown in Figure 2.1 but now assume that all variables in the model are dichotomous with values 0 and 1. In this 1-2 model, the relationship between a single dichotomous explanatory variable Z_{ij} at the individual level and a single dichotomous outcome variable Y_j at the group level is at issue. The type of models that are discussed in this article and of which the model shown in Figure 2.1 is a very basic example, can be seen as a two-level extension of the path models for discrete variables as defined in Goodman's modified path approach (Goodman, 1973). These are extended to include latent variables by Hagenaars (1990) in the modified Lisrel approach. Moreover, the way in which these models allow for the decomposition of joint probability distributions in terms of products of conditional distributions, indicates their resemblance to the directed graph approach as described by, among others, Pearl (2009) for variables measured at a single level.

We opt for a latent class model with the number of latent classes set equal to the number of response categories of the observed individual-level variable, implying that the scores on Z_{ij} are treated as indicators for a dichotomous latent variable ζ_j at the group level (score 0 or 1).[1] The number of latent classes at the group level does not necessarily have to be fixed a priori, but could also be data driven by comparing fit indices for models with varying number of latent classes.[2]

For dichotomous variables, the model can be formulated more formally in terms of two logit regression equations:

$$\text{Logit}(P(Z_{ij}=1|\zeta_j)) = \log\left(\frac{P(Z_{ij}=1|\zeta_j)}{P(Z_{ij}=0|\zeta_j)}\right) = \beta_1 + \beta_2\zeta_j, \tag{2.1}$$

and

$$\text{Logit}(P(Y_j=1|\zeta_j)) = \log\left(\frac{P(Y_j=1|\zeta_j)}{P(Y_j=0|\zeta_j)}\right) = \beta_3 + \beta_4\zeta_j, \tag{2.2}$$

in which β_1 and β_3 are intercepts and β_2 and β_4 slopes. The parameters β_2 and β_4 are log odds ratios indicating the strength of the association between the latent variable ζ_j and the observed variables Z_{ij} and Y_j, respectively. For the general case of K nominal response categories for Z_{ij} and M nominal response categories for Y_j, multicategory logit models can be formulated as described in Agresti (2013).

2.4 Estimation Methods

For continuous outcomes, Croon and van Veldhoven (2007) proposed a stepwise estimation method in which the two parts of the model are estimated separately by what they called an 'adjusted regression analysis'. In this approach the aggregated group means of the variables measured at the individual level are adjusted in such a manner that a regression analysis at the group level using these adjusted group means produces consistent estimates of the regression coefficients. Full information maximum likelihood (FIML) estimates can be obtained by either the 'Persons-as-Variables approach' (Curran, 2003; Mehta & Neale, 2005) or by fitting the model as a two-level structural equation model (Lüdtke et al., 2008) as made possible in software packages such as Mplus (Muthén & Muthén, 1998-2012), LISREL (Jöreskog & Sörbom, 2006), or EQS (Bentler, 2006). These maximum-likelihood methods estimate the parameters from the two parts of the model simultaneously.

Applied to the 1-2 model with discrete data, let $\mathbf{Z}_j$ be the vector containing the I_j individual-level responses for group j. This implies $\mathbf{Z}_j = \{Z_{1j}, Z_{2j}, ..., Z_{I_j j}\}$. The joint

[1] It should be noted that the latent classes at the group level underlying Z_{ij} can not only be interpreted as a measurement model for the items, but also as a group-level discrete random effect since the dependence in the responses is summarized in one random score at the group level. This is how the multilevel structure is taken into account. The predictor X_j, and the outcome Y_j are observed at the group level only, which means that these variables vary only between groups and not within groups.

[2] The number of latent classes could, for example, be determined with the BIC using the number of groups as sample size in the formula (Lukočienė, Varriale, & Vermunt, 2010).

density of $\mathbf{Z}_j$, Y_j and ζ_j equals

$$\begin{aligned} P(\mathbf{Z}_j, Y_j, \zeta_j) &= P(\zeta_j)P(Y_j|\zeta_j)P(\mathbf{Z}_j|\zeta_j) \\ &= \underbrace{P(\zeta_j)P(Y_j|\zeta_j)}_{between} \underbrace{\prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j)}_{within} . \end{aligned} \tag{2.3}$$

Equation 2.3 consists of a product of a between and a within component. In the between component only relations among variables defined at the group level are defined, whereas in the within component the individual-level scores are related to the group-level variable. By taking the product of $P(\mathbf{Z}_j, Y_j, \zeta_j)$ over all groups, the complete-data likelihood is obtained. This is the likelihood function if ζ_j would have been observed. The log-likelihood function for the observed data are then obtained by summing $\log(P(\mathbf{Z}_j, Y_j, \zeta_j))$ over all groups.

Integrating out the latent variable ζ_j from the complete log-likelihood function by summing over its possible values yields the log-likelihood function for the observed data Z_{ij} and Y_j; that is,

$$\log\ L = \sum_{j=1}^{J} \log\left(\sum_{c=1}^{C} \left(P(\zeta_j = c)P(Y_j|\zeta_j = c) \prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j = c) \right) \right), \tag{2.4}$$

in which C represents the number of latent classes, $c = 1, \cdots, C$.

In practice, this incomplete data likelihood function can be constructed in two equivalent ways: with the 'Two-level regression approach' and with the 'Persons-as-Variables approach' (Curran, 2003; Mehta & Neale, 2005). For the first approach, data need to be organized in a 'long file' while for the second approach the data need to be organized in a 'wide file'. More details about these equivalent approaches and the construction of the likelihood accordingly, can be found in Appendix A. The Latent GOLD software (Vermunt & Magidson, 2005a) can be used to estimate the model in both ways.

2.5 Simulation Study 1-2 Model

2.5.1 Aim of the Simulation Study

This section reports the results of a Monte Carlo simulation study which evaluated the (statistical) performance of the latent class approach for analyzing micro-macro relations among dichotomous variables using the 1-2 model. A first aim of the simulation study is to investigate the bias of the estimates of the relevant regression parameters describing the micro-macro relationship. Additionally, the power and observed Type-I error rate of the test of the regression coefficients are determined.

Two different procedures to test for the significance of individual parameters are compared. First, significance is tested by means of the Wald test. This test is easy to implement and only requires the maximum- likelihood estimation of the unrestricted model (leaving the estimation of β free). However, evidence exists that in small samples the likelihood-ratio test may be preferred (Agresti, 2013). The latter testing procedure requires estimating both the unrestricted model and restricted model with $\beta = 0$.

Besides looking at the absolute performance of the latent class approach, its relative performance is assessed by comparing it to three more traditional approaches: mean aggregation, mode aggregation, and disaggregation. The present simulation study investigates how, for all four approaches, the bias in the parameter estimates, their Type-I error rate and the power of the associated tests are affected by (1) the strength of the micro-macro relation, (2) the degree to which the individual-level scores reflect the (latent) group-level score, and (3) the sample sizes at both the individual and group level.

2.5.2 Method

Data were generated according to the 1-2 model shown in Figure 2.1 and formally described by Equations 2.1 and 2.2. In the population model, four factors were systematically varied. First, the micro-macro relation was assumed to be either absent, ($\beta_4 = 0$), moderate ($\beta_4 = 1$), or strong ($\beta_4 = 2$). Second, the individual-level observed variable Z_{ij} was either a poor ($\beta_2 = 1$), a good ($\beta_2 = 3$), or a perfect indicator ($\beta_2 = 200$) of the construct at the group level. In most applications the latter assumption is unrealistic, however it was included in order to compare the other two situations with the perfect situation. Third, the number of groups was set to either 40 or 200, and fourth, the number of individuals within a group was either 10 or 40. Finally, the intercept values β_1 and β_3 were not varied independently, but were chosen such that uniform marginal distributions for Z_{ij} and Y_j were guaranteed. This implies that these marginal distributions were held constant across simulation conditions. Completely crossing the four factors resulted in $3 \times 3 \times 2 \times 2 = 36$ conditions. For each condition, 100 data sets were generated with Latent GOLD (Vermunt & Magidson, 2005a).

Each data set was analyzed in four different ways. First, they were analyzed according to the latent class approach and the estimate of the micro-macro regression coefficient is represented by the term β_4 from Equation 2.2. Second, the same data were analyzed at the group level by aggregating the individual-level predictor scores using the group means, $\bar{Z}_{.j}$, or third, using the group mode, denoted by $\breve{Z}_{.j}$. The logistic regression analyses at the group level are defined by

$$\text{Logit}(P(Y_j = 1|\bar{Z}_{.j}) = \beta_5 + \beta_6 \bar{Z}_{.j}, \tag{2.5}$$

and

$$\text{Logit}(P(Y_j = 1|\breve{Z}_{.j})) = \beta_7 + \beta_8 \breve{Z}_{.j}. \tag{2.6}$$

The estimate of the micro-macro regression coefficient is now represented by β_6 and β_8, respectively. Finally, in the fourth analysis the group-level outcome variable Y_j is disaggregated to the individual level by assigning the group score to every group member as if the score was unique to the individual, so $Y_{ij} = Y_j$ for each individual i in group j. The disaggregated variable Y_{ij} is then regressed on Z_{ij} at the individual level and the corresponding logistic regression equation becomes

$$\text{Logit}(P(Y_{ij} = 1|Z_{ij})) = \beta_9 + \beta_{10} Z_{ij}. \tag{2.7}$$

The estimate of the micro-macro regression coefficient is now represented by β_{10}.

Power was determined with a Wald test by computing the percentage of times that the hypothesis $\beta = 0$ was rejected when in fact there was a non-zero effect present in the population ($\beta = 1$ and $\beta = 2$). The observed Type-I error rate was given by the proportion of significant results for the same hypothesis when there was zero effect in the population ($\beta = 0$). The observed Type-I error rate and power of the likelihood-ratio test were determined in a similar way. In order to assess the main effects of each of the manipulated factors, the results were collapsed over the three other factors.

2.5.3 Results

Bias of the parameter estimates

The means and standard deviations of the estimates of the micro-macro relation are summarized in Table 2.2. When the micro-macro relation was estimated with the latent class approach, the micro-macro effect was estimated without severe bias in all conditions. When Z_{ij} was aggregated to the group level using mean scores, the estimated micro-macro effect was overestimated in all conditions where a micro-macro relation was present, except when the individual-level scores perfectly reflected the construct at the group level. When the mode instead of the mean was used to aggregate the individual-level scores, the bias decreased. This method also seems to work when the individual-level scores were good, and not necessarily perfect, indicators of the construct at the group level. When Y_j was disaggregated to the individual level, the estimated micro-macro effect is estimated with a downwards bias, except when the individual-level scores perfectly reflected the construct at the group level. When the true micro-macro relation was absent in the population, all four approaches estimated the effect unbiasedly.

The results in Table 2.2 indicate that increasing the number of groups from 40 to 200 reduces the bias of the estimates a little and leads to much smaller standard deviations of the estimates for all four approaches. Increasing the number of group members from 10 to 40, improving the quality of the individual-level scores to reflect the group-level construct, or increasing the effect size of the micro-macro relation did not cause large changes in the bias of the mean estimates, nor in the value of their standard deviations.

Power and observed Type-I error rates

The results with respect to power and Type-I error rates were also collapsed for each factor over the three remaining factors and are shown in Table 2.3. The observed power to detect the micro-macro effect could be determined in the 24 conditions in which an effect was present in the population. For the latent class approach, mean aggregation, and mode aggregation, the observed power was, larger than .70 when the true effect was large. A moderate micro-macro effect could only be detected with power larger than .70 in samples with 200 groups. When disaggregating, power is always above .70, except when the individual-level scores are poor indicators of the group-level construct.

The observed Type-I error rates could be evaluated in the 12 conditions with a zero micro-macro effect in the population. In these conditions the observed Type-I error rate was expected to lie between .02 and .09 with a probability of 0.935.[3] When the data were analyzed with the latent class approach, mean aggregation or mode aggregation, all the

[3]This probability is based on a binomial distribution with 100 trials and a success probability equal to .05.

Table 2.2: Means and Standard Deviations of Estimates of Micro-Macro Relationship Estimated with Latent Class Approach, Mean Aggregation, Mode Aggregation, and Disaggregation, after Collapsing

		Latent class	Mean aggregation	Mode aggregation	Disaggregation
	β_4	$b_4(SD)$	$b_6(SD)$	$b_8(SD)$	$b_{10}(SD)$
$L_2 = 40$	0	0.00(0.79)	0.05(1.25)	0.00(0.64)	-0.01(0.43)
	1	1.06(0.77)	1.61(1.25)	0.92(0.63)	0.65(0.42)
	2	2.18(0.84)	3.41(1.45)	1.87(0.70)	1.29(0.48)
$L_2 = 200$	0	0.01(0.32)	0.03(0.50)	0.01(0.28)	0.00(0.19)
	1	1.01(0.33)	1.58(0.56)	0.91(0.28)	0.64(0.19)
	2	2.02(0.38)	3.18(0.62)	1.79(0.31)	1.23(0.20)
$L_1 = 10$	0	0.02(0.61)	0.04(0.76)	0.03(0.46)	0.01(0.32)
	1	1.08(0.60)	1.38(0.80)	0.88(0.44)	0.66(0.30)
	2	2.13(0.65)	2.79(0.86)	1.71(0.48)	1.26(0.33)
$L_1 = 40$	0	-0.01(0.50)	0.03(0.99)	-0.01(0.46)	-0.02(0.30)
	1	0.99(0.50)	1.81(1.00)	0.96(0.48)	0.63(0.30)
	2	2.07(0.56)	3.79(1.20)	1.96(0.53)	1.27(0.35)
$\beta_2 = 1$	0	0.04(0.74)	0.14(1.47)	0.05(0.46)	0.01(0.16)
	1	1.05(0.74)	2.25(1.55)	0.71(0.47)	0.23(0.16)
	2	2.21(0.78)	4.79(1.79)	1.43(0.49)	0.47(0.14)
$\beta_2 = 3$	0	0.00(0.47)	0.00(0.71)	0.00(0.47)	0.00(0.31)
	1	1.02(0.46)	1.50(0.70)	1.00(0.45)	0.65(0.29)
	2	2.07(0.50)	3.06(0.77)	2.03(0.49)	1.24(0.27)
$\beta_2 = 200$	0	-0.02(0.45)	-0.02(0.45)	-0.02(0.45)	-0.02(0.46)
	1	1.03(0.46)	1.03(0.46)	1.03(0.46)	1.05(0.47)
	2	2.03(0.54)	2.03(0.54)	2.03(0.54)	2.07(0.60)

observed Type-I error rates lay between these boundaries. When Y_j is disaggregated to the individual level, the observed Type-I error rates were unacceptably high, ranging from .18 to .60, indicating that this approach leads to an unacceptably liberal significance test for the micro-macro effect.

Increasing the sample sizes, the quality of the individual-level scores to reflect the construct at the group level, or the effect size all lead to increased power, regardless of the manner in which the micro-macro relation is modeled. The observed Type-I error rates do not seem to vary as a function of the four manipulated factors. The results reported above are very similar for the Wald and the likelihood-ratio test.

2.5.4 Conclusion

Overall the latent class approach obtains unbiased parameters even when the individual-level scores poorly reflect the (latent) group-level score with reasonable power and Type-I error rate. Aggregation only works with perfect (mean aggregation) or good (mode aggregation) indicators, which are however rather unrealistic conditions in practice. Using disaggregation, the observed Type-I error rates were unacceptably high so this approach should be avoided anyhow. Since the latent class approach performed better than the other 3 approaches, only this approach is evaluated in a more complex model.

Table 2.3: Power and Observed Type-I Error Rates of Micro-Macro Relationship Estimated with Latent Class Approach, Mean Aggregation, Mode Aggregation, and Disaggregation, after Collapsing

		Latent class		Mean aggregation		Mode aggregation		Disaggregation	
	β_4	Wald	LR	Wald	LR	Wald	LR	Wald	LR
$L_2 = 40$	0	.04	.05	.04	.05	.03	.04	.43	.43
	1	.24	.28	.25	.29	.25	.27	.71	.71
	2	.72	.78	.74	.78	.74	.76	.93	.93
$L_2 = 200$	0	.05	.06	.05	.06	.05	.06	.41	.41
	1	.86	.87	.84	.85	.85	.85	.93	.93
	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$L_1 = 10$	0	.03	.05	.04	.05	.04	.05	.30	.30
	1	.50	.54	.51	.53	.51	.52	.77	.77
	2	.81	.87	.85	.88	.83	.85	.94	.94
$L_1 = 40$	0	.06	.06	.05	.06	.05	.06	.54	.54
	1	.60	.61	.58	.60	.59	.60	.87	.87
	2	.90	.91	.89	.90	.90	.90	.99	.99
$\beta_2 = 1$	0	.04	.06	.04	.05	.04	.04	.18	.18
	1	.40	.45	.41	.43	.40	.42	.59	.60
	2	.73	.82	.77	.81	.76	.77	.90	.90
$\beta_2 = 3$	0	.05	.06	.06	.06	.05	.06	.48	.48
	1	.61	.62	.59	.62	.61	.62	.92	.92
	2	.94	.95	.94	.95	.95	.95	1.00	1.00
$\beta_2 = 200$	0	.04	.05	.04	.05	.04	.05	.60	.60
	1	.64	.65	.64	.65	.64	.65	.95	.95
	2	.89	.91	.89	.91	.89	.91	1.00	1.00

2.6 Adding a Level-2 Predictor to the 1-2 Model

The 1-2 model can be extended to a 2-1-2 model by adding a predictor X_j at the group level as shown in Figure 2.2. In the present discussion X_j is assumed to be dichotomous, but the extension to the general case of Q response categories or to continuous variables is straightforward.

At the group level two logistic regression equations are defined and a latent class model is used to link the individual and group level, so that for dichotomous data the model can be formulated in terms of three logit regression equations:

$$\text{Logit}(P(\zeta_j = 1|X_j)) = \beta_1 + \beta_2 X_j, \tag{2.8}$$

$$\text{Logit}(P(Y_j = 1|X_j, \zeta_j)) = \beta_3 + \beta_4 X_j + \beta_5 \zeta_j + \beta_6 X_j \cdot \zeta_j, \tag{2.9}$$

and

$$\text{Logit}(P(Z_{ij} = 1|\zeta_j)) = \beta_7 + \beta_8 \zeta_j, \tag{2.10}$$

in which β_1, β_3, and β_7 are intercepts and β_2, β_4, β_5, β_6, and β_8 slopes. The regression model for Y_j contains the main effects of ζ_j and X_j and their mutual interaction effect represented by the product variable $X_j \times \zeta_j$. Furthermore, ζ_j itself is regressed on X_j.

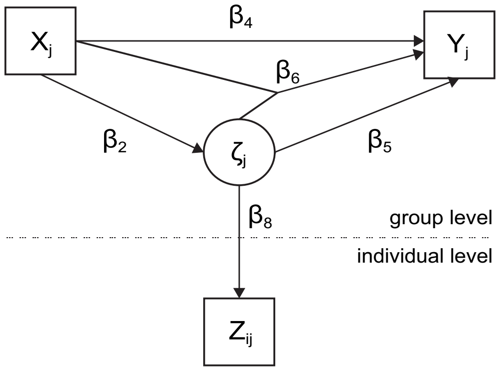


Figure 2.2: Graphical Representation 2-1-2 Model

The joint probability density of X_j, $\mathbf{Z}_j$, Y_j, and ζ_j for an arbitrary group j is defined as

$$\begin{aligned} P(X_j, \mathbf{Z}_j, Y_j, \zeta_j) &= P(X_j)P(\zeta_j|X_j)P(Y_j|X_j,\zeta_j)P(\mathbf{Z}_j|\zeta_j) \\ &= \underbrace{P(X_j)P(\zeta_j|X_j)P(Y_j|X_j,\zeta_j)}_{between}\underbrace{\prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j)}_{within}, \end{aligned} \tag{2.11}$$

while the observed or incomplete data log-likelihood function is

$$\begin{aligned} \log\ L = \sum_{j=1}^{J} \log\Bigg(\sum_{c=1}^{C}\Big(P(X_j)P(\zeta_j = c|X_j)P(Y_j|X_j,\zeta_j = c) \\ \prod_{i=1}^{I} P(Z_{ij}|\zeta_j = c)\Big)\Bigg), \end{aligned} \tag{2.12}$$

in which C represents the number of latent classes, $c = 1, \cdots, C$. The likelihood function can be maximized in the same two ways as described for the 1-2 model in Appendix A, namely the Persons-as-Variables approach and the Two-level regression approach, requiring the data to be appropriately structured. The model can again be estimated with the Latent GOLD software (Vermunt & Magidson, 2005a).

2.7 Simulation Study 2-1-2 Model

2.7.1 Aim of the Simulation Study

The latent class approach, which seems to work well for a simple micro-macro relation with dichotomous variables, is now evaluated in the slightly more complex 2-1-2 model. The Monte Carlo simulation study reported in this section intends to investigate how the bias in parameter estimates, the Type-I error rates, and the power of tests for individual regression coefficients are influenced by (1) the strength of the true relations, (2) the degree to which the individual-level scores reflect the latent group-level score, and (3) the sample sizes at both the individual and group level. As in the previous simulation study, the significance of the parameters is evaluated with both Wald and likelihood-ratio tests.

2.7.2 Method

Data are generated according to the 2-1-2 Model shown in Figure 2.2 and formally described by Equations 2.8, 2.9, and 2.10. In the population models, all three main effects at the macro level were assumed to be either absent ($\beta = 0$), moderate ($\beta = 1$), or strong ($\beta = 2$). The interaction effect between X_j and ζ_j was either negative ($\beta_6 = -1$), absent ($\beta_6 = 0$), or positive ($\beta_6 = 1$). The scores on Z_{ij} were either poor indicators ($\beta_8 = 1$), good indicators ($\beta_8 = 3$), or perfect indicators ($\beta_8 = 200$) of the latent group score ζ_j. The number of groups was set to either 40 or 200, and the number of individuals within a group to either 10 or 40. The intercept values β_1 and β_3, and β_7 were determined in such a way that the marginal distributions of Z_{ij}, ζ_j, and Y_j were uniform. The marginal probability of X_j was made uniform. Crossing the 7 factors resulted in $3 \times 3 \times 3 \times 3 \times 3 \times 2 \times 2 = 972$ conditions.

Again 100 data sets were generated for each condition using Latent GOLD (Vermunt & Magidson, 2005a) and the data sets were analyzed with the latent class approach. Power and observed Type-I error rates were determined for both the Wald and likelihood-ratio tests as described in the method section of the previous simulation study. The power for the main effects of X_j and ζ_j on Y_j was only determined in those conditions in which there was no interaction between X_j and ζ_j in the population. In order to assess the (main) effect of a particular factor in the simulation study, the results obtained in the different conditions were collapsed over the other factors.

2.7.3 Results

Bias in the parameter estimates

A summary of the estimated effects at the group level is given in Table 2.4. First, the results in Table 2.4 indicate that there is some bias in the estimates. Moreover, the magnitude of the bias seems to be proportional to the value of true effect since there is no bias when the true effect equals zero. Bias slightly decreases when the number of groups is increased, but remains similar when the number of individuals within a group is increased, or when the quality of the individual-level scores reflecting the latent group-level score is improved. The standard deviations of the estimates are quite large. Consistent with the first simulation study, increasing the number of groups reduces the standard deviations. Increasing the number of group members and the quality of the indicators have only small effects on the standard deviations.

Power and observed Type-I error rates

The results with respect to power and Type-I error rates are summarized in Table 2.5. The power of the macro-level effects could be observed in the 646 conditions in which there was a non-zero effect in the population. The results can be summarized as follows. First, the power of the test $H_0 : \beta_2 = 0$ for the main effect of X_j on ζ_j is larger than 0.70 when the true value of the effect is strong. When the true effect is moderate, power

Table 2.4: Means and Standard Deviations of Estimates of Group-level Effects 2-1-2 Model Estimated with Latent Class Approach, after Collapsing

	β_2	$\bar{b}_2(SD)$	β_4	$\bar{b}_4(SD)$	β_5	$\bar{b}_5(SD)$	β_6	$\bar{b}_6(SD)$
$L_2 = 40$	0	0.00(0.75)	0	-0.02(1.40)	0	-0.03(1.47)	-1	-1.23(2.10)
	1	1.05(0.78)	1	1.15(1.48)	1	1.19(1.52)	0	-0.03(2.15)
	2	2.08(0.86)	2	2.38(1.60)	2	2.35(1.64)	1	1.17(2.30)
$L_2 = 200$	0	0.00(0.32)	0	-0.02(0.53)	0	-0.01(0.55)	-1	-1.05(0.80)
	1	1.01(0.34)	1	1.02(0.54)	1	1.03(0.57)	0	0.01(0.82)
	2	2.03(0.37)	2	2.06(0.60)	2	2.09(0.63)	1	1.07(0.92)
$L_1 = 10$	0	0.00(0.58)	0	-0.02(1.01)	0	-0.02(1.09)	-1	-1.14(1.55)
	1	1.04(0.61)	1	1.08(1.07)	1	1.12(1.12)	0	0.00(1.58)
	2	2.06(0.67)	2	2.20(1.14)	2	2.21(1.21)	1	1.12(1.72)
$L_1 = 40$	0	-0.01(0.49)	0	-0.03(0.92)	0	-0.02(0.94)	-1	-1.14(1.35)
	1	1.02(0.51)	1	1.09(0.95)	1	1.11(0.97)	0	-0.03(1.38)
	2	2.05(0.57)	2	2.24(1.05)	2	2.23(1.07)	1	1.13(1.50)
$\beta_8 = 1$	0	0.00(0.68)	0	-0.02(1.12)	0	0.02(1.24)	-1	-1.17(1.77)
	1	1.06(0.70)	1	1.11(1.17)	1	1.17(1.27)	0	-0.04(1.80)
	2	2.07(0.75)	2	2.22(1.24)	2	2.26(1.35)	1	1.10(1.92)
$\beta_8 = 3$	0	-0.01(0.47)	0	-0.03(0.90)	0	-0.04(0.91)	-1	-1.12(1.29)
	1	1.02(0.49)	1	1.08(0.93)	1	1.07(0.92)	0	0.01(1.34)
	2	2.04(0.55)	2	2.22(1.03)	2	2.20(1.05)	1	1.14(1.47)
$\beta_8 = 200$	0	-0.01(0.46)	0	-0.02(0.88)	0	-0.03(0.89)	-1	-1.12(1.29)
	1	1.02(0.49)	1	1.07(0.93)	1	1.09(0.94)	0	-0.01(1.31)
	2	2.04(0.55)	2	2.22(1.03)	2	2.20(1.02)	1	1.12(1.45)

is above .70 when the number of groups is 200 but for the other factors the power to detect an moderate effect of X_j on ζ_j lies between .26 and .63. The results are similar for the Wald and likelihood-ratio tests. Second, the power to test $H_0 : \beta_4 = 0$ for the main effect of X_j on Y_j and the power of the test $H_0 : \beta_5 = 0$ for the main effect of ζ_j on Y_j are above .70 when the true effects are strong except for $\beta_5 = 2$ with 40 groups. Moderate main effects can again only be detected with sufficient power when the number of groups is 200. For the other factors the power to detect moderate main effects lies between .22 and .61 and although the obtained power is a bit larger with a likelihood-ratio test compared to a Wald test, the difference is rather small. Third, the power of the test $H_0 : \beta_6 = 0$ for the interaction effect of X_j and ζ_j on Y_j is very low but higher for the likelihood-ratio test than for the Wald test, especially when there are only 40 groups. For the Wald test the power to detect an interaction effect lies between .02 and .30 while for the likelihood-ratio test power lies between .11 and .32.

The observed Type-I error rates could be evaluated in the 324 conditions in which the macro-level effect was absent. As before, the Type-I error rate was expected to lie between .02 and .09. This holds in all conditions, except that the observed Type-I error rates were too low for the test of β_6 when determined with a Wald test in the conditions with 40 groups. The observed Type-I error rates seem to be independent of the manipulated factors. Within acceptable boundaries, the Wald test seems to be slightly too conservative while the likelihood-ratio test seems to be slightly too liberal.

Table 2.5: Power and Observed Type-I Error Rates of Group-level Effects 2-1-2 Model with Wald Test and Likelihood-Ratio Test, after Collapsing

	β_2	Wald	LR	β_4	Wald	LR	β_5	Wald	LR	β_6	Wald	LR
$L_2 = 40$	0	.04	.05	0	.03	.06	0	.03	.07	-1	.04	.12
	1	.26	.30	1	.26	.29	1	.22	.29	0	.01	.06
	2	.73	.78	2	.72	.73	2	.61	.69	1	.02	.11
$L_2 = 200$	0	.05	.06	0	.05	.05	0	.05	.05	-1	.30	.32
	1	.87	.88	1	.87	.86	1	.80	.81	0	.05	.06
	2	1.00	1.00	2	1.00	.99	2	.99	.99	1	.24	.28
$L_1 = 10$	0	.04	.05	0	.04	.06	0	.04	.07	-1	.15	.20
	1	.52	.55	1	.54	.54	1	.46	.52	0	.03	.06
	2	.82	.86	2	.84	.83	2	.75	.81	1	.11	.17
$L_1 = 40$	0	.05	.05	0	.04	.06	0	.04	.06	-1	.19	.24
	1	.61	.62	1	.59	.60	1	.55	.58	0	.03	.06
	2	.91	.92	2	.88	.89	2	.85	.88	1	.15	.21
$\beta_8 = 1$	0	.04	.06	0	.04	.06	0	.03	.07	-1	.11	.17
	1	.46	.51	1	.52	.52	1	.39	.46	0	.02	.06
	2	.75	.82	2	.82	.79	2	.66	.75	1	.08	.14
$\beta_8 = 3$	0	.05	.05	0	.05	.06	0	.04	.06	-1	.20	.24
	1	.62	.63	1	.58	.60	1	.57	.59	0	.03	.06
	2	.92	.93	2	.88	.90	2	.87	.89	1	.16	.22
$\beta_8 = 200$	0	.05	.05	0	.04	.05	0	.04	.06	-1	.20	.25
	1	.62	.63	1	.59	.61	1	.57	.59	0	.03	.06
	2	.92	.92	2	.88	.90	2	.87	.89	1	.16	.22

2.7.4 Conclusion

From this second simulation study, it can be concluded that the latent class approach produces almost unbiased parameters in the 2-1-2 model but standard deviations are quite high and can be reduced by using a large number of groups. Especially for the interaction effect, the power is low in most conditions but can be improved by using a likelihood-ratio test instead of a Wald test. The Type-I error rates seem correct with both the Wald and the likelihood-ratio tests.

2.8 Empirical Data Example

The discrete latent variable approach is illustrated with an empirical application on data from personal networks, in which individuals (egos) are interviewed together with persons from their network (alters). Up till now only research questions could be answered when the dependent variable was defined at the micro level, so at the level of the alters or ties (Snijders et al., 1995; van Duijn, van Busschbach, & Snijders, 1999). The latent variable approach allows to answer research questions with a dependent variable at the higher ego level, so providing new possibilities for investigating a broad range of research questions in studies of personal networks. More specifically, in the current example the effect of belonging to a particular type of personal network on the behavior of the ego himself is explored.

2.8.1 Data

The data come from the Netherlands Kinship Panel Study (NKPS), which is a large-scale database on Dutch families that yields information for individual respondents (the egos) and some of their family members and friends (the alters). The data are publicly available and can be retrieved from http://www.nkps.nl. For the present example, data were available for 8161 egos with maximally six alters nested within each ego: the parents in law, two siblings, two children, and a friend.

Kalmijn and Vermunt (2007) used these data to investigate whether selection in networks is based on age and marital status. In the present paper a different perspective is chosen. Instead of expecting that persons choose the persons in their network based on their marital status, we assume that egos are members of a network in which either many or few people are married. The latent variable ζ_j then represents latent class membership of an ego's network: $\zeta_j = 0$ when the ego belongs to a network in which few members are married versus $\zeta_j = 1$ when the ego belongs to a network in which many members are married. The marital status of the alters ($Z_{ij} = 0$ for unmarried alters and $Z_{ij} = 1$ for married alters) are taken as exchangeable indicators of the type of network an ego belongs to. The dependent level-2 variable in this analysis is the dichotomous variable Y_j indicating whether an ego is married or not ($Y_j = 0$ when the ego is not married versus $Y_j = 1$ when the ego is married). The religiosity of the ego ($X_j = 0$ when the ego j is not religious and $X_j = 1$ when the ego is religious) is treated as the level-2 explanatory variable that affects the probability of an ego to belong to a particular type of network. Eggebeen and Dew (2009) already pointed out that religion is a very important factor in family formation during young adulthood. In the present analysis it is expected that non-religious persons rather belong to the latent class with few married members than to the class with many married members. For religious people, we expect the opposite. Furthermore, we allow for an interaction effect of type of network and religiosity on the dependent variable, implying that the effect of the network on being married can be different for religious and non-religious) persons. The model as formulated here can be extended in several ways. First, the exchangeability assumption, stating that all alters are equivalent indicators of the type of network, can eventually be relaxed if the parents in law, siblings, children, and friend to the network provide (partly) different network information. Second, if necessary, a model with more than two latent classes at the network level could be considered. These extensions will not be further discussed here.

2.8.2 Method

The model, shown in Figure 2.2, is defined by Equations 2.8-2.10 and the model parameters can be estimated with the software package Latent GOLD (Vermunt & Magidson, 2005a) by applying either the Two-level regression or the Persons-as-Variables approach as described in Appendix B.

2.8.3 Results

Since the Persons-as-Variables approach and the Two-level regression approach yield the same results, only the results of the Two-level regression approach are presented here. Looking at the regression coefficients in the second and third column of Table 2.6, it can be seen that the Wald tests for the interaction effect of X_j and ζ_j on the level-2

Table 2.6: Regression Coefficients Empirical Example

	Interaction		No interaction	
Independent variable	β	SE	β	SE
Dependent variable: Network ego				
Intercept (β_1)	-1.16**	0.16	-1.18**	.16
Religion ego (β_2)	4.49**	0.34	4.59**	.33
Dependent variable: Married ego				
Intercept (β_3)	-3.77**	1.16	-3.61**	.96
Religion ego (β_4)	-0.01	2.92	-3.10**	.30
Network ego (β_5)	6.47**	1.16	6.30**	.97
Religion ego * Network ego (β_6)	-3.10	2.94		
Dependent variable: Married alter				
Intercept (β_7)	-0.49**	0.05	-0.49**	.05
Network ego (β_8)	0.91**	0.05	0.91**	.05

* $p < .05$, ** $p < .01$

outcome variable Y_j is not significant. Therefore, a model without this interaction term is presented in the last two columns of the table.

By substituting the estimated parameter values in the logit regressions equations 2.8, 2.9, and 2.10 and transforming them into the probability scale, the probabilities as given in Table 2.7(a), 2.7(b), and 2.7(c) are obtained.
As can be seen from Table 2.7(a), alters in the two network classes have a probability of being married of .38 and .60, respectively. So, the latent classes can be interpreted in terms of the egos belonging to a network with either a minority or a majority of married alters.

Second, Table 2.7(b) indicates that when an ego is not religious, the probability of having a network in which the majority of the persons is married is .23 while it is .97 for an ego that is religious.

Third, Table 2.7(c) shows that the probability of an ego being married depends on whether he is religious or not, and on the type of network the ego belongs to. Since only the main effect of the ego network is significant, only this effect is interpreted. Egos that have a network in which a majority of alters is married, have a higher probability of being married than egos that have a network in which a minority of alters is married and egos that are religious have a lower probability of being married than non religious egos.

2.8.4 Conclusion

An interesting aspect emerging from this analysis is that there is a direct effect of whether an ego is religious on the probability that the ego is married and an indirect effect that runs through the network of the ego. The direct effect is counter intuitive since the probability of being married is higher for non religious than for religious egos. The indirect effect is more intuitive since religious egos have a higher probability to have a network in which the majority of alters is married than a network in which the minority of the alters is married. Egos that belong to a network in which the majority of alters is married have a higher probability to be married themselves compared to egos that belong to a network in which the minority of the alters is married.

Table 2.7: Estimated Probabilities Empirical Data Example

(a)

Network ego	P(Married alter = 1 | Network ego) (SE)
0	.38 (0.01)
1	.60 (0.00)

(b)

Religion ego	P(Network ego = 1 | Religion ego) (SE)
0	.23 (0.03)
1	.97 (0.01)

(c)

Religion ego	Network ego	P(Married ego = 1 | Religion ego, Network ego) (SE)
0	0	.03 (0.03)
0	1	.94 (0.01)
1	0	.00 (0.00)
1	1	.40 (0.03)

2.9 Discussion

Although a wide variety of research questions in the social and behavioral sciences involve micro-macro relations, specific methods to analyze such relationships are not yet fully developed. The current article is contributing to this development by showing how a latent variable approach which was originally proposed for continuous outcomes (Croon & van Veldhoven, 2007) can be modified for the application to discrete outcomes.

We showed that, in a simple 1-2 model, the latent variable approach outperforms more traditionally aggregation and disaggregation strategies with respect to bias with reasonable power and correct observed Type-I error rates. In a more complex 2-1-2 model, there is small bias and standard deviations are a little higher. These can be reduced by using a larger number of groups. Power is acceptable for the main effects but relatively low for the interaction effect, while the observed Type-I error rates are correct. The low power for the interaction effect could be due to general power problems associated with detecting interaction effects by including product terms in the regression equation (McClelland & Judd, 1993; Whisman & McClelland, 2005). Using a likelihood-ratio test instead of a Wald test increases power. Overall, the latent variable approach seems to work well for analyzing micro-macro relations with discrete variables and this enables investigating research questions that could not be addressed appropriately before.

The current research was restricted to models with only one micro-level predictor. Further research should be devoted to models with multiple level-1 variables. In this context it could be explored whether it is more practical to use three-step estimation procedures as described by Bakk, Tekle, and Vermunt (2013), instead of the currently suggested one-step estimation procedure. this would disentangle the aggregation from the group-level analysis. Furthermore, in the current article the focus was set at two-level situations in which the predictors and outcome variable were observed variables. It would be interesting to explore the possibilities to extend the model to the situation in which the outcome variable and/or predictors are latent constructs measured with multiple indicators.

CHAPTER 3

Micro-Macro Analysis with Multiple Micro-Level Variables

Abstract

An existing micro-macro method for a single individual-level variable is extended to the multivariate situation by presenting two multilevel latent class models in which multiple discrete individual-level variables are used to explain a group-level outcome. As in the univariate case, the individual-level data are summarized at the group level by constructing a discrete latent variable at the group level and this group-level latent variable is used as a predictor for the group-level outcome. In the first extension, that is referred to as the Direct model, the multiple individual-level variables are directly used as indicators for the group-level latent variable. In the second extension, referred to as the Indirect model, the multiple individual-level variables are used to construct an individual-level latent variable that is used as an indicator for the group-level latent variable. This implies that the individual-level variables are used indirectly at the group level. The within and between components of the (co)variation in the individual-level variables are independent in the 'Direct model', but dependent in the 'Indirect' model. Both models are discussed and illustrated with an empirical data example.

This chapter is submitted for publication as Bennink, M., Croon, M. A., Kroon, B. & Vermunt, J. K. (2014). *Micro-macro multilevel latent class models with multiple discrete individual-level variables.*

3.1 Introduction

In many research areas, data are collected on individuals (micro-level units) who are nested within groups (macro-level units) (Goldstein, 2011). For example, data can be collected on children nested in schools, on employees nested in organizations, or on family members nested in families. The variables involved may be either measured at the individual level or at the level of the groups. Following Snijders and Bosker (2012), one can distinguish between macro-micro and micro-macro situations. In a macro-micro situation, the outcome or dependent variable is measured at the individual level, while in a micro-macro situation, the outcome variable is measured at the group level. The current article focuses on the latter type of multilevel analysis that is needed when, for example, characteristics of household members are related to household ownership of financial products, or when psychological characteristics of employees are related to organizational performance outcomes. Furthermore, attention is focused on micro-macro analysis for discrete data.

In micro-macro analysis, the individual-level data need to be aggregated to the group level, so the aggregated scores can be related to the group-level outcome. When a group mean or mode is used for aggregation, measurement and sampling error in the individual scores is not accounted for and Croon and van Veldhoven (2007) showed that this neglect of random fluctuation in the individual scores causes bias in the estimates of the group-level parameters. Moreover, this type of aggregation ignores all individual differences within the groups. It is well known that the variability of the group means and modes not only represents between-group variation but also partly reflects within-group variation. Therefore, the analysis of observations from micro-macro designs requires an appropriate methodology that takes into account the measurement and sampling error of the individual scores and separates the within- and between-group association among the variables (Preacher, Zyphur, & Zhang, 2010).

Such techniques have been developed by using a group-level latent variable for the aggregation. For continuous data, Croon and van Veldhoven (2007) provide a basic example of this methodology. The scores of the individuals i from group j on an explanatory variable Z_{ij} are interpreted as exchangeable indicators of an unobserved group score on the continuous latent group-level variable ζ_j. Furthermore, the latent variable is treated as a group-level mediating variable between a group-level predictor X_j and a group-level outcome Y_j. Figure 3.1 represents this model graphically. Any hypothesis in which a group-level intervention is not only expected to influence a group-level (performance) measure directly, but also indirectly through a characteristic of the group members, can be tested with this model.

The model belongs to the general framework of generalized latent variable models described by Skrondal and Rabe-Hesketh (2004) and can also be formulated for categorical data (Bennink, Croon, & Vermunt, 2013), by using a latent class model instead of a factor-analytic model that was used for continuous variables. The latent variable ζ_j then becomes a categorical variable with C categories, $c = 1, \cdots, C$. The scores Z_{ij} of the I_j individuals in group j (collected in the vector $\mathbf{Z}_j$) are treated as 'unreliable' indicators of the group score ζ_j. For an arbitrary group j, the relevant conditional probability distribution for the manifest variables Y_j and $\mathbf{Z}_j$ given X_j is:

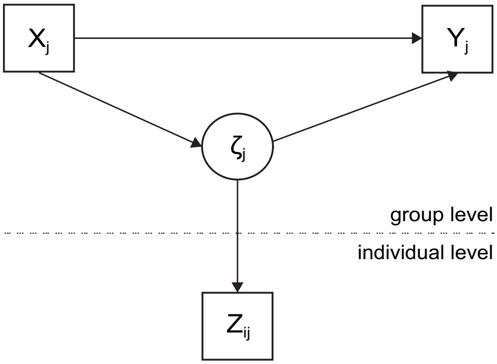


Figure 3.1: Micro-Macro Latent Variable Model with One Micro-Level Variable

$$P(Y_j, \mathbf{Z}_j | X_j) = \sum_{c=1}^{C} P(Y_j, \zeta_j = c | X_j) P(\mathbf{Z}_j | \zeta_j = c). \tag{3.1}$$

The terms on the right hand side of the equation are the between and within part that can be further decomposed as

$$P(Y_j, \zeta_j = c | X_j) = P(\zeta_j = c | X_j) P(Y_j | X_j, \zeta_j = c), \tag{3.2}$$

and

$$P(\mathbf{Z}_j | \zeta_j = c) = \prod_{i=1}^{I_j} P(Z_{ij} | \zeta_j = c). \tag{3.3}$$

In the social and behavioral sciences it is very common to use multiple individual-level variables instead of only a single one. Therefore, two multilevel latent class models are presented that extend the univariate case to the situation with multiple Z_{ij}-variables. As in the existing method, the Z_{ij}-variables are summarized by a single discrete latent variable at the group level (ζ_j). In the first model, that is referred to as the Direct model, the Z_{ij}-variables are directly used as indicators for ζ_j. In the second model, that is referred to as the Indirect model, this is done indirectly through an individual-level latent variable (η_{ij}). The Direct model can, for example, be used to construct a latent classification of households based on the age, gender, and educational level of the household members to predict household ownership of financial products. The Indirect model can, for example, be used when multiple individual-level items on the satisfaction of employees with respect to their relationships at work are used to construct the individual-level latent variable η_{ij} that is used as an indicator for ζ_j to predict organizational performance measures, such as the level of organizational conflicts. In the remainder of the article, both methods and their estimation procedures are discussed and applied to empirical data examples.

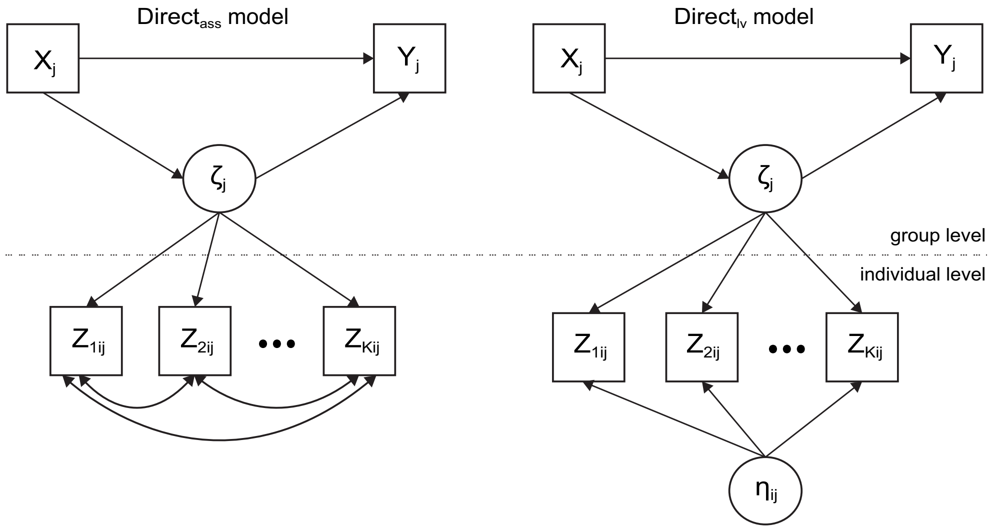


Figure 3.2: Direct Models with Multiple Micro-Level Variables

3.2 Direct Model

Figure 3.1 is extended to a situation with K individual-level variables. These individual-level variables $Z_{1ij}, Z_{kij} \cdots Z_{Kij}$, can be directly used as indicators of the discrete latent group-level variable ζ_j, as done in the model with a single Z_{kij}. In this way, a (latent) typology of groups is constructed based on the multiple individual-level variables. For example, the age, gender, and educational level of household members can be used to construct a classification of households. This classification of groups is used as a predictor for the observed group-level outcome Y_j, for example, the household ownership of a financial product. Also other (observed) group-level predictors represented by X_j, can be included in the model. For instance, the household income can be used as an additional group-level predictor.

Although not necessarily in a model with a single Z_{kij}, in a model with multiple Z_{kij}-variables it needs to be accounted for that the individual-level variables can be dependent within individuals. It is not reasonable to assume that all of the association between the individual-level indicators is explained by ζ_j. This can be done in two ways. As a first alternative, all two-way within associations among the Z_{kij}-variables can be incorporated in the model as shown in the left panel of Figure 3.2. This model is referred to as the 'Direct$_{\text{ass}}$ model'. A second alternative consists of defining a discrete individual-level latent variable η_{ij} with D categories, $d = 1, \cdots, D$, as shown in the right panel of Figure 3.2. This model is referred to as the 'Direct$_{\text{lv}}$ model'.[4]

As in Equation 3.1, the probability distribution of an arbitrary group j contains a between and a within term. For both models, the between part is still represented by Equation 4.3, but they differ with respect to the within part. For the Direct$_{\text{ass}}$ model, the

[4] η_{ij} does not necessarily need to be discrete, but can be defined continuous as well.

within part is

$$P(\mathbf{Z}_j|\zeta_j = c) = \prod_{i=1}^{I_j} P(Z_{1ij}, Z_{kij}, \cdots Z_{Kij}|\zeta_j = c), \tag{3.4}$$

whereas for the Direct$_{\text{lv}}$ model, the within part is

$$P(\mathbf{Z}_j|\zeta_j = c) = \prod_{i=1}^{I_j} \sum_{d=1}^{D} P(\eta_{ij} = d) \prod_{k=1}^{K} P(Z_{kij}|\zeta_j = c, \eta_{ij} = d). \tag{3.5}$$

The group members are used as exchangeable indicators, this implies that $P(Z_{1ij}, Z_{kij}, \cdots Z_{Kij}|\zeta_j = c)$ in the Direct$_{\text{ass}}$ model and $P(Z_{kij}|\zeta_j = c, \eta_{ij} = d)$ in the Direct$_{\text{lv}}$ model, are identical for all individuals. In the Direct$_{\text{ass}}$ model, there is by definition local dependency among the indicators given ζ_j, but in the Direct$_{\text{lv}}$ model, the indicators are locally independent given η_{ij} and ζ_j. It is also important to note is that η_{ij} and ζ_j are assumed to be independent.

3.3 Indirect Model

When the K individual-level variables were intended in the first place to measure an individual-level construct, the relationship between the group-level latent variable and the individual-level items is specified indirectly rather than directly. For example, suppose that the satisfaction of employees with their relationships at work is measured by three indicators: (1) their satisfaction with the relation with their supervisor, (2) the satisfaction with their relation with other coworkers, and (3) the degree in which they experience a family culture at their working environment. These three Z_{kij}-variables may be treated as indicators of an underlying latent construct at the individual level (η_{ij}). In the current article η_{ij} is a discrete variable with D categories, $d = 1, \cdots, D$.[5] Since there may exist group differences on η_{ij}, a group-level latent variable (ζ_j) may be invoked to represent these between-group differences on η_{ij}. This model is graphically shown in Figure 3.3 and referred to as the 'Indirect model'.

Referring to the formal general description in Equation 3.1, the between part of this model is represented again by Equation 4.3, but the within part is now:

$$P(\mathbf{Z}_j|\zeta_j = c) = \prod_{i=1}^{I_j} \sum_{d=1}^{D} P(\eta_{ij} = d|\zeta_j = c) \prod_{k=1}^{K} P(Z_{kij}|\eta_{ij} = d). \tag{3.6}$$

The group members are again treated as exchangeable, so that $P(Z_{kij}|\eta_{ij} = d)$ has the same form for all individuals. The individual-level variables are locally independent given η_{ij} and the two latent variables are dependent since the distribution of η_{ij} depends on ζ_j. In this model there is no immediate need to allow for residual association among the individual indicators since η_{ij} is assumed to account for all of the associations that exist among the indicators.

[5] Varriale and Vermunt (2012) proposed a similar model with a continuous η_{ij} and no group-level outcome.

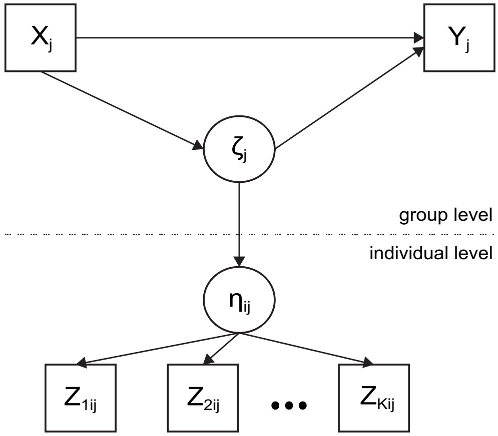


Figure 3.3: Indirect Model with Multiple Micro-Level Variables

3.4 Estimation, Identification, and Model Selection

The micro-macro models presented above are extended versions of the multilevel latent class model proposed by Vermunt (2003). The extension involves that, in addition to having discrete latent variables at two levels, these models contain an outcome variable at the group level. Vermunt (2003) showed how to obtain maximum-likelihood estimates for multilevel latent class models using an EM algorithm, and a very similar procedure can be used here. The log-likelihood to be maximized equals:

$$\begin{aligned}\log\ L &= \sum_{j=1}^{J} \log P(Y_j, \mathbf{Z}_j | X_j) \\ &= \sum_{j=1}^{J} \log \left(\sum_{c=1}^{C} P(\zeta_j = c | X_j) P(Y_j | X_j, \zeta_j = c) \right. \\ &\quad \left. \prod_{i=1}^{I_j} \sum_{d=1}^{D} P(\eta_{ij} = d | \zeta_j = c) P(\mathbf{Z}_{ij} | \zeta_j = c, \eta_{ij} = d) \right). \end{aligned} \tag{3.7}$$

The expected complete-data log-likelihood, which is computed in the E-step and maximized in the M-step, has the following form:

$$
\begin{aligned}
E(\log\ L_{comp}) \quad = \quad & \sum_{j=1}^{J}\sum_{c=1}^{C}\pi_j^c \log\ P(\zeta_j = c|X_j) \\
& + \sum_{j=1}^{J}\sum_{c=1}^{C}\pi_j^c \log P(Y_j|X_j, \zeta_j = c) \\
& + \sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{c=1}^{C}\sum_{d=1}^{D}\pi_{ij}^{cd} \log P(\eta_{ij} = d|\zeta_j = c) \\
& + \sum_{j=1}^{J}\sum_{i=1}^{I_j}\sum_{c=1}^{C}\sum_{d=1}^{D}\pi_{ij}^{cd} \log P(\mathbf{Z}_{ij}|\zeta_j = c, \eta_{ij} = d). \qquad (3.8)
\end{aligned}
$$

Here, π_j^c and π_{ij}^{cd} denote the posterior class membership probabilities $P(\zeta_j = c|Y_j, \mathbf{Z}_j, X_j)$ and $P(\zeta_j = c, \eta_{ij} = d|Y_j, \mathbf{Z}_j, X_j)$, respectively. These posterior probabilities can be obtained in an efficient manner using an upward-downward algorithm. In the upward step we obtain π_j^k and in the downward step we obtain π_{ij}^{cd} as $\pi_j^c P(\mathrm{J}_{ij} = d|\zeta_j = c, Y_j, \mathbf{Z}_{ij}, X_j)$. This algorithm is implemented in the Latent GOLD program (Vermunt & Magidson, 2013) that we used for parameter estimation in the empirical examples presented in the next section.

Since the four sets of model probabilities are parameterized using logit models, the M step involves updating the estimates of a set of logistic parameters in the usual way. Note the three special cases of the micro-macro model are all restricted versions of the general model for which we defined the expected complete-data log-likelihood. The Direct$_{\text{ass}}$ model does not contain a lower-level latent variable, which can be specified by setting $D = 1$. In this model, the joint distribution of $\mathbf{Z}_{ij}$ is modeled with a multivariate logistic model containing the two-variable associations between the responses. In the Direct$_{\text{lv}}$ model and the Indirect model, we assume responses Z_{kij} to be locally independent, meaning that the associations between the responses are fixed to zero. Moreover, in the former Z_{kij} is assumed to be independent of ζ_j and in the latter η_{ij} is assumed to be independent of ζ_j, which are restrictions that can be obtained by fixing the logistic parameters concerned to zero.

As regards the identifiability of the models proposed in this article, similar conditions apply as for regular latent class models. A sufficient condition for identification is that both the individual- and the group-level part of the model are identified latent class models (Vermunt, 2005). For the individual-level model this means that we need at least three Z_{kij}-variables ($K \geq 3$), whereas for the group-level model this means that most groups should have at least three individuals ($I_j \geq 3$). However, also when these conditions are not fulfilled, the micro-macro model concerned may be identified. For example, The Direct$_{\text{ass}}$ model, which contains only a group-level latent variable, is also identified with two individuals per group when $K \geq 2$, and the Indirect model is also identified with $K = 2$ and $I_j \geq 3$. A formal procedure to check identification is to determine the rank of the Jacobian matrix, which can be done using Latent GOLD.

Another important issue concerns the selection of the number of classes at the individual and the group level. For multilevel latent class models, Lukočienė et al. (2010) recommended based on simulation studies to use either the BIC (with the number of groups as sample size in the formula) or the AIC3 for making this decision. In the $\text{Direct}_{\text{ass}}$ model, there is only a group-level latent variable, meaning that we can simply select the model with the number of group-level classes that provides the best fit. For the $\text{Direct}_{\text{lv}}$ model and the Indirect model, on the other hand, the number of classes at both levels have to be determined simultaneously. Here, we follow the suggestion by Lukočienė et al. (2010) to first determine the number of classes at the individual level (D), keeping the number of group-level classes fixed to one ($C = 1$). The second step is then to fix D at this value to determine the number of group-level classes (C). In the final step, the number of individual-level latent classes (D) is reconsidered again while fixing C at the previously determined value.

3.5 Empirical Data Examples

In this section, the $\text{Direct}_{\text{ass}}$ model and the Indirect model are applied to empirical data. In the first example, data on Italian households are used to investigate how demographic characteristics of the household members affect household ownership of financial products. Contrarily to Figure 3.2, this example does not contain an additional group-level predictor X_j. In the second example, data on small firms are used to investigate how the perceived quality of employees of their relationships at work affects organizational performance measures, and whether this relationship is moderated by organizational size. All analyses are carried out in Latent GOLD 5.0 (Vermunt & Magidson, 2013).

3.5.1 Example Direct Model

From the 2010 Survey of Italian Household Budgets (Bank of Italy, 2012), information is available on the ownership of financial products by 7951 Italian families. Three such financial products are taken here as group-level outcomes: the number of postal and bank accounts (ACC), the number of postal and bank savings accounts (SAV), and the number of credit cards (CRD). In the same survey, information is available on various demographic characteristics, such as age (AGE), educational level (EDU), and sex (SEX) of the 19836 individual family members. These individual-level variables are used to construct a latent typology of the families (ζ_j). The research question of interest is whether these different types of households show significant differences with respect to ownership of the three financial products.

For the analysis, the variables on ownership of the financial products were categorized into two categories: either the family owned the financial product (score = 1) or it did not (score = 0). For the variables measured at the individual level, age and educational level were categorized into five categories (1=<30, 2=30-40, 3=41-50, 4=51-65, 5=>65; 1=none, 2=elementary school, 3=middle school, 4=high school, 5=bachelor or higher) and sex was coded: 1=male and 2=female.

For the Latent GOLD analyses, six multinominal logit equations were defined, one for each group-level outcome and one for each individual-level variable. In all equations, a discrete group-level variable ζ_j was used as a predictor. All two-way associations among the group-level outcomes and all two-way associations among the individual-level variables were specified as well. Both the selection criteria BIC (based on the number of groups) and AIC3 suggested a model with at least eighteen household-level classes. This large number of latent classes required to obtain an acceptable statistical fit is probably a

Table 3.1: Class Proportions and Class-Specific Probabilities Direct Model

Class ζ	1	2	3
Class size	.36	.32	.32
AGE = 1	.02	.46	.28
AGE = 2	.06	.17	.05
AGE = 3	.05	.29	.07
AGE = 4	.16	.06	.47
AGE = 5	.71	.02	.12
EDU = 1	.10	.19	.01
EDU = 2	.49	.13	.06
EDU = 3	.30	.40	.31
EDU = 4	.09	.21	.41
EDU = 5	.02	.07	.20
SEX = 1	.43	.50	.49
SEX = 2	.57	.50	.51
ACC = 0	.28	.15	.03
ACC = 1	.72	.85	.97
SAV = 0	.75	.81	.84
SAV = 1	.25	.19	.16
CRD = 0	.92	.62	.47
CRD = 1	.08	.38	.53

consequence of the size of the sample on which the analyses were carried out (n=7951), but it simply precludes a straightforward and illuminative interpretation of the results. For illustrative purposes, the solution with three classes is interpreted here. These classes are well separated as indicated by the Entropy R-squared measure (Vermunt & Magidson, 2005b), $R^2_{entr} = .74$, that is in general labeled to be good when it is larger than $.70$.

The estimates of the logit parameters of the fitted model are all significant at the 1% significance level. The corresponding class-specific response probabilities together with the class proportions are given in Table 3.1. The first group-level class contains 36% of the households. The household members in this class are relatively old, lowly educated and a small majority of the family members is female. The second group-level class contains 32% of the households. The members from this class are relatively young, moderately educated with an equal balance between males and females. Finally, the third group-level category contains also 32% of the households. The members are relatively old, highly educated and gender is again equally distributed.

Compared to the other two classes, the households from the first class have a lower probability to own bank accounts (.72), a higher probability to own savings accounts (.25), and the lowest probability to own credit cards (.08). The households from the second class have a higher probability to own bank accounts than the households from the first class but a lower probability than the households from the third class (.85). They have a lower probability to own savings accounts than the first class but a higher probability than the third class (.19). With regard to credit cards, the second type of households is in the middle of the other two classes as well (.38). The households from the third class have the highest probability to own bank accounts (.97) and credit cards (.53) but the lowest probability to own savings accounts (.19).

To conclude, our analysis yielded a classification of the households in three types that especially differ in composition with respect to age and educational level of the family members. Moreover, the different types of households show clear differences with respect to ownership of financial products. The households with older, lower educated members have a higher probability of owning savings accounts than the other two types of households, but a lower probability of owning bank accounts or credit cards. The households with relatively young and moderately educated members have the highest probability to own savings accounts and is located between the other two classes with respect to owning bank accounts and credit cards. The households with relatively old and highly educated members have the highest probability to own bank accounts and credit cards, and are located in between the other two classes with respect to savings accounts.

3.5.2 Example Indirect Model

In the literature on small-firm Human Resource Management (HRM), it is often assumed that working in a small firm is either fantastic or gruesome (Wilkinson, 1999). This assumption is tested on data collected by dr. B. Kroon by administering two questionnaires. In the first questionnaire, 91 HR managers of small organizations provided information about their HR system and other organizational characteristics. In the second questionnaire, 463 employees provided information about their perceptions of work-related issues, such as their experience of positive relationships at work. The research question of interest is how the perception of employees on their relationships at work affects two organizational performance measures: the level of absenteeism and the amount of conflict in the organization. At the same time, it is investigated whether this relationship is moderated by organizational size.

Organizational size (SIZE) was measured by the total number of employees in the organization, including working owners and part-time employees, as reported by the HR manager. The variable is dichotomized into two categories; one with firms having less than 10 employees (very small firms) and one with firms having 11-50 employees (small firms). This categorizations is adopted from the European Commission (2005), although they refer to firms from the first category as micro organizations. The level of absence (ABS) and industrial conflict (CON) was originally measured on a five point Likert scale ranging from very low to very high (Guest & Peccei, 2001). Since the scores reported by the HR managers were very skewed, the variables are dichotomized to organizations that have very low levels (Cat=1) and low to very high levels (Cat=2) of absenteeism or conflict.

At the individual level, the perception of work relationships were measured by three indicators: (1) satisfaction with the direct supervisor (SUP), (2) satisfaction with colleagues (COL), and (3) the perception of the degree in which the individual experience a family culture at work (FAM). These three indicators were originally measured with multiple items. However, to keep the illustration simple and as close as possible to Figure 3.3, the mean scale scores of each of the three scales is used as an indicator variable in the latent class analysis. Satisfaction with the direct supervisor was originally measured by nine items on a four point Likert scale ranging from never to always (Van Veldhoven, Meijman, & Broersen, 2002). An example item is: "Can you count on your supervisor when you come across difficulties in your work?". Satisfaction with colleagues was originally measured with the same four answer categories on six items (Van Veldhoven et al., 2002). An example item is: "If necessary, can you ask your colleagues for help?". The perception of a family culture at work was originally measured by three items on a

Table 3.2: Class Proportions and Class-Specific Probabilities Indirect Model

(a)

Class η	1	2
Class size	.53	.47
SUP = 1	.67	.01
SUP = 2	.30	.46
SUP = 3	.03	.53
COL = 1	.58	.05
COL = 2	.34	.41
COL = 3	.08	.55
FAM = 1	.40	.15
FAM = 2	.36	.32
FAM = 3	.24	.53

(b)

Class ζ	1	2	3	4	5
Class size	.17	.13	.20	.39	.10
$\eta = 1$	.81	.65	.53	.39	.18
$\eta = 2$	.19	.35	.47	.61	.82
SIZE = 1	.09	.10	.17	.56	.07
SIZE = 2	.28	.18	.23	.17	.14
ABS = 1	.00	1.00	.00	1.00	.00
ABS = 2	1.00	.00	1.00	.00	1.00
CON = 1	.00	.00	1.00	1.00	.00
CON = 2	1.00	1.00	.00	.00	1.00

five point scale ranging from totally disagree to totally agree (Goss, 1991). An example item is: "People here are like family to me".

The model can be formally described with seven multinomial logit models: (1) two for the group-level outcomes in which the main effect of ζ_j, the main effect of organizational size and their interaction effect are used as predictors, (2) one for the group-level latent variable ζ_j in which organizational size is used as a predictor, (3) one for the individual-level latent variable η_{ij} for which ζ_j is a predictor, and (4) three for the individual-level variables for which η_{ij} is a predictor. Furthermore, a two-variable association among the two firm-level outcomes is added to the model. The number of classes for the two latent variables are determined following the stepwise procedure of Lukočienė et al. (2010) using BIC based on the number of groups. This resulted in two classes at the individual level and five classes at the group level. The class separation of the latent variables is sufficient to good ($R^{\eta}_{entr} = .67$ and $R^{\zeta}_{entr} = .92$).

All effects were significant at the 5% level, except the main effect of organizational size and its interaction effect with ζ_j on both group-level outcomes. Therefore, these effects were removed from the model. The class proportions and class-specific response probabilities based on the final fitted model are given in Table 3.2. Table 3.2(a) shows that at the individual level, there is one class that contains 53% of the employees and these employees are not very satisfied with their relationships at work. The second class of individuals contains 47% of the employees that are satisfied with their relationships at work.

Table 3.2(b) provides the conditional probabilities of the discrete categories of the indicators given the discrete categories of the group-level latent variable ζ_j, and the conditional probabilities of the latent categories of ζ_j given the categories of the group-level predictor organizational size. From the first two rows of the table can be seen that at the group level, the five classes differ with respect to the composition of employees from the two individual-level classes. The group-level latent classes are ordered from the lowest probability of an employee belonging to the satisfied individual-level class (.19) through the highest (.82). The first and second group-level classes contain firms with employees from the unsatisfied individual-level classes (.81 and .65, respectively). The class sizes are 17% and 13%. The fourth and fifth group-level classes contain firms that have the highest probability of employees from the satisfied individual-level class (.61 and

.82, respectively). These classes contain 39% and 10% of the firms. The remaining 20% of the firms belong to the third group-level class. In this class a mixture of employees from the two individual-level classes is found.

In the third and fourth row of the table is shown that, the very small firms with maximum 10 employees (SIZE=1), have the highest probability to belong to the fourth group-level class (.56) and the small firms with 11-50 employees (SIZE=2) have the highest probability to belong to the first group-level class (.28). The very small firms have a higher probability to belong to the fourth class than the small firms, though for the remaining four classes it is the other way around. The fourth group-level class contains firms with very low probabilities of absenteeism (.00) and conflict (.00). The second and third group-level classes have, respectively, high probabilities on either absenteeism (1.00) or conflict (1.00). The first and fifth group-level classes have high probabilities to encounter both (1.00 and 1.00).

To conclude, at the individual level, the assumption that working at a firm with less than 50 employees is either fantastic or gruesome is supported, since the two individual-level classes could be interpreted as a satisfied and an unsatisfied class of employees. At the group level the situation is more complex. Although about half of the organizations contain mostly employees from the satisfied individual-level class, these organization belong either to a group-level class that encounters low or high levels of absenteeism and conflict. Thus at the group level, there is no clear positive effect of having satisfied employees on organizational levels of absenteeism and conflict. Organizational size matters in this context, since very small firms have a higher probability to belong to the group-level class without a lot of absenteeism and conflict than small firms.

3.6 Discussion

In the current article, two latent class models, referred to as the Direct model and the Indirect model, are presented that can be used to predict a group-level outcome by means of multiple individual-level variables by extending an existing method for micro-macro analysis with a single individual-level variable to the multivariate case. Both models involve the construction of a group-level latent class variable based on the individual-level variables to summarize the individual-level information at the group level. The group-level latent variable can then be related to other group-level variables, such as a group-level outcome. In the Direct model, the group-level latent classes affect the individual-level variables directly, while in the Indirect model these are affected indirectly via an individual-level latent variable. The Direct model seems most appropriate when the aim of the research is to construct a typology of groups that affect one or more group-level outcomes. In this situation the within and between component of the individual-level variables are independent. The Indirect model seems more appropriate when the individual-level variables are intended to measure an individual-level construct and groups are allowed to differ on the individual-level variable. The within and between component of the individual-level variables are now dependent. Both methods are applied to real data examples.

In the models with a discrete latent variable at each level, the number of classes of the latent variables had to be decided simultaneously since the full model was estimated at once. Although Lukočienė et al. (2010) provided guidelines on how to make this decision, further research should be devoted to study whether their approach is also optimal in the current context. Especially when the latent variables are dependent, one might prefer to determine the number of latent classes of the two variables independently. A stepwise procedure to do this without introducing bias in the group-level parameter estimates, is presented in Bolck, Croon, and Hagenaars (2004), Vermunt (2010), and Bakk et al. (2013). A further limitation of the current method is that the group-level outcome functions as an additional indicator of the latent group-level variable. This implies that the formation of the group-level classes is affected by the outcome variable. This may be counter intuitive since the latent variable is intended to predict the outcome. An additional advantage of using the stepwise procedure just referred to, is that the latent classes can not only be defined independent of each other, but also independent of the group-level outcome.

CHAPTER 4

Stepwise Micro-Macro Analysis

Abstract

Explaining group-level outcomes from individual-level predictors requires aggregating the observed scores on these predictors to the group level and accounting for the measurement error in the aggregated scores to prevent biased estimates. However, it is not clear yet how to perform the aggregation and the correction for measurement error when the individual-level predictors are discrete variables. It is shown how to overcome this problem by a stepwise latent class analysis. In the first step, a latent class model is estimated in which the scores on a discrete individual-level predictor are used as indicators for a group-level latent class variable. In the second step, this latent class model is used to aggregate the individual-level predictor to the group level by assigning the groups to the latent classes. In the final step, a group-level analysis is performed in which the aggregated measures are related to the remaining group-level variables while correcting for the measurement error in the class assignments. The proposed stepwise model is compared to existing methods in a simulation study and extended to a situation with multiple group-level latent variables. Finally, the approach is applied to an empirical data example.

This chapter is submitted for publication as Bennink, M., Croon, M. A., & Vermunt, J. K. (2014). *Stepwise latent class models for explaining group-level outcomes using discrete individual-level predictors.*

4.1 Introduction

Though typically multilevel models attempt to explain an individual-level dependent variable by means of individual- and group-level predictors (Goldstein, 2011; Hox, 2010; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012), the prediction of group-level outcomes is equally important. For example, a health psychologist might be interested in whether students' attitudes (micro-level predictor) affect teacher's stress (macro-level outcome), or a developmental psychologist wishes to investigate whether parenting practices (macro-level predictor) affect their children's study habits (micro-level mediator), which in turn affects children's achievement (micro-level mediator), which subsequently affects parental stress (macro-level outcome) (Bovaird & Shaw, 2012). Snijders and Bosker (2012) refer to this type of multilevel analysis as micro-macro analysis since a micro (or individual)-level predictor is assumed to affect a macro (or group)-level outcome. These micro-macro relationships cannot be addressed within the mainstream multilevel framework (Preacher et al., 2010).

Traditionally, data from micro-macro designs are analyzed by aggregation; that is, the group means of the individual-level variables are assigned to the groups and subsequently a group-level analysis is performed using the aggregated individual scores and group-level variables. Note that this is in fact a stepwise procedure since the aggregation (measurement model) is separated from the group-level analysis (structural model). An example from group-performance research is provided by van Veldhoven (2005), who studied the relationships between perceived human resource practices, work climate, and job stress on the one hand, and prospective and retrospective financial performance on the other hand. Because the financial performance indicators are only available at the business level, individual survey scores were aggregated to mean scores to perform a single-level analysis at the business level.

Although this seems intuitive, the above aggregation approach has various serious drawbacks. One of them is that it is implicitly assumed that the group members provide perfect information about their group, while in practice it is more realistic to expect that the group means contain measurement error. Croon and van Veldhoven (2007) showed that ignoring this measurement error causes a bias in the estimates of the parameters from the structural model. Another problem of the aggregation approach is that it is unclear how to aggregate categorical predictors to the group level. For example, for nominal variables with more than two categories, the group mean has no substantive meaning. It might be more appropriate to use the group modes instead, but even then the problem of ignoring measurement error remains.

To overcome the measurement error issue, Croon and van Veldhoven (2007) proposed using a latent variable model for two-level data in which the individual-level responses serve as indicators for a continuous latent variable which in turn is used as a predictor of the group-level outcome variable (Croon & van Veldhoven, 2007). In this way, the multilevel structure of the data is correctly accounted for and the measurement error of the group-level scores is incorporated in the model. Bennink et al. (2013) extended this latent variable approach to the situation in which the individual predictors are categorical variables. They proposed performing micro-macro multilevel analysis with categorical variables using latent class models.

The most important downside of the latent variable approaches is that these require the measurement and structural part of the corresponding latent variable model to be estimated simultaneously. This is less intuitive and less practical than the traditional procedures in which the aggregation and the group-level analysis are separate steps. In the current article, we demonstrate how to bridge the gap between the traditional and latent variable approaches by showing how to use the latent class model for discrete micro-macro relations in a stepwise manner. More specifically, in the first step, a latent class model is estimated in which the scores on the discrete individual-level predictor are used as indicators for a group-level latent class variable (measurement model). In the second step, this latent class model is used to aggregate the individual-level predictor to the group level by assigning the groups to the latent classes. In the final step, a group-level analysis is performed in which the aggregated measures are related to the remaining group-level variables (structural model), while adjusting for the measurement errors in the class assignments. The latter adjustments, which are based on earlier work by Bakk et al. (2013), Bolck et al. (2004), and Vermunt (2010), are applied to the current research context and tested in a simulation study.

The organization of the article is as follows. First, the latent class model for discrete micro-macro analysis is introduced using a model that contains a single individual-level variable. Second, it is shown how this model can be applied in a stepwise manner. Third, a simulation study is presented to evaluate the performance of the proposed stepwise procedure. Fourth, the stepwise procedure is applied to a more complex model with two individual-level variables and applied to a real data example in which enriched job design (macro-level predictor) affects team productivity (macro-level outcome) directly and indirectly through job control (micro-level mediator) and job satisfaction (micro-level mediator).

4.2 Micro-Macro Latent Class Model

To illustrate the multilevel latent class model for micro-macro analysis, let us consider a simple model that contains a single group-level outcome Y_j, a single group-level predictor X_j, and a single individual-level predictor Z_{ij}. Subscript j is used to denote a particular group and subscript i to denote the individuals within a group. The group-level predictor is expected to affect the group-level outcome directly and indirectly via the individual-level predictor. Any theory in which a group-level intervention is not only expected to influence a group-level (performance) measure directly, but also indirectly through a characteristic of the group members, can be tested with this model. These kinds of models are sometimes referred to as 2-1-2 models since the effect of the level-2 independent variable on the level-2 dependent variable is mediated by a level-1 variable. Also, the term 'bathtub model' is in use here because of the shape of the conceptual model that is shown in Figure 4.1. In this conceptual model, the latent variable is presented in a circle and the manifest variables in rectangles.

The model of interest is a latent class model for two-level data in which the scores of the individual-level units i within group j on the micro-level predictor Z_{ij} are treated as indicators of a discrete latent class variable defined at the group level, ζ_j. Thus, the number of indicators of the latent variable equals the number of individuals within a group. This part of the model is referred to as the measurement part. All group members are treated as equivalent sources of information about the group-level variable; therefore, no one is considered as providing more accurate judgments in this respect than his co-members. This implies that the relationship between the individual-level variable and the

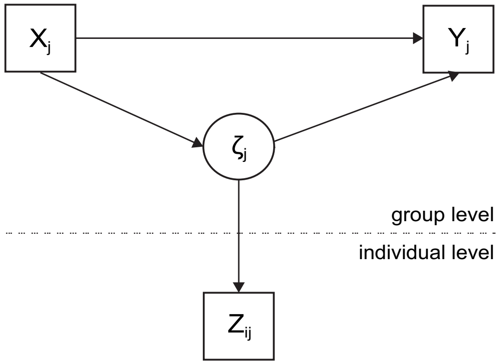


Figure 4.1: Micro-Macro Latent Variable Model with One Individual-Level Predictor

group-level latent variable can be assumed to be the same for all individuals within a group. According to the local independence assumption commonly made in latent class analysis, the individual responses of group members are independent given the score of their group on the latent variable. In the structural part of the model, the group-level latent classes are related to the group-level independent variable X_j and the group-level dependent variable Y_j.

For the general case where all variables in the latent class model are discrete, the model can be formally described with three multi-category logit models (Agresti, 2013). Let there be P, Q, R, and S nominal response categories for, respectively, Z_{ij}, ζ_j, X_j, and Y_j and a particular category is denoted by p, q, r, and s. Then there are $P-1$, $Q-1$, and $S-1$ different logit equations defined for, respectively, Z_{ij}, ζ_j, and Y_j in which each category is compared to an arbitrary baseline category. When the first categories are used as baselines, the multinomial logit equations are:

$$\log\left(\frac{P(Z_{ij}=p|\zeta_j=q)}{P(Z_{ij}=1|\zeta_j=q)}\right)=\beta_p^{Z}+\beta_{p\,q}^{Z\,\zeta}, \tag{4.1}$$

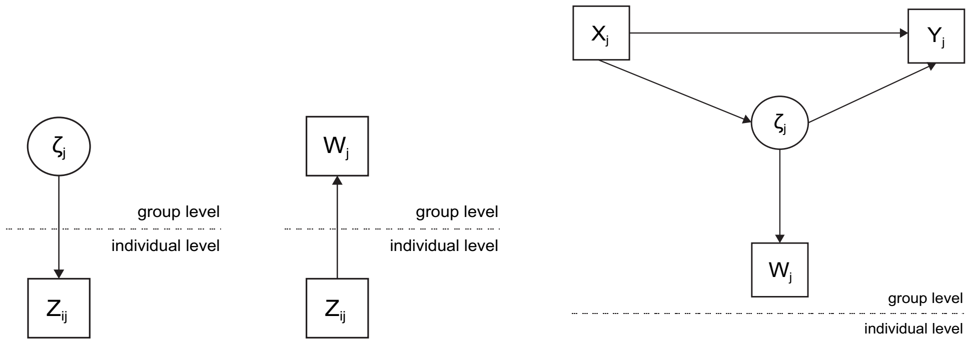
$$\log\left(\frac{P(\zeta_j=q|X_j=r)}{P(\zeta_j=1|X_j=r)}\right)=\beta_q^{\zeta}+\beta_{q\,r}^{\zeta\,X}, \tag{4.2}$$

and

$$\log\left(\frac{P(Y_j=s|\zeta_j=q,X_j=r)}{P(Y_j=1|\zeta_j=q,X_j=r)}\right)=\beta_s^{Y}+\beta_{s\,q}^{Y\,\zeta}+\beta_{s\,r}^{Y\,X}. \tag{4.3}$$

Each equation contains an intercept term (β_p^{Z}, β_q^{ζ}, and β_s^{Y}) and an effect for each of the predictor variables ($\beta_{p\,q}^{Z\,\zeta}$, $\beta_{q\,r}^{\zeta\,X}$, $\beta_{s\,q}^{Y\,\zeta}$, $\beta_{s\,r}^{Y\,X}$).

Equation 4.1 describes the measurement part of the model in which the scores of the individual-level units on the micro-level predictor Z_{ij} are treated as exchangeable indicators of a discrete latent class variable defined at the group level, ζ_j. The structural part of the model is defined in Equations 4.2 and 4.3 in which the group-level latent variable is



(a) First step (b) Second step (c) Third step

Figure 4.2: Graphical Representation of Stepwise Procedure

related to the other group-level variables: ζ_j is regressed on X_j, and the outcome Y_j is regressed on ζ_j and X_j.

The parameters of this latent class model can simultaneously be estimated by full information maximum likelihood estimation and, therefore, this method is further referred to as the one-step approach. This approach mainly has two drawbacks. First, it is less intuitive compared to a stepwise approach that separates the aggregation and group-level analysis. Second, the definition of the latent group-level variable is not only determined by the micro-level indicators, but also by the remaining variables in the structural model. Thus, when the structural part of the model is adapted, say a level-2 covariate or outcome is added or removed, the full model has to be re-estimated and the measurement model may thus change. Especially, the fact that the meaning and possibly also the number of the latent classes is affected by the outcome variable is very undesirable since the latent classes were theoretically intended to predict this outcome. These problems associated with the simultaneous estimation of the model are circumvented with a stepwise approach.

4.3 Stepwise Estimation

A stepwise estimation procedure of the micro-macro latent class model consists of the following three steps:

- First step: Estimate the measurement model (i.e., relate the micro-level predictor to the latent group-level variable).
- Second step: Aggregate the micro-level predictor to the group level by assigning groups to latent classes.
- Third step: Estimate the structural part of the model while correcting for the classification errors that were made in the second step.

So, as graphically illustrated in Figure 4.2(a), a measurement model is defined for ζ_j and the corresponding latent class model is estimated in the first step of the analysis:

$$P(\mathbf{Z}_j) = \sum_{q=1}^{Q} P(\zeta_j = q) \prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j = q). \tag{4.4}$$

Here, the vector $\mathbf{Z}_j$ contains the I_j responses Z_{ij} of the members of group j. The model parameters are the class proportions $P(\zeta_j = q)$ and the conditional response probabilities $P(Z_{ij}|\zeta_j = q)$, which, as shown in Equation 4.1, are typically parameterized using a logit equation. As in the one-step model, the responses of the individual group members on Z_{ij} are assumed to be exchangeable and independent given the latent classes, but the meaning of the latent classes is now only determined by the individual-level scores on the micro-level predictor and no longer by the scores on the group-level variables. The number of latent classes of ζ_j is also determined during this step.

In the second step, the groups are assigned to the Q latent classes based on their scores $\mathbf{Z}_j$. We denoted the assigned class for group j by W_j. The assignment process in which the new variable W_j is constructed is graphically illustrated in Figure 4.2(b). As in a standard latent class analysis, the class assignments are obtained using the posterior class membership probabilities $P(\zeta_j = q|\mathbf{Z}_j)$ from the first step. Several types of deterministic and probabilistic assignment rules have been proposed. In the current article, we focus on modal and proportional assignment. With modal assignment, each group is assigned to the latent class for which the posterior probability is largest. With proportional assignment, a group is assigned to each of the Q classes with a probability equal to the posterior membership probability for the class concerned.

Unless the micro-level predictor is a perfect indicator for the group-level latent class variable, classification errors will be made during the assignment. In order to account for these classification errors (in step three), we use the $Q \times Q$ classification table with entries $P(W_j = t|\zeta_j = q)$. Note that $P(W_j = t|\zeta_j = q)$ is the conditional probability that a group belonging to class q is assigned to class t of W_j ($t = 1, \cdots Q$). The off-diagonals of this table represent classification error probabilities. In Appendix C, we show how these probabilities can be obtained from the latent class model parameters.

In the third step, the structural model is estimated; that is, the assigned scores W_j are related to the other group-level variables, in this case X_j and Y_j. As shown by Bolck et al. (2004) biases are caused by the classification errors introduced in the second step; that is, by the fact that we have W_j instead of ζ_j. However, they also indicated that it is possible to adjust for the classification errors, which prevent biases. Key for their adjustment method is the following relationship between $P(Y_j, W_j|X_j)$ and $P(Y_j, \zeta_j|X_j)$:

$$\begin{aligned} P(Y_j, W_j|X_j) &= \sum_{q=1}^{Q} P(Y_j, \zeta_j = q|X_j) P(W_j|\zeta_j = q) \\ &= \sum_{q=1}^{Q} P(\zeta_j = q|X_j) P(Y_j|X_j, \zeta_j = q) P(W_j|\zeta_j = q). \end{aligned} \tag{4.5}$$

This equation shows that a model for $P(Y_j, \zeta_j|X_j)$ is obtained by estimating a model for $P(Y_j, W_j|X_j)$ and correcting for the probabilities $P(W_j|\zeta_j = q)$. Note that the $P(W_j|\zeta_j = q)$ were computed in step two. The resulting model is shown graphically in Figure 4.2(c).

The model defined in Equation 4.5 can be estimated by maximum likelihood (ML). This involves performing a latent class analysis in which W_j is used as a single indicator.

The probabilities $P(\zeta_j = q|X_j)$ and $P(Y_j|X_j, \zeta_j = q)$ (see Equation 4.2 and 4.3) are freely estimated and $P(W_j|\zeta_j = q)$ is treated as known and thus fixed. We refer to this approach as the ML three-step method (Bakk et al., 2013; Vermunt, 2010). An alternative proposed by Bolck et al. (2004) - and that we therefore call the BCH three-step approach - involves reformulating the problem into a weighted analysis (see also Vermunt (2010)). More specifically, by weighting the data points by the inverse of the classification probabilities $P(W_j|\zeta_j)$, we adjust for the fact that the W_j contain classification errors. The reweighted data can be used as observed data to estimate the structural parameters of interest.

From previous research on three-step latent class analysis, it is known that the bias adjusted stepwise approaches work very well in situations encountered in practice. However, this approach may fail when a small sample size is combined with a very large proportion of classification errors, where the latter can also be quantified as class separation (Bakk et al., 2013; Vermunt, 2010). In such situations, the maximum-likelihood estimates of the first step latent class model will tend to yield classes being more different than they truly are (Galindo-Garre & Vermunt, 2006). Consequently, the amount of classification errors is underestimated and the structural parameters are not adjusted to a sufficient degree.

The proportion of classification errors and the class separation is mainly a function of the number of indicator variables (in micro-macro analysis, this equals the number of individuals within each group), the number of classes, and the response probabilities for the most likely response. Since all group members are assumed to be exchangeable, the response probabilities for the most likely response are the same for individuals within the same group, which makes the measurement model of a micro-macro model more parsimonious than a regular latent class analysis. In the following simulation study, we investigate under which conditions class separation is large enough to perform an unbiased stepwise analysis for the current model.

4.4 Simulation Study

In this section, the performance of the stepwise approaches is first evaluated and compared to manifest aggregation with a group mode, one-step latent aggregation, and stepwise latent aggregation without correcting for measurement error. Second, the lower boundary of the separation between classes is explored at which still unbiased results are obtained with the stepwise approaches. All analyses were carried out with Latent GOLD 5.0 (Vermunt & Magidson, 2013).

Data are generated according to the model shown in Figure 4.1 with all dichotomous variables. An average situation is created by fixing the between-group effects from the structural part of the model to .40 on a logistic scale using effect coding ($\beta^{YX} = \beta^{\zeta X} = \beta^{Y\zeta} = .4$). The number of groups was fixed to 100 ($J = 100$) and the number of individuals within a group to 10 ($n_j = 10$) which are minimum sample sizes for this type of analysis (Bennink et al., 2013). It is expected that larger samples provide slightly better results. The relationship between the scores on the micro-level predictor and the latent variable is varied from weak to strong, again using effect coding ($\beta^{Z\zeta}$ =.20, .40, .60, or .80). This corresponds to class separations, measured with the entropy based R-square, of $R^2_{entr} = .24, .64, .88$, and .97. For each of the four conditions, 500 datasets were generated and analyzed with manifest mode aggregation, the latent variable one-step

Table 4.1: Estimates Between Effects Simple Micro-Macro Model

	Mode	One-step	Three-step					
			None		BCH		ML	
			Modal	Prop	Modal	Prop	Modal	Prop
	Estimates of $\beta^{YX}(SD)^1$							
True $\beta^{Z\zeta}$ =.20	.49(.12)	.41(.19)	.49(.12)	.50(.11)	.45(.17)	.43(.18)	.45(.17)	.45(.13)
True $\beta^{Z\zeta}$ =.40	.46(.12)	.41(.14)	.46(.12)	.47(.12)	.42(.14)	.41(.14)	.42(.14)	.42(.14)
True $\beta^{Z\zeta}$ =.60	.42(.12)	.40(.13)	.42(.12)	.42(.12)	.41(.13)	.41(.13)	.41(.13)	.41(.13)
True $\beta^{Z\zeta}$ =.80	.41(.13)	.41(.14)	.41(.13)	.41(.13)	.41(.14)	.41(.14)	.41(.14)	.41(.14)
	Estimates of $\beta^{\zeta X}(SD)^1$							
True $\beta^{Z\zeta}$ =.20	.18(.10)	.45(.25)	.19(.14)	.12(.08)	.36(.28)	.42(.26)	.37(.26)	.34(.18)
True $\beta^{Z\zeta}$ =.40	.31(.11)	.42(.14)	.32(.11)	.28(.09)	.41(.15)	.41(.14)	.41(.16)	.41(.14)
True $\beta^{Z\zeta}$ =.60	.40(.12)	.40(.12)	.37(.12)	.36(.11)	.40(.13)	.40(.12)	.40(.12)	.40(.12)
True $\beta^{Z\zeta}$ =.80	.40(.11)	.41(.11)	.40(.11)	.40(.11)	.41(.11)	.41(.11)	.41(.11)	.41(.11)
	Estimates of $\beta^{\zeta Y}(SD)^1$							
True $\beta^{Z\zeta}$ =.20	.17(.12)	.50(.33)	.19(.16)	.12(.09)	.38(.34)	.43(.30)	.40(.33)	.34(.21)
True $\beta^{Z\zeta}$ =.40	.30(.13)	.42(.17)	.31(.13)	.27(.11)	.42(.19)	.42(.18)	.41(.18)	.41(.16)
True $\beta^{Z\zeta}$ =.60	.37(.12)	.41(.13)	.38(.12)	.36(.11)	.41(.13)	.41(.13)	.41(.13)	.41(.13)
True $\beta^{Z\zeta}$ =.80	.39(.13)	.40(.13)	.39(.13)	.39(.12)	.40(.13)	.40(.13)	.40(.13)	.40(.13)

$J = 100$, $n_j = 10$, True value between-group effects $= .40$
[1] The estimates are averaged over 500 replications
ζ should be replaced by the manifest group mode of Z in case of mode aggregation

approach, and the latent variable three-step approaches. The three-step procedures were applied with both modal and proportional assignment.[6]

The average estimated structural parameters in each condition are presented in Table 4.1 and should be compared with their true value of .40. As expected, the one-step and the corrected three-step procedures provide unbiased results, regardless whether the modal or proportional assignment rule was used. Only when the quality of the indicators is extremely poor ($\beta^{Z\zeta} = .20$), both methods perform less well. This occurs, as shown below, in the situation in which classes are estimated as being more different than they truly are. Furthermore, as can be seen, the standard deviations of the estimates decrease when the quality of the indicators increases.

The other methods fail. When the uncorrected three-step method is used, the estimates of the group-level relationships in which ζ_j is involved are underestimated in line with Bolck et al. (2004), regardless whether the modal or proportional assignment rule was used. Because the indirect effect is underestimated, the direct effect is overestimated. The bias decreases when the strength of the relationship between Z_{ij} and ζ_j increases. When Z_{ij} is aggregated to the group level using the manifest group mode, the estimates of the between-group parameters are biased and this bias decreases when the quality of the indicators improves. In line with previous research (Bennink et al., 2013), the parameter estimates are only unbiased when the strength of the relationship between the individual-level predictor and the corresponding group-level variable is very good ($\beta^{Z\zeta} = .80$). The standard deviations of the estimates obtained with mode aggregation and the uncorrected

[6] In the one-step and the three-step ML methods, we used weakly informative priors for the model probabilities to prevent boundary estimates for the logit parameters. Because this does not work in the three-step BCH method, 82 datasets with the modal assignment rule and 73 datasets with the proportional assignment rule did not converge and were excluded from the further analysis.

Table 4.2: True and Estimated Proportion of Classification Errors

		Modal		Proportional	
R^2_{entr}	True $\beta^{Z\zeta}$	True	Estimated $(SD)^1$	True	Estimated $(SD)^1$
.24	.20	.27	.21(.08)	.35	.29(.09)
.35	.25	.22	.18(.05)	.30	.25(.06)
.45	.30	.18	.15(.03)	.24	.22(.04)
.55	.35	.14	.13(.03)	.19	.18(.04)
.64	.40	.11	.10(.02)	.15	.15(.03)
.88	.60	.04	.03(.01)	.05	.05(.01)
.97	.80	.01	.01(.00)	.01	.01(.01)

J=100, $n_j = 100$, True value between-group effects $= .40$
[1]The estimates are averaged over 500 replications

three-step procedure are stable over the true quality of the indicators.

As shown above, the corrected three-step approaches perform less well with poor indicators, the situation corresponding with an extremely low class separation of .24.[7] To illustrate why this occurs, the true and estimated proportion of classification errors are shown in Table 4.2 for each condition. As can be seen, the proportion of classification errors is underestimated in the condition in which $R^2_{entr} = .24$ resulting in third-step parameters which are not sufficiently corrected. To explore what would be a sufficiently high class separation, also indicators with $\beta^{Z\zeta}$ values of .25, .30, and .35, corresponding to R^2_{entr} values of .35, .45, and .55, are added to the table. As can be seen, with a class separation of .45, the estimated proportion of classification errors gets close to the actual proportion. This applies to both modal and proportional assignment. It can also be seen, that the variability (the standard deviation) of the estimated proportion of classification error decreases when the true quality of the indicators increases.

To conclude, the adjusted three-step approaches provide estimates that are as good as the estimates from a one-step analysis, as long as the parameters as the class separation is sufficient ($R^2_{entr} = .45$). It does not matter whether a BCH or ML correction is used or whether modal or proportional assignment is used. When the relationship between the micro-level predictor and the group-level variable is very strong, all methods provide unbiased estimates, but this is not a realistic situation in practice. When the relationship between the micro-level predictor and the group-level variable is moderate, a latent variable should be used for the aggregation since the manifest mode aggregation provides biased estimates. When the relationship between the micro-level predictor and the group-level variable is extremely weak, all methods may provide biased estimates for the group-level sample size investigated in the current simulation study.

4.5 Multiple Macro-Level Latent Variables

Since stepwise modelling could be especially useful in large and complex models, the simple model discussed so far is extended to a more complex model. As shown in Figure 4.3, this

[7]As stated before, class separation is not only a function of the quality of the indicators but also of the number of indicators which in our application equals the number of individuals within a group. For example, an indicator with a $\beta^{Z\zeta}$-value of .20 yields a $R^2_{entr} = .42$ when $n_j = 20$ and $R^2_{entr} = .55$ when $n_j = 30$.

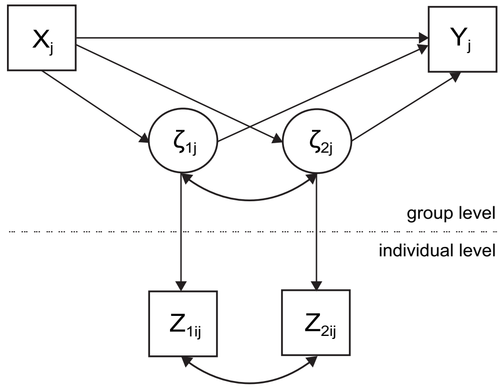


Figure 4.3: Micro-Macro Latent Variable Model with Multiple Macro-Level Latent Variables

model contains two micro-level predictors, Z_{1ij} and Z_{2ij}, and thus two latent variables, ζ_{1j} and ζ_{2j}. The corresponding multinomial logit equations for Z_{1ij} and Z_{2ij} are the same as described in Equation 4.1, and the ones for ζ_{1j} and ζ_{2j} are the same as described in Equation 4.2.

When X_j, ζ_{1j} ζ_{2j}, and Y_j contain, respectively, R, W, Q, and S categories and a particular category is denoted by r, w, q, and s, the multinomial logit equation for Y_j is:

$$\log\left(\frac{P(Y_j = s|\zeta_{1j} = w, \zeta_{2j} = q, X_j = r)}{P(Y_j = 1|\zeta_{1j} = w, \zeta_{2j} = q, X_j = r)}\right) = \beta_s^Y + \beta_{s\ w}^{Y\zeta_1} + \beta_{s\ q}^{Y\zeta_2} + \beta_{s\ r}^{YX}. \tag{4.6}$$

The first categories are used as baseline categories, β_s^Y is the intercept of the response variable, and $\beta_{s\ w}^{Y\zeta_1}$, $\beta_{s\ q}^{Y\zeta_2}$, and $\beta_{s\ r}^{YX}$ are the regression parameters of the predictor variables.

When the two individual-level predictors would be continuous variables, it would be common practice to include both their between- and within-group correlation in the model. In the case of discrete variables, this concept is translated by incorporating an association between Z_{1ij} and Z_{2ij} ($a_{Z_1Z_2}$), and between ζ_{1j} and ζ_{2j} ($a_{\zeta_1\zeta_2}$). Thus, Z_{1ij} and Z_{2ij} are not expected to be independent given the latent group-level variables. Instead, while keeping the latent group-level variables constant, there may still be some residual within-group association between the micro-level predictors. At the between level, it is also expected that there is some residual association between ζ_{1j} and ζ_{2j}, after controlling for X_j.

While estimating this model in a single step is straightforward, when estimating it stepwise, it has to be decided how to define the first-step model(s). The first option would be to formulate a separate measurement model for ζ_{1j} and ζ_{2j} as described in Equation 4.4. By formulating two measurement models, the meaning of the latent classes is only influenced by the individual-level scores on the corresponding micro-level predictors. The number of latent classes for ζ_{1j} and ζ_{2j} can be determined without being influenced by the variables from the structural part of the model, but the eventual residual within-group association among Z_{1ij} and Z_{2ij} is ignored.

An alternative would be to formulate a single simultaneous measurement model for the two latent variables which includes the residual within-group association between Z_{1ij}

and Z_{2ij}:

$$P(\mathbf{Z}_{1j}, \mathbf{Z}_{2j}) = \sum_{w=1}^{W} \sum_{q=1}^{Q} P(\zeta_{1j} = w, \zeta_{2j} = q) \prod_{i=1}^{I_j} P(Z_{1ij}, Z_{2ij} | \zeta_{1j} = w, \zeta_{2j} = q). \quad (4.7)$$

By doing so, the meaning of the latent classes is still not influenced by the group-level variables from the structural part of the model, but the number of latent classes for ζ_{1j} and ζ_{2j} should be determined simultaneously.

An analysis is carried out to explore whether the misspecification of the measurement model arising from ignoring the residual within-group association among Z_{1ij} and Z_{2ij} affects the estimates of the between-level parameters. Since sampling fluctuation is not of primary interest here, one very large data set $(J = 10000, I_j = 100)$ is generated in each of the investigated conditions. If it turns out that ignoring the within-group association provides biased estimates in such very large samples, the estimates in smaller samples can be expected to be even worse because of sampling fluctuation.

The population models varied in the strength of the relationship between the latent variables and the corresponding micro-level predictors (indicators) and the strength of the within-group association among the micro-level predictors (within-group association):

- indicators ($\beta^{Z_1\zeta_1}$ and $\beta^{Z_2\zeta_2}$): .20, .40, or .60
- within-group association ($a_{Z_1Z_2}$): .00, .20, .40, or .60

Similar to the previous simulation study, all variables, including the latent ones, are dichotomous and the between-group effects from the structural part of the model are fixed to .40 on the logistic scale using effect coding ($\beta^{\zeta_1 X} = \beta^{\zeta_2 X} = a_{\zeta_1\zeta_2} = \beta^{Y\zeta_1} = \beta^{Y\zeta_2} = \beta^{YX} = .40$). Note that because of the large number of individuals within a group, the R^2_{entr} value will be very high (>.90) in all conditions.

The generated datasets were analyzed with the one-step and the bias adjusted three-step approaches. In the one-step procedure, we used both the correct specification containing the residual within-group association and the incorrect model excluding this association. In the first step of the BCH and ML bias adjusted stepwise procedures, we used either a single joint measurement model with the association between Z_{1ij} and Z_{2ij} or two separate measurement models which ignore this association. Both modal and proportional assignment rules were used to assign the groups to latent classes in the second step of the analysis. The manifest mode aggregation and uncorrected three-step procedures were not used because the previous simulation study showed that these methods already fail in a simpler model.

Table 4.3 presents the results for the conditions in which $a_{Z_1Z_2}$ is varied and $\beta^{Z_1\zeta_1}$ and $\beta^{Z_2\zeta_2}$ are fixed to .40. The reported results concern the between-level parameter which is most strongly affected by ignoring the within-group association; that is, the association between ζ_{j1} and ζ_{j2} ($a_{\zeta_1\zeta_2}$), which has a true value of .40. As can be seen, when the within-group association among Z_{1ij} and Z_{2ij} is correctly modeled, both the one-step and the bias adjusted three-step methods provide unbiased estimates. However, when this within-group association is not taken into account, the between association estimate is biased with all estimation procedures. The larger the value of the ignored within-group association among Z_{1ij} and Z_{2ij}, the more the between-group association among ζ_{1j} and ζ_{2j} is overestimated. Note that the estimates obtained with the one-step and the various types of bias adjusted three-step estimates are all very similar. The

Table 4.3: Estimates Between Association ζ_{1j} and ζ_{2j} ($a_{\zeta_1\zeta_2}$)

	One-step		BCH modal		BCH prop		ML modal		ML prop	
$a_{Z_1Z_2}$	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
.00	.40	.40	.39	.40	.40	.40	.39	.40	.40	.40
.20	.40	.44	.40	.43	.40	.44	.40	.43	.40	.44
.40	.40	.50	.40	.48	.40	.50	.40	.48	.40	.50
.60	.41	.60	.42	.56	.43	.59	.42	.56	.43	.59

$J = 10000$, $n_j = 100$, True value $a_{\zeta_1\zeta_2} = .40$
Yes = $a_{Z_1Z_2}$ is incorporated in measurement model
No = $a_{Z_1Z_2}$ is not incorporated in measurement model

Table 4.4: Estimates Between Effects when Within-Group Association is not Modelled

	One-step		BCH modal		BCH prop		ML modal		ML prop	
$\beta^{Z\zeta}$	.60	.20	.60	.20	.60	.20	.60	.20	.60	.20
$a_{Z_1Z_2}$	.20	.60	.20	.60	.20	.60	.20	.60	.20	.60
β^{ζ_1X}	.41	.34	.41	.35	.41	.35	.41	.35	.41	.35
β^{ζ_2X}	.39	.33	.39	.34	.39	.34	.39	.34	.39	.34
$a_{\zeta_1\zeta_2}$	.38	.91	.38	.81	.38	.88	.38	.81	.38	.87
$\beta^{Y\zeta_1}$	.40	.32	.40	.34	.40	.33	.40	.33	.40	.33
$\beta^{Y\zeta_2}$	.40	.30	.40	.32	.40	.31	.40	.32	.40	.31
β^{YX}	.44	.45	.44	.44	.44	.44	.44	.45	.44	.44

$J = 10000$, $n_j = 100$, True value between-group effects $= .40$

estimates of the remaining between-group effects are not as much biased as the between-group association. When there is bias in the remaining between-group parameters it is a downwards bias that probably compensates the overestimation of the between-group association among ζ_{1j} and ζ_{2j}.

Table 4.4 shows how the bias that is caused by ignoring the within-group association among the micro-level predictors interacts with the quality of the micro-level scores as indicators for the group-level latent variables. With bad indicators and a strong ignored within-group association ($\beta^{Z_1\zeta_1} = \beta^{Z_2\zeta_1} = .20$ and $a_{Z_1Z_2} = .60$), the estimates of the between-group association among ζ_{1j} and ζ_{2j} are very biased, while with good indicators and a small ignored within-group association ($\beta^{Z_1\zeta_1} = \beta^{Z_2\zeta_1} = .60$ and $a_{Z_1Z_2} = .20$) the estimates are still good.

Altogether, these results show that the bias adjusted three-step procedures can be used for this micro-macro model without introducing more bias compared to the one-step procedure, as long as the within-group association among the micro-level predictors is modeled in the first step. The within-group association can only be ignored when the micro-level scores are very good indicators of the group-level latent variables and the within-group association is small.

4.6 Data Example

The stepwise micro-macro model with two micro-level predictors is now applied to a real data example. Since the BCH and ML correction procedures and the modal and proportional assignment rules provided similar results in the simulation study, only one method is applied here, namely the ML three-step approach with modal assignment. The inspiration for the current data example comes from a paper by Croon, van Veldhoven, Peccei, and Wood (2014). These authors show the relevance of bathtub multilevel mediation models, such as the one discussed in this paper, for research on human resource management (more specifically on job design) and organizational performance by using an example from the Workplace Employment Relations Survey 2004 (WERS2004). This is a publicly available large-scale dataset from the United Kingdom with representative sampling at both the employee and the workplace level. More information about the survey can be found at www.wers2004.info. Croon et al. (2014) investigated to what extent the relationship between the adoption of enriched job designs at the level of the workplace, on the one hand, and workplace labor productivity, on the other hand, was mediated, at the individual level, by employees' experienced sense of job control and job satisfaction. For the current application, all measures from Croon et al. (2014), *enriched job design*, *job control*, and *job satisfaction*, were categorized into variables with three categories of approximately equal size (low, medium, and high), while *labor productivity* was transformed into a variable with two approximately equally sized categories (low and high). To keep the discussion simple, the current illustration ignores that the variables were originally measured with multiple items. The analyses were performed with Latent GOLD 5.0 (Vermunt & Magidson, 2013) on 18,505 employees nested within 1,455 workplaces. More information about the syntax used in this analysis is given in Appendix D. Also, the syntax for performing a very similar analysis in Mplus (Muthén & Muthén, 1998-2012) is provided in Appendix E.

In the first step of the procedure, a measurement model is formulated in which the individual-level scores on *job control* (Z_{1ij}) and *job satisfaction* (Z_{2ij}) were used as exchangeable indicators for two latent variables at the group level (ζ_{1j} and ζ_{2j}) with a residual within-group association ($a_{Z_1Z_2}$) among the individual-level measures on *job control* and *job satisfaction* included as well. The optimal number of latent classes for the two latent variables was determined simultaneously by comparing fit indices of models with all possible combinations of one to five classes for each latent variable. Because the decision is about the number of classes at the group level, the fit indices that incorporate the sample size are based on the number of groups (Lukočienė et al., 2010). BIC, AIC3, CAIC, and SABIC were lowest for the model with three latent classes for both *job control* and *job satisfaction*. Contradictorily, AIC was lowest when both latent variables contained four latent classes. Since most fit indices point towards this direction and the individual-level observed variables had three response categories, the three-class solutions for both group-level latent variables were retained.

Table 4.5 displays the class sizes and the class-specific response probabilities for both latent variables. These can be used to interpret the latent classes. Table 4.5(a) shows that the first latent variable has a class that contains 35% of the groups, and these workplaces contain mostly employees with a high probability of scoring low ($p = .70$) on *job control*. There is a second latent class that contains 56% of the groups and in these workplaces employees have the highest probability to score medium on *job control* ($p = .84$). The

Table 4.5: Class Sizes and Class-Specific Response Probabilities Measurement Model First Step

(a) Job control

	Group-level latent classes 1	2	3	Overall
Size	.35	.56	.09	
Job control				
Low	.70	.13	.01	.32
Medium	.30	.85	.66	.64
High	.00	.02	.33	.04

(b) Job satisfaction

	Group-level latent classes 1	2	3	Overall
Size	.61	.23	.16	
Job satisfaction				
Low	.88	.38	.34	.69
Medium	.08	.62	.00	.19
High	.04	.00	.66	.12

final class of workplaces contains 9% of groups that contain individuals that, like the ones from the second latent class, have the highest probability to score medium ($p = .66$), but they have a larger probability to score high ($p = .33$) on *job control*. From Table 4.5(b) it can be seen that the second latent variable has a class that contains 61% of the groups and these workplaces have the highest probability to score low on *job satisfaction* ($p = .88$). The second class contains 23% of the workplaces with employees having the highest probabilities to score medium ($p = .62$) on *job satisfaction*. The last class contains 16% of the workplaces in which employees have the highest probability to score high on *job satisfaction*. The groups from the third class also have quite a high probability to score low ($p = .34$), but this is probably caused by the fact that overall most groups score low on *job satisfaction* ($p = .69$). Based on these posterior probabilities, all workplaces are assigned to a particular latent class using a modal assignment rule. The class separation was .39 for the latent variable *job control* and .43 for *job satisfaction* which is relatively low for a three-step analysis.

In the third step, the assigned latent class variables were related to the group-level measures of *enriched job design* (X_j) and *labor productivity* (Y_j) in a latent class model in which the assigned class membership scores from the second step were used as single indicators with known measurement errors, namely the classification errors from the second step. The group-level parameters obtained with the ML bias adjusted three-step approach are presented in Table 4.6, from which the estimates of the parameters for the main effects of the response variables are omitted. For all effects, dummy coding was used with the first categories as reference categories. Only the significance of the global effects are reported since the significance of the category-specific parameters depends on coding.

The overall effect of *enriched job design* on ζ_{1j} (job control) is significant ($\chi^2 = 23.71,\ df = 4,\ p < .001$) and the category-specific parameters are presented in Table 4.6(a). All category-specific parameters are positive implying that the reference category scores lower than the other categories. Both $\beta_{2\ 3}^{\zeta_1 X}$ and $\beta_{3\ 3}^{\zeta_1 X}$ differ more from the reference category than $\beta_{2\ 2}^{\zeta_1 X}$ and $\beta_{3\ 2}^{\zeta_1 X}$. The overall effect of *enriched job design* on ζ_{2j} (job satisfaction) is also significant ($\chi^2 = 29.82,\ df = 4,\ p < .001$). The category-specific parameters from Table 4.6(b) show that the $\beta_{2\ 2}^{\zeta_2 X}$ and $\beta_{2\ 3}^{\zeta_2 X}$ are positive while $\beta_{3\ 2}^{\zeta_2 X}$ and $\beta_{3\ 3}^{\zeta_2 X}$ are negative. Hence, the first two categories score higher than the reference group and the latter two score lower than the reference group.

The overall association among ζ_{1j} and ζ_{2j} is significant ($\chi^2 = 13.92,\ df = 4,\ p =$

Table 4.6: Bias-Adjusted ML Parameters Structural Model

(a)

$\beta_{l\ q}^{\zeta_1 X}$	b	se
$\beta_{2\ 2}^{\zeta_1 X}$	0.18	0.29
$\beta_{3\ 2}^{\zeta_1 X}$	1.54	0.57
$\beta_{2\ 3}^{\zeta_1 X}$	0.98	0.31
$\beta_{3\ 3}^{\zeta_1 X}$	1.96	0.58

(b)

$\beta_{n\ q}^{\zeta_2 X}$	b	se
$\beta_{2\ 2}^{\zeta_2 X}$	0.67	0.34
$\beta_{3\ 2}^{\zeta_2 X}$	-0.79	0.45
$\beta_{2\ 3}^{\zeta_2 X}$	1.31	0.32
$\beta_{3\ 3}^{\zeta_2 X}$	-0.32	0.44

(c)

$a_{l\ n}^{\zeta_1 \zeta_2}$	b	se
$a_{2\ 2}^{\zeta_1 \zeta_2}$	-0.79	0.49
$a_{2\ 3}^{\zeta_1 \zeta_2}$	-1.23	0.77
$a_{3\ 2}^{\zeta_1 \zeta_2}$	-0.15	1.04
$a_{3\ 3}^{\zeta_1 \zeta_2}$	1.68	0.87

(d)

β_p^Y	b	se
$\beta_{2\ 2}^{Y \zeta_1}$	0.26	0.25
$\beta_{2\ 3}^{Y \zeta_1}$	0.18	0.41
$\beta_{2\ 2}^{Y \zeta_2}$	1.08	0.28
$\beta_{2\ 3}^{Y \zeta_2}$	-0.20	0.34
$\beta_{2\ 2}^{Y X}$	0.27	0.15
$\beta_{2\ 3}^{Y X}$	0.30	0.15

Note: X = *Enriched job design*, ζ_1 = *Job control*, ζ_2 = *Job satisfaction*, and Y = *Labor productivity*

$.008$), but the category-specific variables from Table 4.6(c) are not significant. It is, therefore, difficult to interpret the association.

The overall effect of ζ_{1j} (job control) on *labor productivity* is not significant ($\chi^2 = 1.37,\ df = 2,\ p = .51$) , while the overall effects of ζ_{2j} (job satisfaction) and *enriched job design* are significant ($\chi^2 = 22.87,\ df = 2,\ p < .001$ and $\chi^2 = 4.97,\ df = 2,\ p = .08$), although the latter only at a significance level of 10% and not at 5%. The category-specific parameters are presented in Table 4.6(d). All categories except $\beta_{2\ 3}^{Y \zeta_2}$ score higher than the reference group.

To conclude, there is a significant direct effect of *enriched job design* on *labor productivity*. The two paths of the indirect effect of *enriched job design* (macro level) on *labor productivity* (macro level) through *job satisfaction* (micro level) are significant, while only the first part of the indirect path through *job control* (micro level) is significant and the second part is not significant.

4.7 Discussion

A stepwise multilevel latent class model was proposed to predict group-level outcomes by means of discrete individual- and group-level predictors. In the first step, a latent class model was estimated in which the individual-level predictor was used as an indicator for a group-level latent class variable (measurement model). In the second step, the individual-level predictor was aggregated to the group level based on the latent class model from the first step. This had two important advantages. First, the measurement error in the aggregated scores is known. Second, it is a more elegant way of aggregating a discrete variable than using a mean or mode. Next, the aggregated scores are related to the remaining group-level variables while correcting for the known measurement error (structural model). It is shown that the bias adjusted stepwise procedures work without introducing bias since the results of the stepwise approaches were very similar to the parameters that are obtained when the measurement and structural model are simultaneously estimated in a one-step analysis. Since researchers are used to working in a stepwise manner when they aggregate with a manifest mode, they can continue to work in the way they are used to while accounting for measurement error in the aggregated scores and still get unbiased results.

Two issues are of importance when the stepwise procedures are applied. First, in case the model contains multiple macro-level latent variables, the within-group association among the micro-level indicators needs to be included in the first step of the stepwise procedure, unless the within-group association is small and the micro-level scores are very good indicators of the group-level latent variables. Conceptually, it would suit the philosophy of stepwise estimation better to formulate two separate measurement models in the first step, one for each latent variable, but the simulation study showed that ignoring this within-group association provides biased estimates of the between-group association among the latent variables. It is unrealistic to assume that there is no residual within-group association among the predictors since that would imply that all associations among the micro-level predictors can be explained through the group-level latent variables. Second, class separation needs to be sufficiently high ($R^2_{entr} = .45$) since results with poorly separated classes are only correct with large sample sizes. In practice, this is no problem, since it is of little use to aggregate a variable that will not or only weakly be related to the group level.

The stepwise ML procedure is applied to a real data example in which labor productivity (group-level outcome) is explained by enriched job design (group-level predictor), job control (individual-level predictor), and job satisfaction (individual-level predictor). All variables from the example were constructed from multiple items but were used as single variables in the model. For the continuous version of the current application, Croon et al. (2014) found that the factor analytic model was better equipped to detect bathtub-type linkages than a model using scale scores. An interesting direction for further research would be to see whether this is also the case for discrete variables, thus study micro-macro models in which especially the micro-level predictors, but also the group-level outcome and predictor, are latent variables measured with multiple items.

Finally, in the current article, only attention is paid to the parameter estimates of the model and not to the standard errors of these parameters. Theory suggest that when fixed parameter estimates, obtained in the first step of the stepwise procedure, are plugged into the likelihood function, the effect of their sampling variability on the uncertainty about the estimates in the third step should be accounted for (Murphy & Topel, 1985). Fortunately, an easily accessible correction method for the standard errors is already made available by Bakk, Oberski, and Vermunt (2014). In situations with a large sample size, such as the current simulation study and data example, correction of the standard errors is not needed because the uncertainty about the estimates from the first step is very small (Bakk et al., 2014).

CHAPTER 5

Micro-Macro Analysis with a Latent Macro-Level Outcome

Abstract

In educational measurement, responses of students on items are not only used to measure the ability of students, but also to evaluate and compare the performance of schools. Analysis should ideally account for the multilevel structure of the data, and school-level processes not related to ability, such as working climate and administration conditions, need to be separated from student and school ability. However, in educational studies such as PISA, TIMMS and COOL$^{5-18}$, this is seldomly done. This study presents a model that simultaneously accounts for the nested structure, controls student ability for processes at school level, classifies schools to monitor and compare schools, and tests for school-level item bias.

This chapter is accepted as Bennink, M., Croon, M. A., Keuning, J. & Vermunt, J. K. (2014). Measuring student ability, classifying schools, and detecting item bias at school level based on student-level dichotomous items. *Journal of Educational and Behavioral Statistics, 39*(3), 180-201.

5.1 Introduction

A growing number of studies aim at the monitoring of student achievement across schools and countries. In the United States, for example, students are tested in grades 3 through 8, and at one grade in high school as required by the *No Child Left Behind Act* of 2002. Other examples are (a) the *Programme for International Student Assessment* (PISA) in which 15-year-old students from approximately 70 countries are tested to evaluate and compare educational systems, (b) the *Trends in International Mathematics and Science Study* (TIMSS) in which the mathematics and science achievements of fourth- and eighth-grade students from the United States are compared to those of students in other countries, and (c) the *Progress in International Reading Literacy Study* (PIRLS) that reports every five years on the reading achievements of fourth-grade students worldwide.

The different studies share several characteristics. First, data are collected on individual students that are nested within higher level units such as classrooms, schools or countries. Analyses of these data should take this multilevel structure into account (Goldstein, 2011; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Moreover, in many studies of this kind, the student level data are not only used to measure the ability of students, but also to evaluate and compare the performances of higher level units (see, for example, Goldstein et al. (1993) and Leckie and Goldstein (2009)). Finally, in all these studies the item responses for the students might not only be considered as a reflection of ability but also as a reflection of processes that take place at the higher level. Some of these higher level processes are related to ability differences among the higher-level units, but other processes may be related to nonability differences between the units (Borghans, Meijers, & ter Weel, 2008). At classroom level, for instance, the working climate might, either positively or negatively, affect the item responses of the students. At school level, the administration condition might affect the students' behavior. If it was a low-stakes administration in which the test results have no severe consequences for the students, the data might not only reflect ability but also a lack of motivation by the students taking the test. At country level, the political environment might affect the students' performances to some extent.

These characteristics are not covered when the student responses are related to ability with the (one-level) item response theory models that are commonly used (Embretson & Reise, 2000). Student ability should ideally be modeled in relation to variables such as the overall ability at schools, working climate, administration conditions, and political environment to be able to disentangle student ability from school-level ability and school-level processes not related to ability. By doing so, the student ability measures are not confounded with school-level processes when student performance is evaluated, and schools can be compared based on processes that are not related to ability. This is all possible within an existing general multilevel latent variable framework (Muthén & Asparouhov, 2011; Skrondal & Rabe-Hesketh, 2004; Vermunt, 2008) by formulating a model with latent variables at two levels: a student level and a higher level such as school or country. The model that is presented in this study fits within this framework and has several advantages over common item response theory models. First, the model explicitly accounts for the multilevel structure of the data. Second, the model controls student ability for processes that may occur at higher levels, and finally, the model allows for a comparison between schools as it classifies schools into groups that can be interpreted as school types, according to their performances and characteristics not related to ability. A

Table 5.1: Nine Fold Classification of Multilevel Latent Variable Models

Student-level latent variable(s)	School-level latent variable(s)		
	Continuous	Discrete	Combination
Continuous	A1	A2	A3
Discrete	B1	B2	B3
Combination	C1	C2	C3

potential additional advantage of the model is that both uniform and nonuniform higher-level item bias can be studied. This can be useful in improving school performance as it clarifies which items functioned differently at which type of school, regardless of the ability level of the students at the schools. For instance, schools from school types in which some items function worse than in schools from other school types, could devote more time to teaching the topics covered by these items. In the model it is assumed that ability is unidimensional and therefore a continuous latent variable is used to measure ability. The nonability component is most probably multidimensional since it captures various types of school-level processes. Therefore, this part is modeled with a discrete latent variable.

The general latent variable framework is first described, providing the statistical background of the approach proposed in this article and illustrating the flexibility and general applicability of the general framework. The method is then applied to data obtained from the Dutch study COOL$^{5-18}$ in which the achievements of 5- to 18-year-old students are studied on the basis of a test with dichotomous educational items. In the current context, the main goal of the analysis was to classify schools but also to discuss uniform and nonuniform school level item bias. Finally, the necessity of using a complex (multilevel) model is demonstrated by comparing the fit of the model to the fit of less complex alternative models such as the two-parameter item response theory model.

5.2 General Framework

The general multilevel latent variable framework as referred to in the present study was described by Skrondal and Rabe-Hesketh (2004). It was later extended by, among others, Vermunt (2008) and Muthén and Asparouhov (2011). The framework allows for a definition of latent variables at the student level and/or the higher level. These variables can be either continuous or discrete, or even a combination of a continuous and a discrete latent variable at one or both levels. These three alternatives at both levels result in nine possible models presented in Table 5.1. Model C3 can be considered the most general model, whereas the eight remaining models are special cases of this more general model.

The nine-fold classification was already presented in Vermunt (2008), Palardy and Vermunt (2010) and Varriale and Vermunt (2012), and there are too many modelling options to discuss them all. Therefore, the framework is not discussed extensively again but only shortly presented to be able to place the A3 model from the current article into a broader context. In this model schools are classified into unobserved groups because they are homogeneous with respect to their item response theory models. This is combined with continuous latent variables at both the student and the school level.

Most models formulated within the framework can be estimated with existing software such as WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), GLLAMM (Rabe-Hesketh, Skrondal, & Pickles, 2004), Latent GOLD (Vermunt & Magidson, 2005a), or Mplus (Muthén & Muthén, 1998-2012). More information about the estimation procedures can be found in Fox and Glas (2001), Goldstein, Bonnet, and Rocher (2007), Vermunt (2008), Fox (2010), Palardy and Vermunt (2010), Muthén and Asparouhov (2011), and Varriale and Vermunt (2012). In the present study, the software package Latent GOLD (Vermunt & Magidson, 2013) was used to estimate the models.

5.3 Method

5.3.1 Data

The general multilevel latent variable framework was applied to the Dutch cohort study COOL$^{5-18}$ in which, over three waves, measurements are conducted on various academic subjects from children from four grades on different academic tracks (more information about COOL$^{5-18}$, although in Dutch, can be found at www.cool5-18.nl and in Zijsling, Keuning, Naayer, and Kuyper (2012)). For the present study, only the data collected during the second measurement wave in Grade 9 on the subject of English was used. The test was administered to 3,458 students from 60 different schools and included a total of 24 dichotomous scored multiple choice items. Each item contains a short English text and in a single multiple choice question it was tested whether the student globally understood the text. An unofficial translation of the items that students found easiest (item 1) and most difficult (item 17) is presented in Appendix F.

The first part of the test (Item 1-10) was designed to be easier than the second part (Item 11-24). To provide a sense of the degree of school-level clustering in the data, the mean of the aggregated number of correct answers is 13 out of 24, with a standard deviation of 1.4. The latter standard deviation is not an unbiased estimate of the between-school variation, since it is also a function of the within-school variability.

The Grade-nine students were recruited from different school tracks. Slightly less than half of the students were recruited from pre-vocational secondary education while the remaining students were recruited from either senior general secondary education or pre-university education. The students were not equally divided across the participating schools. Some schools participated with a very small number of students (< 10) in the most recent wave of data collection, while other schools participated with several hundreds of students. The reason for this is that some secondary schools chose for a relatively small participation with only the students who also participated in grade 6 of elementary school (i.e., 'individual participation'), while other schools chose to participate at a larger scale with not only the students who had already been involved in COOL$^{5-18}$ during a previous wave of measurement but also with their classmates (i.e., 'collective participation'). The schools were located in different regions of the Netherlands, covering all 12 provinces. The urbanization level for the schools varied from *rural* (1) to *moderately urbanized* (3) to *very strongly urbanized* (5).

5.3.2 Model for Uniform Item Bias at School Level

The items in COOL$^{5-18}$ were designed to measure a continuous latent ability trait. From a substantive point of view it thus seemed natural to model the students' responses on the test items by an item response theory (IRT) model. Basically, all IRT models could be used but in the present application the two-parameter logistic model (2PLM) (Embretson & Reise, 2000) was chosen. In this model, the continuous latent ability score for student i nested within school j is labeled θ_{ij} and the responses of the students on the P test items are used as indicators. For the current application, there is no need to define latent classes at the level of the students.

The school-level processes are captured by a discrete latent variable at the school level, C_j, which represents the clustering of schools into one of K (latent) school types based on the responses of the students on the test items. Measurement error in the classification is accounted for because a latent variable is used.

Modelling the heterogeneity between groups as discrete has two important advantages. First, a model-based grouping of schools is best for the purpose of comparison. It is more feasible to compare a manageable number of groups of schools than all individual schools. Second, the effect of C_j on the item responses can also be interpreted as uniform item bias at the school level since the probabilities of a correct response for students going to different types of schools, keeping their ability levels constant, are allowed to differ. Therefore, C_j is a discrete random effect that assumes there are homogeneous groups of schools with similar intercepts.

The relations between y_{pij}, θ_{ij} and C_j can formally be described by the following logit equation:

$$\text{Logit}(P(y_{pij} = 1|\theta_{ij}, C_j)) = b_{1p}\theta_{ij} + \sum_{k=1}^{K} b_{2pk} \times I_{C_j=k}, \tag{5.1}$$

in which $I_{C_j=k}$ is an indicator function that equals 1 if $C_j = k$ and 0 otherwise, and θ_{ij} and C_j are assumed to be statistically independent.

As can be seen, the logit of the conditional probability of student i from school j to answer item p correctly is related to the latent ability θ_{ij} of the student and the latent class C_j the school belongs to. The slope parameter b_{1p} is the discrimination parameter for item p, which is controlled for the effect of the latent classes at the level of the schools.[8] In the uniform bias model, this slope parameter does not vary across the latent classes indicating that the effect of the individual latent trait θ_{ij} on the item responses remains the same in all classes and only differs with respect to the overall response tendency. For identification purposes, b_{1p} is fixed to 1 for the first item. The intercept parameter b_{2pk} is related to the difficulty parameter for item p for a student from a school in class k, $C_j = k$. The intercept parameter is controlled for the effect of the latent ability of students.

Item Response Function (IRF) graphs can be constructed by plotting the response probabilities on the y-axis and θ_{ij} on the x-axis. The response probabilities can be obtained by transforming Logit$(P(y_{pij} = 1|\theta_{ij}, C_j))$ into a probability by

$$P(y_{pij} = 1|\theta_{ij}, C_j) = \frac{\exp(b_{1p}\theta_{ij} + \sum_{k=1}^{K} b_{2pk} \cdot I_{C_j=k})}{1 + \exp(b_{1p}\theta_{ij} + \sum_{k=1}^{K} b_{2pk} \cdot I_{C_j=k})}. \tag{5.2}$$

A separate IRF graph can be constructed for each item for each of the K latent classes. Because only the intercept is class specific, the joint representation of all IRFs for the

[8] b_{1p} is assumed to multiply both the level-1 and level-2 parts of theta shown in Equation 5.3. This common IRT assumption can be relaxed as discussed in Muthén and Asparouhov (in press).

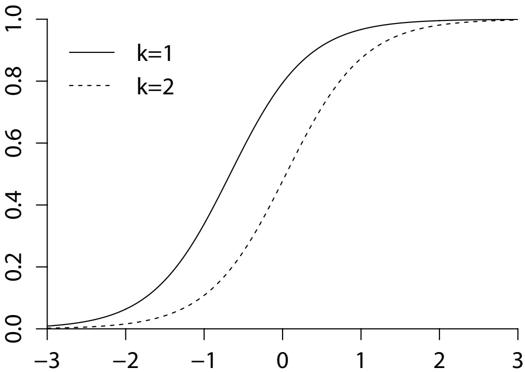


Figure 5.1: IRF Item 10 (Uniform School-Level Item Bias)

same item shows a set of K nonintersecting parallel curves. An example IRF graph is shown in shown in Figure 5.1.

In the model described so far, the latent class variable is assumed to capture all relevant differences among the schools. Some of these differences may pertain to a general ability level, whereas others may relate to higher-level (school) characteristics - such as working climate and administration conditions - which are independent of the general ability level of the schools. Therefore, the ability and the nonability components of the school differences are separated by postulating a continuous latent variable at the school level, θ_j, to represent the ability component of the between-school differences.

The relationship between θ_j and θ_{ij} is given by

$$\theta_{ij} = \theta_j + e_{ij} \tag{5.3}$$

in which both latent variables are treated as continuous normally distributed variables. In this way, a student's individual ability is expressed as a deviation from the average ability of the school, implying that the ability differences among students are decomposed into two components. One component captures the differences in ability levels between schools (θ_j) while the other component captures the ability differences within schools (e_{ij}) and their variances express the strength of both components. In this way, school ability is measured indirectly, while an alternative would be to allow θ_j to affect the student-level item responses directly. The latter implies that individual-level responses can be broken down directly into a between and a within part. Since the current items were designed to measure an individual-level trait, the first option was chosen.

The linear relationship between θ_j and θ_{ij} is a special case of a more general linear relationship between the two latent variables: $\theta_{ij} = b_3 + b_4\theta_j + e_{ij}$, with the intercept b_3 and the slope b_4 fixed to 0 and 1 respectively. These restrictions are needed for identification purposes. Without these restrictions, alternative identification constraints would be required, such as fixing both the within-school variance of θ_{ij} and the between-school variance of θ_j to 1. Opting for the constraints on the parameters of the linear relation between θ_{ij} and θ_j instead of on the variances, allows for a direct comparison of the variation in the student's abilities within the schools and the variation at the school level.

A common problem in comparing schools is that school effects on student learning are confounded with differences between schools in their intake achievements. Therefore, the latent variable θ_j is assumed to be independent of the latent class variable C_j to separate the ability and nonability components of the between school variation. It is possible to relax or test this assumption by allowing an association between θ_j and C_j. This could be done by adding a direct effect of θ_j on C_j or by adding a direct effect of C_j on θ_j.[9] Both options would make the interpretation of the latent variables dependent on each other while in the current application ability and nonability need to be separated to control the classification for general ability. Otherwise it cannot be disentangled whether ability is caused by learning or intake effects. That C_j now represents nonability gives an additional reason to model C_j as discrete rather than continuous since it is not realistic to assume that nonability is unidimensional.

To summarize, in a 'regular' one-level IRT model, a continuous latent trait (θ_{ij}) is assumed to underlie the item responses of students and this latent trait is interpreted as student ability. When responses are collected from students from different schools, a multilevel model is needed to adjust for the dependency among students from the same school. This is done by decomposing θ_{ij} into its within and between components. It is expected that all variability in the item responses that remains after controlling for the relations discussed so far can be attributed to nonability issues. These are modeled as discrete because the main goal was to classify schools, but keep ability out of this classification. Besides, it is not realistic to assume that all processes that are related to nonability are unidimensional.

Two manifest school-level variables were included in the model as explanatory variables for latent class membership. The first variable, X_{1j}, is dichotomous and represents the form of the school's participation, i.e., individually or collectively. This variable was included in the analysis because it might have affected the motivation for the schools to participate in COOL$^{5-18}$. For the collective participants, the report they receive upon completion of the measurement was most likely the primary reason to participate, while for the individual participants the social and scientific relevance of the study was probably the decisive factor. This difference might (unintentionally) have had an impact on the administration conditions. The second variable, X_{2j}, represents the urbanization level for the schools. This variable was included in the model because it is found to have an impact on student ability on a rather regular basis (Tekwe et al., 2004). As mentioned before, the urbanization level for a school was defined on a 5-point Likert scale. In the present analysis it was treated as a continuous variable. Given the two explanatory variables, the probabilities of latent class membership are given by:

$$\log\left(\frac{P(C_j = k|X_{1j}, X_{2j})}{P(C_j = K|X_{1j}, X_{2j})}\right) = b_{5k} + b_{6k}X_{1j} + b_{7k}X_{2j}. \tag{5.4}$$

As can be seen, the equation relates the manifest school-level predictors, X_{1j} and X_{2j}, to the latent school-level classes, C_j. Category K of C_j is used as a reference category. With only two categories, this multinomial logit equation simplifies to a binary logit equation. The intercept is denoted by b_{5k} while the effects of X_{1j} and X_{2j} are captured by b_{6k} and b_{7k}, respectively.

According to the nine-fold classification from Table 5.1 the model as presented so far is an A3 model as it includes one continuous variable at the student level (θ_{ij}) and both

[9] In the current application, both directs effects proved not significant.

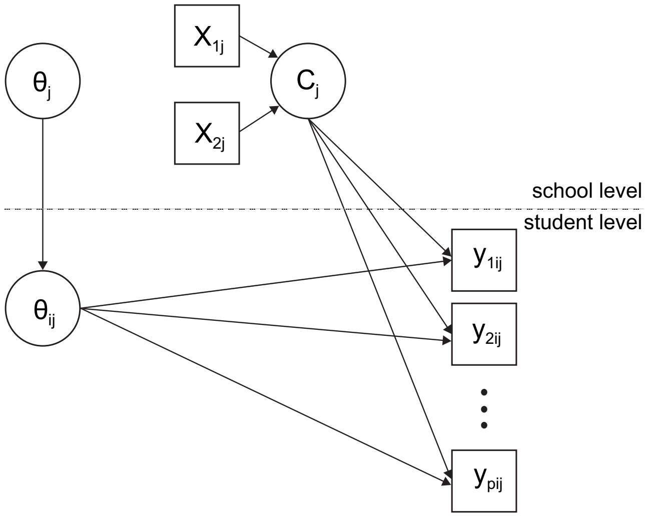


Figure 5.2: Conceptual Model (Uniform School-Level Item Bias)

a continuous (θ_j) and a discrete (C_j) latent variable at the school level. The model is graphically illustrated in Figure 5.2 in which the rectangles represent manifest variables and the circles latent variables.

After fitting the model, schools may be assigned to one of the K latent classes of C_j. Several deterministic and probabilistic classification methods have been proposed such as modal, random, and proportional assignment (Dias & Vermunt, 2008; Goodman, 1974a, 1974b; McLachlan & Peel, 2000). We would recommend modal assignment for the purpose of classifying schools, since it minimizes the total probability of misclassification.

5.3.3 Model for Nonuniform Item Bias at School Level

Up to now, only uniform item bias at the school level was allowed. In order to examine the occurrence of nonuniform item bias at school level, an interaction effect between θ_{ij} and C_j is added to the item equations. Such a model can be defined in the following manner:

$$\text{Logit}(P(y_{pij} = 1|\theta_{ij}, C_j)) = \sum_{k=1}^{K} (b_{8pk}\theta_{ij} + b_{9pk}) \cdot I_{C_j=k}, \tag{5.5}$$

$$\theta_{ij} = \theta_j + e_{ij}, \tag{5.6}$$

and

$$\log\left(\frac{P(C_j = k|X_{1j}, X_{2j})}{P(C_j = K|X_{1j}, X_{2j})}\right) = b_{12k} + b_{13k}X_{1j} + b_{14k}X_{2j}. \tag{5.7}$$

This is represented graphically in Figure 5.3. The relations from the model for uniform item bias at school level are shown in gray and the added interaction effects are shown in black.

The interaction effects can be interpreted in two equivalent ways. The first interpretation is that the item bias at school level can be stronger (or weaker) for students

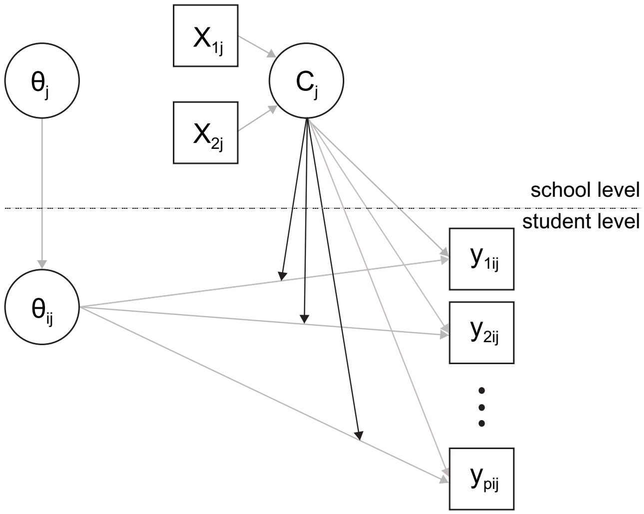


Figure 5.3: Conceptual Model (Nonuniform School-Level Item Bias)

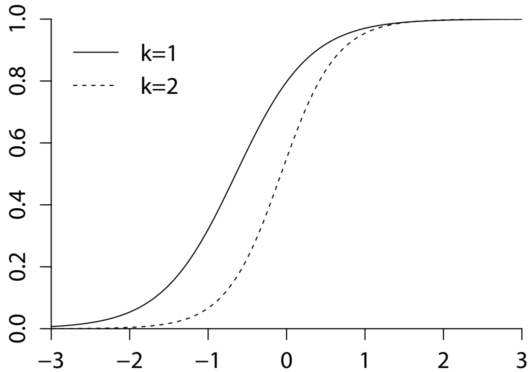


Figure 5.4: IRF Item 10 (Nonuniform School-Level Item Bias)

with higher (or lower) individual latent ability. The second interpretation is that the association between the individual latent ability of students and the item responses can be stronger or weaker depending on the class to which the school of the student belongs. In both cases it implies that both the discrimination and the difficulty parameter are class dependent, while only the difficulty parameter is class dependent in the model for uniform item bias at school level. For students from a school in class k, the discrimination parameter for item p is represented by b_{8pk} and the difficulty parameter is represented by b_{9pk}. The interpretation of the other parameters in the model is equivalent to the interpretation of the parameters included in the previous model, which only allowed for the detection of uniform item bias. The IRF graphs can be constructed in the same way as described for the uniform item bias model, but they are no longer parallel because the slopes are class specific, as shown in an example in Figure 5.4.

5.3.4 Less Complex Alternative Models

The two models for the analysis of the COOL$^{5-18}$ data presented above are both rather complex A3 models (see Table 5.1). In order to evaluate the explanatory power of these models and to ascertain that their complexity is justified, they should fit better than less complex models. In this article, three simplified models derived from the basic starting model are considered.

A first simplification consists of removing the continuous latent variable at the school level θ_j from the model. The model then reduces to a so-called A2 model (see Table 5.1). However, ignoring a random effect at the higher level could result in an overextraction of the number of latent classes (Palardy & Vermunt, 2010). This is due to the fact that the classification of the schools is no longer controlled for the overall ability level of the schools. This may thwart the substantive interpretation of the school classifications.

A second simplification of the original model consists of the removal of the discrete latent variable C_j from the model. The A1 model (see Table 5.1) that we obtain in this way is a two-level item response theory model. This model no longer allows classifying schools in discrete classes, or to study school-level item bias. The individual scores of the students, however, are still controlled for the overall ability within the schools.

A one-level item response model, finally, can be obtained by removing both the continuous and the discrete latent variable at the school level. This model could be labeled as 'A0 model' as no single latent variable at the school level is included in the model anymore but only a continuous student-level latent variable is used. This type of model is currently used in studies like PISA, TIMMS, PIRLS and COOL$^{5-18}$. In contrast to the models presented in this study, these models do not control the latent ability for the students in any way for processes that might occur at the level of the classroom, school or country.

The fit of these three less complex models will be compared to that of the two original A3 models in which the between-school differences are modeled by a continuous and a discrete latent variable.

5.4 Results

5.4.1 Uniform Item Bias at School Level

Applying the A3 uniform school-level item bias model to the data first required the determination of the optimal number of latent school types (i.e. the optimal number of categories for C_j). This was achieved by comparing BIC and CAIC values of models with various numbers of latent classes. Lukočienė et al. (2010) recommend using BIC and CAIC values based on the number of schools if the number of latent classes at the school level has to be determined. When in Table 5.2 the fit indices for the A3 model with uniform school-level item bias are compared, the model with two latent classes at the school level proved to fit best because both the BIC and CAIC values were relatively low for this number of latent classes.

A way of getting around the decision of whether fit indices should be based on the number of students or on the number of schools, is to use AIC or AIC3 values as these are not a function of sample size. In the present analysis, however, these fit indices led to the same conclusion as BIC and CAIC. Given the results from Table 5.2, it was decided to continue

Table 5.2: Fit Indices for Models Fitted on the COOL$^{5-18}$ Data

Model	Classes	Item bias	LL	BIC	CAIC	AIC	AIC3	Parameters
A3	1	Nonuniform	-50998	102197	102246	102094	102143	49
A3	2	Nonuniform	-50864	102138	102238	101928	102028	100
A3	3	Nonuniform	-50828	102275	102426	101958	102109	151
A3	4	Nonuniform	-50777	102380	102582	101957	102159	202
A3	1	Uniform	-50998	102197	102246	102094	102143	49
A3	2	Uniform	-50912	102135	102211	101976	102052	76
A3	3	Uniform	-50891	102204	102307	101988	102091	103
A3	4	Uniform	-50907	102346	102476	102073	102203	130
A2	1	Uniform	-51157	102511	102559	102411	102459	48
A2	2	Uniform	-50968	102244	102319	102086	102161	75
A2	3	Uniform	-50912	102242	102344	102028	102130	102
A2	4	Uniform	-50860	102249	102378	101979	102108	129
A1	-	-	-50998	102197	102246	102094	102143	49
A0	-	-	-52086	104363	104410	104265	104312	47

BIC and CAIC are based on the number of schools, LL=log-likelihood

with the two-class model.

In this two-class model, the majority of the schools was classified in the first latent class (87%), whereas only a small minority of the schools was classified in the second class (13%). The entropy R^2 is .82, indicating a good class separation. The regression parameters for the items can be found in Table 5.3.
The discrimination parameters, b_{1p}, were all positive and significant at the 1% level. This demonstrates that higher individual latent ability increases the conditional probability of answering an item correctly. The intercept parameters, b_{2pk}, are higher for the first part of the test compared to the second part of the test, for both classes. As designed, the first part of the test was easier than the second part.

Moreover, for all but one item the intercept parameters for the students from schools in the second class (b_{2p2}) were lower than the difficulty parameters estimates for the students from schools in the first class (b_{2p1}). The differences in intercepts proved significant for almost all items, which means that uniform item bias was detected for almost all items. No significant school-level item bias was detected for Item 15, 21, and 22.

To give an impression of the substantive importance of these differences, the expected number of correct answers for an average student ($\theta_{ij} = 0$) attending a school from the first class is 13.6, while it is 10.6 for an average student from a school of the second class. Furthermore, the largest difference in intercept is -1.44, implying that the odds of answering item 10 correctly is $exp(1.44) \approx 4$ times smaller for schools in the second class compared to schools from the first class.

The relative importance of ability and the school classification for the item responses can be assessed by comparing the absolute effects of θ_{ij} and C_j on the same scale or, in other words, to compute the standardized regression coefficients. The standard deviation of θ_{ij} is fixed to 1 for identification purposes and the standard deviation of C_j can be transformed to 1 using the class proportions:

$$SD(C_j) = \sqrt{P(C_j = 1) \times (1 - P(C_j = 1))} = \sqrt{.87 \times (1 - .87)} = 0.34. \qquad (5.8)$$

Therefore, $\frac{(b_{2pk} - b_{1pk})}{0.34}$ is on the same scale as b_{1p} and both are presented in Table 5.3. For an average student ($\theta_{ij} = 0$), during the first part of the test, school type is more

Table 5.3: Regression Parameters Items Uniform Item Bias

Item (p)	P(correct)	b_{1p}	SE	b_{2p1}	SE	b_{2p2}	SE	$b_{2p2} - b_{2p1}$	SE	$b_{2p2} - b_{2p1}$ [1]
1	.83	1.00	NA	1.82*	.06	0.50*	.12	-1.32*	.13	-3.88
2	.77	1.13*	.19	1.34*	.05	0.87*	.13	-0.46*	.14	-1.35
3	.61	1.05*	.17	0.51*	.04	0.04	.12	-0.47*	.13	-1.38
4	.70	1.24*	.20	1.02*	.05	0.11	.13	-0.91*	.14	-2.68
5	.70	1.67*	.25	1.02*	.06	0.17	.14	-0.85*	.15	-2.50
6	.77	0.92*	.17	1.32*	.05	0.81*	.13	-0.51*	.13	-1.50
7	.69	2.70*	.39	1.15*	.07	-0.01	.17	-1.16*	.19	-3.41
8	.71	0.74*	.14	0.98*	.04	0.22	.12	-0.76*	.13	-2.24
9	.55	1.61*	.24	0.24*	.05	-0.11	.14	-0.36**	.15	-1.06
10	.74	2.02*	.30	1.35*	.06	-0.09	.15	-1.44*	.16	-4.24
11	.49	1.45*	.22	0.03	.05	-0.78*	.14	-0.81*	.15	-2.38
12	.41	1.94*	.28	-0.42*	.06	-0.96*	.16	-0.54**	.17	-1.59
13	.54	2.54*	.36	0.23*	.06	-0.30	.17	-0.54**	.19	-1.59
14	.60	1.66*	.24	0.53*	.05	-0.25	.14	-0.79*	.15	-2.32
15	.45	1.06*	.17	-0.20*	.04	-0.39**	.12	-0.19	.13	-0.56
16	.58	1.68*	.25	0.37*	.05	0.05	.14	-0.31**	.15	-0.91
17	.33	1.27*	.20	-0.76*	.05	-1.10*	.14	-0.34**	.14	-1.00
18	.42	1.44*	.21	-0.36*	.05	-0.65*	.13	-0.29**	.14	-0.85
19	.39	1.43*	.21	-0.49*	.05	-0.65*	.13	-0.16	.14	-0.47
20	.35	0.99*	.16	-0.64*	.04	-0.91*	.13	-0.27**	.14	-0.79
21	.42	1.30*	.20	-0.36*	.05	-0.53*	.13	-0.17	.14	-0.50
22	.33	0.81*	.14	-0.76*	.04	-0.60*	.12	0.16	.13	0.48
23	.42	1.19*	.19	-0.32*	.04	-0.81*	.13	-0.49*	.14	-1.44
24	.45	0.83*	.14	-0.19*	.04	-0.55*	.12	-0.35**	.12	-1.03

b_{1p} = slope and b_{2pk} = intercept, * = $p < .001$, ** = $p < .05$, [1] = standardized difference.

important than ability. As the difficulty of the items increases in the second part of the test, ability becomes more important than school type.

Although there were reasons to believe that the form of participation and urbanization level could eventually predict latent class membership, neither the effect of X_{1j} nor that of X_{2j} was significant: for participation, the regression coefficient was estimated as $b_6 = -0.37$ $(SE = 2.20,\ p = .87)$; for urbanization level the coefficient was $b_7 = -0.45$ $(SE = 0.35,\ p = .20)$.

Finally, it was found that the between-school variance of θ_j was equal to 0.02 $(SE = 0.01,\ p < .001)$ and the within-school variance of θ_{ij} was equal to 0.20 $(SE = 0.05,\ p < .001)$. As could be expected, the differences on the latent ability trait among students within schools were much larger than the differences across the mean school levels since only 9% of the variability in individual latent ability is due to differences between schools $(\frac{.02}{.02+.20} = .09)$.

The main conclusion of this analysis is that the two-class model indicates a small group of schools that performed worse than the other schools and that there is severe uniform item bias at the level of the schools. This bias could not have been detected with the item response theory models that are normally used in comparable studies.

5.4.2 Nonuniform Item Bias at School Level

In a further analysis of the same data, the model with nonuniform item bias was fitted. Looking at the fit statistics of the A3 model with nonuniform item bias in Table 5.2, all fit indices point to a two-class solution. The fit of this model was compared to the fit of the previous model with uniform item bias using a likelihood-ratio test. The test showed the interaction effect to improve the fit of the uniform item bias model significantly ($X^2 = 96,\ df = 100 - 76 = 24,\ p < .001$). Therefore, based on this test, the model with nonuniform item bias at school level is to be preferred over the model with uniform item bias. Table 5.2 also shows that the AIC and AIC3 model selection criteria favored this model, but the BIC and CAIC values based on the number of schools pointed towards the model with uniform item bias. Overall, these results indicate that the model with nonuniform item bias can be preferred over the model with uniform item bias at school level.

Under the nonuniform item bias model, more schools were classified into the group of poorly performing schools. The size of the second class increased from 13% to 21% and the entropy R^2 increased to .87. The regression parameters for the items can be found in Table 5.4. The discrimination parameters for item p are represented by b_{8p1} and b_{8p2} for, respectively, the first and the second class. As can be seen in Table 5.4, b_{8p1} is positive and significant at the 1% level for all items. This means that higher individual latent ability increases the conditional probability of answering an item correctly for students from schools in the first class. The estimates of b_{8p2} were not significant for Item 20-22 and 24, implying that for these items ability is not related to the probability of answering an item correctly in schools from the second latent class. These items were maybe too difficult for students from schools of the second school type so that these students might have simply guessed the answers.

For thirteen items (Item 1-5, 9-10, 12, and 14-18) the discrimination parameters were not significantly different across classes. This means that nonuniform item bias was not present for these items. The nonuniform bias was detected, however, for some of the other items. Whereas for Item 6, 8, and 11 the discrimination parameter was significantly larger for students from schools in the second class, the discrimination parameter was significantly smaller for these students for Item 7, 13, and 19-24.

The intercept parameter for item p is b_{9p1} for a student from a school in the first class and b_{9p2} for a student from a school in the second class. As before, the intercepts were higher for the first part of the test than for the second in both classes. Moreover, b_{9p2} was significantly smaller than b_{9p1} for sixteen items, which means that these items were more difficult for students from schools in the second class. Please note that this difference in intercepts between the classes can only be interpreted for an average student ($\theta_{ij} = 0$) since the IRFs are nonparallel. The intercept parameters for the remaining eight items proved not significantly different between the two classes.

The expected number of correct answers for an average student attending a school from the first class is 13.6, while it is 11.3 for an average student from a school of the second class. The relative importance of ability and school type is not shown in Table 5.4, but the same pattern as for the model for uniform item bias emerged. For an average student, in the first part of the test, school type is more important than ability, but as the difficulty of the items increases in the second part of the test, ability becomes more important than school type.

The conclusions about the effects of the school-level predictors and about the

Table 5.4: Regression Parameters Items Nonuniform Item Bias

Item (p)	b_{8p1}	SE	b_{8p2}	SE	$b_{8p2} - b_{8p1}$	SE	b_{9p1}	SE	b_{9p2}	SE	$b_{9p2} - b_{9p1}$	SE
1	1.00	NA	1.59*	.40	0.59	.40	1.84*	.06	0.76*	.12	-1.08*	.13
2	1.20*	.25	1.60*	.42	0.40	.37	1.35*	.05	1.01*	.12	-0.34**	.14
3	1.16*	.23	1.01**	.32	-0.15	.31	0.54*	.05	0.05	.10	-0.49*	.11
4	1.39*	.27	1.49*	.38	0.11	.32	1.04*	.05	0.31**	.11	-0.73*	.12
5	1.78*	.33	2.42*	.54	0.64	.40	1.02*	.05	0.41**	.13	-0.61*	.15
6	0.91*	.20	1.90*	.47	0.99**	.41	1.31*	.05	1.10*	.13	-0.21	.14
7	3.16*	.56	2.18*	.50	-0.98**	.43	1.18*	.08	0.20	.12	-0.99*	.15
8	0.77*	.18	1.63*	.41	0.85**	.36	0.97*	.04	0.54*	.11	-0.43*	.12
9	1.83*	.33	1.44*	.37	-0.39	.32	0.25*	.05	-0.05	.10	0.30**	.12
10	2.12*	.39	2.84*	.62	0.73	.45	1.37*	.06	0.20	.14	-1.17*	.16
11	1.51*	.28	2.77*	.60	1.26**	.46	0.03	.04	-0.54*	.14	-0.57*	.15
12	2.19*	.39	2.09*	.49	-0.10	.37	-0.41*	.06	-0.83*	.13	-0.42**	.14
13	3.03*	.54	2.23*	.50	-0.81**	.41	0.22**	.07	-0.06	.12	-0.28	.15
14	1.80*	.33	2.21*	.50	0.41	.38	0.54*	.05	-0.02	.13	-0.56*	.14
15	1.21*	.23	1.08*	.32	-0.13	.29	-0.20*	.04	-0.33*	.10	-0.13	.11
16	1.88*	.34	1.82*	.43	-0.06	.34	0.37*	.05	0.14	.11	-0.23	.13
17	1.50*	.28	0.87**	.32	-0.63	.33	-0.75*	.05	-1.07*	.11	-0.32**	.12
18	1.65*	.30	1.44*	.37	-0.21	.32	-0.36*	.05	-0.56*	.11	-0.20	.12
19	1.77*	.32	0.70**	.29	-1.07**	.34	-0.49*	.05	-0.65*	.10	-0.16	.11
20	1.19*	.23	0.55	.28	-0.64**	.32	-0.63*	.05	-0.89*	.10	-0.26**	.11
21	1.69*	.31	0.27	.26	-1.42*	.37	-0.35*	.05	-0.56*	.09	-0.21	.11
22	1.09*	.22	-0.06	.25	-1.15*	.34	-0.77*	.05	-0.69*	.10	0.08	.11
23	1.48*	.28	0.64**	.28	-0.83**	.32	-0.32*	.05	-0.67*	.10	-0.36*	.11
24	1.08*	.21	0.06	.25	-1.01**	.32	-0.19*	.04	-0.52*	.09	-0.33**	.10

b_{8pk} = slope and b_{9pk} = intercept, * = $p < .001$, ** = $p < .05$

differences between the between- and within-school variances were the same as in the first analysis. Latent class membership at the school level could not be predicted from the predictors at the school level as both the effect of X_{1j} and X_{2j} were not significant: for participation it was found that $b_{13} = -1.43$ $(SE = 1.11,\ p = .20)$, while for urbanization level $b_{14} = -0.58$ $(SE = 0.31,\ p = .07)$ was obtained. Since the predictors were already not significant in the previous model, they could obviously have been removed from the present analysis. They were nevertheless included in the model in order to ensure that the two models differ only in the manner the item bias at school level was modeled (i.e., uniform or nonuniform).

The variances for θ_j and θ_{ij} were equal to 0.02 $(SE = 0.01,\ p < .001)$ and 0.16 $(SE = 0.05,\ p < .001)$ respectively, which means there was less variation in the overall latent ability of schools than in the latent ability of the students within the schools. In fact, 11% of the variability in individual latent ability is due to differences among schools $(\frac{.02}{.02+.16} = .11)$.

5.4.3 Less Complex Alternative Models

In addition to the two models that were proposed in this study, three other - less complex - models were estimated, including the standard one-level item response theory model that is usually used in comparable studies (A0 model). As can be concluded from Table 5.2, all fit indices show that the most complex A3 models provide the best fit. Table 5.2 also shows that more classes would be needed if the random effect of θ_j was ignored as in the A2 model. As indicated by all fit indices except CAIC, the number of latent classes at the school level increases from two to three (BIC) or even four (AIC and AIC3) if θ_j had not been included in the model. The results of these alternative analyses show that the more complex A3 models provide a better fit to the data than the simplified models considered here.

5.5 Discussion

A model was presented to analyze data sets that are typical for large-scale studies such as PISA, TIMMS, and COOL$^{5-18}$. This model has several advantages over the models that are currently in use because it allows for a simultaneous modelling of student performance and school performance. First, the nested structure of the data has been handled correctly. Second, student ability is controlled for processes that may occur at the school level. Third, schools are classified after controlling for ability and this can be useful monitoring and comparing schools. Fourth, schools can improve their education by focusing on specific topics that are covered by items that induce school-level item bias.

The current models can also be applied when teachers (or classrooms) instead of schools are used as higher-level units to measure and inform teachers which parts of the curriculum they teach well and which parts they teach poorly, as evidenced by their students' achievements. Then teachers are classified, but this classification is controlled for the ability of students and the average ability of the children being educated by the teachers. At the same time, teacher-level item bias is studied. With large samples, it is even possible to consider models with three nested levels: students within teachers within schools.

As the model fits within the general framework of multilevel latent variable models, software is already available to estimate the model. The relative complexity of the model should not be an impediment to its application in multilevel situations as considered here. The present study clearly shows that the complex model provides a better fit than the simpler alternatives. In the light of the similarities between the research designs employed in the field of educational research, it is unlikely that the present results are specific to the COOL$^{5-18}$ study. Although the current model is compared to less complex alternatives, it has not been compared to even more complex alternatives such as using a three-parameter IRT model instead of a two-parameter IRT model. Another more complex model would take into account the longitudinal features of the COOL$^{5-18}$ data. With these models, change in student and school ability could be studied providing even more information about school performance. It would especially provide the possibility to disentangle intake and learning effects, which is not possible in a cross-sectional dataset.

In studies where latent variables are used, it is often rather difficult to establish the meaning of the variables. In the present model, the interpretation of θ_{ij} and θ_j as student ability and the overall ability of a school are rather straightforward. However, the interpretation of the latent classes at the school level is less clear. The two manifest school-level variables that were included in the model to predict class membership - that is, the form of participation and the urbanization level for the schools - proved unrelated to latent class membership of the schools. A level-2 sample size of 60 may not have provided enough power to demonstrate these between-school effects.

Motivation could also have been related to the performances of students. This would mean that there is one large latent class of schools that was motivated to participate (highstakes schools) and a small latent class of schools that was less motivated to participate in the study (low-stakes schools). The negative uniform item bias at school level could then easily be explained: students from schools in the second class had lower probabilities of answering the items correctly because they were less stimulated to do well by their school. This could be an interesting direction for future research.

CHAPTER 6

Discussion

In micro-macro multilevel analysis, a group-level outcome is explained by individual- and group-level predictors. The aim of the current dissertation was to contribute to the development of statistical methods for micro-macro analysis by making a latent variable approach for continuous data (Croon & van Veldhoven, 2007) applicable to discrete data. To reach this goal, latent class models were presented in which a latent discrete group-level variable was used to aggregate the discrete individual-level data to the group level. This group-level latent class variable can be related to other group-level predictors and group-level outcomes in a group-level analysis.

Chapter 2 showed that, for nominal data, this latent class approach outperforms more traditional methods such as aggregating the individual-level data with a manifest mean or mode. The simulation study in this chapter showed that the estimates of the group-level parameters were unbiased when a latent class model was used. When the manifest group means or modes were used, these estimates were biased, because measurement and sampling error in the aggregated scores was not accounted for. Measurement error can be caused by group members providing imperfect information about their group and sampling error can occur when information is not gathered from all group members. The more error the group means and modes contained, the more biased the group-level estimates were. When the unrealistic situation occurs that there is no error at all, the manifest approaches provided unbiased results as well. It was also shown that the power and Type-I error rates of the group-level estimates were reasonable good with the latent variable approach. This was not the case when the group-level outcome was disaggregated to the individual level and therefore this method should be avoided at any time. Besides providing unbiased estimates for the group-level parameters, the latent class approach solves another specific issue with respect to discrete data, namely that it is unclear how to aggregate a discrete variable. Using the group means has only a substantive meaning for dichotomous data since the mean can be interpreted as a proportion. For discrete data with more than two categories, the group mode seems more appropriate than the group mean, but still measurement and sampling error is not accounted for. Altogether, Chapter 2 showed that, in basic models, the latent class approach provides an elegant way to aggregate discrete data to the group level while taking measurement and sampling error into account.

In Chapter 3, more advanced models were proposed to simultaneously aggregate multiple individual-level variables to the group level by means of a discrete latent group-level variable. In the first model, the group-level latent variable affected the multiple individual-level variables directly, but in the second method this was modelled indirectly via an individual-level latent variable. The key difference between the methods is that the within and between part of the (co)variation in the individual-level variables are independent in the first method and dependent in the second method.

Instead of estimating the full models in one step, a more practical stepwise estimation procedure was proposed in Chapter 4 that separates the aggregation from the group-level analysis. In the first step, a latent class model was estimated in which the scores on the individual-level variable served as indicators for a discrete latent group-level variable. In the second step, the individual data were aggregated by assigning the groups to the group-level latent classes based on the model from the first step. In the final step, the assigned group scores were related to the group-level outcome, while correcting for the classification errors made during the aggregation step. These classification errors were known from the previous steps. As long as the correction in the final step was used, the group-level parameters were unbiased. When multiple latent group-level variables were specified, unbiased estimates were only obtained when a single model for both group-level latent variables was estimated in the first step of the analysis since then a residual within-group association among the individual-level variables could be included.

In Chapter 5, an application from educational testing was presented. Schools were classified into discrete latent classes based on student-level attainment items and this can be interpreted as a micro-macro situation in which the group-level outcome is latent and the individual-level variables are observed. Student ability was modelled and controlled for school-level processes and, at the same time, the multilevel structure of the data was correctly handled. An additional benefit of the models presented is that the latent group-level classes could also be used to detect uniform and nonuniform school-level item bias. This could be useful for schools to improve their teaching on specific topics covered by the items showing school-level item bias.

To conclude, a very flexible framework for micro-macro multilevel analysis for discrete data was developed that is able to handle multiple individual-level predictors and multiple latent group-level variables. The estimation can be done in either one or three steps and the applications are very diverse. Since the model fits within a more general framework, software is already available and all analysis can be performed in Latent GOLD 5.0 (Vermunt & Magidson, 2013).

A limitation of the method is that class separation, measured by the entropy R-square value, should be sufficient ($R^2_{entr} \geq .45$). Otherwise classes are estimated to be more different than they truly are and this causes bias in the group-level estimates. This is not only an issue when stepwise estimation is used, but a general issue in latent class analysis (Galindo-Garre & Vermunt, 2006). Besides the practical issues related to low class separation, it should also be considered whether it is warranted from a theoretical point of view to aggregate individual-level data to the group level when the individual-level data are only weakly related to the group-level latent classes. Since class separation is a function of the number of indicators and in micro-macro analysis the number of indicators equals the number of individuals within a group, class separation is never a problem when the sample size at the individual-level is large.

Five points are important to discussion. First, in the simulation study in Chapter 2 on basic micro-macro models, the number of groups seemed more important to obtain sufficient power for the estimates of the group-level effects than the number of individuals within a group. In the conditions with 200 groups, power was sufficient. Since the simulation study was limited, future research could be devoted to power calculations to obtain minimum sample sizes for the current type of analysis. Muller, LaVange, Landesman Ramey, and Ramey (1992), Tu, Kowalski, Zhang, Lynch, and Crits-Christoph (2004), and Snijders (2005) already started to study power in (multilevel) latent class models, but did not include a group-level outcome.

Second, in the current chapters, all individuals were treated as equivalent sources of information about the group-level variable. This exchangeability assumption is warranted when all group members play similar or identical roles in the group but is probably less vindicated when the group members differ with respect to their functioning in the group. For example, when data from multiple family members are aggregated to the family level, the information from the parents and siblings could also be treated differently. This implies that all family members do not contribute equally to the (latent) family-level score.

Third, most models discussed in the dissertation contain an indirect effect of a group-level predictor (X_j) on a group-level outcome (Y_j) via a group-level latent variable (ζ_j), but no formal test for this indirect effect was provided nor available. The two paths that form the indirect effect were only tested separately. Future research could be devoted to develop statistical tests for an indirect effect with categorical variables. This line of research is already started by Pearl (2012) and Vansteelandt (2012).

Fourth, only within-group models with discrete variables were discussed and future research could look into hybrids of continuous and discrete (latent) variables in this part of the model as discussed in Chapter 2. For example, one might assume that a discrete latent variable at the group level underlies a continuous observed variable at the individual level by using a latent profile model (Bartholomew & Knott, 1999). In situations in which the observed explanatory variables at the individual level are discrete, also an item response model (Embretson & Reise, 2000) with a continuous latent group-level variable might be considered for the aggregation. Examples are provided in Fox and Glas (2003) and Fox (2005). A more general framework for combining discrete and continuous latent variables at different levels of an hierarchical model is discussed in Vermunt (2008), Palardy and Vermunt (2010) and Varriale and Vermunt (2012).

Last, only two-level models were studied, while it might be interesting to add time as a third level so longitudinal micro-macro research questions can be answered. For example, individual-level data can be measured over time so that time points are nested within individuals and the individuals are nested within groups. Another example would be to observe the group-level outcome over time. This makes it possible to study whether micro-macro relationships change over time. At the moment, there are no such methods available.

APPENDIX A

Two Equivalent Estimation Procedures

This appendix shows how the likelihood function of the 1-2 model as defined in Equation 2.4 can be constructed in two equivalent ways, that is, with the 'Two-level regression approach' and with the 'Persons-as-Variables approach' (Curran, 2003; Mehta & Neale, 2005).

Two-Level Regression

The Two-level regression approach is illustrated in Figure 2.1. In practice, the group-level variables are treated as individual-level variables but the group-level score of a particular group is assigned to a single individual from that group, while the scores of the other individuals within that group on this variable are defined as missing. Note that this is not the same as disaggregating the group-level variable since that would come down to assigning the group score to each and every group member. Since the individuals within the same group are exchangeable, it does not matter to which individual the group-level score is assigned, but for convenience it will be assumed here that assignment is to the first individual in a group.

The data are stored in a long file in which each row of the data matrix corresponds to an individual, but an additional group identification variable is defined that indicates to which group an individual belongs. The group-level outcome defined as an individual-level variable is denoted by Y_{ij}^*, so that $Y_{ij}^* = Y_j$ for $i = 1$, and Y_{ij}^* is missing for $i \neq 1$. The variables originally measured at the individual level are simply reproduced in the data matrix. Table A.1 provides an example data matrix with three groups, the first two groups consisting of three individuals, and the third group of two individuals.

The joint density of $\mathbf{Z}_j$, $\mathbf{Y}_j^*$ and ζ_j equals

Table A.1: Example Data Matrix Two-Level Regression Approach

Groupid	Y^*_{ij}	Z_{ij}
1	Y_1	Z_{11}
1	.	Z_{21}
1	.	Z_{31}
2	Y_2	Z_{12}
2	.	Z_{22}
2	.	Z_{32}
3	Y_3	Z_{13}
3	.	Z_{23}

$$\begin{aligned} P(\mathbf{Z}_j, \mathbf{Y}^*_j, \zeta_j) &= P(\zeta_j)\left\{\prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j)P(Y^*_{ij}|\zeta_j)\right\} \qquad \text{(A.1)} \\ &= P(\zeta_j)\left\{\prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j)\right\}\left\{\prod_{i=1}^{I_j} P(Y^*_{ij}|\zeta_j)\right\} \\ &= P(\zeta_j)\left\{\prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j)\right\} P(Y^*_{1j}|\zeta_j)\left\{\prod_{i=2}^{I_j} P(Y^*_{ij}|\zeta_j)\right\}. \end{aligned}$$

Aggregating over the missing values $Y^*_{2j}, Y^*_{3j}, ..., Y^*_{I_j j}$ in applying full information maximum likelihood yields

$$\begin{aligned} P(\mathbf{Z}_j, Y^*_{1j}, \zeta_j) &= P(\zeta_j)\left\{\prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j)\right\} P(Y^*_{1j}|\zeta_j)\left\{\prod_{i=2}^{I_j}\sum_{Y^*_{ij}} P(Y^*_{ij}|\zeta_j)\right\} \\ &= P(\zeta_j)\left\{\prod_{i=1}^{I_j} P(Z_{ij}|\zeta_j)\right\} P(Y^*_{1j}|\zeta_j). \qquad \text{(A.2)} \end{aligned}$$

The latter simplification follows from the fact that $\sum_{Y^*_{ij}} P(Y^*_{ij}|\zeta_j) = 1$. Since $Y^*_{1j} = Y_j$, this is equivalent to the log-likelihood function described Equation 2.4.

Persons-as-Variables

The Persons-as-Variables approach is illustrated in Figure A.1 for the case that each group consists of maximum three members.

Each of the three individuals within a group defines a different variable at the group level and as a consequence, there are as many 'person variables' as there are individuals in the groups. A separate equation is needed to describe the relationship between each person variable and ζ_j. Since the individuals from the same group are assumed to be

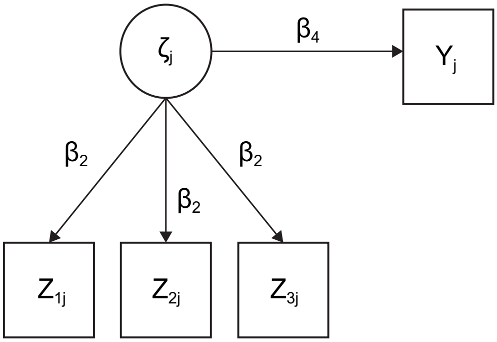


Figure A.1: Persons-as-Variables Approach

Table A.2: Example Data Matrix Persons-as-Variables Approach

Y_j	Z_{1j}	Z_{2j}	Z_{3j}
Y_1	Z_{11}	Z_{21}	Z_{31}
Y_2	Z_{12}	Z_{22}	Z_{32}
Y_3	Z_{13}	Z_{23}	.

exchangeable, the relationships between the different person variables and ζ_j are required to be completely identical. As a consequence of these exchangeability constraints, it does not matter who is assigned to Z_1, who to Z_2, et cetera. This approach can still be applied with unequal group sizes: the number of person variables needed is equal to the largest group size and smaller groups have missing scores on the unused person variables. In this way, the log-likelihood function described Equation 2.4 is obtained.

The Persons-as-Variables approach requires the data matrix to be structured in a wide file format in which each of the rows represent a group, while its columns correspond to the person variables as defined above. As an example Table A.2 shows the data matrix with three groups, the first two groups consisting of three individuals and the third group of two individuals.

APPENDIX B

Latent GOLD Syntax Empirical Example

This appendix explains how the 2-1-2 model from the Latent GOLD software (Vermunt & Magidson, 2005a) by either the 'Persons-as-Variables approach' or the 'Two-level regression approach'. Estimation by the Two-level regression approach requires that the data are structured in a long file format with 8161 (# level-2 units) $\times$ 6 (# level-1 units) = 48966 rows. An indicator variable `egoid` is needed for identifying the different egos within which the alters are nested.

The relevant parts of the syntax for this approach are:

```
options
  bayes categorical=1
  missing includeall;
variables
  caseid      egoid;
  dependent   z nominal, y nominal;
  independent x nominal;
  latent      zeta nominal 2;
equations
  zeta        <- (b1)1 + (b2)x;
  y           <- (b3)1 + (b4)x + (b5)zeta + (b6)x*zeta;
  z           <- (b7)1 + (b8)zeta;
```

In the `options` section of syntax, all default settings can be accepted with two exceptions. First, `bayes categorical=1` is declared to prevent boundary solutions. Second, by default Latent GOLD applies listwise deletion of cases with missing data. For obtaining maximum-likelihood estimates with missing data, `missing includeall` should be declared. In the `variables` section the `egoid` variable should be defined as the `caseid`. In the same section a list of the dependent, independent, and latent variables should be provided. For nominal latent variables, the number of latent classes is specified after the definition of the scale type. The regression equations defining the model are formulated in the `equations` section of the syntax.

When the same model is estimated with the Persons-as-Variables approach, the results will be the same but the data file is constructed as a wide file with 8161 (level-2 units) rows corresponding to the different egos and with the columns corresponding to the variables defined on the egos and their alters. A separate equation has to be specified for each alter as shown in the part of the syntax that differs from the syntax of the Persons-as-Variables approach:

```
variables
  dependent   z1 nominal, z2 nominal, z3 nominal,
              z4 nominal, z5 nominal, z6 nominal,
              y nominal;
  independent x nominal;
  latent      zeta nominal 2;
equations
  zeta <- (b1)1 + (b2)x;
  y    <- (b3)1 +  (b4)x + (b5)zeta + (b6)x*zeta;
  z1   <- (b7)1 + (b8)zeta;
  z2   <- (b7)1 + (b8)zeta;
  z3   <- (b7)1 + (b8)zeta;
  z4   <- (b7)1 + (b8)zeta;
  z5   <- (b7)1 + (b8)zeta;
  z6   <- (b7)1 + (b8)zeta;
```

To establish that the alters are exchangeable indicators of the latent variable at the ego level, the regression coefficients are restricted to be equal using the arbitrary chosen value labels `(b7)` for the intercepts and `(b8)` for the indicator loadings. It is also possible to substitute the last six equations by `z1-z6 <- (b7)1 + (b8)zeta`.

APPENDIX C

Technical Details on the Computation of $P(W = t|\zeta = q)$

This Appendix shows how $P(W = t|\zeta = q)$ is computed using the classification information from the second step of the stepwise analysis. A more detailed description is provided by Bakk et al. (2013) and Vermunt (2010).

Let $P(\zeta_j = q|\mathbf{Z}_j)$ denote the posterior class membership probability for group j and $P(W_j = t|\mathbf{Z}_j)$ denote the probability by which a group is assumed to belong to class t given the applied assignment rule. Using the modal class assignment rule, also called modal a posterior assignment (MAP), groups are assigned to that category of W_j for which $P(\zeta_j = q|\mathbf{Z}_j)$ is largest:

$$P(W_j = t|\mathbf{Z}_j) = \begin{cases} 1 & \text{if } P(\zeta_j = t|\mathbf{Z}_j) > P(\zeta_j = s|\mathbf{Z}_j)\ \forall s \neq t \\ 0 & \text{otherwise} \end{cases} \tag{C.1}$$

Using the proportional assignment rule, each group is assumed to belong to a particular latent class with a probability equal to the posterior membership probability for the class concerned implying

$$P(W_j = t|\mathbf{Z}_j) = P(\zeta_j = t|\mathbf{Z}_j). \tag{C.2}$$

The probability of being assigned to class t conditional on belonging to the true class q, $P(W = t|\zeta = q)$, is theoretically defined as:

$$P(W = t|\zeta = q) = \frac{\sum_{\mathbf{Z}} P(\mathbf{Z})P(\zeta = q|\mathbf{Z})P(W = t|\mathbf{Z})}{P(\zeta = q)}. \tag{C.3}$$

Note that the sum is taken over all possible patterns of $\mathbf{Z}$. Because the number of possible patterns can be very large, it is more practical to take the sum over the data pattern of the groups present in the available data sets, which yields:

$$P(W = t|\zeta = q) = \frac{\sum_{j=1}^{J} P(\zeta_j = q|\mathbf{Z}_j = z_j)P(W_j = t|\mathbf{Z}_j = z_j)\frac{1}{J}}{P(\zeta_j = q)}. \tag{C.4}$$

As shown by Vermunt (2010), when the specified model is correct, the theoretical and empirical definition of $P(W = t|\zeta = q)$ provide very similar results.

APPENDIX D

Latent GOLD 5.0 Syntax Empirical Data Example

To perform a bias adjusted stepwise analysis on a micro-macro model with two micro-level predictors in Latent GOLD 5.0 (Vermunt & Magidson, 2013), the data need to be structured in a long file format with the number of rows equal to the number of individuals. An identifier variable, here labeled `id`, is needed to identify which individuals belong to which group. The scores on the group-level variables are only assigned to a single group member, for convenience the first group member. Therefore, $Y_{ij} = Y_j$ for $i = 1$ and Y_{ij} is missing for $i \neq 1$, and $X_{ij} = X_j$ for $i = 1$ and X_{ij} is missing for $i \neq 1$.

The relevant parts of the syntax for the first-step measurement model are:

```
options
   <default settings>
   outfile 'step3data.txt' classification keep y, x;
variables
   caseid id;
   dependent z1 nominal, z2 nominal;
   latent zeta1 nominal 3, zeta2 nominal 3;
equations
   zeta1 <- 1;
   zeta2 <- 1;
   z1 <- 1 + zeta1;
   z2 <- 1 + zeta2;
   z1 <-> z2;
```

To save the posterior class membership probabilities to a data file, one has to add the command `outfile 'datastep3.txt' classification` to the `options` section. The command `keep` is used to add the variables from the structural part of the model, that are not used in the first-step model, to the output dataset `datastep3.txt` as well. For the remaining part, one can use the default settings for the `options`.

In the `variables` section, the `id` variable should be defined as the `caseid`. In the same section, a list of the dependent and latent variables should be provided. For nominal latent variables, the number of latent classes is specified after the definition of the scale type.

The regression equations of the first step model are formulated in the `equations` section of the syntax. These include the equations defining the measurement part of the model (`z1 <- 1 + zeta1` and `z2 <- 1 + zeta2`), together with the intercepts for the latent variables (`zeta1 <- 1` and `zeta2 <- 1`), and an equation describing the within-group association among the micro-level predictors (`z1 <-> z2`).

The relevant parts of the syntax to estimate the bias corrected third-step structural model are:

```
options
   <default settings>
   step3 ml modal simultaneous;
variables
   dependent y nominal;
   independent x nominal;
   latent zeta1 nominal posterior=(zeta1#1 zeta1#2 zeta1#3),
          zeta2 nominal posterior=(zeta2#1 zeta2#2 zeta2#3);
equations
   zeta1 <- 1 + x;
   zeta2 <- 1 + x;
   zeta1 <-> zeta2;
   y <- 1 + zeta1 + zeta2 + x;
```

This syntax needs to be run on the data file `datastep3.txt`, that was created in the previous step of the analysis. By default, records with missing values are excluded from the analysis, which results in keeping only the first record of each group, and thus ensures that the analysis is performed at the group level. The `step3` command specifies the options to be used in the third-step analysis. These concern the correction method (either `none`, `ml`, or `bch`) and the assignment rule (`modal` or `prop`). The command `simultaneous` is needed to make sure that all equations from the `equations` section are estimated at once rather than one by one.

In the `variables` section, the dependent, independent and latent variables from the structural part need to be specified. The two latent variables are connected to the stored posterior membership probabilities from the data file using the commands `posterior=(zeta1#1 zeta1#2 zeta1#3)` and `posterior=(zeta2#1 zeta2#2 zeta2#3)`. Finally, all equations from the structural part of the model are specified under `equations`.

APPENDIX E

Mplus Syntax Empirical Data Example

A similar bias adjusted stepwise analysis can also be done in Mplus (Muthén & Muthén, 1998-2012), although results will not be identical since a slightly different measurement model needs to be used in the first step and a slightly different structural model needs to be used in the third step of the analysis. Additionally, only the ML procedure can be applied since the BCH correction is currently not available in Mplus.

To start, the data need to be structured in a long file format. However, it does matter whether the group-level variables are assigned to a single group member or to all group members since a new group-level data set needs to be created manually for the third step of the analysis. The relevant parts of the syntax to run the first-step measurement model on the data file just described are:

```
ANALYSIS:
TYPE = TWOLEVEL MIXTURE;

VARIABLE:
NAMES ARE id x z1 z2 y;
USEVARIABLES  id x z1 z2 y;
CLASSES = zeta1(3) zeta2(3);
CATEGORICAL = z1 z2;
BETWEEN = zeta1 zeta2;
CLUSTER = id;
AUXILIARY = x y;

MODEL:
\%WITHIN\%
\%OVERALL\%
e by z2_r@1 z3_r@1;
e*1;
```

```
\%BETWEEN\%

MODEL zeta1:
\%BETWEEN\%
\%zeta1#1\%
[z1$1* z1$2*];
\%zeta1#2\%
[z1$1* z1$2*];
\%zeta1#3\%
[z1$1* z1$2*];

MODEL zeta2:
\%BETWEEN\%
\%zeta2#1\%
[z2$1* z2$2*];
\%zeta2#2\%
[z2$1* z2$2*];
\%zeta2#3\%
[z2$1* z2$2*];

SAVEDATA: FILE=inddatastep3.dat;
SAVE = CPROB;
```

The first-step model is a `TWOLEVEL MIXTURE` analysis. In the `VARIABLE` section, the latent variables and the number of latent classes are defined after the command `CLASSES` and the indicators are defined after the command `CATEGORICAL`. It did not seem possible to define the indicators as nominal; therefore, a cumulative logit model is used instead of a multinomial model. It needs to be stated that the latent variables are group-level variables after the command `BETWEEN`. The `id` variable is defined as the `CLUSTER` variable, and with the command `AUXILIARY`, the structural variables are saved in the file `inddatastep3.dat`.

The equations from the measurement model are stated in the `MODEL` section. It is not possible to specify a within association among the discrete micro-level predictors; therefore, a continuous latent variable at the individual level is defined with the two micro-level predictors as indicators. Separate models are specified for the group-level latent variables in which the last categories of the indicator variables are used as reference categories.

Last, the commands `SAVEDATA: FILE=inddatastep3.dat;` and `SAVE = CPROB;` create a disaggregated individual-level dataset that contains the observed group-level variables needed in the structural model, the group-level posterior probabilities, and the group-level modal assignments. Since the posterior probabilities and the assigned scores are disaggregated, a new group-level data file needs to be created. Two dummy variables for the three categories of `x` that are to be used as independent variables in the structural model are added to this dataset. Also, the classification table needs to be constructed manually in the way that is explained in Appendix C. We used the following script in R (version 3.0.1) to do this:

```
###Make a group-level dataset for third-step analysis
```

```
inddata<-as.matrix(read.table("inddatastep3.dat"))
# colnames: z1 z2 y x p11 p12 p13 p21 p22 p23 p31 p32 p33 w1 w2 w g

groupdata<-matrix(0,length(table(inddata[,17])),ncol(inddata))
groupdata[1,]<-inddata[1,]
z<-1

for (i in 2:nrow(inddata)){
if(inddata[i,17]!=inddata[(i-1),17]){
z<-z+1
groupdata[z,]<-inddata[i,]}}

#Make two dummy variables for x with last category as reference category

dummies<-matrix(0,nrow(groupdata),2)

for (i in 1:nrow(groupdata)){
ifelse(groupdata[i,4]==1,dummies[i,1]<-1,dummies[i,1]<-0)
ifelse(groupdata[i,4]==2,dummies[i,2]<-1,dummies[i,2]<-0)}

groupstep3<-cbind(groupdata[,c(3,4,14,15,17)],dummies)
# colnames: y x w1 w2 g xd1 xd2
write.table(groupstep3,"step3.dat", row.names = FALSE,col.names = FALSE)

### Construct classification tables (based on modal assignment)

##zeta1
dat1<-matrix(0,nrow(groupdata),16)
# P(zeta1|z1)
dat1[,1]<-(groupdata[,5]+groupdata[,6]+groupdata[,7])
dat1[,2]<-(groupdata[,8]+groupdata[,9]+groupdata[,10])
dat1[,3]<-(groupdata[,11]+groupdata[,12]+groupdata[,13])
#w1
dat1[,4]<-groupdata[,14]
# P(w1|z1)
for (i in 1:nrow(groupdata)){
ifelse(dat1[i,4]==1,dat1[i,5]<-1,dat1[i,5]<-0)
ifelse(dat1[i,4]==2,dat1[i,6]<-1,dat1[i,6]<-0)
ifelse(dat1[i,4]==3,dat1[i,7]<-1,dat1[i,7]<-0)}

dat1[,8:10]<- c(dat1[,1]*dat1[,5], dat1[,2]*dat1[,5], dat1[,3]*dat1[,5])
dat1[,11:13]<-c(dat1[,1]*dat1[,6], dat1[,2]*dat1[,6], dat1[,3]*dat1[,6])
dat1[,14:16]<-c(dat1[,1]*dat1[,7], dat1[,2]*dat1[,7], dat1[,3]*dat1[,7])

t<-matrix(0,3,3)
t[1:3,1]<-c(sum(dat1[,8]),sum(dat1[,9]),sum(dat1[,10]))
```

```
t[1:3,2]<-c(sum(dat1[,11]),sum(dat1[,12]),sum(dat1[,13]))
t[1:3,3]<-c(sum(dat1[,14]),sum(dat1[,15]),sum(dat1[,16]))
tt<-cbind(t,as.matrix(apply(t,1,sum)))

Dzeta1<-matrix(0,3,3)
Dzeta1[1,]<-tt[1,1:3]/tt[1,4]
Dzeta1[2,]<-tt[2,1:3]/tt[2,4]
Dzeta1[3,]<-tt[3,1:3]/tt[3,4]

res1<-c(log(Dzeta1[1,1]/Dzeta1[1,3]), log(Dzeta1[1,2]/Dzeta1[1,3]),
log(Dzeta1[2,1]/Dzeta1[2,3]), log(Dzeta1[2,2]/Dzeta1[2,3]),
log(Dzeta1[3,1]/Dzeta1[3,3]), log(Dzeta1[3,2]/Dzeta1[3,3]))

##zeta2
dat2<-matrix(0,nrow(groupdata),16)
dat2[,1]<-(groupdata[,5]+groupdata[,8]+groupdata[,11])
dat2[,2]<-(groupdata[,6]+groupdata[,9]+groupdata[,12])
dat2[,3]<-(groupdata[,7]+groupdata[,10]+groupdata[,13])
dat2[,4]<-groupdata[,15]
for (i in 1:nrow(groupdata)){
ifelse(dat2[i,4]==1,dat2[i,5]<-1,dat2[i,5]<-0)
ifelse(dat2[i,4]==2,dat2[i,6]<-1,dat2[i,6]<-0)
ifelse(dat2[i,4]==3,dat2[i,7]<-1,dat2[i,7]<-0)}

dat2[,8:10]<- c(dat2[,1]*dat2[,5], dat2[,2]*dat2[,5], dat2[,3]*dat2[,5])
dat2[,11:13]<-c(dat2[,1]*dat2[,6], dat2[,2]*dat2[,6], dat2[,3]*dat2[,6])
dat2[,14:16]<-c(dat2[,1]*dat2[,7], dat2[,2]*dat2[,7], dat2[,3]*dat2[,7])

q<-matrix(0,3,3)
q[1:3,1]<-c(sum(dat2[,8]),sum(dat2[,9]),sum(dat2[,10]))
q[1:3,2]<-c(sum(dat2[,11]),sum(dat2[,12]),sum(dat2[,13]))
q[1:3,3]<-c(sum(dat2[,14]),sum(dat2[,15]),sum(dat2[,16]))
qq<-cbind(q,as.matrix(apply(q,1,sum)))

Dzeta2<-matrix(0,3,3)
Dzeta2[1,]<-qq[1,1:3]/qq[1,4]
Dzeta2[2,]<-qq[2,1:3]/qq[2,4]
Dzeta2[3,]<-qq[3,1:3]/qq[3,4]

res2<-c(log(Dzeta2[1,1]/Dzeta2[1,3]), log(Dzeta2[1,2]/Dzeta2[1,3]),
log(Dzeta2[2,1]/Dzeta2[2,3]), log(Dzeta2[2,2]/Dzeta2[2,3]),
log(Dzeta2[3,1]/Dzeta2[3,3]), log(Dzeta2[3,2]/Dzeta2[3,3]))
```

The vectors res1 and res2 contain the logit parameters needed in the third-step analysis for which the relevant parts of the syntax are:

```
DATA:
FILE IS step3.dat;
```

```
ANALYSIS:
TYPE = MIXTURE;

VARIABLE:
NAMES ARE y x w1 w2 g xd1 xd2;
USEVARIABLES  y w1 w2 xd1 xd2;
CLASSES = zeta1(3) zeta2(3);
NOMINAL = y w1 w2;

MODEL:
\%OVERALL\%
y on zeta1 zeta2 xd1 xd2;
zeta1 on xd1 xd2;
zeta2 on xd1 xd2;
zeta2 on zeta1;

MODEL zeta1:
\%zeta1#1\%
[w1#1@1.8941224];
[w1#2@-2.5387260];

\%zeta1#2\%
[w1#1@4.4133758];
[w1#2@3.6361397];

\%zeta1#3\%
[w1#1@-0.8973495];
[w1#2@-10.3883285];

MODEL zeta2:
\%zeta2#1\%
[w2#1@6.6811118];
[w2#2@6.2972301];

\%zeta2#2\%
[w2#1@0.4932635];
[w2#2@2.8635253];

\%zeta2#3\%
[w2#1@-5.1415522];
[w2#2@-0.2323258];
```

This syntax needs to be run on the data file `step3.dat` that was created with the R-script presented and defines a single-level `MIXTURE` analysis. In the `VARIABLES` section, the nominal dependent variable `y` and the assigned scores that function as indicators with known measurement error `w1` and `w2` need to be specified as `NOMINAL`. In the `MODEL` section, the structural equations are defined.

Due to differences in software implementations, there are two differences with the third-step model used in Latent GOLD. First, the model for `y` is saturated and contains the interaction effect of `zeta1` and `zeta2`, while in Latent GOLD only main effects are included. Second, it is not possible to include an association among two categorical variables; therefore, `zeta1` is regressed on `zeta2`.

APPENDIX F

Example Items

1. At what age did David start playing chess?
A 5
B 9
C 11

KING DAVID
David Howell, now 11, was nine when he broke the world record for the youngest player to beat a grandmaster in an official game. The youngster beat Dr John Nunn in a game of 'blitz' chess. David started playing at the age of five when he got a second-hand chess set. Soon he began beating his father and became British under-9, under-10 and under-11 champion.
Source: Funday Times

17. What does this advertisement point out?
The fact that
A the export of tropical wood from Ghana needs to be confined.
B Ghana needs a lot of development funds to renew the forests.
C Ghana handles the cutting and export of wood in a responsible manner.
D Ghana puts a lot of effort into countering illegal cutting down of wood.

OUR FORESTS ARE OUR CHILDREN'S FUTURE
Ghana's permanent Forest Reserves provide an annual sustainable timber Harvest of 1.2 million cubic meters. Selective logging systems and fallow periods offer renewable wood supplies as well as environmental and social benefits.

Whilst some forest has to give way to other needs, our permanent reserves are a principal source of jobs, supporting over 250,000 people; they are part of Ghana's own economic and social development.

Wood is in our homes, schools, hospitals, and offices; it is used on land, sea, rivers and for fuel.

Wood experts are important too. They go to many African countries with less forest areas.

Ghana believes in the future of its forest and is making its own decisions about land use.

Some people in the developed world want to stop importing tropical wood. Why? For sure, bans do not help under-resourced tropical countries and they certainly do not ensure good forest management.

For full details of our forest management systems, please contact:
Ghana timber export development board
P.O. Box 515, Takoradi, GHANA
Tel: 239 31 2921-6
Fax: 233 31 4690

102 Park Street, London, W1Y 3RJ
UNITED KINGDOM
Tel: 0171 493 4901-4
Fax: 0171 493 9923

References

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating latent class assignments to external variables: standard errors for correct inference. *Political Analysis*.

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias adjusted three-step approaches. *Sociological Methodology*, *43*(1), 272-311.

Bank of Italy. (2012). Historical database of survey of household income and wealth, 1977-2010.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London, United Kingdom: Arnold.

Bennink, M., Croon, M. A., & Vermunt, J. K. (2013). Micro-macro multilevel analysis for discrete data: A latent variable approach and an application on personal network data. *Sociological Methods and Research*, *42*(4), 431-457.

Bentler, P. M. (2006). *EQS structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Bolck, A., Croon, M. A., & Hagenaars, J. A. P. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3-27.

Borghans, L., Meijers, H., & ter Weel, B. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, *46*(1), 2-12.

Bovaird, J. A., & Shaw, L. H. (2012). Multilevel structural equation modeling. In B. P. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (p. 501-518). New York, NY: The Guilford Press.

Croon, M. A., & van Veldhoven, M. J. P. M. (2007). Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model. *Psychological Methods*, *12*(1), 45-57.

Croon, M. A., van Veldhoven, M. J. P. M., Peccei, R., & Wood, S. (2014). Researching individual well-being and performance in context: Multilevel mediational analysis for bathtub models. In M. J. P. M. van Veldhoven & R. Peccei (Eds.), *Well-being and performance at work: The role of context.* London, UK: Psychology Press.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*(4), 529-569.

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple-goal multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology*, *89*(6), 1035-1056.

Dias, J. G., & Vermunt, J. K. (2008). A bootstrap-based aggregate classifier for model-based clustering. *Computational Statistics*, *23*(4), 643-659.

Eggebeen, D., & Dew, J. (2009). The role of religion in adolescence for family formation in young adulthood. *Journal of Marriage and Family*, *71*(1), 108-121.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.

European Commission. (2005). *The new SME definition: user guide and model declaration*. Brussels, Belgium: Publication Office.

Fox, J.-P. (2005). Multilevel IRT using dichotomous and polytomous items. *British Journal of Mathematical and Statistical Psychology*, *58*(1), 145-172.

Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. New York, NY: Springer.

Fox, J.-P., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*(2), 271-288.

Fox, J.-P., & Glas, C. A. W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*, *68*(2), 169-191.

Galindo-Garre, F., & Vermunt, J. K. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika*, *33*(1), 43-59.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, United Kingdom: John Wiley & Sons.

Goldstein, H., Bonnet, G., & Rocher, T. (2007). Multilevel structural equation models for the analysis of comparative data on educational performance. *Journal of Educational and Behavioural Statistics*, *32*(3), 252-286.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., & Thomas, S. (1993). A multilevel analysis of school examination results. *Oxford Review of Education*, *19*(4), 425-433.

Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika*, *60*(1), 179-192.

Goodman, L. A. (1974a). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*(2), 215-231.

Goodman, L. A. (1974b). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I - A modified latent structure approach. *American Journal of Sociology*, *79*(5), 1179-1259.

Goss, D. (1991). *Small business and society*. London, UK: Routledge.

Guest, D. E., & Peccei, R. (2001). Partnership at work: Mutuality and the balance of advantage. *British Journal of Industrial Relations*, *39*(2), 207-236.

Hagenaars, J. A. P. (Ed.). (1990). *Categorical longitudinal data: log-linear panel, trend, and cohort analysis*. Newbury Park, CA: Sage Publications.

Hagenaars, J. A. P., & MacCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. Cambridge, United Kingdom: Cambridge University Press.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Hox, J. J., & Roberts, J. K. (Eds.). (2011). *Handbook of advanced multilevel analysis*. New York, NY: Routledge.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.8 for windows*. Lincolnwood, IL: Scientific Software International, Inc.

Kalmijn, M., & Vermunt, J. K. (2007). Homogeneity of social networks by age and marital status: A multilevel analysis of ego-centered networks. *Social Networks*, *29*(1), 25-43.

Keith, T. Z. (2006). *Multiple regression and beyond*. Boston, MA: Pearson Education, Inc.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*(4), 418-444.

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A*, *172*(4), 835-851.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 x 2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*(4), 444-467.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203-229.

Lukočienė, O., Varriale, R., & Vermunt, J. K. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, *40*(1), 247-283.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325-337.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Erlbaum.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bullitin*, *114*(2), 376-390.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: John Wiley & Sons.

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259-284.

Muller, K. E., LaVange, L. M., Landesman Ramey, S., & Ramey, C. T. (1992). Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association*, *87*(420), 1209-1226.

Murphy, K. M., & Topel, R. H. (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics*, *3*(4), 88-97.

Muthén, B. O., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced mulilevel analysis* (p. 15-40). New York, NY: Taylor and Francis.

Muthén, B. O., & Asparouhov, T. (in press). Item response modeling in Mplus: A multidimensional, multi-level, and multi-timepoint example. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of item response theory: Models, statistical tools, and applications.* Boca Raton, FL: Chapman & Hall/CRC Press.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Palardy, G., & Vermunt, J. K. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics*, *35*(5), 532-565.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.

Pearl, J. (2012). The mediation formula: A guide to the assessment of causal pathways in nonlinear models. In C. Berzuini, P. Dawid, & L. Bernardinelli (Eds.), *Causality: Statistical perspectives and applications* (p. 151-179). West Sussex, United Kingdom: John Wiley & Sons, Ltd.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209-233.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*(2), 167-190.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2005). *MLwiN version 2.02*. Bristol, United Kingdom: Centre for Multilevel Modelling, University of Bristol.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Rutter, M., & Maughan, B. (2002). School effectiveness findings 1979-2002. *Journal of School Psychology*, *40*(6), 451-475.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC Press.

Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, p. 1570-1573). Hoboken, NJ: John Wiley & Sons, Inc.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, United Kingdom: Sage Publications.

Snijders, T. A. B., Spreen, M., & Zwaagstra, R. (1995). The use of multilevel modeling for analysing personal networks: Networks of cocaine users in an urban area. *Journal of Quantitative Anthropology*, *5*(2), 85-105.

Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., . . . Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, *29*(1), 11-36.

Tu, X. M., Kowalski, J., Zhang, J., Lynch, K. G., & Crits-Christoph, P. (2004). Power analyses for longitudinal trials and other clustered designs. *Statistics in Medicine*, *23*(18), 2799-2815.

van Duijn, M. A. J., van Busschbach, J. T., & Snijders, T. A. B. (1999). Multilevel analysis of personal networks as dependent variables. *Social Networks*, *21*(2), 187-210.

van Veldhoven, M. J. P. M. (2005). Financial performance and the long-term link with

HR practices, work climate and job stress. *Human Resource Management Journal*, *15*(4), 30-53.

Van Veldhoven, M. J. P. M., Meijman, T., & Broersen, S. (2002). *Handleiding VBBA: onderzoek naar de beleving van psychosociale arbeidsbelasting en werkstress met behulp van de vragenlijst beleving en beoordeling van de arbeid*. Amsterdam, the Netherlands: Stichting Kwaliteitsbevordering Bedrijfsgezondheidszorg.

Vansteelandt, S. (2012). Estimation of direct and indirect effects. In C. Berzuini, P. Dawid, & L. Bernardinelli (Eds.), *Causality: Statistical perspectives and applications* (p. 126-150). West Sussex, United Kingdom: John Wiley & Sons, Ltd.

Varriale, R., & Vermunt, J. K. (2012). Multilevel mixture factor models. *Multivariate Behavioral Research*, *47*(2), 247-275.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, *33*(1), 213-239.

Vermunt, J. K. (2005). Mixed-effects logistic regression models for indirectly observed discrete outcome variables. *Multivariate Behavioral Research*, *40*(3), 281-301.

Vermunt, J. K. (2008). Multilevel latent variable modeling: An application in educational testing. *Austrian Journal of Statistics*, *37*(3&4), 285-299.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450-469.

Vermunt, J. K., & Magidson, J. (2005a). *Latent GOLD 4.0 user's guide*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2005b). *Technical guide for Latent GOLD 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2013). *LG-syntax user's guide: Manual for Latent GOLD 5.0 syntax module*. Belmont, MA: Statistical Innovations.

Waller, M. J., Conte, J. M., Gibson, C. B., & Carpenter, M. A. (2001). The effect of individual perceptions of deadlines on team performance. *The Academy of Management Review*, *26*(4), 586-600.

Whisman, M. A., & McClelland, G. H. (2005). Designing, testing, and interpreting interactions and moderator effects in family research. *Journal of Family Psychology*, *19*(1), 111-120.

Wilkinson, A. (1999). Employment relations in SME's. *Employee Relations*, *21*(3), 206-217.

Zijsling, D., Keuning, J., Naayer, H., & Kuyper, H. (2012). *Cohortonderzoek COOL5-18: Technisch rapport meting VO-3 in 2011.* (Tech. Rep.). Groningen, the Netherlands: GION.

Summary

This dissertation deals with multilevel models for predicting outcomes at the group level from predictors measured at the individual level. This form of multilevel analysis, which is rather common in social sciences, is referred to as micro-macro analysis. Croon and Van Veldhoven already proposed a latent variable model for micro-macro analysis with continuous (latent) variables. The aim of the current project was to generalize their approach so that it can also be applied to discrete (latent) variables, by using a latent class model instead of a factor-analytic model.

Characteristic of the latent class approach is that the scores of the group members on an individual-level variable are used as exchangeable indicators of a discrete latent class variable at the group level. This part of the model is referred to as the within-group model and makes it possible to aggregate a discrete individual-level variable to the group level while accounting for measurement and sampling error in the aggregated scores. In the between-group model, the group-level latent classes are related to other (manifest) group-level variables. For example, a group-level outcome can be regressed on the group-level latent classes.

In Chapter 2, this approach was presented in more detail by applying it to a simple model with a single dichotomous individual-level predictor and a single dichotomous group-level outcome. In a simulation study, the method was compared to two more traditional approaches, namely aggregation of the individual-level scores with a (manifest) mean or mode and disaggregation of the group-level outcome. Both would imply to perform a single-level analysis at either the group or individual level. Mean aggregation is only substantively meaningful when the individual-level variable is dichotomous since the mean can be interpreted as a proportion. For nominal variables with more than two categories, a group mean has no substantive meaning and using the group mode instead might be an alternative. A simulation study showed that the latent class approach performed better than the more traditional procedures, since the latent class approach obtained unbiased group-level parameters even when there was measurement and sampling error in the latent group-level scores. Also the power and Type-I error rate of the tests of the estimates of the group-level parameters were reasonably good. Aggregation with a mean or mode only worked when there was no measurement and sampling error, which is rather unrealistic in practice. Disaggregation provided Type-I error rates that were unacceptably high. The simple model was extended to a more complex model by adding a manifest group-level predictor to the model. A second simulation study showed that the proposed latent variable approach still worked well in this more complex situation, but a larger

number of groups was needed to retain sufficient power. Especially for an interaction effect between the group-level predictor and the group-level latent classes, power was low in most conditions. The latent variable approach was illustrated with an empirical example in which data from a personal network were used.

The models from Chapter 2 are limited in the sense that they contain only a single individual-level variable while in social research it is very common to use multiple observed individual-level variables or to use individual-level constructs that are indirectly measured with multiple items. In Chapter 3, two latent class models were presented that can be used for micro-macro analysis with multiple individual-level variables. In the first model, the individual-level variables were used directly as indicators for the group-level latent class variable. In this way, a typology of groups was constructed based on the individual-level variables. It was unrealistic to assume that the individual-level variables only contained between-group variation. Therefore, associations among the individual-level variables were used to capture the residual within-group association among the individual-level variables. In the second model, the multiple individual-level variables intended to measure an individual-level construct. In this situation, it is not very intuitive to model the within-group association among the micro-level variables as a residual. It is more appropriate to treat the micro-level variables as indicators for an individual-level latent variable on which groups can differ. This implies that the group-level latent variable had an indirect effect on the individual-level variables. The most important difference between the two models is that the within and between component of the (co)variation in the individual-level variables were independent in the first situation and were dependent in the second. A real data example was provided to illustrate both situations. For the first situation an example from marketing research was used and for the second situation an example from research to small firm human resource practices.

Thus far, all models were estimated in one step using maximum-likelihood estimation. In Chapter 4, a more practical stepwise estimation procedure of the micro-macro latent class model was proposed in which the latent class model that was used to aggregate the discrete individual-level variables (within-group model) and the group-level analysis (between-group model) can be estimated separately without introducing bias in the estimates of the group-level parameters. In the first step, a latent class model was estimated in which the scores on a discrete individual-level predictor were used as indicators for a group-level latent class variable (within-group model). In the second step, this latent class model was used to aggregate the individual-level predictor to the group level by assigning the groups to the latent classes. In the final step, a group-level analysis was performed in which the aggregated measures were related to the remaining group-level variables while correcting for the classification error in the class assignments (between-group model). Two methods were presented to do so, both using the same underlying model. These methods were compared to the one-step latent class approach presented in Chapter 2 and provided estimates that were as good as the estimates from the one-step analysis, as long as the class separation was sufficient. A second simulation study showed that when a model contains two group-level latent variables with one individual-level variable as an indicator each, the within-group association among these individual-level predictors needed to be modelled in the within-group model. The within-group association could only be ignored when the individual-level scores were very good indicators of the group-level latent variables and the within-group association was small. At the end of the chapter, the stepwise approach was applied to an empirical data example in which *team* *productivity* (macro level) was explained by *job control* (micro level), *job satisfaction* (micro level), and *enriched job design* (macro level).

Chapter 5 showed an application from educational measurement in which test data on students were used to classify schools. This is a micro-macro situation since individual-level data are used to explain a group-level characteristic, but the group-level variable of interest is a latent class variable instead of an observed variable. As in a regular one-level Item Response Theory model, a continuous latent trait was assumed to underlie the item scores of the students and this latent trait was interpreted as student ability. Since the scores were collected from students from different schools, a multilevel model was needed to adjust for the dependency among students from the same school. This was done by decomposing student ability into its within and between components. It was expected that all variability in the item scores that remained after controlling for student and school ability could be attributed to the latent classification of schools. At the same time, the latent group-level classes could be used to study uniform and nonuniform school-level item bias. This can be useful to improve school performance as it clarifies which items functioned differently at which type of school, regardless of the ability level of the students at the schools. For instance, schools from school types in which some items functioned differently, could devote more time to teaching the topics covered by these items. Because the models simultaneously (1) accounted for the nested structure of the data, (2) controlled student ability for processes at school level, (3) classified schools to monitor and compare schools, and (4) tested for school-level item bias, the model was rather complex. When applying the models to data from the Dutch study COOL$^{5-18}$, in which the achievements of five- to eighteen-year-old students are studied on the basis of a test with dichotomous educational items, the complex models provided better global fit than simpler alternative models.

To conclude, a flexible method that handles micro-macro situations with discrete data was provided. A broad range of applications was shown and when not straightforward, syntax was provided to stimulate the use of the method.

Samenvatting (Summary in Dutch)

Dit proefschrift gaat over multilevelmodellen die toegepast kunnen worden om uitkomsten op groepsniveau (bijvoorbeeld teamprestatie) te voorspellen op basis van predictors die gemeten zijn op het niveau van individuen (bijvoorbeeld de motivatie en vaardigheden van werknemers). Dit veelvoorkomende type multilevelanalyse wordt micro-macroanalyse genoemd. Croon en van Veldhoven hebben een latentevariabelemodel voorgesteld voor micro-macroanalyse met continue (latente) variabelen. Het doel van dit proefschrift is hun methode te generaliseren zodat deze ook toegepast kan worden op discrete (latente) variabelen. Hiervoor wordt een latenteklassenmodel in plaats van een factormodel gebruikt.

De scores van de groepsleden op een variabele die is gemeten op het individuele niveau worden gebruikt als inwisselbare indicatoren van een discrete latenteklassenvariabele op groepsniveau. Dit deel van het model wordt het 'binnengroepenmodel' genoemd en zorgt ervoor dat het mogelijk is om een discrete variabele te aggregeren naar het groepsniveau, terwijl rekening wordt gehouden met de meet- en steekproeffouten in de geaggregeerde scores. In het 'tussengroepenmodel' worden de latente klassen op groepsniveau gerelateerd aan andere (geobserveerde) variabelen op groepsniveau. Er kan bijvoorbeeld een regressieanalyse uitgevoerd worden met de uitkomst op het groepsniveau als afhankelijke variabele en de latente klassen op het groepsniveau als onafhankelijke variabele.

In Hoofdstuk 2 wordt de methode in meer detail besproken en toegepast op een eenvoudig model met een dichotome predictor op het individuele niveau en een dichotome uitkomst op het groepsniveau. In een simulatiestudie wordt de methode vergeleken met twee meer traditionele methoden, namelijk het aggregeren van de scores op het individuele niveau met een (geobserveerd(e)) gemiddelde of modus en het disaggregeren van de groepsuitkomst. Beide methoden impliceren een analyse op één niveau, namelijk of op groepsniveau of op het individuele niveau. Aggregeren met een gemiddelde is alleen inhoudelijk zinvol wanneer de individuele variabele dichotoom is, omdat het gemiddelde dan geïnterpreteerd kan worden als een proportie. Voor nominale variabelen met meer dan twee categorieën heeft het gemiddelde geen inhoudelijke betekenis en is het gebruiken van de modus een alternatief. De simulatiestudie laat zien dat de latenteklassenmethode beter presteerde dan de meer traditionele methoden, omdat zuivere schatters van de parameters op het groepsniveau verkregen werden, zelfs als er meet- en steekproeffouten in de latente klassen op het groepsniveau zaten. Ook de power en Type-I fouten van de testen voor de schatters van de groepsparameters waren goed. Aggregeren met een gemiddelde of modus werkte alleen goed wanneer er geen meet- en steekproeffouten waren, wat in de praktijk zeer onrealistisch is. Disaggregeren gaf onacceptabel hoge Type-I fouten en kan daarom het beste vermeden worden. Het eenvoudige model is uitgebreid naar een complexer model door een geobserveerde predictor op het groepsniveau toe te voegen. Een tweede simulatiestudie laat zien dat de voorgestelde latentevariabelemethode nog steeds goed werkte in deze complexere situatie, maar er waren meer groepen nodig om voldoende power te hebben. Vooral voor een interactie-effect tussen de predictor op groepsniveau en de latente klassen op groepsniveau, was de power laag in de meeste condities. De latenteklassenmethode wordt geïllustreerd met een empirisch voorbeeld waarin data over persoonlijke netwerken werden gebruikt.

De modellen uit voorgaand hoofdstuk zijn beperkt in die zin dat ze maar één variabele op het individuele niveau bevatten, terwijl het binnen de sociale wetenschappen vaak voorkomt dat er gebruik wordt gemaakt van meerdere geobserveerde variabelen op het individuele niveau of van individuele constructen die indirect gemeten zijn door middel van meerdere items. In Hoofdstuk 3 worden twee latenteklassenmodellen gepresenteerd die gebruikt kunnen worden voor micro-macroanalyse met meerdere variabelen op het individuele niveau. In het eerste model worden de individuele variabelen direct gebruikt als indicatoren voor de latenteklassenvariabele op groepsniveau. Op deze manier werd een typologie van groepen geconstrueerd die gebaseerd is op de individuele variabelen. Het is onrealistisch om aan te nemen dat de variabelen op het individuele niveau alleen (co)variatie op het individuele niveau bevatten. Daarom werd de (co)variatie binnen de individuele variabelen opgenomen door middel van associaties op het individuele niveau. In het tweede model waren de individuele variabelen ontworpen om een construct op het individuele niveau te meten. In deze situatie is het minder intuïtief om de samenhang tussen de individuele variabelen te modelleren als een residu. Het is gepaster om de individuele variabelen te gebruiken als indicatoren voor een individueel construct, waar groepen op kunnen verschillen. Dit houdt in dat de variabele op het groepsniveau een indirect effect had op de variabelen op het individuele niveau. Het belangrijkste verschil tussen de twee modellen is, dat de samenhang tussen groepen en de samenhang binnen groepen van de variabelen op het individuele niveau onafhankelijk zijn in het eerste model en afhankelijk in het tweede. Een empirisch voorbeeld wordt gegeven om beide situaties te illustreren. Voor het eerste voorbeeld werd data uit marktonderzoek gebruikt en voor het tweede voorbeeld data uit onderzoek naar humanresourcepraktijken in kleine organisaties.

Tot zo ver werden alle modellen geschat in één enkele stap door middel van maximumlikelihoodschattingen. In Hoofdstuk 4 wordt een meer praktische en stapsgewijze schattingsmethode voor het micro-macro latenteklassenmodel voorgesteld waarin het latenteklassenmodel dat gebruikt werd om de discrete individuele variabelen te aggregeren (het binnengroepenmodel) en de analyse op het groepsniveau (het tussengroepenmodel) afzonderlijk van elkaar geschat kunnen worden, zonder dat de schattingen van de groepsniveauparameters onzuiver worden. In de eerste stap werd een latenteklassenmodel geschat waarin de scores op de discrete predictor op het individuele niveau gebruikt werden als indicatoren voor een latenteklassenvariabele op groepsniveau (het binnengroepenmodel). In de tweede stap werd dit latenteklassenmodel gebruikt om de predictor op het individuele niveau te aggregeren door de groepen toe te wijzen aan de latente klassen. In de laatste stap werd een groepsniveauanalyse gedaan waarin de geaggregeerde metingen werden gerelateerd aan de overige groepsniveauvariabelen, terwijl rekening gehouden werd met de meetfouten in de klassentoewijzigingen (het tussengroepenmodel). Er werden twee methoden gepresenteerd om dit te doen, die beide gebruik maken van hetzelfde onderliggende model. Deze methoden werden vergeleken met de eenstaps latenteklassenmethode uit Hoofdstuk 2 en daaruit bleek dat ze schattingen produceerden die net zo goed waren als de schattingen van de eenstapsanalyse, zolang de klassen goed van elkaar konden worden onderscheiden. Een tweede simulatie laat zien dat in een model met twee latente variabelen op groepsniveau (met elk een variabele op het individuele niveau als indicator), de samenhang tussen de predictors op het individuele niveau opgenomen moet worden in het binnengroepenmodel dat geschat wordt in de eerste stap. De samenhang tussen de predictors op het individuele niveau kon alleen genegeerd worden wanneer de individuele scores zeer sterke indicatoren van de latente variabelen op groepsniveau waren en de samenhang tussen de predictors klein was. Aan het einde van het hoofdstuk wordt de stapsgewijze methode toegepast op een empirisch datavoorbeeld waarin *team productivity* (groepsniveau) verklaard werd door *jobcontrol* (individuele niveau), *job satisfation* (individuele niveau) en *enriched jobdesign* (groepsniveau).

Hoofdstuk 5 is een toepassing vanuit onderwijskundig onderzoek waarin data die verzameld werden onder leerlingen, werden gebruikt om scholen te classificeren. Dit is een micro-macrosituatie omdat data op het individuele niveau gebruikt werden om een groepskenmerk te voorspellen, maar de groepsuitkomst waarin men geïnteresseerd is, is een latenteklassenvariabele in plaats van een geobserveerde variabele. Zoals in een standaard Item-Response-Theory-model op één niveau, werd verondersteld dat er een continue latente variabele aan de itemscores van de leerlingen ten grondslag ligt en deze werd geïnterpreteerd als de vaardigheid van de leerlingen. Omdat de scores verzameld waren bij leerlingen van verschillende scholen, was een multilevelmodel nodig om de afhankelijkheid tussen leerlingen van dezelfde school op te nemen. Dit werd gedaan door de vaardigheid van de leerlingen te splitsen in een gedeelte dat verklaard werd binnen scholen en een gedeelte dat verklaard werd tussen scholen. Aangenomen werd dat alle variatie in de itemscores die overbleef na de constanthouding van de vaardigheid van de leerlingen en scholen, kon worden toegewezen aan de latente classificering van scholen. Tegelijkertijd konden de schoolniveauklassen gebruikt worden om uniforme en niet-uniforme item bias op het niveau van de scholen te onderzoeken. Dit kan nuttig zijn bij het verbeteren van schoolprestaties omdat het duidelijk maakt welke items anders gemaakt worden op welk type school, ongeacht de vaardigheden van de leerlingen. Bijvoorbeeld, scholen van het schooltype waarin sommige items slechter werden gemaakt, kunnen meer tijd besteden aan het onderwijzen van de onderwerpen die in deze items aan bod komen. De modellen zijn vrij complex omdat ze gelijktijdig (1) rekening houden met de geneste structuur van de data, (2) de vaardigheid van de leerlingen controleren voor processen die zich afspelen op het schoolniveau, (3) de scholen classificeren om deze scholen te kunnen volgen en te kunnen vergelijken en (4) het bestuderen van item bias op schoolniveau mogelijk maakt. Wanneer de modellen werden toegepast op data uit de Nederlandse COOL$^{5-18}$ studie, waarin vijf- tot achttienjarige leerlingen bestudeerd worden op basis van een test met dichotome vaardigheidsitems, bleek dat deze complexere modellen de data beter verklaarden dan eenvoudigere alternatieven.

Alles samengenomen is er een flexibele methode ontwikkeld om te kunnen omgaan met micro-macrosituaties met discrete data. Er is een breed scala aan toepassingen getoond en om het gebruik van de methode te stimuleren werd de benodigde syntax beschreven wanneer deze niet evident was.

Dankwoord (Acknowledgements in Dutch)

Na vier jaar is mijn proefschrift klaar, oftewel, is mijn badkuip vol! Dat heb ik uiteraard niet in mijn badeendje gedaan en daarom wil ik hierbij graag de mensen bedanken die direct of indirect aan dit resultaat hebben bijgedragen.

Allereerst hebben mijn (co)promotoren Jeroen Vermunt en Marcel Croon ervoor gezorgd dat het water in deze badkuip een aangename temperatuur heeft gekregen door regelmatig aan de warme of koude kraan te draaien. Marcel, jouw geduld, didactische vaardigheden en kennis zijn van onschatbare waarde geweest voor mij en dit proefschrift. Jeroen, jouw scherpe inzichten, directheid en enthousiasme zijn een bron van inspiratie geweest. Beiden, heel erg bedankt voor de goede begeleiding en de prettige samenwerking!

Ik bedank mijn promotiecommissie voor het lezen en becommentariëren van het proefschrift. Ook bedank ik Jos Keuning en Brigitte Kroon voor hun bijdrage aan de desbetreffende hoofdstukken. Verder hebben de leden van de *extended vici group* een grote bijdrage geleverd door hun eerlijke commentaren, die soms als een koude douche voelden: even schrikken maar erg verfrissend. Bedankt daarvoor! Carri, bedankt voor het corrigeren van de Nederlandse teksten in mijn proefschrift. Lianne, bedankt voor het zetten van de puntjes op de i's.

Werken bij het MTO departement is als een warm bad voor mij geweest. Daarvoor wil ik al mijn (oud)collega's en (oud)roomies bedanken! In het bijzonder Zsuzsa en Daniël, maar ook Guy, Hendrik, Johan, John, Joris, Lianne, Marcel, Marieke, Meike, Peter, Pieter, Renske, Ruud en Wilco. Mijn muzikale collega's bedank ik dat ze met hun band *the significant others* willen spelen op mijn promotiefeest!

Zonder mijn vrienden zouden de afgelopen jaren niet zo leuk zijn geweest. Een paar vriendinnen wil ik in het bijzonder bedanken. Ellen, bedankt voor je steun tijdens de afgelopen jaren. Evelyne, bedankt voor onze bijzondere vriendschap. Ik ben heel dankbaar voor onze gesprekken over familie, loopbanen en heel veel meer. Marie-Anne, wij gaan van $s\ t\ \alpha\ t\ s$ - rappers tot carnavalskrakers en van roddels tot serieuze gesprekken. Dat maakt onze vriendschap zo waardevol. Laura, wij kennen elkaar al vanaf dat we samen de robotdans dansten toen we vier jaar oud waren en ik waardeer de kwaliteit van onze vriendschap heel erg. Evi, het mooiste aan Brabant ben jij! Jij hebt er (mede) voor gezorgd dat mijn tijd in Tilburg zo leuk was. Ook bedankt dat je mijn paranimf bent.

Verder heb ik een grote patchworkfamilie (zoals oma Erie dat zo mooi zou zeggen) die erg belangrijk voor mij is. De familie Sijssens bedank ik dat ze mij (weer) met open armen in hun familie hebben opgenomen. De vakanties, feestdagen, verjaardagen, fietstochten, shop-uitjes, appjes en bommetjes waren een welkome afleiding tijdens het schrijven van

mijn proefschrift. Ook de Meijers en Klaassens behoren inmiddels tot mijn familie en ook jullie wil ik graag bedanken voor jullie steun. In het bijzonder bedank ik Irene voor de mentale steun door plaats te nemen in het cortège, Vivian voor het ontwerpen van de prachtige omslag van dit boekje, Bas voor al je 'niceness', Kieran voor je perfecte 'lekkere koffies' en Marjolijn voor het laten zien hoe je 'even doorpakt'. Gonnie en Henk, bedankt voor jullie liefdevolle steun, interesse, nuchterheid, gezelligheid en humor. Met de familie Snijdoodt deel ik ontelbaar veel mooie en leuke momenten, in het bijzonder met Mariken (alias Bokkie). Ik denk dat het kiezen voor 10-10 als datum voor mijn verdediging, iets is wat ik van jullie kant van de familie heb meegekregen. Boudewijn, bedankt dat je altijd zo attent bent en voor de leuke potjes Monopoly deal en Machiavelli. Laurienemans, bedankt voor alle heugie meugies, prachtige nagels en girlpower op de motor. Mijn ouders, broertje en zusje nemen een centrale plaats in het patchwork in. Pap, bedankt voor het vertrouwen dat je altijd in me hebt. Be happy! Mam, bedankt voor het laten zien van de (vele) wereld(en) buiten de wetenschap. Bram, da dude rules! Vi'tje, kleintje, bedankt voor al je wijsheid en dat je mijn paranimf bent.

Last but not least: Wouter, bedankt voor alles! Bij jou vind ik m'n rust en samen voel ik me overal ter wereld thuis.