NETHERLANDS JOURNAL OF PSYCHOLOGY

VOLUME 67, NUMBER 4, JANUARY 2013

SPECIAL ISSUE – PART 2

Non-standard structural equation modelling

GUEST EDITORS

Suzanne Jak

Annemarie Zand Scholten

Frans J. Oort

Volume 67/number 4/January 2013
Netherlands Journal of Psychology

Editorial Policy

The *Netherlands Journal of Psychology* publishes original articles of high quality, including empirical articles, review essays on selected books, theoretical and methodolological papers in any area of psychology, as well as ongoing commentaries and discussion, and short reports on the validation of psychodiagnostic instruments and related methodological tools.

The *Journal's* focus is on empirical and conceptual studies that contribute to the theoretical explanation of human behaviour and experience. Manuscripts can deal with human development, social processes and the social perspective of human behaviour, psychopathology, forensic psychology and psychiatry neuroscience, psychophysiology, philosophy of mind, the computational approach, emotion, cognition (including attention, perception, and memory), decision-making, human performance, educational psychology, health, selection and assessment and human behaviour in organisations, selection and assessment. Any other manuscripts that may be of importance to those involved with psychological research are welcomed. Manuscripts should be written to the professional and academic readership.

It is the *Netherlands Journal of Psychology's* policy to publish about recent developments in psychology and related fields. It is also the editors' policy to regularly focus on recent developments in psychology in the Netherlands. In particular, authors may be invited to contribute to special issues.

The *Netherlands Journal of Psychology* will help to focus the interest of psychologists, and others professionally interested in the field, on information about developments in a wide area of specialist research activities and results of empirical and theoretical work in the field of psychology.

The *Netherlands Journal of Psychology* offers specialists the opportunity to publish their findings to a broad audience that is scientifically interested in psychology.

The *Netherlands Journal of Psychology* aims at offering a continuous forum for intellectual discussion and transfer of knowledge and information about developments in psychology.

Categories of articles to be published and manuscripts to be welcomed are:

Review articles
Manuscripts should be limited to a maximum of 8000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Original reports of empirical research
Manuscripts should be limited to a maximum of 5000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Thematic issues
Authors with plans for planning a thematic issue are encouraged to contact the editor in an early stage. Consultation will be necessary about the number of manuscripts, number of words per manuscript, review procedure, covering of the field, etc.

Developments in the field (short)
Manuscripts should be limited to a maximum of 1000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Book review essays (short)
Manuscripts with a maximum of 1000 words need no preliminary contact with the editor. Authors planning larger contributions must contact the editor.

Letters to the editor
Manuscripts should be limited to a maximum of 500 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript.

Technical notes
Manuscripts should be limited to a maximum of 2000 words. Authors planning contributions that exceed this maximum must contact the editor prior to submitting their manuscript. Authors reporting (in English) on, e.g. validation of tests and other instruments in the Dutch language or in the Netherlands are encouraged to report in *The Netherlands Journal of Psychology*. The editors want to inform non-Dutch researchers of developments and of the use of instruments in the Dutch language.

Editor's address
Professor René van Hezewijk
Faculty of Psychology, Open University of the Netherlands, PO Box 2960, 6401 DL Heerlen, the Netherlands
tel + 31 45 576 23 99
All correspondence by e-mail: rene.vanhezewijk@ou.nl

Publisher: Nederlands Instituut van Psychologen, PO Box 2085, 3500 GB Utrecht, the Netherlands

Subscription rates
Personal rate: € 125.–
Institutional rate: € 226.–
Student rate: € 62.50
All prices are per calendar year and include value added tax (VAT)
Price per issue: € 31.95 (incl. VAT)

Subscription administration
Performis Media, PO Box 2396, 5202 CJ 's Hertogenbosch, the Netherlands, tel: + 31 73 689 58 89. For information and orders, please consult www.performis.nl

Change of address
Please notify any changes in addressee and/or address via www.performis.nl.

Payment
Please use the payment/accept giro form if possible as this simplifies the administrative process.

Advertisements
Performis Media, PO Box 2396, 5202 CJ 's Hertogenbosch, the Netherlands, tel: + 31 73 689 58 89.
The *Netherlands Journal of Psychology* is published four times a year.

ISSN 1872-552x

Contents

VOLUME 67, NUMBER 4, JANUARY 2013

Non-standard structural equation modelling (Part 2)

This issue of the *Netherlands Journal of Psychology* is the second of two special issues that give an account of non-standard applications of Structural Equation Modelling (SEM), as presented at the 2012 Meeting of the Working Group SEM (Amsterdam, 22-23 March 2012). Since the foundation of the Working Group SEM in 1986, advanced structural equation modelling has been discussed in annual meetings at various locations in Germany, the Netherlands, and Switzerland. The presentation and discussion of methodological problems and developments in structural equation modelling are the main objectives of the Working Group.

In this second special issue, Fischer, Brandt, Schermelleh-Engel, Moosbrugger, and Klein give an introduction of moderated linear and moderated quadratic effects in regression models. They present a screening procedure to identify potential moderated quadratic effects in regression models, which can also be used to distinguish moderated quadratic effects from moderated linear effects. Barendse, Oort, Jak, and Timmerman use exploratory factor analysis to investigate the dimensionality in multilevel discrete data. With data from educational research, they illustrate two approaches to investigate the dimensionality, one with the restriction that there is measurement invariance across clusters, and one without this restriction. Reinecke describes two models that can be used to handle missing data in panel designs, due to dropout. He combines the dropout models with a latent growth model. Using panel data from a study of the development of delinquent behaviour, he shows how these models provide information about the dropout processes in the study. In the last article of this issue, Jak, Oort, Roorda, and Koomen extend the two-stage approach to meta-analytical SEM by proposing a method to treat studies that do not fully report all correlation coefficients. The method is illustrated with meta-analytic path modelling of data about teacher-student relations, attachment and achievement at school.

We believe that these two special issues of the *Netherlands Journal of Psychology* together give a fine impression of the possibilities of non-standard SEM, and of the variety of substantive research questions that can benefit from an advanced SEM approach.

Suzanne Jak
Annemarie Zand Scholten
Frans J. Oort
University of Amsterdam, Amsterdam, the Netherlands
E-mail: S.Jak@uva.nl

References

Barendse, M. T., Oort, F. J., Jak, S., & Timmerman, M. E. (2013). Multilevel exploratory factor analysis of discrete data. *Netherlands Journal of Psychology, 67*, 114-121.

Fischer, K. A., Brandt, H., Schermelleh-Engel, K., Moosbrugger, H. A. G. (2013). Detecting the type of moderated nonlinearity: A screening procedure. *Netherlands Journal of Psychology, 67*, 106-113.

Jak, S., Oort, F. J., Roorda, D. L., Koomen, H. M. Y. (2013). Meta-analytic structural equation modelling with missing correlations. *Netherlands Journal of Psychology, 67*, 132-139.

Reinecke, J. (2013). Growth curve models and panel dropouts: Applications with criminological panel data. *Netherlands Journal of Psychology, 67*, 122-131.

SUZANNE JAK
PhD candidate in Methods and Statistics, Research Institute of Child Development and Education, University of Amsterdam. Her research interests include structural equation modelling, measurement bias, multilevel structural equation modelling and meta-analytical structural equation modelling.

ANNEMARIE ZAND SCHOLTEN
Assistant professor with the Methods and Statistics research group in the Department of Child Development and Educational Sciences at the University of Amsterdam. Her research interests include the quantitative structure of psychological measurement and the meaningfulness and stability of inference based on statistical and modelling techniques under different types of measurement level assumption violations.

FRANS J. OORT
Full professor of Methods and Statistics and Director of the Research Institute of Child Development and Education, University of Amsterdam. His research interests include measurement and measurement bias, and statistical modelling, especially non-standard applications of structural equation modelling.

Detecting the type of moderated nonlinearity: A screening procedure

While in social and behavioural sciences several theories include moderated linear (interaction) effects, moderated quadratic effects are usually not hypothesised. However, moderated quadratic relationships have gained increasing interest in different research areas during the last years. Unfortunately, power to detect this type of effect is often quite low, because a complex nonlinear regression model which includes all linear terms as well as all second-order and third-order terms has to be analysed. In this article we first develop a screening procedure analytically, which can be performed before a more complex nonlinear multiple regression model is selected and does not require the functional form of the nonlinear relation. Then we demonstrate that the screening procedure is able to differentiate between moderated linear and moderated quadratic effects using two artificial datasets as a first test of the procedure's performance. Advantages and limitations of the screening procedure are discussed.

Where: Netherlands Journal of Psychology, Volume 67, 106-113

Received: 27 June 2012; Accepted: 20 November 2012

Keywords: Moderator effect; Moderated quadratic effect; Nonlinear regression; Interaction

Authors: Kevin A. Fischer, Holger Brandt, Karin Schermelleh-Engel, Helfried Moosbrugger and Andreas G. Klein

Goethe University Frankfurt, Germany

Correspondence to: Kevin A. Fischer, Goethe University, Department of Research Methods and Evaluation, Mertonstrasse 17, 60054 Frankfurt am Main, Germany email: kefische@stud.uni-frankfurt.de

Over the last 40 years, interest in nonlinear relationships between variables has substantially increased. There have been numerous empirical studies aiming to detect moderator (also called interaction) effects in order to enhance the validity of prediction. Moderator effects have mostly been modelled by adding product terms to a linear multiple regression equation (see, e.g., Aiken & West, 1991; Allison, 1977; Blalock, 1965; Cohen & Cohen, 1975; Wright, 1976). A moderator effect is present when the size or direction of the effect of a predictor variable X on a criterion variable Y varies systematically according to the level of a second predictor variable Z and when the addition of the product term XZ to the multiple regression equation enhances the proportion of explained variance in the criterion variable beyond the proportion of explained variance due to the linear effects. In the following we will distinguish between two different types of moderator effects, a moderated linear effect and a moderated quadratic effect.

Moderated linear effect

Typically, a regression model with a moderated linear effect is given by the following equation (e.g., Cohen & Cohen, 1975)

$$Y = b_0 + b_1X + b_2Z + b_3XZ + e \tag{1}$$

where b_0 is the intercept, b_1 and b_2 are linear effects, b_3 is the moderated linear effect, and e is a residual variable.

Rearranging equation (1) shows that if b_3 is nonzero the regression of Y on X depends on the values of Z, i.e., the slope of Y on X increases (or decreases) linearly with increasing Z (*Eq. 2a*), as well as the regression of Y on Z depends on the values of X (*Eq. 2b*):

$$Y = b_0 + (b_1 + b_3Z)X + b_2Z + e \tag{2a}$$

$$Y = b_0 + (b_2 + b_3X)Z + b_1X + e \tag{2b}$$

Thus the regression equation with the characteristic product term (XZ) is symmetrical in the predictor X and in the predictor Z. The regression of Y on X is linear at every value of Z, and the regression of Y on Z is also linear at every value of X.

A typical example for a moderated linear effect is included in the theory of planned behaviour (TPB; Ajzen, 1991; Ajzen & Madden, 1986), a further development of the theory of reasoned action (Ajzen & Fishbein, 1980; Fishbein & Ajzen 1975). In short, the TPB predicts that planned behaviour is determined by *behavioural intentions* which in turn are influenced by an individual's *attitude toward the behaviour, subjective norms* related to the execution of the behaviour, and the individual's *perception of the control* over this behaviour. The TPB also postulates a moderator effect: volitional control, i.e., perceived behavioural control, is expected to moderate the intention-behaviour relationship in such a way that the effect of intention on behaviour is stronger when actual control is high. Thus, there is a moderated linear effect.

Moderated quadratic effect

A moderated quadratic effect is present when a quadratic relationship between the predictor variable X and the criterion variable Y varies systematically according to the level of a second predictor (moderator) variable Z (see Aiken & West, 1991).

In analogy to the product term XZ in the regression model with a moderated linear effect, the moderated quadratic term is formed as XZ^2 for a potential moderator variable X. A regression with a moderated quadratic effect is given by

$$Y = b_0 + b_1X + b_2Z + b_3XZ^2 + e \qquad (3)$$

Rearranging equation (3) shows that if b3 is nonzero the regression of Y on X depends on the values of the moderator variable, which may be Z or X, depending on theory:

$$Y = b_0 + (b_1 + b_3Z^2)X + b_2Z + e \qquad (4a)$$

$$Y = b_0 + (b_2 + b_3XZ)Z + b_1X + e \qquad (4b)$$

The regression equation with the moderated quadratic term (XZ^2) is not symmetric as can be seen in Equation (4a) and Equation (4b). This means that the effect of X on Y represents a gradual steady linear change when Z changes. Instead, the effect of Z on Y represents a gradual steady quadratic change when X changes.

An example of a moderated quadratic effect is given by Baer and Oldham (2006). They investigated the relationship between *experienced creative time pressure* and *creativity* at work and postulated that the strength of this quadratic relationship is dependent on the moderator variable *support for creativity*. The hypothesis for a moderated quadratic effect could be confirmed: For employees receiving high support for creative behaviour, the relation between experienced time pressure and creativity followed an inverted U-shaped function, while the relation was almost linear for employees receiving low support (see Figure 1).

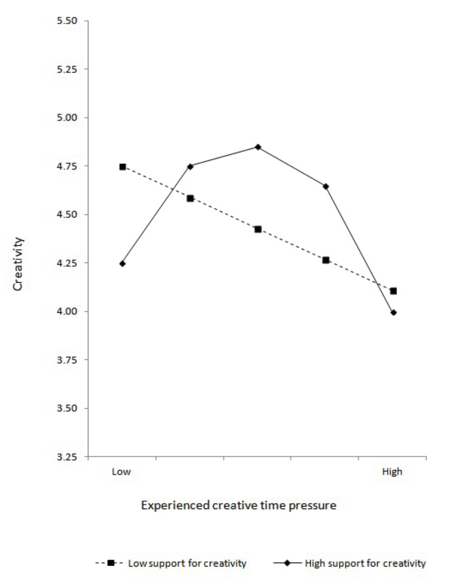

Figure 1 Illustration of a moderated quadratic effect based on a study by Baer and Oldham (2006) using an artificial dataset. It can be seen that the relation between creativity and experienced creative time pressure is moderated by support for creativity. For persons receiving high support, the dependent variable creativity first increases and then declines with increasing values of experienced creative time pressure and reaches a maximum value of creativity for intermediate experienced creative time pressure. The relation is linear only for persons receiving low support, i.e., creativity declines with increasing values of experienced creative time

While in social and behavioural sciences several theories exist that include moderated linear effects, moderated quadratic effects are usually not hypothesised. However, moderated quadratic relationships have gained increasing interest in different research areas during the last years and several empirical studies have investigated this type of nonlinear effects, although it seems to be difficult to detect them (e.g., Baer & Oldham, 2006; De Bruijn et al., 2007; Le et al., 2011; Salamin & Hom, 2005). Le et al. (2011), for example, investigated several nonlinear effects of personality traits on job performance dimensions. Their results confirmed the hypothesised inverted U-shaped relations between two personality traits, *conscientiousness* and *emotional stability*, and the criterion variables *task performance* and *organisational citizenship behaviour*. Additionally, they investigated four moderated quadratic effects with *job complexity* as moderator variable, but only one of these effects reached significance.

Generally, higher-order nonlinear effects, such as moderated quadratic effects, only seldom reach significance in empirical research. But even if significant effects are found, these may be spurious due to methodological problems. Therefore it would be of crucial interest for researchers to be able to decide whether moderated quadratic effects or moderated linear effects are actually in the data without forming complex nonlinear terms. In the following, we will present a new screening method developed to differentiate between moderated quadratic and moderated linear effects without including complex nonlinear terms.

The article is structured as follows: First, we briefly review the methodological problems in identifying moderated linear and moderated quadratic effects. Second, we introduce a new screening method which allows differentiating between a moderated quadratic effect and a moderated linear effect. Third, we illustrate the performance of the screening method using an artificial dataset. Finally, we discuss the limitations and the future perspectives of the new method.

Methodological problems

In general there are two main methodological problems in identifying moderated effects in regression analysis: multicollinearity and spurious effects.

Multicollinearity

The first problem of multiple regression, multicollinearity, occurs when predictors in the regression equation are highly correlated. A high correlation between the predictors X and Z implies that the two variables are very similar, and it becomes difficult to determine which of the variables accounts for variance in the dependent variable Y. In case of severe multicollinearity, standard errors are inflated and inferential tests based on the standard errors have a low power.

The problem of multicollinearity becomes more serious for nonlinear regression models. In general, product terms (e.g., XZ, X^2, Z^2) are highly correlated (e.g., $r = .90$ or above) with the variables that are used to form the product terms, independent of whether X and Z themselves are correlated or not. If the predictor variables X and Z are correlated, then the product terms are also correlated. Although centring predictor variables (converting the scores of the predictor variables into deviation form) is a convenient method for reducing multicollinearity, the product terms may still be correlated with the linear predictors, given that there is correlation between X and Z. This implies that with increasing numbers of product terms the problem of multicollinearity also increases. As a consequence, the power to detect effects associated with product terms decreases with the number of simultaneously tested nonlinear effects.

Spurious effects

The second problem of nonlinear regression is concerned with spurious effects. Several authors have claimed that for the analysis of interaction (moderated linear) effects quadratic effects should always be included in the nonlinear regression model (e.g., MacCallum & Mar, 1995; Klein et al., 2009). This is due to the fact that if a true ‘quadratic’ effect is present in the population but only an interaction effect is being analysed (as in equation 1), then this interaction effect could become significant and would lead to a wrong conclusion: What appears to be a significant interaction effect might actually be a significant quadratic effect. Thus, a model for analysing interaction effects should always include quadratic effects, too. Equation (1) should therefore be extended as follows:

$$Y = b_0 + b_1X + b_2Z + b_3XZ + b_4X^2 + b_5Z^2 + e \qquad (5)$$

Obviously, the inclusion of the quadratic terms leads to an increased multicollinearity (if X and Z are correlated) and thus to a lower power to detect an interaction effect. The recommended model is in this sense a conservative procedure as it reduces the probability for detecting spurious interaction effects. Klein et al. (2009) showed that all product terms of the same order (here: order 2) and lower-order terms (here: order 1) need to be included in the model in order to prevent spurious effects. They used a Taylor polynomial of degree 2 to approximate the nonlinear regression function and proved that the inclusion of all product terms of the same order as well as all lower-order terms is necessary to avoid spurious effects.

This line of reasoning for the moderated linear regression can be transferred to a moderated quadratic regression which includes a product term of order 3 (XZ^2). Consequently, in addition to the predictor variables X and Z, all product terms of order 2, XZ, X^2, and Z^2 (see equation 5), and all product terms of order 3, X^3, Z^3, XZ^2, and X^2Z, should be included in the regression equation:

$$Y = b_0 + b_1X + b_2Z + b_3XZ + b_4X^2 + b_5Z^2 + b_6X^2Z + b_7XZ^2 + b_8X^3 + b_9Z^3 + e \quad (6)$$

Using regression equation (6) instead of equation (3) leads to a very conservative procedure for testing a moderated quadratic effect.
The low power for detecting moderated quadratic effects seldom leads to significant results. Nonetheless, a complex model for testing moderated quadratic effects is necessary from a statistical perspective in order to prevent spurious effects. A non-significant moderated quadratic effect in such a complex model as given in equation (6) could be interpreted in two different ways: Either there is no effect in the population or there is an effect in the population but the power is too low to detect the effect in the sample.

Method

Screening procedure

In this section we present a screening procedure which can be used prior to estimating a complex regression model. This procedure should enable a researcher to identify potential moderated quadratic effects and differentiate these effects from moderated linear effects.

First, we describe the differences between moderated linear and moderated quadratic effects and the implications on how to distinguish between these two types of effects. Second, we introduce the screening procedure based on this distinction.
In general, with an infinite number of data points, one could distinguish moderated linear effects from moderated quadratic effects by using conditional effects. A conditional effect means that the relationship between a criterion variable *Y* and a predictor variable *X* depends on the level of a moderator variable *Z* and may be different across values of *Z*.
In Figure 2 the two different types of effects are illustrated for uncorrelated predictors *X* and *Z*. To simplify the function the linear effects were set to zero.

For a *moderated linear* term (*XZ*), the increase in slope of the criterion *Y* on the predictor *X* is constant across levels of the predictor *Z* (see Figure 2, a1). For a *moderated quadratic* effect (XZ^2), the slope of the criterion *Y* on the predictor *X* changes linearly across levels of the predictor *X* (Figure 2, b1). Formally, the conditional effect γ_z of *Y* on *X* given *Z* = z is defined as

$$\gamma_z = \frac{Cov(Y, X \mid Z = z)}{Var(X \mid Z = z)} \quad (7)$$

where γ_z is the conditional effect of *X* on *Y* for the subgroup with value *Z* = z, $Cov(Y,X \mid Z = z)$ is the conditional covariance between *Y* and *X*, and $Var(X \mid Z = z)$ is the conditional variance of *X*. As can be seen from equation (7) the conditional covariance is proportional to the conditional effect. Based on the type of change of γ_z across different values of *Z*, an inference can be made regarding the type of moderated effect. For a model including a moderated *linear* effect (equation 1), it follows that

$$\gamma_z = (\beta_1 + \beta_3 z)X \quad (8)$$

Thus, if γ_z depends linearly on *Z*, this is evidence for a moderated linear effect.
For a moderated *quadratic* effect as in equation (2), the following equality holds:

$$\gamma_z = (\beta_1 + \beta_3 z^2)X \quad (9)$$

Thus, if γ_z depends quadratically on Z, then this would be evidence for a moderated quadratic effect. In general, the conditional covariance or variance cannot be computed for continuous data without

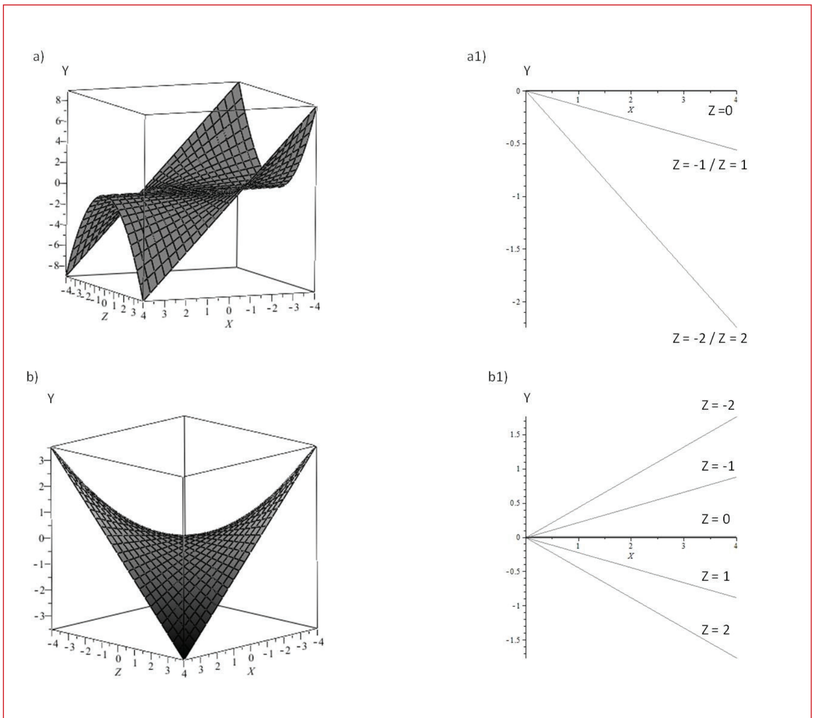

Figure 2 Regression surface of a model with a moderated quadratic effect (a) and a model with a moderated linear effect (b). In a1) and b1) the conditional effect of X on Y is illustrated for five levels of Z. The main distinctive features between the two moderated effects are depicted in a1) and b1): The moderated quadratic effect shows an increase in slope from Z=-2 to Z=0 and a decrease in slope from Z=0 to Z=2, while the moderated linear effect is symmetric and shows a constant decrease of the slope from Z=-2 to Z=2

additional assumptions. But, it can be computed when Z is categorical. Therefore, we use an approximation of the conditional covariance matrix by using a categorised version of Z. For the categorisation, the sample is split into five subgroups by collapsing scale values into more coarse classifications (cf. Butts & Ng, 2009). Compared with a moderated regression that uses continuous variables, the categorisation of a predictor variable leads to a loss of power and a reduction of effect size in this procedure (cf. MacCallum, Zhang, Preacher, & Rucker, 2002). Even so, for detecting moderated nonlinear effects by means of moderated regression analysis with all product terms of order 3 included, a very large sample size would be needed to detect the nonlinear effects. Such a very large sample size may be unrealistic to achieve in many research settings. Therefore, when sample size is relatively small, the proposed method may be the preferred choice for detecting moderated nonlinear effects.
Here, we categorise the variable Z into five subgroups using thresholds at specific values $z_1 < z_2 < \ldots < z_4$ (see Figure 3).

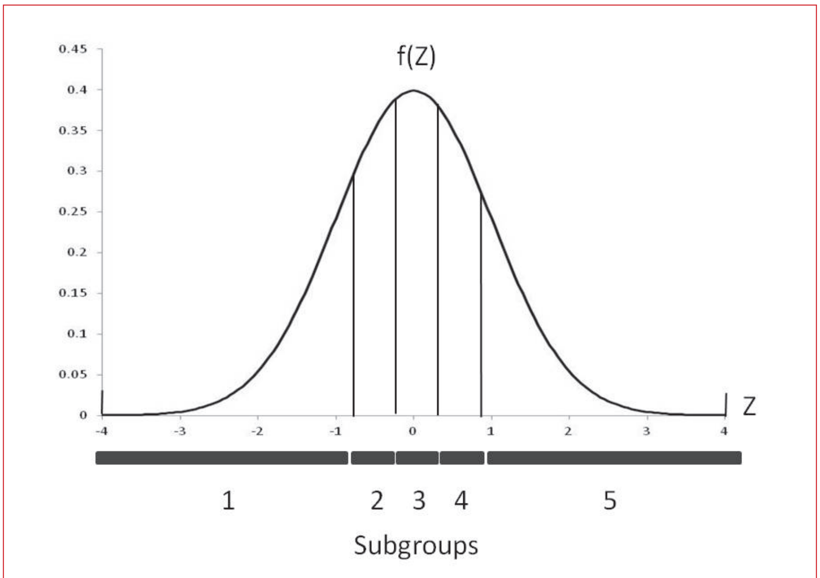

Figure 3 Categorisation of a standardised normally distributed variable. The sample is split into five subgroups which all contain the same percentage of subjects

For each of the five subgroups the conditional effect of X on Y given $Z = z$ is estimated.

$$\gamma_{z_{i,i+1}} = \frac{Cov(Y, X \mid z_i < z \le z_{i+1})}{Var(X \mid z_i < z \le z_{i+1})} \tag{10}$$

The interpretation of the conditional effect is the linear approximation of the average true effect for subjects in the subgroup $z_i \le z < z_{i+1}$. For the example with five subgroups in Figure 3, five conditional effects can be estimated.
The varying slopes of the g different conditional effects across the five subgroups are an indication of the type of relationship as described above. Again, if the increase in slopes is quadratic, evidence for a moderated quadratic effect exists, while a linear slope is indicative of a moderated linear effect.

Models for data generation

Our example is based on a study by Baer and Oldham (2006). For our purposes, we used a subset of the predictors of the original study. We examined the relationship between creativity (*creat*) and creative time pressure (*time*) that is moderated by support for creativity (*sup*). Following the results of the study we generated data for two datasets.

The first dataset ($N = 400$) was generated using the following population model which includes a *moderated quadratic* effect ($sup \times time^2$):

$$creat = 0.01 - .03sup - .24time - .07sup \times time - .03time^2 - .14sup \times time^2$$

In order to keep the example as simple as possible we fixed the following effects to zero: sup^2, $sup^2 \times time$, sup^3, and $time^3$. The variables *creative time pressure* (*time*) and *support for creativity* (*sup*) were generated as standardised normally distributed variables with zero means and variances of one. The correlation between *time* and *sup* was set to $\varrho = -0.24$. The variance of *creativity (creat)* was chosen such that the variance of *creat* was also one. Thus, the regression coefficients can be interpreted as standardised coefficients.

The second dataset was generated based on the following population model which includes no moderated quadratic effect, but a significant *moderated linear* effect ($sup \times time$):

$$creat = -0.02 - .03sup - .24time - .22sup \times time - .03time^2$$

Here, we fixed the quadratic effect of *support for creativity* (sup^2) to zero. This model which includes as a nonlinear effect only the moderated linear effect $sup \times time$ served as a comparison standard. For reasons of comparability, the original non-significant moderated linear effect was resized such that its unique contribution to the variance accounted for was similar to the unique contribution of the moderated quadratic effect.

Procedure

For both datasets, we categorised the moderator variables in order to receive five subgroups. We used the variable *creative time pressure (creat)* as grouping variable, and in the second trial we used the moderator variable *support for creativity (sup)* as grouping variable. The categorisation of the data was performed such that each subgroup contained 20% of the values of the grouping variable. As a result, we

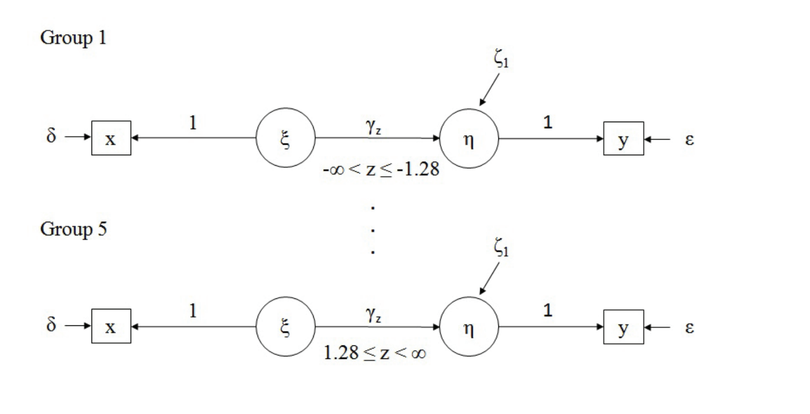

Figure 4 MSA model for the simulation study, where the latent predictor variable ξ equals x and the dependent latent variable η equals y. The linear relationship between η and ξ and therefore the relationship between y and x is analysed simultaneously along different intervals of the z variable. Here in five subgroups

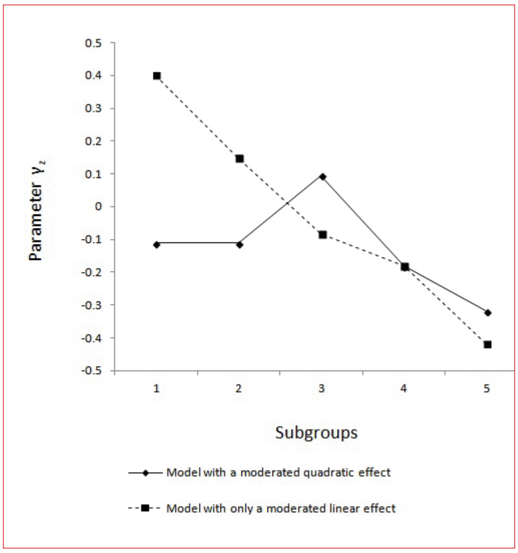

Figure 5 Graphical representation of the relationship between γ_z across the five subgroups of creative time pressure (Z) for a model containing a moderated quadratic effect and a model containing only a moderated linear effect. The groups were built by categorising the predictor variable creative time pressure

obtained five subgroups with a comparable number of observations. The five subgroups were then used for data analyses. A multi-sample analysis (MSA) was performed using the M*plus* program (Muthén & Muthén, 1998-2010). We simultaneously regressed the criterion variable *creativity (creat) on creative time pressure (time)* in all five subgroups. Figure 4 illustrates the MSA model for the conducted study. The estimates of the conditional regression coefficient across all five subgroups were collected and used for the screening procedure.

Hypothesis 1: For the first generated dataset which includes the moderated quadratic term *support for creativity × creative time pressure*^{2} *(XZ*^{2}*)* we hypothesised that the relation between the criterion variable *creativity (Y)* and the predictor variable support for *creativity (X)* at different levels of the grouping variable *creative time pressure* (*Z* = z) follows a quadratic trend.

Hypothesis 2: For the second generated dataset which includes the moderated linear term *support for creativity × creative time pressure* we hypothesised that the relation between the criterion variable *creativity (Y)* and *support for creativity (X)* on the levels of the grouping variable *creative time pressure* (*Z* = z) follows a linear trend.

Results

The quadratic change of the parameter γ_z across the five subgroups of *creative time pressure* that was postulated in hypothesis 1, could be confirmed for the relationship between the criterion variable *Y* and the moderator variable *X* as depicted in Figure 5. The value of the parameter γ_z increases from the first to the third subgroup, but decreases for groups 4 and 5. In this case the parameter γ_z describes a quadratic relationship across the five subgroups. Hypothesis 2 could also be confirmed for the model which only contained a moderated linear effect. The value of the parameter γ_z decreases linearly from the first to the fifth subgroup.

Discussion

In this article our goals were threefold. First, we wanted to give an introduction to the concepts of moderated linear (interaction) and moderated quadratic effects as higher-order effects are quite rarely investigated in empirical research. Two methodological problems associated with moderated quadratic effects were identified. The problem of multicollinearity between product terms, which occurs when the predictors *X* and *Z* are highly correlated, has the consequence that large standard errors and low power results. Another problem is concerned with spurious effects, which can occur when not all nonlinear terms necessary for the analysis of a complex nonlinear regression model are included in the regression equation. As pointed out by several authors (e.g., Klein et al., 2009), models with moderated linear effects should also include all additional nonlinear terms of the same order (second-order), i.e. both quadratic effects, and all lower-order terms, i.e. all linear effects. For models with third-order nonlinear effects, we argued that all effects of the same order, both moderated quadratic effects and two cubic effects, should be included in the regression equation together with all second-order and linear effects. This helps researchers to avoid spurious effects that would occur when, for example, a quadratic effect is present in the data but a moderator model is being investigated. Unfortunately, including all nonlinear effects leads to a low power, and large sample sizes are required to detect these effects.

Second, we described analytically how moderated linear and moderated quadratic effects are distinguishable by using the conditional effects of X on Y across values of Z. Our semi-parametric screening procedure can be performed before a more complex nonlinear multiple regression model is selected. It does not require the functional form of the nonlinear relation.

Third, we investigated the screening procedure using artificial datasets based on empirical research by Baer and Oldham (2006) as a first test of the procedure's performance. The results confirmed that a moderated quadratic effect between *support for creativity* × *experienced creative time pressure*^{2} is rediscovered in our artificial dataset. The parameter γ_z reflects the slope between *support for creativity* and *creativity* in every subgroup of *experienced creative time pressure*. If the parameter γ_z shows a parabolic relation over the five subgroups of Z this is an indication for a moderated quadratic effect. To show that this relation was not caused randomly we formulated additionally a model with a moderated linear effect. In this model, the relation between the parameter γ_z over the subgroups of *experienced creative time pressure* was linear. This is an indication for a moderated linear effect in the dataset. For researchers this can in some circumstances be a helpful feature, because often there is only limited knowledge on a specific nonlinear effect. In this case the interpretation of the result depends on which nonlinear product variable was included in the analysis. If a true moderated quadratic effect exists in the data, but incorrectly only a moderated linear effect is included in the regression equation, this can lead to misleading interpretations of the relationship between two variables. The screening approach enables a researcher to decide whether to include a moderated nonlinear term or not.

Before applying the procedure researchers should decide how many subgroups they want to use for detecting moderated nonlinear effects. The number of subgroups primarily depends on the original sample size of the study. On the one hand, each subgroup should consist of approximately 80 subjects to produce unbiased estimates for the maximum-likelihood estimator. On the other hand, in order to identify a possible nonlinear trend in the data, at least five subgroups should be used. The number of subgroups needs to be an odd number. This has the advantage that there is a reference group with the mean of the unconditioned regression parameter z. Using only three subgroups would require large sample sizes, which are often not available; therefore, based on our experience, at least five groups appear necessary to produce reliable results.

Nevertheless there are some limitations of our research that need to be considered. Up to now the procedure is a descriptive procedure. Even if a specific moderated nonlinear effect was found in the data it cannot be assured that this effect will also be present in the population. Furthermore, the procedure is only concerned with moderated nonlinear effects; additive nonlinear effects cannot be detected by this method, because the slope of a conditional effect between X on Y does not vary across levels of Z (for a further discussion see Klein et al., 2009). As is usually done in analysing MSA, a chi-square difference test is applied that tests an equality constraint for the five conditional regression coefficients. With this constraint the difference between a nonlinear and a linear model is tested. For different types of nonlinear models, however, one cannot apply the MSA technique, because the different nonlinear models are not nested relative to each other.

A limitation of regression analyses in general is that all variables are treated as if they were perfectly reliable (cf. Cohen, Cohen, West, & Aiken, 2003). Often, this assumption is violated in practice. The problem is even aggravated when nonlinear terms are added to the regression equation, because product terms are even less reliable than the variables that are used to form the products. The consequence is that the population regression coefficients are usually underestimated. Another limitation of regression analyses is that each construct is only measured by a single variable. Because constructs are generally quite heterogeneous they cannot be represented by a single indicator variable. Therefore structural equation modelling (SEM) could have an advantage. Over the last 20 years, several methods for the analysis of latent nonlinear models have been developed (cf. Jöreskog & Yang, 1996; Klein & Moosbrugger, 2000; Marsh, Wen, & Hau, 2004). These methods are able to estimate nonlinear effects on the level of latent variables without bias. However, larger sample sizes are needed to detect small nonlinear effects, especially when several nonlinear effects are included in the model. The consequence is that power is quite low such that large sample sizes are needed. Therefore nonlinear SEM may benefit greatly from a screening procedure that could be performed before analysing complex models. The development of a screening procedure for nonlinear SEM is our next project.

References

Aiken, L. S., & West, S. G. (1991). *Multiple Regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs: NJ: Prentice-Hall.

Ajzen, I., & Madden, T. (1986). Prediction of goal directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology*, 22, 453-474.

Allison, P. D. (1977). Testing for interaction in multiple regression. *American Journal of Sociology*, 83, 144-153.

Baer, M., & Oldham, G. R. (2006). The curvilinear relation between experienced creative time pressure and creativity: Moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology*, 91, 963-970.

Blalock, H. M., jr. (1965). Theory building and the concept of interaction. *American Sociological Review*, 30, 374381.

Butts, M. M., & Ng, T. W. H. (2009). Chopped liver? OK. Chopped data? Not OK. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 563-585). New York: Routledge.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analyses for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences*. (3rd ed.). Mahwah, NJ: Erlbaum.

De Bruijn, G. J., Kremers, S. P. J., de Vries, H., van Mechelen, W., & Brug, J. (2007). Associations of social-environmental and individual-level factors with adolescent soft drink consumption: Results from the SMILE study. *Health Education Research, 22*, 227-237.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Jöreskog, K., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 57-87). Mahwah; NJ: Lawrence Erlbaum Associates.

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, 65, 457-474.

Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Kelava, A. (2009). Assessing spurious interaction effects. In T. Timothy & M. S. Khine (Eds.), *Structural equation modeling in educational research* (pp. 13-28). Rotterdam: Sense Publishers.

Le, H., Oh, I. S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, 96, 113-133.

MacCallum, R., & Mar, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin, 118*, 405-421.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19-40.

Marsh, H., Wen, Z., & Hau, K. T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9,* 275-300.

Muthén, LK., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th Ed.). Los Angeles, CA: Muthén & Muthén.

Wright, G. C., jr. (1976). Linear models for evaluating conditional relationships. *American Journal of Political Science, 20*, 349-373.

Salamin, A., & Hom, P. W. (2005). In search of the elusive U-shaped performance-turnover relationship: Are high performing Swiss bankers more liable to quit? *Journal of Applied Psychology, 90*, 1204-1216.

KEVIN FISCHER
A student in the Department of Psychology at Goethe University Frankfurt, Germany. His research interests include nonlinear structural equation modelling and evaluation of goodness of fit measures.

HOLGER BRANDT
PhD candidate in the Department of Psychology at Goethe University Frankfurt, Germany. His research interests include nonlinear structural equation modelling, latent longitudinal models, and causal models for the identification of intermediate variables.

KARIN SCHERMELLEH-ENGEL
Professor of the Department of Psychological Research Methods and Evaluation, at Goethe University Frankfurt, Germany. Her research focuses on nonlinear and linear structural equation modelling. Her research interests include confirmatory factor analyses, evaluation of model fit, multivariate statistics, construct validation, and psychological assessment.

HELFRIED MOOSBRUGGER
Professor emeritis of the Department of Psychology at Goethe University, Frankfurt, Germany. His research focuses on test theory and psychological assessment as well as nonlinear structural equation modelling. His current research interests concern evaluation of model fit, evaluation of academic achievement, construct validation, and computer-based psychological assessment.

ANDREAS KLEIN
Professor of the Department of Psychology at Goethe University Frankfurt, Germany. His research interests include statistical developments in the areas of nonlinear structural equation modelling, model fit, latent longitudinal models, and models for the detection of surrogate and mediator variables in causal relationships.

Multilevel exploratory factor analysis of discrete data

Exploratory factor analysis (EFA) can be used to determine the dimensionality of a set of items. When data come from clustered subjects, such as pupils within schools or children within families, the hierarchical structure of the data should be taken into account. Standard multilevel EFA is only suited for the analysis of continuous data. However, with the robust weighted least squares estimation procedures that are implemented in the computer program M*plus*, it has become possible to easily conduct EFA of multilevel discrete data. In the present paper, we show how multilevel EFA can be used to determine the dimensionality in discrete two-level data. Measurement invariance across clusters implies equal dimensionality across levels. We describe two procedures, one with and one without measurement invariance restrictions across clusters. Data from educational research serve as an illustrative example.

Where: Netherlands Journal of Psychology, Volume 67, 114-121

Received: 30 August 2012; Accepted: 16 January 2013

Keywords: Exploratory factor analysis; Dimensionality; Discrete data; Multilevel data; Weighted least squares estimation

Authors: Mariska (M.T.) Barendse*, Frans J. Oort**, Suzanne Jak** and Marieke E. Timmerman*

*Heymans Institute of Psychology, University of Groningen, the Netherlands, **Research Institute of Child Development and Education, University of Amsterdam, the Netherlands

Correspondence to: F.J. Oort, Research Institute of Child Development and Education, University of Amsterdam; Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, the Netherlands, e-mail: f.j.oort@uva.nl

Introduction

The dimensionality of a set of items can be defined as the minimum number of underlying unobserved (latent) variables that is needed to describe all relationships between all item responses (Lord & Novick, 1968; Zhang & Stout, 1999). If we restrict ourselves to linear relationships, then exploratory factor analysis (EFA) can be used to assess how many latent variables (or common factors) are needed to explain all item responses (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999; Conway & Huffcutt, 2003). EFA is an appropriate technique to determine dimensionality because the EFA model is unconstrained, so that any misfit can only be attributed to the number of factors being too small. However, ordinary EFA is only suited for the analysis of normally distributed continuous item responses.

Item responses are generally discrete. Test items are often scored as ‘right’ or ‘wrong’, with binary codings 1 and 0. Or respondents give judgements on, for example, a three-point response scale with ‘not applicable to me’, ‘somewhat applicable to me’, and ‘applicable to me’ scored as 1, 2, 3. Wirth and Edwards (2007) give an overview of estimation methods that can be used with discrete item responses. Some of these have been implemented in structural equation modelling computer programs such as M*plus* (Muthén & Muthén, 2010), and so it has become feasible to conduct factor analysis of discrete variables. Barendse, Oort, and Timmerman (2012) conducted a simulation study of EFA of discrete variables and found that robust weighted least squares estimation with polychoric correlations worked well in assessing dimensionality.

In social and behavioural research, we often encounter hierarchically structured data, such as data from students in schools, children in families, or patients sharing the same physicians. Mixed model or multilevel analysis accounts for the dependencies in multilevel data (Snijders & Bosker, 1999). In the case of two-level data, the first level pertains to within-cluster variation (e.g., differences between students within schools) and the second level to between-cluster variation (e.g., differences between schools). Due to the work of Asparouhov and Muthén (2007), the robust weighted least squares estimation implemented in M*plus* (Muthén & Muthén, 2010) can handle multilevel discrete data.

The purpose of this paper is to show how multilevel EFA analysis can be used to assess the dimensionality of a set of discrete responses. We will describe two procedures. In the first procedure we separately assess the dimensionality of within-cluster variance and between-cluster variance, without any restrictions across levels. In the second procedure we assume measurement invariance across clusters, to make sure that the common factors have the same interpretation across clusters. Jak, Oort, and Dolan (2012a) have shown that this measurement invariance restriction implies measurement invariance across levels as well.

Both procedures will be illustrated with data from educational research on student-teacher relationships.

Methods

Below we briefly describe the two-level EFA model, the identification and estimation of its parameters, the evaluation of fit, the two procedures to assess dimensionality, and the rotation of a two-level EFA solution. We currently apply two-level EFA to discrete item responses, but the approach can also be applied to other variables (e.g., continuous scores, counts), and be extended to more than two levels.

With discrete data we assume that the observed discrete item responses are representations of continuous unobserved responses. That is, the vector of observed discrete item responses x_{ij} of individual i in cluster j is considered to be a representation of a vector of underlying continuous response variables y_{ij}, with associated thresholds that determine the x_{ij} values (e.g., Olsson, 1979; Muthén, 1984).

Model

In multilevel models, the underlying continuous variables y_{ij} are decomposed into cluster means μ_j, and individual deviations from the cluster means η_{ij}:

$$y_{ij} = \mu_j + \eta_{ij} \,. \tag{1}$$

The individual deviations η_{ij} are assumed to be independent of the cluster means μ_j so that variance-covariance matrix of y, denoted Σ_{TOTAL} (with variances and covariances across all clusters), is the sum of the variance-covariance matrix of μ, denoted Σ_{BETWEEN} (with variances and covariances between clusters), and the variance-covariance matrix of η, denoted Σ_{WITHIN} (with variances and covariances within clusters),

$$\Sigma_{\text{TOTAL}} = \Sigma_{\text{BETWEEN}} + \Sigma_{\text{WITHIN}} \,. \tag{2}$$

In two-level factor analysis, the between and within variance-covariance matrices can be separately modelled as

$$\Sigma_{\text{BETWEEN}} = \Lambda_B \Phi_B \Lambda_B' + \Theta_B \,, \tag{3}$$
$$\Sigma_{\text{WITHIN}} = \Lambda_W \Phi_W \Lambda_W' + \Theta_W \,. \tag{4}$$

In Equation 3, Φ_B is the variance-covariance matrix of the common between factors of the cluster means μ, Λ_B is the matrix of factor loadings of the cluster means on these common between factors, and Θ_B is the (diagonal) matrix with residual variances of the cluster means. In Equation 4, Φ_W is the pooled-within variance-covariance matrix of the common within factors of the individual deviations from the cluster means, Λ_W is the pooled-within matrix of factor loadings of the individual deviations on these common within factors, and Θ_W is the (diagonal) pooled-within matrix with residual variances of the individual deviations.

Measurement invariance

If we want to make sure that the interpretation of the common within factors is the same in all clusters, then we have to assume measurement invariance across clusters (i.e., in factor analysis of mean and covariance structures, intercepts and factor loadings of y-variables are the same across clusters; Muthén, 1994; Rabe-Hesketh, Skrondal, & Pickles, 2004; Jak, Oort, & Dolan, 2012a, 2012b). Jak et al. (2012a) explain that measurement invariance across clusters implies equal factor loadings across levels ($\Lambda_W = \Lambda_B = \Lambda$), yielding the following two-level model:

$$\Sigma_{\text{BETWEEN}} = \Lambda \Phi_B \Lambda' \,, \tag{5}$$
$$\Sigma_{\text{WITHIN}} = \Lambda \Phi_W \Lambda' + \Theta_W \,, \tag{6}$$

where Λ is a matrix of factor loadings that is equal across all clusters and across the within and between levels, implying that common factors do have the same interpretation across all clusters and across levels. In addition, there is no residual variance at the between level ($\Theta_B = 0$), implying that no other factors than the common factors are affecting the between-level responses (no 'cluster bias', Jak et al., 2012a).

Identification

In ordinary EFA, the (single level) model is identified with sufficient and necessary scaling and rotation constraints such as an identity matrix for the variance-covariance matrix of common factors ($\Phi = I$) and echelon form for the matrix of factor loadings (Λ elements $\lambda_{pk} = 0$ if $p < k$). In two-level EFA (Equations 3 and 4), sufficient constraints are $\Phi_W = I$, $\Phi_B = I$, and echelon form for both Λ_W and Λ_B. However, if we assume measurement invariance ($\Lambda_W = \Lambda_B = \Lambda$, and $\Theta_B = 0$), then we can estimate the variances of the common factors at the between level (i.e., diagonal(Φ_B) free instead of $\Phi_B = I$). In addition, we can choose either

- to estimate the full factor loading matrix instead of having an echelon form (Λ full free instead of Λ echelon), or

- to estimate correlations between the common factors at the within level (diagonal(Φ_W) = I instead of Φ_W = I), or
- to estimate covariances between the factors at the between level (Φ_B symmetrical free instead of Φ_B diagonal free).

Estimation

The computer program M*plus* provides various estimation methods for SEM with discrete data (Muthén & Muthén, 2010), such as the so-called weighted least squares estimation method with a robust mean-and-variance corrected chi-square fit criterion (WLSMV; Muthén, du Toit, & Spisic, 1997), which has been advocated in previous simulation studies (e.g., Beauducel & Herzberg, 2006; Barendse, et al., 2012). Asparouhov and Muthén (2007) developed a method for multilevel data that can be applied to discrete data, using polychoric correlations. To compare nested multilevel models, one should use the estimation method with a mean-corrected chi-square fit criterion (denoted WLSM; rather than the mean-and-variance corrected WLSMV), as only WLSM provides a valid chi-square statistic to test the difference in fit of nested multilevel models (Muthén, 1998-2004; Satorra & Bentler, 2001).

Evaluation of fit

As the evaluation of fit of multilevel models for discrete data is still subject to study, we resort to fit criteria that are commonly applied in structural equation modelling. A significant chi-square test of overall goodness-of-fit indicates that the model does not fit the data (i.e., the hypothesis of exact population fit is rejected). In addition to the chi-square test of exact fit, we can use the root mean square error of approximation (RMSEA) as an index of approximate fit. RMSEA values below 0.08 and 0.05 indicate satisfactory and close fit, respectively (Browne & Cudeck, 1992). We will also report the standardised root mean square residual (SRMSR) and its weighted counterpart (WRMSR), which indicate the difference between the polychoric correlations and the correlations implied by the EFA model. SRMSR values below 0.05 (e.g., Sivo, Fan, Witta, & Willse, 2006) and WRMSR values below 1.0 (Yu & Muthén, 2002) are considered acceptable.

The difference in fit of two hierarchically related models (or nested models) can be tested with the chi-square difference test. We should note, however, that with WLSM estimation, this chi-square difference is subject to a scaling correction and cannot be calculated by simply taking the difference of the two chi-square values that are associated with the fit of two models (Muthén, 1998-2004; Satorra & Bentler, 2001).

Dimensionality assessment

We describe two procedures to determine the dimensionality of two-level data.

Procedure 1. The first procedure has two steps. In the first step, we leave Σ_{BETWEEN} free to be estimated, impose an exploratory factor model on Σ_{WITHIN} (Equation 4), and fit a series of models with increasing numbers of common within factors to determine the minimum number of common within factors that provides good fit. In the second step, we retain the minimum number of common within factors (determined in the first step), and fit a series of models with increasing numbers of common between factors to determine the minimum number of common between factors that provides good fit.

Procedure 1 may yield a different number of between factors than the number of within factors. So, the dimensionality of the between structure may be different from the dimensionality of the within structure. Still, even if the dimensionality is the same across levels, the interpretation of the between factors is different from the interpretation of the within factors as Λ_W and Λ_B are different. Moreover, the interpretation of the factors across clusters is not the same either, as the values of the Λ_W elements are pooled within values. Matrix Λ_W can be interpreted as the average of as many cluster specific Λ matrices as there are clusters. So, in theory, the Λ_W interpretation may not apply to any of the individual clusters at all.

Procedure 2. In Procedure 2 we require measurement invariance across clusters, which implies $\Lambda_W = \Lambda_B$ and $\Theta_B = 0$ (Jak et al., 2012a). With these restrictions, we fit a series of two-level EFA models to Σ_{WITHIN} and Σ_{BETWEEN} as given by Equations 5 and 6, with increasing numbers of common factors, to determine the minimum number of common factors that provides good fit. Due to the measurement invariance restriction, the common factors have the same number and the same interpretations across all clusters and across both levels.

Rotation

Just as in ordinary (single level) EFA, the solution can be rotated to facilitate interpretation. If the solution is obtained through Procedure 1, using the two-level EFA given by Equations 3 and 4, with both Φ_W and Φ_B equal to identity and both Λ_W and Λ_B having echelon form, then the within and between solutions can be rotated separately, in the same way as in ordinary EFA (Browne, 2001; Oort, 2011). If the solution is obtained through Procedure 2, using the two-level EFA given by Equations 5 and 6, with Φ_W identity, Φ_B free, and Λ echelon, then we preserve the identical interpretation of within and between factors by rotating the within and between structures together. Application of a rotation criterion as desired to the echelon Λ yields a transformation matrix T, and rotated factor loadings Λ^* and variance-covariance matrices Φ_W^* and Φ_B^*,

$$\Lambda^* = \Lambda\, T\,, \tag{7}$$

$$\Phi_W^* = (T^{-1})(T^{-1})' = (T'\,T)^{-1}\,, \tag{8}$$

$$\Phi_B^* = (T^{-1})\,\Phi_B\,(T^{-1})'\,. \tag{9}$$

See Browne (2001) for a comprehensive explanation of rotation in EFA.

Illustration

As an illustrative example, we apply multilevel EFA to data that were gathered with the student-teacher relationship scale (STRS; Koomen, Verschueren, Van Schooten, Jak, & Pianta, 2011). We have complete data from 649 teachers who reported about their relationships with two or three children each, 1493 children in total, aged 3 to 12. The 28 items of the STRS are hypothesised to capture three aspects of the student-teacher relationship: closeness, conflict, and dependency. The items have five-point response scales, ranging from 1 (‘definitely does not apply’) to 5 (‘definitely does apply’).

Preliminary analysis

First we check whether the between-level variances and covariances are sufficiently large to warrant a multilevel analysis. Intra-class coefficients of the item responses vary between 0.15 and 0.49. Furthermore, we fitted a Null Model ($\Sigma_{BETWEEN} = 0$, Σ_{WITHIN} free) to test whether there is between-level variance, and an Independence Model ($\Sigma_{BETWEEN}$ diagonal, Σ_{WITHIN} free) to test whether there is between-level covariance. Neither model fits the data: Null Model chi-square = 4547.4, df = 389, $p < 0.001$, RMSEA = 0.085; Independence Model chi-square = 4195.0, df = 378, $p < 0.001$, RMSEA = 0.082. As the intra-class coefficients are high and the Null Model and Independence Model do not fit the data, we conclude that these data require a model that accounts for the two-level hierarchical structure of the data.

Procedure 1 within-level results

Table 1 gives the fit results (chi-square, RMSEA,

Table 1 Series of multilevel exploratory factor analyses to determine the dimensionality

Number of within factors	Number of between factors	DF	Chi-square	RMSEA	SRMSR Within	SRMSR Between	WRMSR	Chi-square difference test: Chi-square	DF	Prob.
Series 1										
1	n/a	350	13737.605	0.160	0.169	-	3.122	-	-	-
2	n/a	323	2302.291	0.064	0.059	-	1.061	16647.295	27	0.000
3	n/a	297	827.289	0.035	0.033	-	0.591	619.279	26	0.000
4	n/a	272	576.248	0.027	0.028	-	0.480	181.013	25	0.000
5	n/a	248	464.731	0.024	0.024	-	0.421	88.807	24	0.000
Series 2										
3	1	647	3086.883	0.050	0.033	0.180	1.182	-	-	-
3	2	620	1332.493	0.028	0.033	0.077	0.685	338.595	27	0.000
3	3	594	1300.727	0.028	0.033	0.067	0.665	43.458	26	0.017
3	4	569	1258.966	0.028	0.033	0.061	0.650	44.247	25	0.010
3	5	545	1223.605	0.029	0.033	0.058	0.639	37.789	24	0.036
Series 3										
1	1	755	14252.961	0.109	0.170	0.589	3.483	-	-	-
2	2	726	4018.500	0.055	0.062	0.304	1.677	2004.806	29	0.000
3	3	697	2705.744	0.044	0.048	0.291	1.236	325.442	29	0.000
4	4	668	1923.310	0.035	0.034	0.278	0.920	198.966	29	0.000
5	5	639	1590.571	0.032	0.032	0.244	0.760	125.307	29	0.000

Note: 1493 pupils are rated by 649 teachers.

SRMSR, WRMSR) for three series of two-level EFA models. In the first series, Σ_{BETWEEN} is unrestricted and Σ_{WITHIN} conforms a one-, two-, three-, four-, or five-factor model (as in Equation 4). The chi-square test is consistently significant, indicating that none of the models fits the data exactly. However, the RMSEA indicates satisfactory fit of the two-factor model and close fit of the three-factor model. The SRMSR and WRMSR indices also suggest acceptable fit of the three-factor model. We therefore continue Procedure 1 with three factors at the within level.

Procedure 1 between-level results

In the second series, Σ_{WITHIN} is restricted to a three-factor model (Equation 4), and Σ_{BETWEEN} is restricted to either a one-, two-, three-, four-, or five-factor model (Equation 3). For each of these models, the chi-square test of exact fit is significant, but due to the gain in degrees of freedom, the relative fit is much better than in the first series of models. According to the RMSEA we would select the EFA model with three within factors and two between factors.

The chi-square difference test indicates that exact fit keeps improving with each additional between factor, but only if we test a 5% level of significance. When testing at a 1% level of significance, we would also select the EFA model with three within factors and two between factors, because at 1%, an additional between factor does not significantly improve exact fit. The same model is also suggested by the WRMSR, but the between-level SRMSR does not fall below 0.05 for any of the models.

Procedure 2 measurement invariance results

In the third series of models we impose measurement invariance restrictions and fit two-level EFA models as given by Equations 5 and 6, with increasing numbers of factors. All chi-square tests are significant, thereby rejecting exact fit. The three-factor model is the first model that meets the RMSEA criterion of close fit (RMSEA < 0.05). The same model also meets the SRMSR criterion (SRMSR < 0.05), but only for the within part. The WRMSR criterion (WRMSR < 1.0) suggests a four-factor model, but Φ_B estimates for this model have unreasonably high standard errors.

Relying on the RMSEA index of fit and on the substantive argument that the STRS is supposed to cover three aspects of student-teacher relationships, we prefer the three-factor model.

The three-factor EFA model with measurement invariance restrictions is nested under the three-within three-between factor model without measurement invariance restrictions in the second series. According to the Satorra and Bentler (2001) chi-square difference test, the hypothesis of measurement invariance should be rejected (chi-square difference = 582.7, df = 103, p < 0.001). However, as the RMSEA nevertheless indicates close fit for the restricted model as well, we still prefer the measurement invariant EFA model.

Rotation results

A substantive interpretation of the common factors that is valid across all clusters requires measurement invariance. To facilitate the interpretation of the three-factor two-level EFA model with measurement invariance (Equations 5 and 6), we use the oblimin criterion to rotate the solution (Browne, 2001). As student-teacher relationship factors are likely to be correlated, we opted for oblique rotation, rather than orthogonal. Rotation results are given in Table 2.

From Table 2 it appears that almost all conflict, dependency, and closeness items have their highest loadings on the first, second, and third factor. We have therefore named these factors ‘Conflict’, ‘Dependency’, and ‘Closeness’. Oblique rotation yields correlated factors. The correlations between the factors Conflict and Dependency (0.39 within level and 0.76 between level), and between Conflict and Closeness (-0.40 within level and -0.64 between level) are substantial. Conspicuously, the within-level correlation between Dependency and Closeness is positive (0.17), albeit small, whereas the between-level correlation is negative (-0.23), showing a difference in the sign of the correlations between judgements of pupils on the one hand and judgements by teachers on the other hand. We note that Koomen et al. (2011) found a zero correlation between Dependency and Closeness, but they neglected the two-level structure of the data and conducted a confirmatory factor analysis with simple structure.

Discussion

In this paper, we have proposed and illustrated two EFA procedures to determine the dimensionality of multilevel discrete data. The first procedure does not involve any across level restrictions, leaving room for different within-level and between-level factor solutions. In that case, the within-level factor loadings (Λ_W) should be interpreted as a summary of all possible individual cluster factor loadings. In the second procedure we assume measurement invariance, to make sure that factors have the same interpretation across all clusters. This assumption entails across-level invariance of within-level and between-level factor loadings ($\Lambda_W = \Lambda_B$).

Without the measurement invariance restriction, common factors may not have the same interpretation across clusters, or across levels, giving room to so-called 'cluster bias' (Jak et al., 2012a, 2012b). In the presence of cluster bias, differences between test scores are not completely attributable to differences in the trait(s) one intended to measure. In our student-teacher relationships example, different STRS item scores should be fully explained by differences in scores on the common factors that

Table 2 Exploratory factor analysis of 28 items of the Student-Teacher Relationship Scale (STRS; 649 teachers and 1493 pupils): Oblimin rotation of a three-factor two-level model with measurement invariance

Within- and between-factor loadings ($\Lambda_W = \Lambda_B$)

Items	Conflict	Dependency	Closeness
Closeness items			
I share an affectionate, warm relationship with this child	-0.485	0.050	**1.415**
If upset, this child will seek comfort from me	0.080	0.052	**1.131**
This child is uncomfortable with physical affection or touch from me	0.013	-0.038	**0.601**
This child values his/her relationship with me	-0.206	-0.045	**1.200**
When I praise this child, he/she beams with pride	0.190	-0.066	**0.780**
This child is overly dependent on me	-0.442	0.379	**0.706**
This child tries to please me	-0.009	0.014	**1.031**
It is easy to be in tune with what this child is feeling	0.113	0.076	**1.315**
This child openly shares his/her feelings and experiences with me	-0.514	0.107	**1.104**
This child allows himself/herself to be encouraged by me	0.014	0.147	**0.746**
This child seems to feel secure with me	-0.466	-0.067	**1.225**
Conflict items			
This child and I always seem to be struggling with each other	**1.369**	-0.071	-0.127
This child easily becomes angry with me	**1.293**	0.059	0.106
This child feels that I treat him/her unfairly	**1.313**	0.024	-0.118
This child sees me as a source of punishment and criticism	**0.943**	0.240	-0.423
This child remains angry or is resistant after being disciplined	**1.477**	0.022	0.134
Dealing with this child drains my energy	**1.957**	-0.014	0.067
When this child is in a bad mood, I know we're in for a long and difficult day	**1.583**	0.176	0.163
This child's feelings toward me can be unpredictable or can change suddenly	**1.572**	0.133	-0.117
Despite my best efforts, I'm uncomfortable with how this child and I get along	**1.187**	0.176	**-0.838**
This child whines or cries when he/she wants something from me.	**0.644**	**0.713**	-0.160
This child is sneaky or manipulative with me	**0.918**	0.099	-0.390
Dependency items			
This child reacts strongly to separation from me	0.050	**0.672**	0.069
This child is overly dependent on me	-0.373	**1.609**	-0.203
This child asks for my help when he/she really does not need help	0.241	**0.619**	0.099
This child expresses hurt or jealousy when I spend time with other children	0.525	**0.650**	-0.022
This child fixes his/her attention on me the whole day long	0.039	**0.898**	0.209
This child needs to be continually confirmed by me	0.156	**0.685**	0.028

	Within-factor correlations (Φ_W)			Between-factor correlations (Φ_B)		
	Conflict	Dependency	Closeness	Conflict	Dependency	Closeness
Conflict	1.000			0.201 (1.000)		
Dependency	0.391	1.000		0.314 (0.757)	0.857 (1.000)	
Closeness	-0.402	0.173	1.000	-0.197 (-0.644)	-0.114 (-0.228)	0.466 (1.000)

* *Correlations are given within parentheses; factor loadings > 0.6 are in bold type set; residual variances (Θ_W) not shown.*

we named Conflict, Dependency, and Closeness. If there is cluster bias then apparently other between factors, such as the sex of the teacher or size of the class, also directly affect the STRS item scores. Cluster bias in item responses would then invalidate comparisons of groups that differ in, for example, teacher sex or class size.

In the illustrative analysis of the STRS data, the hypothesis of measurement invariance in the three-factor two-level EFA is rejected by the chi-square difference test (WLSM chi-square difference = 582.7, df = 103, $p < 0.001$). With higher dimensional models, the hypothesis is rejected as well (four-factor WLSM chi-square difference = 562.8, df = 124, $p < 0.001$; five-factor WLSM chi-square difference = 603.4, df = 143). This suggests that measurement invariance does not really hold (in the population). However, considering the fit criteria that indicate close fit, we still prefer the three-factor measurement invariant EFA model, especially because the measurement invariance restriction is substantively important. Without this restriction we cannot validly interpret the within-level EFA results, and therefore we are willing to sacrifice exact fit for interpretability.

The evaluation of fit of multilevel models to discrete data is still subject to study, with inconclusive results, both in the structural equation modelling of discrete data and in the structural equation modelling of multilevel data. Fit measures of multilevel models express the combined (mis)fit at multiple levels. As there are many more observations at the within level than at the between level, the within level has more influence on the overall fit than the between level. Ryu and West (2009) and Boulton (2011) proposed level-specific fit measures for multilevel structural equation modelling (e.g., SRMSR within and SRMSR between). As yet, it is most sensible not to rely on a single fit criterion, and to take the within- and between-level sample sizes into account.

In the present study we combined the challenges of multilevel data and discrete data. Our example analysis shows that it is possible to conduct EFA with multilevel discrete data, that it yields interpretable results, but that the evaluation of fit is partly subjective.

Acknowledgements

This publication is supported by MAGW open competition grant 400-09-084 from the Netherlands Organization for Scientific Research (NWO). The authors thank Helma Koomen for making her data available for secondary analysis.

References

Asparouhov, T. & Muthén, B. (2007). *Computationally efficient estimation of multilevel high-dimensional latent variable models*. Proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology.

Barendse, M. T., Oort, F. J. & Timmerman, M. E. (2012). Using exploratory factor analysis to determine the dimensionality of discrete responses. Manuscript submitted for publication.

Beauducel, A. & Herzberg P. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least square estimation in confirmatory factor analysis. *Structural Equation Modeling, 13,* 186-203.

Boulton, A. (2011). *Fit Index Sensitivity in Multilevel Structural Equation Modeling*. Unpublished doctoral dissertation, University of Kansas-Kansas.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research, 21,* 230-258.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36,* 111-150.

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods 6,* 147-168.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272-299.

Jak, S., Oort, F. J., & Dolan, C. V. (2012a). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling,* in press.

Jak, S., Oort, F. J., & Dolan, C. V. (2012b). Measurement bias in multilevel data. *Structural Equation Modeling,* in press.

Koomen, H. M. Y., Verschueren, K., van Schooten, E., Jak, S., & Pianta R. (2011) Validating the Student-Teacher Relationship Scale: Testing Factor Structure and Measurement Invariance across Child Gender and Age in a Dutch Sample. *Journal of School Psychology, 50,* 215-234.

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

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115-132.

Muthén, B. (1994), Multilevel covariance structure analysis. *Sociological Methods & Research 22,* 376-398.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript technical report. Retrieved

April 2, 2011, from http://pages.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

Muthén, B. (1998-2004). *M-plus Technical Appendices*. Los Angeles, CA: Muthén & Muthén. Retrieved April 5, 2011, http://statmodel.com/download/techappen.pdf

Muthén, B. O., & Muthén, L. K. (2010). *M-plus user's guide: statistical analysis with latent variables*. Los Angeles, CA: Muthén & Muthén.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44,* 443-460.

Oort, F. J. (2011). Likelihood-based confidence intervals in exploratory factor analysis. *Structural Equation Modeling, 18*, 383-396.

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

Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling, 16,* 583-601.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507-514.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for 'optimal' cut-off properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education, 74,* 267-288.

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

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12,* 58-79.

Yu, C. Y., & Muthén, B. (2002). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes* (Technical Report) Los Angeles: University of California at Los Angeles, Graduate School of Education & Information Studies.

Zhang, J., & Stout, W. F. (1999). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika, 64*, 129-152.

MARISKA (M.T.) BARENDSE
PhD candidate at the Department of Psychometrics and Statistics; Heymans Institute of Psychology, University of Groningen. Research interests include: structural equation modelling, exploratory factor analysis, dimensionality of discrete data, and measurement bias.

FRANS J. OORT
Full professor of Methods and Statistics and Director of the Research Institute of Child Development and Education, University of Amsterdam. His research interests include measurement and measurement bias, and statistical modelling, especially non-standard applications of structural equation modelling.

SUZANNE JAK
PhD candidate in Methods and Statistics, Research Institute of Child Development and Education, University of Amsterdam. Her research interests include structural equation modelling, measurement bias, multilevel structural equation modelling and meta-analytical structural equation modelling.

MARIEKE E. TIMMERMAN
Associate professor in the Psychometrics and Statistics group of the Heymans Institute for Psychological Research, University of Groningen. Her research interest is multivariate modelling of data from complex designs, with a special interest in latent variable models and data reduction methods.

Growth curve models and panel dropouts: Applications with criminological panel data

Several statistical models exist to study panel dropouts with respect to the underlying missing data mechanism. The paper discusses two models which extend the classical growth curve model: the selection model and the pattern mixture model. Specific variants of these models are applied with five-wave data from a criminological panel study. The selection model shows that the observed variable has an influence on the rate of panel dropouts within the same panel wave when the average delinquency rate has its peak. In this case the dropout process is not missing at random. With decreasing delinquency the results suggest that the dropout process is missing at random. In addition, the pattern mixture model is able to identify the class of respondents with the highest amount of dropouts which are also those ones which reach the highest delinquency rates in the entire time range of the study.

Where: Netherlands Journal of Psychology, Volume 67, 122-131

Received: 30 June 2012; Accepted: 20 November 2012

Keywords: Missing data; Growth curve model; Selection model; Pattern mixture model; Criminological panel study

Author: Jost Reinecke

Introduction

Several research attempts has been made to cope with missing data for different survey designs. The methodological literature has mostly discredited simple and easy to use methods that discard incomplete cases from the substantive analysis (listwise or pairwise deletion of missing data) or techniques that replace the missing values with a single set of values (e.g., mean imputation). More advanced techniques such as multiple imputation or full information maximum likelihood have been proposed as appropriate under certain conditions relating to the mechanisms that produce the missing data (for an overview see Enders, 2010).

Rubin (1987) distinguishes three different missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR). Following the notation of Little and Rubin (2002) the vector of observed data is Y_{obs}; the vector of missing data is Y_{mis}: $Y = (Y_{obs}, Y_{mis})$.

Correspondence to:
Jost Reinecke, Faculty of Sociology, University of Bielefeld, Germany, e-mail: jost.reinecke@uni-bielefeld.de

A missing data matrix $M(=M_{ij})$ provides the information if, for a person i, the value of a variable j is missing (M_{ij}=1) or not (M_{ij}=0). ψ denotes the parameter vector influencing the probability of missing data. θ contains the parameters of substantive interest.

The missing data mechanism is MCAR when the probability of missing values on a variable does not depend either on the observed values Y_{obs} or the missing values Y_{mis}: $f(M|Y, \psi) = f(M|\psi)$. The assumption of MCAR is required with listwise deletion since observed values for cases with missing values on variables under study are discarded. Under MCAR complete cases are a simple random sample of all cases. The missing data mechanism is MAR when the probability of missingness depends on the observed values, but is unrelated to the missing values: $f(M|Y, \psi) = f(M|Y_{obs}, \psi)$. Additionally it is required that θ and ψ are distinct parameters. MAR is a weaker assumption than MCAR. If the missing data mechanism is MAR likelihood-based conclusions about the parameters of substantive interest (θ) are possible without information about the parameters ψ that govern the missing data process (Little & Rubin, 2002, p. 119). Full information maximum likelihood and multiple imputation produce unbiased parameters under MAR. In contrast to listwise deletion (LD) available data are employed. Finally, the process of missingness is not missing at random (NMAR) when the process of missingness also depends on the missing values themselves: $f(M/Y, \psi) = f(M/Y_{mis}, Y_{obs}, \psi)$. The missing data mechanism cannot be ignored because the missing data contain information about the substantive parameters theta.

In case of NMAR it is necessary to explore and model the missing data mechanism together with the substantive application. Selection models and more advanced variants of pattern mixture models for panel data were recently discussed within the context of Psychology and Methodology (Enders, 2010; Enders, 2011; Muthén, Asparouhov, Hunter, & Leuchter, 2011). These models are extensions of the classical growth curve model which is able to separate intra- and interindividual development of observed measurements (Meredith & Tisak, 1990).

Based on these discussions this paper focuses on applications of a selection and a pattern mixture model for criminological panel data to explore whether permanent and temporary panel attritions are due to MAR or NMAR mechanisms. The following section starts with a brief description of the latent growth curve model followed by the extensions to selection and pattern mixture models. Then the criminological panel data used for the applications are described, followed by the results of the models estimated with the program M*plus* (Muthén, 1998-2010). Finally, a discussion and concluding remarks will be provided.

Models

Growth curve models

The possibility that the individual trajectories of a dependent variable can vary is one of the main advantages of the growth curve model. The formal representation of a growth curve model can be seen either as a multilevel, random-effects model or as a latent variable model, where the random effects are latent variables (Meredith & Tisak, 1990, p. 108; Willett & Sayer, 1994, p. 369):

$$y_i = \Lambda \eta \varepsilon_i \tag{1}$$

y_i is a $t \times 1$ vector of repeated measurements for observation i where t is the number of panel waves. η is a $q \times 1$ vector of latent growth factors where q is the number of these factors. ε is a $t \times 1$ vector of time-specific measurement errors, and Λ is the $t \times q$ matrix of factor loadings with fixed coefficients representing the functional form of the individual trajectories. Variations of individual trajectories are captured by q-numbers of latent variables η whereas usually η_1 is the *intercept*, η_2 is the *linear slope* and in case of nonlinear development η_3 represents the *quadratic slope* (cf. Figure 1).[1] If applicable, additional latent variables can be specified. It is assumed that the latent growth factors and measurement errors are independent and multivariate normally distributed:

$$\begin{bmatrix} \eta \\ \varepsilon_i \end{bmatrix} \approx \left(\begin{bmatrix} \alpha \\ 0 \end{bmatrix}, \begin{bmatrix} \Psi & 0 \\ 0 & \Theta \end{bmatrix} \right) \tag{2}$$

where α is a $q \times 1$ vector of growth factor means and Ψ is the respective $q \times q$ covariance matrix. Θ is a $p \times p$ covariance matrix of time-specific measurement errors which are usually constrained to be a diagonal matrix. For estimation a probability density function is used:

$$f(y_i) = \phi\,[y_i; \mu(\theta), \Sigma(\theta)] \tag{3}$$

where ϕ is the probability density function for y_i and θ is the vector of all parameters to be estimated. $\mu(\theta)$ is a $p \times 1$ model-implied mean vector given by

$$\mu(\theta) = \Lambda\alpha \tag{4}$$

and $\Sigma(\theta)$ is a $p \times p$ model-implied covariance matrix given by

$$\Sigma(\theta) = \Lambda\Psi\Lambda' + \Theta \tag{5}$$

Parameters in θ can be estimated by ML, maximising the likelihood that the measurements y_i are drawn from a multivariate normal distribution. The means of the latent growth factors α show the average development of the measurement y_i across p panel waves within a homogenous population.

Selection and pattern mixture models

For regression analyses the selection model of Heckman (1976, 1979) is often applied as a bias correction method with NMAR data on the particular variable under study. This selection model has

[1] I prefer to discuss the growth curve model with three latent variables η because the observed variable y_t in the applications in the next section shows a nonlinear development over time.

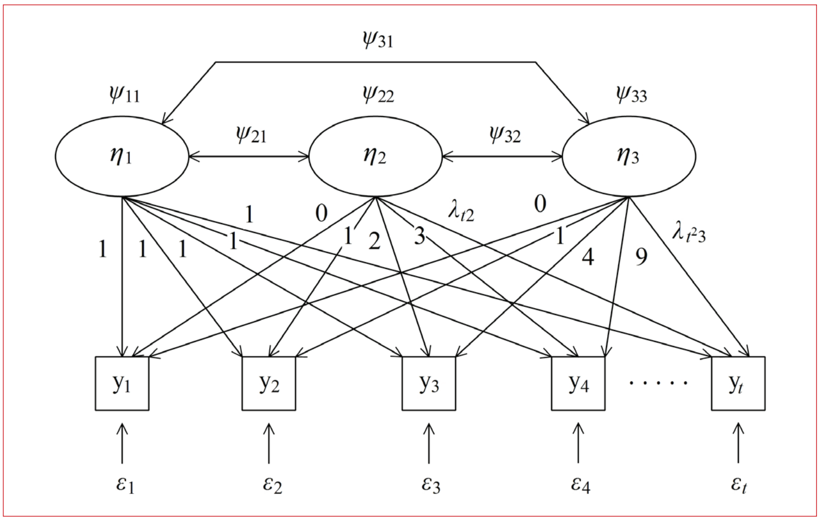

Figure 1 Quadratic growth curve model for t panel waves

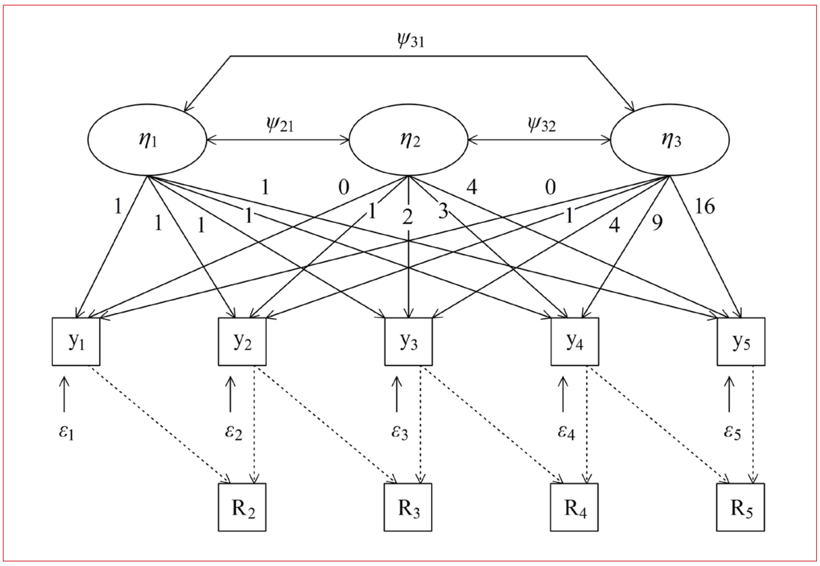

Figure 2 Quadratic growth curve model with indicator variables according to Diggle and Kenward (1994)

two parts: one part consists of the substantive regression equation, the other part predicts the response probabilities with an additional equation. Selection models for longitudinal data also combine a substantive model with additional equations to predict the missingness of the data. In this paper missingness refers to permanent or temporary dropout from a panel study. The substantive part of the model can be analysed with growth curves whereas the methodological part of the model contains logistic regressions predict the missing data indicators. Numerous model formulations are discussed throughout the statistical literature but two longitudinal selection models have been proposed and applied recently (Enders, 2010, p. 304f.; Enders, 2011, p. 7):

1 The selection model of Wu and Carroll (1988) uses growth curve variables to predict the probability of missing data. This model contains missing data indicators R_t which denote whether the observed variable *y* at a particular panel wave *t* is observed or not. The indicator variables are regressed on the growth curve part of the model via logistic or probit regression equations. If panel data with five waves are used the model contains five observed variables (y_1,……,y_5) and five indicator variables (R_1,……,R_5). Here, the probability to remain in the panel is dependent on the random coefficients of the developmental process. For example, the higher the linear slope the more probable are the dropouts of respondents. Linking the response probabilities to the growth curve variables might be useful when the dropout process depends on the overall trajectory (for applications see Enders, 2011 and Reinecke, 2012). The model assumes that observed variables y_t are uncorrelated with indicator variables R_t. Detection of a specific MAR or NMAR dropout mechanism related to particular panel waves seems to be difficult. In addition, variances of the quadratic terms of the models are often too small to explain the probability of missingness. Therefore, the model of Wu and Carroll (1988) will not be considered for the applications.

2 The selection model of Diggle and Kenward (1994) contains the same variables but the indicator variables R_t are regressed directly on the observed variables y_t as well as on the lagged variable y_{t-1} (Figure 2). The significance of the logistic regression coefficients (dashed lines in the figure) allows conclusion about the missing data mechanism: If there are no relationships between y_t, y_{t-1} and R_t, the dropouts are unrelated to the observed variables which would follow a MCAR mechanism. If the lagged variable y_{t-1} has an potential impact on R_t, the mechanism is MAR. That means, dropout at time *t* is related to the observed values from the previous panel wave. If significant within-wave relationships between R_t and y_t are detected, a NMAR mechanism is plausible. That means dropout at time t is related to the observed values from the same panel wave. The model assumes a multivariate normal distribution for the continuous variables y_t.

An alternative framework to model NMAR dropout mechanism is the pattern mixture model. This approach defines subgroups of cases with the same missing data pattern and estimates the substantive model within each group or pattern. The pattern-specific estimates can be averaged across the groups to get a single set of estimates that account for the NMAR mechanism (cf. Enders, 2010, p. 299). But, some of the parameters are not estimable and identifying restrictions have to be specified. Combining the indicator variables R_t with a quadratic

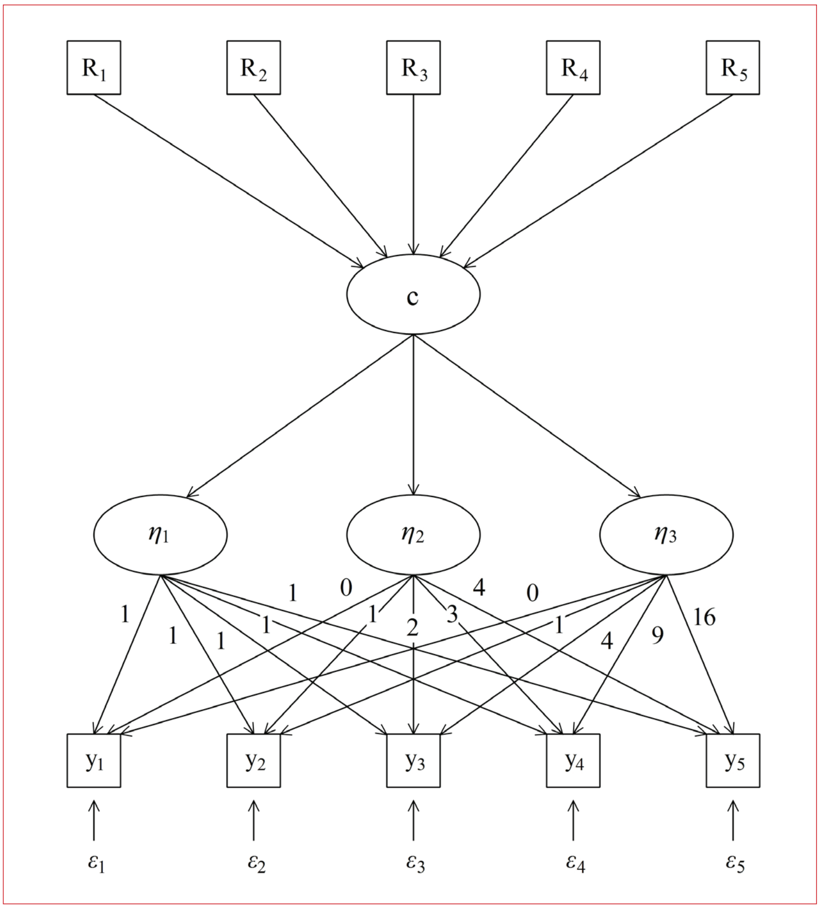

Figure 3 Latent dropout pattern mixture model (Roy, 2003)

growth curve model leads to some non-identified parameters (e. g. the means of the linear and quadratic slope for respondents dropping after t_1 and the mean of the quadratic slope for respondents dropping out after t_2, see Muthén et al., 2011, p 20). Equality restrictions to other dropout patterns can be applied and solve the identification problem. Despite of the fact that wrong restrictions can lead to substantial bias the pattern mixture model assumes that every respondent with the same dropout time has a common distribution, i. e. the sample under study is homogenous in that respect. To overcome this assumption of homogeneity Roy (2003) proposed a *latent dropout pattern mixture model* where a class variable c is influenced by the dropout indicators R_t (Figure 3). Variables R_t and latent class variable c are connected via a multinomial logistic regression model:

$$P(c_i = k \mid R_{1i} \dots R_{Ti}) = \frac{e^{\gamma_{0k}+\sum_{t=1}^{T}\gamma_{tk}R_{ti}}}{\sum_{s=1}^{K} e^{\gamma_{0s}+\sum_{t=1}^{T}\gamma_{tk}R_{ti}}} \qquad (6)$$

Equation 6 estimates the probability that a panel dropout for a particular class is higher or lower than for the reference class. Latent class variable c itself influences the latent growth curve variables and models the unobserved heterogeneity of the development under study. Instead of considering individual variation of single means of the vector η the so-called growth mixture model (GMM) allows different classes of individuals to vary around different means (Muthén & Shedden, 1999):

$$y_{ik} = \Lambda_k \eta_{jk} + \varepsilon_{ik} \qquad (7)$$

Parameters of the model are estimated for $k = 1,\dots,. K$ latent classes. The number of categories of class variable c represent the degree of unobserved heterogeneity in the data. The probability density function for the GMM is a finite mixture of normal distributions:

$$f(y_i) = \sum_{k=1}^{K} \pi_k \phi_k [y_i; \mu_k(\theta_k)\Sigma(\theta_k)] \qquad (8)$$

π_k is the unconditional probability that a measurement belongs to latent class k, ϕ_k is the multivariate probability density function for latent class k. $\mu_k(\theta_k)$ represents the model-implied mean vector given by

$$\mu_k(\theta_k) = \Lambda_k \alpha_k \qquad (9)$$

and $\Sigma_k(\theta_k)$ is the model-implied covariance matrix given by

$$\Sigma_k(\theta_k) = \Lambda_k \Psi_k \Lambda_k^t + \Theta_k. \qquad (10)$$

The mixture model of Roy (2003) makes explicit use of the GMM and proposes that the dropout mechanism is related to the mixture of the growth curves.

The model is estimated by maximising the log likelihood function within the admissible range of parameter values given classes and data. The program M*plus* uses the principle of maximum likelihood estimation and employs the EM algorithm for maximisation (Dempster, Laird, & Rubin, 1977; Muthén & Shedden, 1999).[2] For a given solution, each individual's probability of membership in each class is estimated. Individuals can be assigned to the classes by calculating the posterior probability that an individual i belongs to a given class k. Each

[2] The integration method of M*plus* tests several sets of starting values evaluating the maximum initial stage log likelihood value. The seed number corresponding to that value is used for the final estimation of the model. For re-estimation of the model parameters the optimal seed value of the previous run can be included in the input file (for details see Muthén, 1998-2004).

individual's posterior probability estimate for each class is computed as a function of the parameter estimates and the values of the observed data (Muthén, 1998-2004).

It is always an empirical question how many classes are sufficient to describe the unobserved heterogeneity of the data. By classifying each individual into his most likely class, a table with rows corresponding to individuals classified into a given class can be constructed. The columns of that table show the average conditional probabilities to be in the particular class. Quality of the classification is summarised by the entropy measure E_k (Muthén, 1998-2004), which ranges from zero to one, where values close to one indicate a good classification of the data.

In mixture models a k class model is not nested within a $k + 1$ group model. Therefore, conventional mixture tests like the Akaike Information Criterion (AIC; Akaike, 1987), the Bayesian Information Criterion (BIC; Schwarz, 1978) or the sample-size adjusted Bayesian Information Criterion (SABIC; Sclove, 1987) have to be used for model comparisons.[3] If the k-class model contains a redundant class, the $k - 1$-class model with the smaller AIC, BIC or SABIC value should be chosen. An expansion of the model by adding a class is desirable only if the resulting improvement in the log likelihood exceeds the penalty for more parameters. But accepting or rejecting a model on the basis of the information criteria is more or less descriptive and does not imply any statistical test.

Lo, Mendell, and Rubin (2001) proposed a likelihood ratio-based method for testing $k - 1$ classes against k classes in mixture models. The Lo-Mendell-Rubin likelihood ratio test (LMR- LRT) considers the usual likelihood ratio for testing the $k - 1$ model against a k model but with the correct distribution. The p-value from the test represents the probability that H0 is true, i.e., that the model is sufficient with one less class. Therefore, a low p-value indicates that the $k - 1$ class model has to be rejected and the k-class model can be accepted for substantive interpretations. Information criteria and the LMR-LRT will be used for the selection of the appropriate number of classes on the Roy model.

Data

The selection and pattern mixture model of Diggle and Kenward (1994) and Roy (2003) will be applied with data from the Study 'Crime in the modern City' (CrimoC, see Boers, Reinecke, Mariotti, & Seddig, 2010). This ongoing prospective panel study started in 2002 in Duisburg, an industrial city of 500,000 inhabitants in western Germany, with nine annual data waves having been collected. The annual survey started in 2002 with 3411 pupils from the 7th grade of all school types. Their mean age was 13. 70% of the schools (40 out of 57) agreed to participate. From these, 87% of the 7th-graders participated in the first year, which represented 61% of all 7th-graders in Duisburg. In subsequent years, the rates of participation ranged from 84% to 92%.

The CrimoC study contains panel datasets covering different numbers of panel waves. The following analyses stem from a five-wave panel dataset covering the period from late childhood (conducted in the year 2002) to late adolescence (conducted in the year 2006). A panel dataset with the minimum information from at least two out of five panel waves contains 3909 persons. This dataset has 23 different dropout patterns in total. Table 1 summarises nine of the most frequent dropout patterns. Patterns 1 to 5 contain one-wave dropouts and the patterns 6 to 9 two- and three-wave dropouts. Respondents with cross-sectional data from one time point only are not considered. Comparisons with respondents who participated five years in a row (n=1552) show some differences in the distribution of gender and school type. The incomplete panel dataset (n=3909) contains more females, somewhat fewer respondents from lower junior high schools and more from grammar schools.

Sixteen different types of delinquent behaviour were measured by self-reports due to the period of the last 12 months. These types include violence, aggravated assault with or without a weapon, shoplifting, car and bicycle theft, vandalism, graffiti, scratching, drug consumption and drug dealing.[4] Prevalence of the 16 different offenses are summed up to an index for each panel wave. The index has a range between zero and 16. Higher index values indicate more versatile criminal activity. On average, the mean offense rate increases up to the second panel wave (average age of 15) and decreases thereafter. The curvilinear development (also described as age-crime curve) is typical for adolescents aged between 14 and 18 years and is labelled as an *adolescent-limited* type of delinquent behaviour (e.g. Moffitt, 1993). The index of five panel waves (y_t) will be used for the selection and pattern mixture models in the following section. Substantive analysis with the complete data pattern support the curvilinear development of delinquency and the use of quadratic

[3] The SABIC replaces N in the BIC formula with (N + 2)/24 and is included in the M*plus* output.
[4] Internet crime was not included because measurements were not conducted in the first two panel waves.

Table 1 Sample of missing data patterns

Pattern	2002	2003	2004	2005	2006	n
1	--	t_2	t_v	t_4	t_5	403
2	t_1	--	t_3	t_4	t_5	134
3	t_1	t_2	--	t_4	t_5	109
4	t_1	t_2	t_3	--	t_5	96
5	t_1	t_2	t_3	t_4	--	224
6	t_1	t_2	--	--	--	289
7	t_1	t_2	t_3	--	--	150
8	--	--	t_3	t_4	t_5	275
9	--	--	--	t_4	t_5	114

Table 2 Distributions of the indicator variable R for five panel waves (2002-2006)

Panel wave	Variable	n	%
t_1	R_1 (0)	3308	84.6
	R_1 (1)	601	15.4
t_2	R_2 (0)	3265	83.5
	R_2 (1)	644	16.5
t_3	R_3 (0)	3293	77.2
	R_3 (1)	289	7.4
	R_3 (2)	327	8.4
t_4	R_4 (0)	3510	89.8
	R_4 (1)	226	5.8
	R_4 (2)	173	4.4
t_5	R_5 (0)	3209	82.1
	R_5 (1)	467	11.9
	R_5 (2)	233	6.0

(0): Person is not missing, (1): Person has monotone missing pattern.
(2): Person has non-monotone missing pattern

growth curve models (see Mariotti & Reinecke, 2010).

Indicator variables R_t have three categories, one for respondents with no missing data, one for respondents with a monotone missing pattern and one for respondents with a non-monotone missing pattern. Only monotone patterns can occur for the first and second panel wave. Therefore, indicator variables R_1 and R_2 have two categories whereas R_3, R_4 and R_5 have three categories. The distributions are shown in Table 2. Regarding the last three panel waves 5.8% up to 11.9% of the respondents have a monotone dropout pattern. Between 4.4% and 8.4% show a non-montone dropout pattern for the same time period.

Results

Results of the Diggle and Kenward model (Figure 2) as well as of the Roy model (Figure 3) are presented and discussed as follows. Six variants of the Diggle and Kenward selection model are estimated with the program M*plus* (see Table 3). The models DK4, DK5 and DK6 consider the time-invariant variable gender which are related to the growth curve variables as well as to the indicator variables. Models DK2 and DK5 have a smaller number of parameters than DK1 and DK4 because the relations between observed variables y_t and R_t are restricted to be equal across time for $t = 3, 4, 5$. The relation between y_2 and R_2 is estimated separately. In the same way, the lagged relations between y_{t-1} and R_t are restricted to be equal for $t = 3, 4, 5$ and the lagged relation between y_1 and R_2 is estimated separately. The models postulate that time-related differences for the dropout mechanism occur only between t_1 and t_2 and not between t_2 and further panel waves. In model DK5 the relations between gender and indicator variable R_t are also restricted to be equal over time for $t = 3, 4, 5$.

Models DK3 and DK6 enlarge the time-invariant restrictions to all panel waves. There are no separate estimators for the path between y_2 and R_2 as well as for the path between y_1 and R_2. In model DK6 the relations between gender and indicator variable R_t are also restricted to be equal over all measurements.

Comparisons of the information criteria AIC show nearly equal values between DK1 and DK2 as well as between DK4 and DK5. The BIC values support the specification of models DK2 and DK6 whereas the SABIC values support the models DK2 and DK5. All in all, the models DK3 and DK6 seem to be too restrictive according to their AIC and SABIC values. Therefore, models DK2 and DK5 are accepted for a more detailed description and discussion.

The means of the growth curve variables confirm the curvilinear development of delinquency in model DK2 and model DK5 (see Tables 4 and 5): Intercepts and linear slopes have positive parameter estimates while the estimates of the quadratic slopes are negative. The development confirms the usual trend of the age-crime curve detected in other longitudinal studies (e. g. Wikström, Oberwittler, Treiber, & Hardie, 2012).

The within-time influences of the observed variable y_2 on to the indicator variable R_2 are positive and significant in both models indicating that dropouts of the second panel wave are related to the amount of delinquent behaviour in the same wave. This result provides evidence for an NMAR mechanism. The odds ratio for the significant path in model DK2 is 1.147 reflecting a slightly higher chance of dropouts for people with higher delinquent mean rates within the current panel wave (see Table 4). Nearly the same result is obtained in model DK5 (odds ratio=1.136, see Table 5). In contrast, the lagged

Table 3 Results of the Diggle and Kenward models

Model	Parameter	AIC	BIC	SABIC
Models without gender				
DK1	29	74108	74290	74198
DK2	25	74107	74264	74185
DK3	23	74126	74270	74197
Models with gender				
Model	**Parameter**	**AIC**	**BIC**	**SABIC**
DK4	36	72755	72980	72865
DK5	30	72752	72939	72844
DK6	27	72766	72935	72849

Table 4 Estimated parameters of model DK2

Variable	Intercept	Standard error	*z*-value	
Intercept	0.874	0.034	25.390	
Linear slope	0.354	0.034	10.336	
Quadratic slope	-0.101	0.008	-12.953	
Relation	**Regression**	**Standard error**	***z*-value**	**Odds ratio**
$y_1 \rightarrow R_2$	-0.018	0.031	-0.593	0.982
$y_2 \rightarrow R_2$	0.137	0.022	6.308	1.147
$y_{2,3,4} \rightarrow R_{3,4,5}$	0.079	0.013	6.144	1.083
$y_{3,4,5} \rightarrow R_{3,4,5}$	-0.003	0.018	-0.171	0.997

Table 5 Estimated parameters of model DK5

Variable	Intercept	Standard error	*z*-value	
Intercept	1.102	0.054	20.277	
Linear Slope	0.432	0.054	7.946	
Quadratic Slope	-0.115	0.013	-9.113	
Relation	**Regression**	**Standard error**	***z*-value**	**Odds ratio**
$y_1 \rightarrow R_2$	-0.023	0.031	-0.754	0.977
$y_2 \rightarrow R_2$	0.128	0.022	5.910	1.136
Gender $\rightarrow R_2$	-0.341	0.089	-3.874	0.711
$y_{2,3,4} \rightarrow R_{3,4,5}$	0.066	0.014	4.851	1.068
$y_{3,4,5} \rightarrow R_{3,4,5}$	-0.007	0.018	-0.373	0.933
Gender $\rightarrow R_{3,4,5}$	-0.375	0.064	-5.810	0.688

relationships between y_1 and R_2 are small and not significant in both models. For the subsequent waves the reported relationships have opposite results: The lagged relationships are now positive and significant and the within-time regressions are small and not significant. This result provides evidence for an MAR mechanism. Odds ratios for the significant paths are 1.08 and 1.07 reflecting a slightly higher chance of dropouts for people with higher delinquent mean rates in the particular previous panel wave. It seems to be obvious that with overall increasing delinquency rates in early adolescence the chance of temporary dropouts is higher for people with larger delinquency rates. With overall decreasing delinquency rates the dropout mechanism is changing from NMAR to MAR.

For model DK5 the negative influence of gender toward the indicator variable y_2 (odds ratio=0.71) and the other indicator variables (odds ratio=0.69) shows that female respondents have a lower chance of a temporary dropout than male respondents. Earlier analyses with similar panel data measuring self-reported delinquent behaviour have also shown that female respondents have a higher probability to remain in the study (Reinecke & Weins, in press).

The Roy model (Figure 3) considers unobserved heterogeneity in the sample due to the development of delinquent behaviour. Variation of the class variable c (number of classes) reflects the size of the unobserved heterogeneity. Each class represents a subgroup with different trajectories which can be related to dropout time. The following analysis varies the number of classes between one and six. In line with Equation 6 class specific logistic regressions between the indicator variables R_t and class variable c are estimated. For mixture models with two or more classes all or part of the logistic regression coefficients can be set equal within the classes. This would test the assumption that the dropout process within the classes is not dependent on the particular panel wave.

Table 6 gives an overview about the estimated models. Looking at the LMR-LRT, models with more than four classes produce redundant information in the additional classes (p-value>0.05). That applies for models without restrictions as well as for models with the time- related equality restrictions (Models with labels *eq* and *peq*). All in all, the restricted models have better model fits than the unrestricted ones. Models with three or four latent classes reflect the best mixture of the developmental trajectories. Previous mixture analyses have shown that those four different classes reflect a substantial decomposition of the developmental process in delinquent behaviour (cf. Mariotti & Reinecke, 2010; Reinecke, in press).

Additionally, for these models time-related equality restrictions of the logistic regression coefficients are separated between the first two and the last three panel waves (Models with the label *peq*). Recall, that R_1 and R_2 have only two categories and can only consider monotone dropouts (see Table 2). So, these models consider differences in the dropout process across the time range of the study. Therefore, the partly restricted mixture model with four classes showing a significant LMR-LRT (p-value=0.04) will be discussed in detail.

The trajectories of the four classes can be described

Table 6 Results of the mixture models

Classes	Parameter	E_k	AIC	BIC	SABIC	LMR-LRT	*p*-value
1	22	–	77428	77566	77496	–	–
2	31	0.942	75265	75460	75361	2152	0.00
2(eq)	27	0.941	75262	75432	75345	2125	0.00
3	40	0.944	73632	73883	73756	1629	0.00
3(peq)	34	0.945	73360	73844	73736	1612	0.00
3(eq)	32	0.950	73686	73887	73785	1548	0.06
4	46	0.939	72438	72727	72580	1251	0.05
4(peq)*	37	0.939	72431	72663	72545	1243	0.04
4(eq)	34	0.941	72445	72659	72551	1231	0.05
5	55	0.930	71653	71998	71823	792	0.32
5(eq)	39	0.937	71669	71913	71789	768	0.57
6	64	0.936	71068	71469	71266	595	0.26
6(eq)	44	0.933	71073	71349	71209	591	0.19

*(eq) gives the model results under the restriction that the logistic regression coefficients of all panel waves are set to be equal within the classes.(peq) gives the model results under the restriction that the logistic regression coefficients of the first two panel waves and the last three panel waves are set to be equal separately within the classes. *denotes the model discussed in detail.*

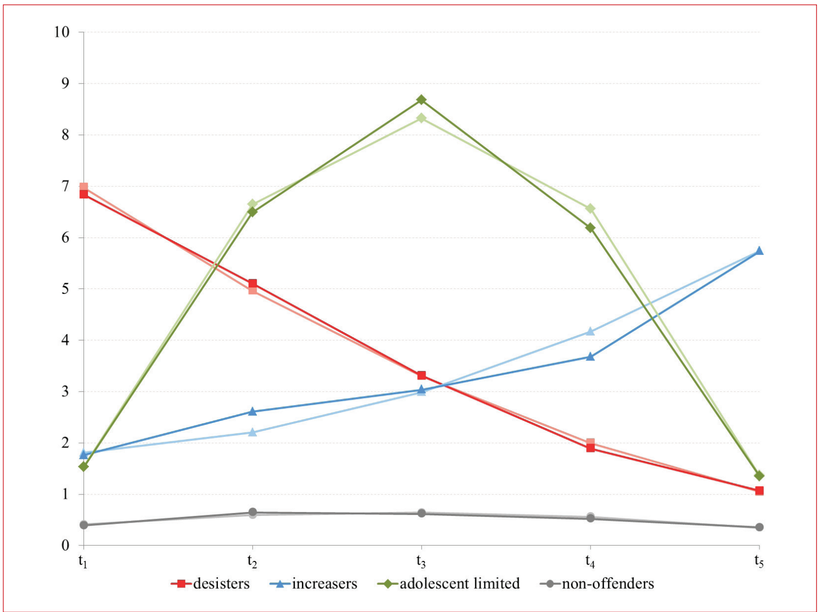

Figure 4 Observed and estimated trajectories of the Roy model with four classes circle = observed means, triangle = estimated means, description of the classes (from bottom to top): non-offenders, increasers, desisters, adolescent limited

as follows (cf. Figure 4): the class of *non-offenders* with almost no delinquency (n=3373), the class of *increasers* with a slow increase of delinquency (n=182), the class of *desisters* with high delinquency at wave one and a continuous decrease thereafter (n=207) and finally a class of people with a curvilinear trajectory reflecting an *adolescent limited* type of delinquent behaviour (n=147). There are only slight differences between observed and estimated means for all four trajectories.

Estimated logistic regression coefficients between indicator variables R_t and class variable c show the influence of the panel dropouts on to the class membership (Table 7). Reference category is class 4 (*non-offenders*). Recall that regression coefficients are set to be equal within classes for the indicator variables R_1 and R_2 and R_3 to R_5 respectively. Most relations are not significant due to the low number of cases. For class 1 (*desisters*) and class 2 (*increasers*) dropouts are not more likely compared with class 4 (*non-offenders*). But for class 3 the chance to drop off the study is much higher than for class 4 (odds ratio=2.243). This result holds for the first two panel waves. There is no similar effect regarding the last three panel waves. Again, the hypothesis that people with larger delinquency rates have a higher chance to drop off the study is supported. For subsequent panel waves the indicator variables have no significant effect on class 3 compared to class 4. With a decrease of delinquency after the third panel wave the chance to drop temporarily off the study is less likely.

Cross-tabulating the distribution of indicator variable R_2 with the distribution of the class variable c indicates that nearly one third of the people with an adolescent limited type of delinquent behaviour dropped temporarily off the study (54 out of 147) whereas the number of dropouts is much lower in the other classes (Table 8). Although relations of the other indicator variables are much smaller the tendency of a relationship between a temporary dropout and the delinquency rate is confirmed. In contrast to the Diggle and Kenward model the Roy model has the advantage that this relationship can be identified for a substantively important group of people, i.e., respondents with high levels of delinquency.

Conclusions

Procedures and techniques to handle missing data in cross-sectional as well as longitudinal designs are well-known and discussed under methodological considerations. Quite often the MAR assumption is reasonable, but in case the missing mechanism is related to the dependent observed variable itself MAR-based techniques would produce biased results. To consider missing data mechanisms in panel designs a substantive model (e.g., latent growth curve model) can be combined with an additional model that describes the dropout process across the panel waves. One possibility is the selection model of Diggle and Kenward (1994) which augments the growth curves with logistic regressions to estimate the probability of missing data at each wave depending on the substantive

Table 7 Logistic regressions of indicator variables R_t

Relation	Coefficient	Standard error	z-value	Odds ratio
R_1, R_2 → Class 1	0.064	0.157	0.407	1.066
R_3, R_4, R_5 → Class 1	0.136	0.090	1.511	1.146
R_1, R_2 → Class 2	0.176	0.289	0.608	1.192
R_3, R_4, R_5 → Class 2	0.008	0.107	0.076	1.008
R_1, R_2 → Class 3	0.808	0.161	5.023	2.243
R_3, R_4, R_5 → Class 3	-0.126	0.104	-1.221	0.881

Class 4 is the reference class (non-offenders).

Table 8 Relation between indicator variable R_2 and classes

R2	Classes				∑
	1	2	3	4	
0	159 76.81%	159 87.36%	93 63.27%	2.854 84.61%	3.265 83.53
1	48 23.19%	23 12.64%	54 36.73%	519 15.39%	644 16.47
∑	207 100.00%	182 100.00%	147 100.00%	3.373 100.00%	3.909 100.00

variable. MAR as well as NMAR mechanisms can be detected. Another possibility is the pattern mixture model of Roy (2003) that makes use of the growth mixture approach (Muthén & Shedden, 1999). The dropout indicators is related to the latent class variable which reflects the degree of unobserved heterogeneity in the data.

Both models have been applied to data from a German panel study which explores the development of adolescents' delinquent behaviour (Boers et al., 2010). Five panel waves are used for the current analyses with respondents participating at least in two of the five waves. The development of delinquent behaviour is curvilinear which requires a latent growth curve model with an intercept, a linear and a quadratic slope. Results of the selection model of Diggle and Kenward (1994) indicate an NMAR mechanism for the first two panel waves while an MAR mechanism is detected for the last three panel waves. This result is also stable when the growth curve model is conditioned on gender. The NMAR mechanism occurs when on average the delinquency rate is increasing while the dropout seems to be MAR when on average the delinquency rate is decreasing. Results of the pattern mixture model of Roy (2003) complement the findings of the selection model. Dropout indicators of the first two waves have a significant influence on the class of respondents which follow an adolescent limited type of delinquent behaviour and have on average the highest delinquency rates across the time range of the study.

The application of the selection and pattern mixture model shows, in principle, the possibility to explore different dropout processes in panel studies due to MAR and NMAR mechanisms. Integration of a submodel that describes the propensity for missing data allows to identify the impact of the missing data mechanism on to the longitudinal results. The applied selection and pattern mixture models are capable to produce accurate parameter estimates when their requisite assumptions hold. But when the assumptions are violated they are also prone to substantial bias. Therefore, the current findings should be interpreted with caution. The logistic regression coefficients are small due to the low number of temporary dropouts. And only one effect is significant in the pattern mixture model for the smallest class of respondents. With further available panel waves the obtained results have to replicated. Other previous analyses in the same substantive context have also detected a significant relationship between the level of delinquency and the possibility to drop off at least temporarily from the longitudinal setting (Reinecke & Weins, in press).

References

Akaike, H. (1987). Factor analysis and the AIC. *Psychometrika*, 52(3), 317-332.

Boers, K., Reinecke, J., Mariotti, L., & Seddig, D. (2010). Explaining the development of adolescent violent delinquency. *European Journal of Criminology*, 7, 499-520.

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

Diggle, P., & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *43*(1), 49-93.

Enders, C. K. (2010). *Applied Missing Data analysis*. New York: Guilford Press.

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16*(1), 1-16.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables

and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4), 120-137.
Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 47*(1), 153-161.
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with Missing Data* (2nd ed.). New York: Wiley.
Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778.
Mariotti, L., & Reinecke, J. (2010). Delinquenzverläufe im Jugendalter: Wachstums- und Mischverteilungsmodelle unter Berücksichtigung unbeobachteter Heterogenität. In *Sozialwissenschaftliche Forschungsdokumentation 21*. Münster: Institut für sozialwissenschaftliche Forschung e.V.
Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122. Moffitt, T. E. (1993). Life-course-persistent and adolescence-limited antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674-701.
Muthén, B. O. (1998-2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.
Muthén, B. O. (1998-2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
Muthén, B. O., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modelling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods, 16*(1), 17-33.
Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463-469.
Reinecke, J. (2012). *Wachstumsmodelle*. Mehring: Rainer Hampp Verlag.
Reinecke, J. (in press). Wachstums- und Mischverteilungsmodelle in der kriminologischen Längsschnittforschung. In S. Eifler & D. Pollich (Eds.), *Empirische Forschung über Kriminalität - Perspektiven und Herausforderungen*. Wiesbaden: VS-Verlag.
Reinecke, J., & Weins, C. (in press). The development of delinquency during adolescence: a comparison of missing data techniques. *Quality and Quantity, online first*, DOI: 10.1007/s11135-012-9721-4.
Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics, 59*(4), 829-836.
Rubin, D. B. (1987). *Multiple Imputation for nonresponse in surveys*. New York: Wiley.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.
Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333-343.
Wikström, P. O., Oberwittler, D., Treiber, K., & Hardie, B. (2012). *Breaking rules: The social and situational dynamics of young people's urban crime*. Oxford: Oxford University Press.
Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381.
Wu, M. C., & Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics, 44*(1), 175-188.

JOST REINECKE
Studied sociology at the University of Duisburg, Germany, and received his Dr. rer. soc. from the University of Giessen, Germany. He is Professor of Quantitative Methods in the Social Sciences at the University of Bielefeld, Germany. His current research areas include statistical techniques of longitudinal data analysis, the handling of missing data in complex survey designs, multilevel structural equation models, latent class analysis, rational choice theories and applications in criminological life-course research.

Meta-analytic structural equation modelling with missing correlations

Cheung and Chan (2005) proposed a two-stage method to conduct meta-analytic structural equation modelling (MASEM). MASEM refers to the technique of fitting structural equation models to pooled correlation or covariance matrices from several studies. Unfortunately, researchers do not always report all correlations between the variables of interest. In this paper, we propose a method to deal with missing correlations in the two-stage approach. We illustrate the proposed model with a meta-analysis of teacher-child relationships variables from 99 studies. In addition, using simulated data, we show that our method leads to more precise parameter estimates than the existing approach.

Where: Netherlands Journal of Psychology, Volume 67, 132-139

Received: 14 August 2012; Accepted: 20 November 2012

Keywords: Meta-analytic structural equation modelling, two-stage approach, missing correlations

Authors: Suzanne Jak*, Frans J. Oort*, Debora L. Roorda** and Helma M. Y. Koomen*

*Research Institute of Child Development and Education, University of Amsterdam, the Netherlands, **School Psychology and Child and Adolescent Development, University of Leuven, Belgium

Correspondence to: Suzanne Jak, Methods and Statistics Department of the Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, the Netherlands, e-mail: S.Jak@uva.nl

Meta-analytic structural equation modelling (MASEM) refers to the technique of fitting structural equation models to correlation or covariance matrices from several studies. A well-known approach to conduct MASEM is the two-stage approach of Cheung and Chan (2005). In the first stage, correlation matrices are tested for homogeneity across studies. If the matrices are not significantly different from each other, they are combined to form a pooled correlation matrix. In the second stage, the pooled correlation matrix is taken as the observed matrix in an SEM analysis. User-friendly software to apply the two-stage method is available in the R-Package metaSEM (Cheung, 2011), which utilises the OpenMx package (Boker et al., 2011). MetaSEM gives parameter estimates with standard errors, a chi-square measure of fit, and likelihood based confidence intervals (see Neale & Miller, 1997) for parameters at both stages of the analysis.

Ideally, researchers always report the correlations between *all* variables in their study. However, often not all correlations between the research variables are given in a paper. Sometimes, the missing correlations can be derived from other statistics that the authors do provide, such as regression coefficients. However, this is not always possible, for example when two variables are both outcome variables in regression analyses. The two-stage approach incorporates studies with missing variables, but a way to handle missing correlations has not yet been proposed. As a consequence, for each missing correlation, one of the two variables associated with the correlation has to be treated as missing. We will refer to this method as the *omitted variables approach* (OV approach).

In the present paper, we propose a method to deal with missing correlations in the two-stage approach. This method involves adding one parameter to the model for each missing correlation. We will refer to this method as the *omitted correlations approach* (OC approach). After outlining the method, we illustrate its use with a meta-analysis of teacher-child relationships variables.

MASEM with missing variables and missing correlations

Meta-analysis combines the results from several studies. For MASEM, correlation matrices of the variables of interest will be collected from several studies. The analysis has two stages.

Stage 1: Pooling correlation matrices from several studies
Let R_g be the p_g x p_g sample correlation matrix and p_g be the number of observed variables in the gth study. Some observed correlation matrices may have missing correlations. Moreover, not all studies may include all variables. The correlation matrices for the first three studies may be:

$$R_1 = \begin{bmatrix} 1 & & \\ r_{21_1} & 1 & \\ r_{31_1} & r_{32_1} & 1 \end{bmatrix} \quad R_2 = \begin{bmatrix} 1 & \\ r_{32_2} & 1 \end{bmatrix} \quad R_3 = \begin{bmatrix} 1 & & \\ r_{21_3} & 1 & \\ r_{31_3} & - & 1 \end{bmatrix}.$$

Here, Study 1 has all variables and correlations, Study 2 misses a variable, and Study 3 has all variables but misses a correlation. The OV approach accounts for missing variables, but not for missing correlations. In the OC approach, we account for missing correlations by adding a new matrix (matrix C_g in Equation 1) to the model. We substitute an arbitrary value (e.g. zero) in the observed matrix R_g for a missing correlation. We obtain an estimate of the population correlation matrix R_{pop} of all p variables by fitting a multigroup SEM model, in which the model for each group (study) is:

$$\Sigma_g = D_g (M_g R_{pop} M_g^t) D_g^t + C_g . \qquad (1)$$

In this model, R_{pop} is the p x p population correlation matrix with diag(R_{pop}) = I, matrix Mg is a p_g x p selection matrix that filters out the missing variables in study g. Matrix M_g is constructed by taking a p x p identity matrix and removing the rows corresponding to the missing variables in study g. D_g is a p_g x p_g diagonal matrix that accounts for differences in variances across the g studies. New in the OC approach is the addition of Matrix C_g, which is used to account for missing correlations. Matrix C_g is a symmetric p_g x p_g correction matrix, with fixed zeros for all present correlations and a free parameter for the missing correlations in study g.

With the identification constraint diag(R_{pop}) = I and a free D_g matrix, the hypothesis that is being tested is equality of covariances (not of correlations) as the variances in each study do not necessarily equal unity. The homogeneity of covariance matrices (covariances and variances) can be tested by constraining the elements of the diagonal matrix D_g to be equal across studies, so that D_g = D for all g (Cheung & Chan, 2005). The unity of variances can be tested by constraining the elements of the diagonal matrix D_g to unity, D_g = I for all g.

The model in Equation 1 is identical to the model in Stage 1 of Cheung and Chan's two-stage approach, except for the correction matrices C_g. In matrix C_g, the free parameter for each missing correlation will take on a value that minimises the difference between the arbitrary chosen value for the missing correlation in the observed matrix R_g (e.g., zero), and the estimate in R_{pop} for the corresponding correlation. A chi-square measure of fit for the model in Equation 1 is obtained by comparing its -2 log likelihood with the -2 log likelihood of the saturated model. The saturated model is given by:

$$\Sigma_g = D_g R_g D_g^t . \qquad (2)$$

The difference between the -2 log likelihoods follows a chi-square distribution with degrees of freedom equal to the difference in numbers of parameters between the two models. When the chi-square test turns out significant, then the hypothesis of homogeneity of covariances is rejected. The chi-square statistic can also be used to calculate approximate fit indices such as the Root Mean Square Error of Approximation (RMSEA, Steiger & Lind, 1980).

When the model fit is not acceptable, then the hypothesis of homogeneity of covariances is not tenable, so that the estimation of R_{pop} is not valid. Researchers may then create clusters of more similar studies, and construct separate pooled correlation matrices for all clusters of studies. Alternatively, a random effects model could be used, which estimates variances (and covariances) between the pooled correlation coefficients across studies. In this paper we do not consider random effects models.

The model from Equation 1 can be fitted using maximum likelihood estimation with any structural equation modelling program. However, writing the syntax can be very laborious in some programs. OpenMx (Boker et al., 2011) is a very flexible R-package, allowing the use of all R functions (R Development Core Team, 2011).

Stage 2: Fitting structural equation models
At Stage 2, the pooled correlation matrix from Stage 1 is used as the input matrix in an SEM analysis. Cheung and Chan (2005) propose using weighted least squares estimation at this stage. Weighted least squares estimation takes the asymptotic covariance matrix of the correlation coefficients from Stage 1 as the weight matrix in the fit function. Some correlation coefficients at Stage 1 are estimated using information from more studies than other correlation coefficients. As a result, coefficients that are based on more studies will have smaller variance in the asymptotic covariance matrix, and thus get more weight in the estimation process than coefficients that are based on less studies.

SEM generally requires the use of covariance matrices, however, the input matrix at Stage 2 is a correlation matrix. Treating the correlation matrix as a covariance matrix leads to incorrect results when estimating confidence intervals or when testing specific hypotheses (Cudeck, 1989). To obtain correct results at Stage 2, we add a so-called estimation constraint. This constraint enforces the diagonal of the model implied correlation matrix to identity.

Illustrative example

Data

Roorda, Koomen, Spilt and Oort (2011) collected 99 studies that reported correlations between positive teacher-student relations and negative teacher-student relations on the one hand and student engagement and student achievement on the other hand. Correlations between positive teacher-student relations and negative teacher-student relations were collected afterwards for the present paper. Of these studies, 63 were conducted at primary schools and 36 at secondary schools. In total, there where 129,184 respondents (sample sizes ranging from 42 to 39,553). Based on leading theories about teacher-student relations (Connell & Wellborn, 1991; Pianta, 1999), teacher-student relations were considered as exogenous variables and engagement and achievement as endogenous variables. Out of the 99 studies, 20 studies missed a correlation between two variables, and 90 studies did not include one or more of the four variables.

Results

Table 1a gives the fit results of the several models we fitted to the 99 correlation matrices, using OpenMx (Boker et al., 2011). Model 1 is a saturated model, meaning that a correlation matrix is estimated for each study, without equality restrictions across studies. This model is used as a baseline model, to obtain fit indices for Models 2 to 4. Model 2 is a model in which we restricted all covariances to be equal, without restrictions on the variances across studies (equal covariances). Model 3 is a model in which we restricted all covariances and variances to be equal across studies (equal variances and covariances). Model 4 is a model in which we additionally restricted all variances to be unity (equal variances across studies).

All three models at Stage 1 had significant chi-square values, indicating that the models do not fit

Table 1 Fit results of models at Stage 1 and Stage 2 using two approaches (N = 129,184)

a) Fit results using the OC approach

Model	-2 log likelihood	df	χ^2	RMSEA + 95% CI
STAGE 1				
1. Saturated	271981.9	0	0	-
2. Equal covariances (D_g = free)	277400.2	214	5418.34	.0137 [.0133 ; .0141]
3. Equal variances and covariances ($D_g = D$)	277957.2	474	5975.35	.0095 [.0092 ; .0097]
4. Variances equal to unity ($D_g = I$)	277957.4	478	5975.53	.0094 [.0092 ; .0097]
STAGE 2				
5. Mediation model (based on Model 2)	144.73	2	144.73	.0235 [.0197 ; .0274]

b) Fit results using the OV approach

Model	-2 log likelihood	df	χ^2	RMSEA + 95% CI
STAGE 1				
1. Saturated	249821.0	0	0	-
2. Equal covariances (D_g = free)	254872.7	193	5051.64	.0140 [.0136 ; .0144]
3. Equal variances and covariances ($D_g = D$)	255420.2	434	5599.22	.0096 [.0093 ; .0099]
4. Variances equal to unity ($D_g = I$)	255420.3	438	5599.28	.0096 [.0093 ; .0098]
STAGE 2				
5. Mediation model (based on Model 2)	178.48	2	78.48	.0261 [.0223 ; .0301]

Table 2 Pooled correlation matrix across 99 studies, N = 129,184

	1.	2.	3.	4.
1. Positive teacher-student relations	1			
2. Negative teacher-student relations	-.373	1		
3. Student engagement	.385	-.345	1	
4. Student achievement	.157	-.152	.284	1

the data exactly. The RMSEAs were all below .05, indicating close approximate fit (Browne & Cudeck, 1992). As we do not have any hypothesis on the variances being equal across studies, we take the result of Model 2 as the estimate of the population correlation matrix. This correlation matrix is given in Table 2, and is used as the input for the Stage 2 analysis.

The Stage 2 model is based on social-motivational theory (Connell & Wellborn, 1991), in which it has been hypothesised that student engagement acts as a mediator in the association between teacher-student relations and student achievement. Empirical studies have provided some support for the mediating role of engagement (e.g., Hughes, Luo, Kwok, & Loyd, 2008). Therefore, the path model we fitted was a mediation model, in which the influence of positive and negative teacher-child relationships on student achievement was mediated by student engagement. This model fitted the population correlation matrix from Stage 1 closely according to the RMSEA. Figure 1 provides a graphical representation of the model, with standardised parameter estimates and 95% confidence intervals. Positive teacher-student relations had a positive effect on Student engagement (β = .296, p < .05). Negative teacher-student relations had a negative influence of about the same size (β = -.255, p < .05). Student engagement had a medium sized positive effect on Student achievement (β = .322, p < .05). The indirect effect of Positive teacher-student relations on Student achievement via Student engagement was small and positive (β = .095, p < .05). Negative teacher-student relations had a similar small-sized negative indirect effect on Student achievement (β = -.082, p < .05). The model explained 20.8 % of the variance in Student engagement, and 10.0% of the variance in Student achievement.

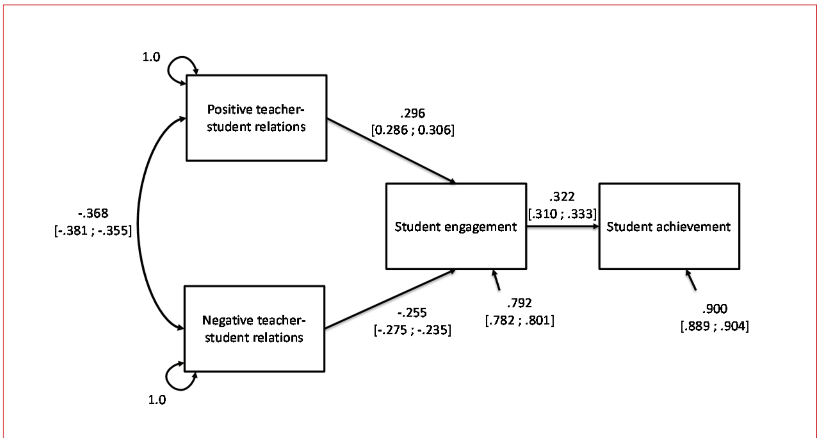

Figure 1 Path model with parameter estimates and their 95% confidence intervals

Results when deleting variables with missing correlations

We compared our results with the results obtained with the OV approach. This involved the deletion of a variable in 20 of the 99 studies. As can be seen in Table 1b, this approach leads to a loss of degrees of freedom and slightly worse model fit. However, the models still fit closely according to the RMSEA. Some parameter estimates are different from the previous analysis, and the likelihood based confidence intervals are somewhat wider. Figure 2 shows a graphical comparison of the parameter estimates and confidence intervals from the two analyses. Each graph pictures the parameter estimate (the dot) with its 95% confidence interval (the line) for the analysis with the OC approach (upper part) and for the analysis with the OV approach (lower part).

Simulation study

In order to investigate the effect of accounting for the missing correlations (OC approach), compared with deleting variables associated with missing correlations (OV approach), we performed an analysis of simulated data. We generated complete data for 100 studies. The pooled correlation matrix was estimated based on the full data, and based on data with missing correlations using the two approaches. The data were simulated under extreme conditions, so that differences between the two analysis methods became more apparent than in the illustration. Our expectation is that the OC approach leads to models with more power, better parameter estimates and smaller confidence intervals than the OV approach.

Data generation

We chose values of the population correlations based on the data from Roorda et al. (see Table 2). For each of the 500 replications, complete raw data were

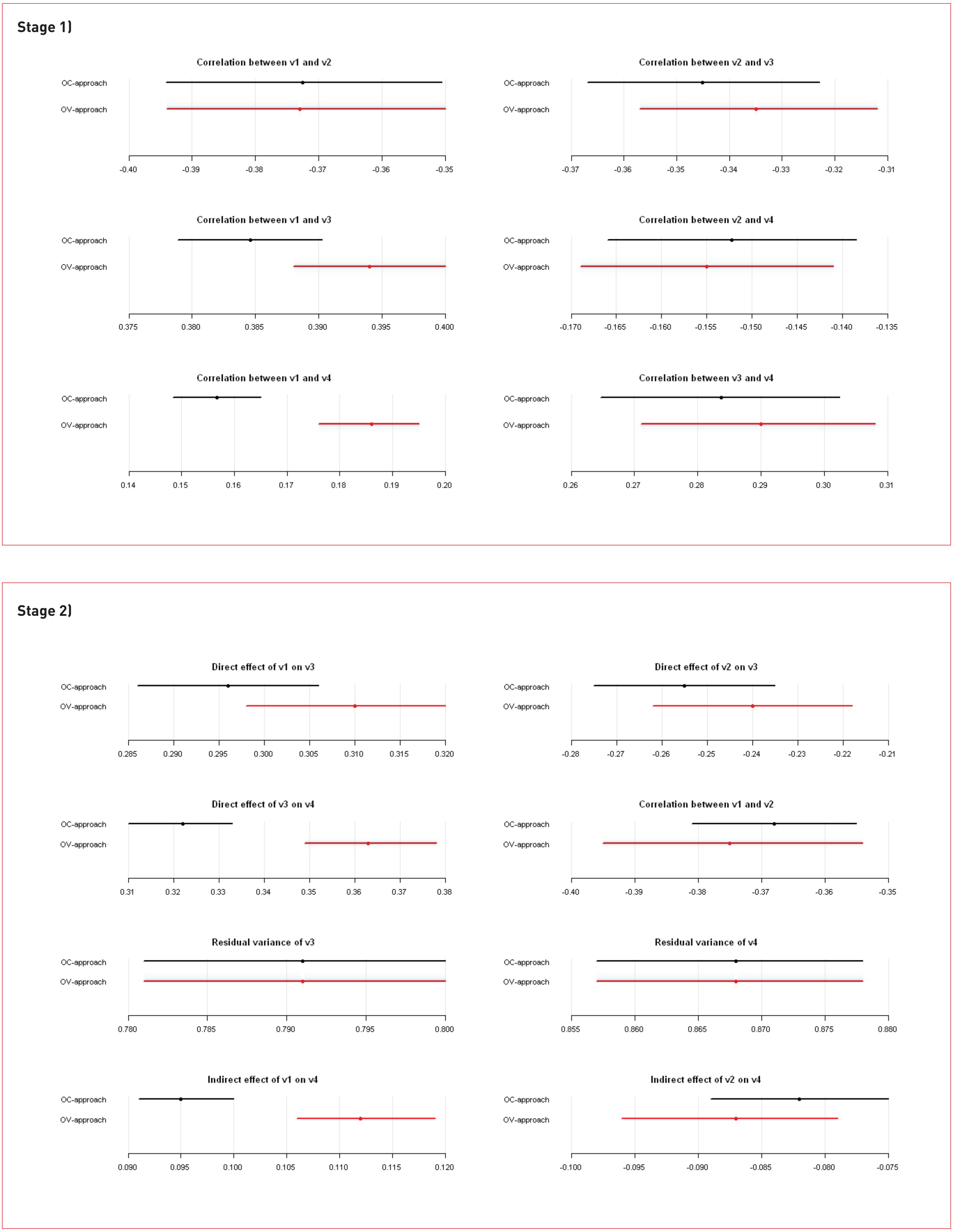

Figure 2 Parameter estimates (dots) with 95% confidence intervals (lines) of all parameters for the analysis with the OC approach (upper dot + line) and the OV approach (lower dot + line)

Note: v1 = Positive teacher-student relations, v2 = Negative teacher-student relations, v3 = Student engagement, v4 = Student achievement

Table 3 Estimated average pooled correlations and bias percentages with a) complete data, b) the OC approach and c) the OV approach

a) Complete data	Estimated correlations				Estimation bias (%)		
	Variable 1	Variable 2	Variable 3	Variable 4	Variable 1	Variable 2	Variable 3
Variable 1	1						
Variable 2	-.371	1			- 0.54%		
Variable 3	.356	-.343	1		- 0.56%	- 0.58%	
Variable 4	.156	-.151	.283	1	- 0.64%	- 0.66%	- 0.35%
b) OC approach	*Estimated correlations*				*Estimation bias (%)*		
	Variable 1	Variable 2	Variable 3	Variable 4	Variable 1	Variable 2	Variable 3
Variable 1	1						
Variable 2	-.373	1			0.00%		
Variable 3	.359	-.349	1		0.28%	1.16%	
Variable 4	.159	-.153	.287	1	1.27%	0.66%	1.06%
c) OV approach	*Estimated correlations*				*Estimation bias (%)*		
	Variable 1	Variable 2	Variable 3	Variable 4	Variable 1	Variable 2	Variable 3
Variable 1	1						
Variable 2	-.371	1			-0.54 %		
Variable 3	.355	-.345	1		-0.84 %	0.00 %	
Variable 4	.156	-.151	.283	1	-0.64 %	-0.66 %	-0.35 %

drawn from the multivariate normal distribution, with means equal to zero, and variance covariance matrix equal to the population correlations. We chose 100 as the number of studies, with each study having a sample size of 100. So, each dataset contained 100 x 100 = 10000 scores on 4 variables. For each study, the correlation matrix was calculated and included in the meta-analysis.

Results with complete data

Fitting the model from Equation 1 (with $D_g = I$) to the complete data in 500 samples led to the average correlations in Table 3a. Percentages of estimation bias in all parameters are calculated as 100 × (mean estimated value – population value) / population value. According to Muthén, Kaplan and Hollis (1987), estimation bias less than 10% can be considered negligible. With complete data, estimation bias was below 1% for all parameters. The model had 994 degrees of freedom, the average of all chi-square values was 294.48 (*SD* = 41.28).

Results with missing correlations

The correlation between variables 1 and 3 was deleted randomly for 80% of the studies. Also for 80% of the studies we randomly deleted the correlation between variables 2 and 3. In this way, about 64% of the studies missed both correlations, and about 32% studies missed one of them, while only about 4% of the studies had complete data.

Using the OC approach, we obtained the pooled correlation matrix shown in Table 3b. The model had 834 degrees of freedom, the average chi-square value was 119.29 (*SD* = 39.60). The estimated correlations based on the data with missing correlations were close to the population correlations. The largest difference was found for the correlation between variables 1 and 4, which deviated 1.27% from the population correlation. The results in Table 3c were obtained by removing one variable for each missing correlation (the OV approach). This model had 618 degrees of freedom, the average chi-square value was 62.85 (*SD* = 29.98). The results do not differ very much from the results in Table 3b. The largest difference was found for the correlation between variables 1 and 3, which deviated -0.84 % from the population value.

The bias in parameter estimates is not very different across the three models (complete data vs. OC approach vs. OV approach). However, a structural

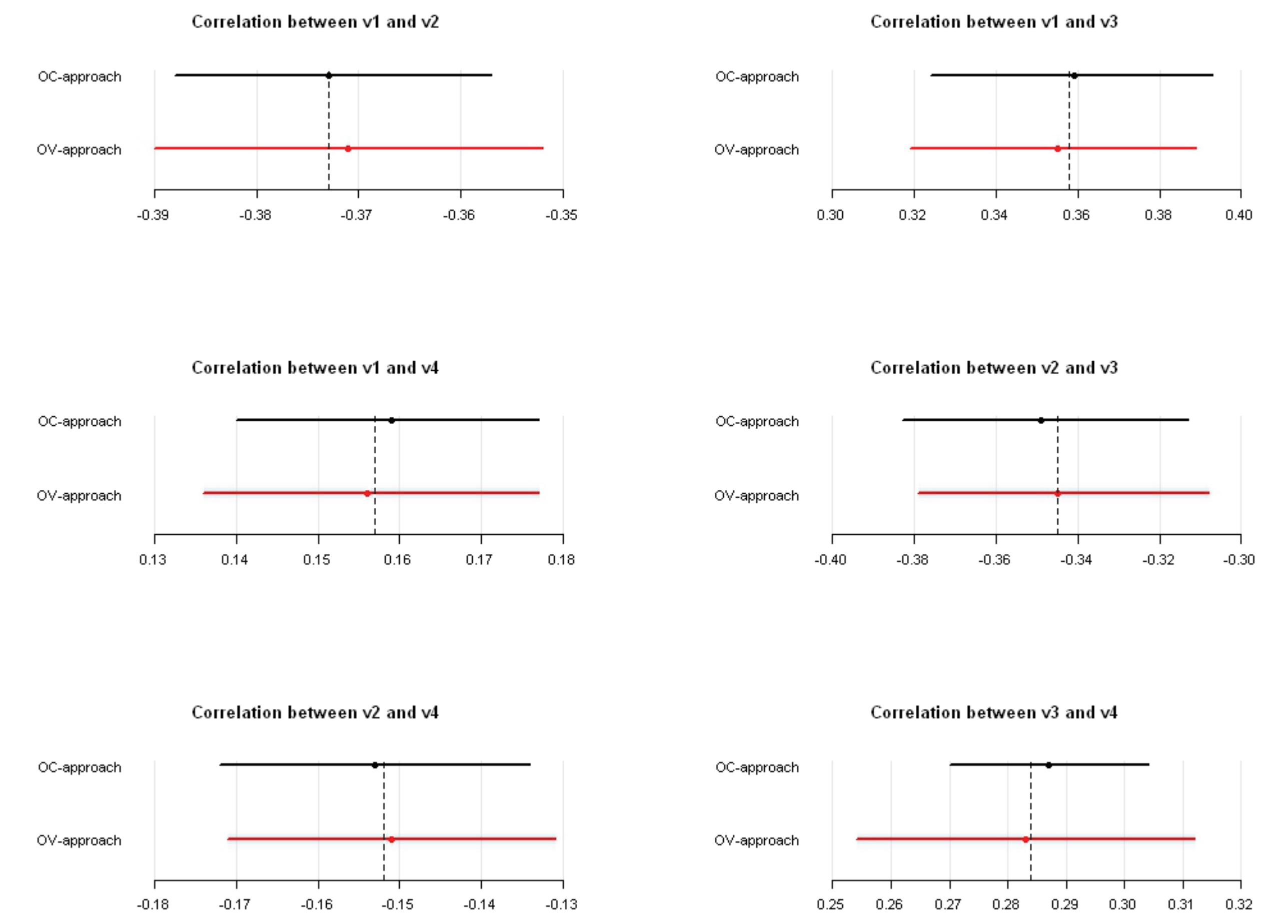

Figure 3. Average parameter estimates (dots) with 95% confidence intervals (lines) of all correlations with the OC approach (upper dot + line) and the OV approach (lower dot + line). The dotted vertical line denotes the population value of the parameter

difference can be seen in width of the likelihood based confidence intervals. These are smaller with the use of the OC approach, compared with the OV approach. Figure 3 gives a graphical representation of the parameter estimates and the associated 95% likelihood based confidence intervals. The upper dot and line denote the average parameter estimate and confidence interval of the analysis with the OC approach, while the lower dot and line denote the average parameter estimate and confidence interval of the analysis with the OV approach. The dotted vertical line shows the population value of the parameter. As expected, omitting information leads to larger confidence intervals. The difference is most clearly seen in the correlations between variables 3 and 4 in the lower right corner of the figure. It is not surprising that the difference is so apparent for the correlation between variables 3 and 4. As we deleted the correlation between variables 1 and 3 and between variables 2 and 3, both methods use equal amounts of information about these correlations. However, where our method still uses information about the correlation between variables 3 and 4, this information is often deleted in the other approach, leading to less precise parameter estimates.

Discussion

In this paper we have proposed a method to incorporate missing correlations in MASEM. The method was demonstrated with an example from teacher-child interactions. Using simulated data, the method was compared with the current practice, which is to delete one of the variables that is associated with the missing correlation. As the OC approach uses more information than the OV approach we expected that our method would lead to better parameter estimates and smaller confidence intervals. Results from the very small simulation study indicated that the OC approach leads to models with larger degrees of freedom and smaller confidence intervals. The parameter estimates were close to the population values for all methods, indicating that deleting one or two variables in 80% of the studies did not really influence parameter recovery.

A possible explanation of the similar results with respect to parameter bias between the methods is that in our study the missingness of the correlation coefficients were introduced randomly. Maximum likelihood estimation with data missing at random is known to lead to unbiased parameter estimates (e.g., Enders & Bandalos, 2001; Newman, 2003). In true meta-analysis, the missing correlations may not be missing at random, and it would be interesting to investigate the effect of not-random missingness. For example, if the correlations between variables 3 and 1 and variables 3 and 2 are mainly missing in studies where the correlation between variable 4 and 3 is high, then if we delete variable 3 in the OV approach, the information about the correlation between variable 4 and 3 is lost, and the parameter will be underestimated. Using the OC approach, all remaining information would be used and the parameter estimate is expected to be closer to the true value.

References

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., et al. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76,* 306-317.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230-258.

Cheung, M. W. L. (2011). 'metaSEM: Meta-analysis: A Structural Equation Modeling Approach', R package version 0.7-0, http://courses.nus.edu.sg/course/psycwlm/internet/metaSEM/ (accessed 12 March 2012).

Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10,* 40-64.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. Gunnar & A. Sroufe (Eds.), *The Minnesota symposium on child development: Self-processes and development* (Vol. 23, pp. 43-77). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*, 317-327.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430-457.

Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology, 100*, 1-14.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 52,* 431-462.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood based confidence intervals in genetic models. *Behavior Genetics, 27*, 113-120.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6,* 328-362.

Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. Washington, DC: American Psychological Association.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. [Computer software]. Vienna, Austria.

Roorda, L. R., Koomen, H. M. Y., Spilt, J. L. & Oort, F. J. (2011). The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research, 81*, 493-529.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. *Paper presented at the annual meeting of the Psychometric Society,* Iowa City, IA.

SUZANNE JAK
PhD candidate in Methods and Statistics, Research Institute of Child Development and Education, University of Amsterdam. Her research interests include structural equation modelling, measurement bias, multilevel structural equation modelling and meta-analytical structural equation modelling.

FRANS J. OORT
Full professor of Methods and Statistics and Director of the Research Institute of Child Development and Education, University of Amsterdam. His research interests include measurement and measurement bias, and statistical modelling, especially non-standard applications of structural equation modelling.

DEBORA L. ROORDA
Postdoc research fellow at the School Psychology and Child and Adolescent Development Department at the University of Leuven, Belgium. Her research interests focus on teacher-student interactions and their influence on children's emotional well-being, school engagement, and achievement.

HELMA M. Y. KOOMEN
An assistant professor in special education at the Research Institute of Child Development and Education at the University of Amsterdam. Her research interests address student-teacher interaction and teachers' pedagogical practices in relation to emotional well-being, social adjustment, task behaviours, and achievement of children with and without special need.