KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Centrum voor Onderwijseffectiviteit en –evaluatie
Centrum voor Methodologie van het Pedagogisch Onderzoek

Advanced techniques in multilevel growth curve modeling

Application to educational effectiveness research

Proefschrift aangeboden tot
het verkrijgen van de graad van
Doctor in de Pedagogische Wetenschappen
Door **Dickson Nkafu Anumendem**
o.l.v. Prof. Dr. Bieke De Fraine
Prof. Dr. Patrick Onghena
Prof. Dr. Jan Van Damme

2011

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Centrum voor Onderwijseffectiviteit en –evaluatie
Centrum voor Methodologie van het Pedagogisch Onderzoek

Advanced techniques in multilevel growth curve modelling

Application to educational effectiveness research

Proefschrift aangeboden tot
het verkrijgen van de graad van
Doctor in de Pedagogische Wetenschappen
Door **Dickson Nkafu Anumendem**
o.l.v. Prof. Dr. Bieke De Fraine
Prof. Dr. Patrick Onghena
Prof. Dr. Jan Van Damme

2011

Dickson Nkafu Anumendem, Advanced techniques in multilevel growth curve models. Application to educational effectiveness research

Doctoral dissertation submitted to obtain the degree of Doctor in Educational Sciences, 2011.

Supervisor: Prof. Dr. Bieke De Fraine

Co-supervisors: Prof. Dr. Patrick Onghena
Prof. Dr. Jan Van Damme

There has been a general belief in school effectiveness research that schools have a larger impact on their students' growth than on their students' outcomes at a certain point in time. This belief emanates mainly from the research results in which the school effect on student initial status for mathematics has been found to be about three times less than the school effect on learning rates or students' progress over time. Several studies have prompted growth in student outcomes over time to gain great acceptance among many educational effectiveness researchers as the most appropriate criterion for assessing school effectiveness. The investigation of such changes in students' outcomes has dramatically boosted the number of longitudinal studies in educational effectiveness research in the last two decades. In addition to this, researchers now understand that cross-sectional designs underestimate the impact of schools and that these designs do not provide the proper framework for studies on school effectiveness.

The use of repeated measures data make multilevel growth curve models an invaluable statistical tool in educational research. This is because a multilevel growth curve model estimates changes in student outcomes more accurately by taking into account the hierarchical nature of the data. Befitting results are not only appealing to researchers but also to policy makers and parents who both want a meticulous education for their citizens and children respectively.

The main aim of this dissertation is to improve the statistical methods applied by educational effectiveness researchers in order to have more credible results. In this context, school effect estimates from traditional methods and the proposed methods of this dissertation are compared to argue persuasively for the need for more advanced techniques when using growth curve models. Such techniques will not only be applicable to educational effectiveness research in but to educational research as a whole and all other research fields interested in growth curve modelling. The school effect estimates on student status and student growth are used for different types of student outcomes like well-being, mathematics, and language achievement.

Manuscript 1 defines clearly how the school effect on students' growth can be estimated using multilevel growth curve models with more than two levels. It also shows how the manner of coding time affects these estimates. Manuscript 2 introduces techniques to properly handle multilevel growth curve models with serial correlation at higher levels beyond level 1, while Manuscript 3 introduces a new multilevel growth curve model which can be used to model growth data with two or more levels of serial correlation simultaneously. Because most studies of school effects on students' growth have focused only on one effectiveness criterion, which is problematic given that school effects are only moderately consistent over different criteria. Moreover, the consistency issue has seldom been studied through multivariate growth curve models; Manuscript 4 introduces a model that can handle multivariate multilevel growth data with an unequal number of measurement occasions.

Data from the LOSO-project (the Dutch acronym for Longitudinal Research in Secondary Education) and the SiBO-project (the Dutch acronym for School Career in Primary School) are used to answer the research questions of this dissertation. The main software used is SAS 9.2, MLwiN 2.02 and Mplus 6.1.

This dissertation shows clearly how the choice of a time coding affects school effect estimates and their interpretation. It also recommends that the choice of a time coding should not only be based on the ease of interpretation and model convergence. The results show that school effects on students' well-being and language achievement in secondary school are greater for student growth than for student status. This work also indicates that the common assumption of serially uncorrelated level 1 residuals usually fails and therefore the need for appropriate modelling of this serial correlation is invaluable. These results demonstrate how modelling of serially correlated residuals at level 1 or level 2 has a huge payoff on school effects estimates. Because of the increasing popularity of multilevel growth curve models as a flexible tool for investigating longitudinal change in students' outcomes, this study investigates some covert issues in methodology resulting from repeated measures data structure. A complex double serial correlation multilevel growth curve model is developed and the results of this model show great improvement in school effects estimates compared to those of models without double serial correlation correction. This dissertation also investigates the school effects on pupils' growth in both mathematics and reading comprehension (and their relation) in primary schools taking previous changes in mathematics into account through a bivariate transition multilevel growth curve model. The results show that stronger growth in mathematics tends to associate with stronger growth in reading comprehension. Earlier growth in mathematics is also found to predict subsequent growth in reading comprehension.

Dickson Nkafu Anumendem, Gevorderde technieken in multiniveau groeicurvemodellen. Toepassing op onderwijseffectiviteitsonderzoek.

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Pedagogische Wetenschappen, 2011.

Promotor: Prof. Dr. Bieke De Fraine

Co-promotoren : Prof. Dr. Patrick Onghena
Prof. Dr. Jan Van Damme

In onderwijseffectiviteitsonderzoek wordt algemeen aangenomen dat scholen een grotere impact hebben op de leerwinst dan op de resultaten van de studenten op één bepaald tijdstip. Deze overtuiging is voornamelijk gebaseerd op onderzoeksresultaten waarin het aandeel van het schooleffect op de aanvangsscore wiskunde ongeveer drie keer minder groot bleek te zijn dan het aandeel op de leerwinst of de vorderingen die de studenten boeken gedurende een bepaalde tijd. Deze en een aantal andere studies hebben ertoe geleid dat de leerwinst van leerlingen binnen het onderwijseffectiviteitsonderzoek beschouwd wordt als het beste criterium om schooleffectiviteit te meten. Het onderzoek van dergelijke veranderingen in de leerlingresultaten heeft het aantal longitudinale studies in het onderwijseffectiviteitsonderzoek drastisch doen stijgen in de laatste twee decennia. Bovendien zien onderzoekers nu in dat cross-sectionele designs de impact van scholen onderschatten en dus geen degelijk kader bieden voor schooleffectiviteitsonderzoek. Het gebruik van gegevens uit herhaalde metingen maakt multiniveau groeicurvemodellen een waardevolle analysetechniek in het onderwijsonderzoek. Dit komt omdat een multiniveau groeicurvemodel veranderingen in leerlingresultaten nauwkeuriger schat door rekening te houden met de hiërarchische aard van de gegevens. Correcte resultaten zijn niet alleen belangrijk voor onderzoekers, maar ook voor beleidsmakers en ouders, die beiden een degelijk onderwijs voor hun burgers, respectievelijk kinderen, wensen.

Het belangrijkste doel van dit proefschrift is het verbeteren van de statistische methoden die onderwijseffectiviteitsonderzoekers gebruiken om tot zo meer correcte resultaten te komen. In dit opzicht zijn de traditionele schattingen van schooleffecten en de voorgestelde methoden van dit proefschrift met elkaar vergeleken om zo de noodzaak voor meer geavanceerde technieken bij het gebruik van groeicurve-modellen in de verf te zetten. Dergelijke technieken zullen niet alleen van toepassing zijn op het onderwijseffectiviteitsonderzoek, maar ook op het onderwijsonderzoek in het algemeen, en op alle andere onderzoeksgebieden die geïnteresseerd zijn in groeicurvenmodellering. De schattingen van het effect van de school op de prestaties en groei van de leerlingen worden gebruikt voor verschillende types van leerlingresultaten, zoals welbevinden, prestaties voor wiskunde en voor Nederlands.

Manuscript 1 definieert duidelijk hoe het effect van de school op de groei van de leerlingen geschat kan worden met behulp van multiniveau groeicurvemodellen met meer dan twee niveaus. Het toont ook aan hoe de tijdscodering van invloed is op deze schattingen. Manuscript 2 introduceert technieken om multiniveau groeicurvemodellen met seriële correlatie op een hoger niveau dan niveau 1 goed te verwerken. Manuscript 3 introduceert een nieuw multiniveau groeicurvemodel dat kan worden gebruikt om longitudinale data met twee of meer niveaus van seriële correlatie tegelijk te modelleren. De meeste studies naar schooleffecten op de groei van studenten zijn slechts gericht op één effectiviteitscriterium, wat problematisch is omdat de effecten van de school slechts matig consistent zijn over verschillende criteria. Bovendien werd deze consistentiekwestie zelden onderzocht door middel van multivariate groeicurvemodellen. Vandaar dat Manuscript 4 een model introduceert dat kan omgaan met longitudinale, multivariate en multiniveau gegevens met een ongelijk aantal meetmomenten.

Gegevens van het LOSO-project (Longitudinaal Onderzoek in het Secundair Onderwijs) en het SiBO-project (Schoolloopbanen in het Basisonderwijs) worden gebruikt om de onderzoeksvragen van dit proefschrift te beantwoorden. De belangrijkste software die werd gebruikt zijn SAS 9.2, MLwiN 2.02 en Mplus 6.1.

Dit proefschrift laat duidelijk zien hoe de keuze van tijdscodering invloed heeft op de schattingen van schooleffecten en hun interpretatie. Er wordt dan ook aanbevolen de keuze van een tijdscodering niet alleen te baseren op het gemak van de interpretatie of de convergentie van het model. De resultaten laten zien dat de schooleffecten op het welbevinden en de taalprestaties van de leerlingen in het secundair onderwijs, groter zijn voor de groei van de studenten dan voor de behaalde scores op een bepaald moment. Dit werk illustreert ook het belang van het modelleren van serieel gecorreleerde residuen op niveau 1 of niveau 2. Omwille van de toenemende populariteit van multilevel groeicurvemodellen als een flexibel instrument voor het onderzoeken van longitudinale verandering in leerlingenresultaten, onderzoekt deze studie een aantal verborgen problemen in de methodologie die voortvloeien uit de datastructuur van herhaalde metingen. We ontwikkelden een complex multilevel groeicurvemodel met dubbele correctie voor seriële correlatie, en de resultaten van dit model laten een grote verbetering in de schattingen van schooleffecten zien vergeleken met die van modellen zonder dubbele seriële correlatie correctie. Dit proefschrift onderzoekt ook de schooleffecten op de groei van leerlingen, zowel voor wiskunde als voor begrijpend lezen (en hun relatie) in het basisonderwijs, rekening houdend met eerdere groei in wiskundeprestaties via een bivariaat transitie multilevel groeicurvemodel. Met behulp van een voorgesteld tweefasig effectiviteitscriterium en het bivariaat transitie multilevel groeicurvemodel, tonen de resultaten aan dat een sterkere groei in de wiskunde samenhang vertoont met een sterkere groei in begrijpend lezen. Eerdere groei in wiskunde blijkt de latere groei in begrijpend lezen te kunnen voorspellen.

ACKNOWLEDGMENT / DANKWOORD

"As long as three different states of the same **H_2O** continue to exist with a triple point, three different persons of the same **God** will continue to exist with a trinity point". Anumendem Dickson

The writing of this dissertation would obviously not have been possible without the invaluable moral, spiritual, physical and financial support of a number of special people. I think of all those people who stopped by to lend a helping hand during those rough, tough and difficult moments that lead to this dissertation.

I wish to thank first and foremost my supervisor Prof. Dr. Bieke De Fraine, co-supervisors Prof. Dr. Patrick Onghena and Prof. Dr. Em. Jan Van Damme for their very inspiring and encouraging support. The wonderful sacrifices you all made went a long way to boost and stimulate my doctoral training, academic development and research experience as a whole. All your friendship, efforts and sacrifice have enable me to complete this doctoral dissertation much earlier than its due date. Prof. Dr. Bieke, you were more than just a supervisor to me. Indeed I say with pride that you are a true friend and my sister too. I am equally very grateful to Prof. Dr. Geert Verbeke for his methodological contributions.

Special thanks also go to all my (ex) colleagues of the Centre for Educational Effectiveness and Evaluation for their friendship, collaboration and hard work in the LOSO- and SiBO-projects which resulted in the creation of the two huge data bases used in this dissertation.

Outside the university environment, I extend my sincere gratitude to my father Mr. Nkafu Mathias (late) and mother Mrs. Nkafu Cecilia for their unceasing love and care to enable me attend this level of education. I am equally very grateful to my uncle Mr. Asaba George and his wife Mrs. Asaba Mary for their moral and financial support all these years. To my brothers Fr. Julius-Simomia and Peter Fornji, I say thank you for always standing by me during all important moments of my life. I also thank all my relatives and friends who in one way or the other have added a stone to increase my global development.

It would be the biggest mistake I ever made if I forgot to thank in a very special way my dear wife Claudine and daughter Marieliz for all their love, care, patience, sacrifice, encouragement and support whenever it was needed. Darling you are wonderful and remain at the centre of this successful dissertation, thank you once again for always being there for me.

Of course this doctoral dissertation would not have been realized without the necessary financial assistance. For this, I thank the FWO (Fonds voor Wetenschappelijk Onderzoek-Vlaaderen) very immensely for their support.

Anumendem Dickson Nkafu
Leuven, May 2011

TABLE OF CONTENTS

INTRODUCTION

The field of educational research has seen in the past three decades a great change in the use and development of sophisticated statistical methodology and techniques for analysing longitudinal data. The very early results of Raudenbush's (1989, 1995) more advanced growth models have challenged researchers over the use of archaic or outdated statistical methodology. For example, cross-sectional analysis, which do not provide the proper framework for studies in school effectiveness processes because many of these processes change with the passing of time. Despite the wide range of statistical methods available for the analysis of multiple wave data, growth curve models seem to have gained great acceptance among many educational effectiveness researchers as an invaluable tool for assessing school effectiveness. The investigation of changes in students' outcomes has necessitated the use of repeated measurements and resulted in the general call for longitudinal studies of school effects (Teddlie & Reynolds, 2000).

One of the most appropriate ways of modelling longitudinal data is by the estimation of individual growth trajectories as advocated by several authors (Rogosa, 1995; Singer & Willett, 2003; Willett, 1997). Growth curve models (GCM) assume that there is a growth curve that has given rise to the scores on the measurement occasions. The model puts smooth curves over the observed measures to estimate continuous trajectories that are believed to underlie the observed outcomes (De

Fraine et al., 2005). These models are also multilevel models, because measurement occasions are nested within students (who themselves are nested within schools). In this dissertation, "student growth" refers to the growth parameters or slope and "student status" refers to the intercept of the growth curve model. The general measure of school effect is the intraclass correlation for random intercept GCM models. In addition, this dissertation introduces the percentage of variance between schools for both criteria (status or growth) as a measure for the school effect in random intercepts and random slopes GCM.

There are a number of studies that have been carried out in this domain but with diverse results. Some authors confirm that the school effect is larger when student growth is used as the effectiveness criterion (May, Supovitz, & Perda, 2004; Raudenbush & Bryk, 2002; Van Damme & De Fraine, 2005); others found exactly the opposite (Reardon, 2003). Some others have even found mixed results when more than one outcome is used (De Fraine et al., 2006; Wilkins & Ma, 2002; 2003).

This dissertation investigates how the choice of a time coding affects school effects estimates and their interpretation. It also gives the circumstances under which time coding would be appropriate while illustrating on the pitfalls of such choices when based only on the ease of model interpretation and convergence. The hypothesis here is that school effects in student's well-being and language achievement in secondary school, are greater for student growth than for student status. This work also looks deeper into the common assumption of serially uncorrelated level 1 residuals and hypothesizes that this assumption usually fails and therefore the need for appropriate modelling of this serial correlation is invaluable. Because of the increasing popularity of multilevel growth curve models as a flexible tool for investigating longitudinal change in students' outcomes, this dissertation makes a critical look into some of the coverts in methodology resulting from repeated multilevel data structure. Multilevel growth curve models are also frequently applied to data with more than two levels of hierarchy as a result necessitating a closer look at more advanced forms of serial correlation. A complex double serial correlation is developed and empirical tested and the results compared with other ordinary growth curve models. This comparison is done to identify how much improvement this new approach with double serial correlation correction can bring on the estimation of school effects for the case of three level data.

Most studies of school effects on student growth have focused only on one effectiveness criterion, which is problematic given that school effects are only moderately consistent over different criteria. Moreover, the consistency issue has seldom been studied through multivariate growth curve models. This study investigated the school effects on pupils' growth in both mathematics and reading comprehension (and their relation) in primary schools taking previous changes in mathematics into account through a bivariate transition multilevel growth curve model. Using the proposed two-stage effectiveness criteria and the bivariate transition multilevel growth curve, it is hoped that the results showed that stronger growth in mathematics tends to associate with stronger growth in reading comprehension.

The cardinal aim of this dissertation is therefore to ameliorate the statistical methods applied by researchers in the field of educational effectiveness in order to have more valid results. To do this, school effect estimates from the traditional methods and the new methods are adequately examined to show the need for more advanced techniques when using growth curve models. Such techniques will not only be applicable to educational research but all research fields interested in modelling growth. The school effect estimates on student status and student growth are used for different types of student outcomes like well-being, mathematics, Dutch language. Manuscript 1 defines clearly how the school effect on student's growth can be estimated using multilevel growth curve models with more than two levels. It also shows how the manner of coding time affects these estimates. Manuscript 2 introduces techniques to properly handle multilevel growth curve models with serial correlation at higher levels beyond level 1 while Manuscript 3 establishes a new multilevel growth curve model which can be used to model growth data with two or more levels of serial correlation simultaneously. Finally Manuscript 4 introduces a model that can handle multivariate multilevel growth data with an unequal number of measurement occasions.

1. Problem statement

There has been a general belief in school effectiveness research that schools have a larger impact on their students' growth than on their students' outcomes at a certain point in time. This belief emanates mainly from the research results of Raudenbush in which the proportion of the school effect on student initial status for mathematics was about 6 times less than that for learning rates or students' progress over time (Raudenbush, 1989, 1995). He found that the school accounts for more than 80% of the variance in mathematics growth over time. The above research and many others have evoked the use of growth in student outcomes over time and this growth gained great acceptance among many researchers as the most appropriate criterion for assessing school effectiveness. The investigation of such changes in students' outcomes has boosted the number of longitudinal studies in educational research. In addition to this, researchers now understand that cross-sectional designs underestimate the impact of schools and that these designs do not provide the proper framework for studies on school effectiveness. The use of repeated measure data makes multilevel growth curve models an invaluable statistical tool in educational research. This is because this statistical method models changes in student outcomes more efficiently by taking into account the hierarchical nature of the data. More accurate results are not only appealing to researchers but also to policy makers and parents.

1.1. Educational effectiveness research

Educational effectiveness research is a domain that studies questions like: Do schools really differ with respect to their impact on pupils? How big is this effect? How can it be assessed? What are the factors that cause the effect? Can this effect be found in all types of educational contexts, such as in primary and secondary schools, in the various countries around the world? What are the theoretical explanations behind research findings in this area? It is of course capital that a criterion (or several criteria) be chosen to connote that one school is more effective than another. The sphere of influence of educational effectiveness has proportionately stretched to assessing the magnitude and stability of these school effects. This dissertation will look into two criteria for educational effectiveness: student status and student growth (in cognitive and non-cognitive outcomes) with great focus on the magnitude and accuracy of school effects estimates for these two criteria.

1.2. Modelling educational effectiveness

Reading through the literature of educational effectiveness research, one finds several definitions and theoretical models proposed by different authors. There are many definitions of the concept "effectiveness", especially because every discipline has its own viewpoint. Scheerens and Bosker (1997) provide a description of three possibilities: the economic, the organization-theoretical and the pedagogical definitions. Most definitions indicate that a school or a class is more effective when it realizes the desired goals to a higher extent. But what is meant by "desired goals" also differs immensely across studies and fields. This makes the choice of the effectiveness criterion a very crucial issue. The current study will address two main criteria: student status and student growth, for both students' cognitive and non-cognitive outcomes.

However, educational effectiveness can be broken down into at least two major parts: instructional effectiveness and school effectiveness. *Instructional effectiveness* is used to describe educational effectiveness at the classroom level while *school effectiveness* is used for the effectiveness of the school as an organization. Scheerens and Creemers (1989) define *educational effectiveness* as "the effectiveness of the educational system in general comprising all modes of schooling". With the passing of time, there has also been a transition of school effectiveness to educational effectiveness. The main reason for this being the fact that school effectiveness has been broadened to include different modes and different organizational levels of schooling.

A model is used to specify or visualize complex phenomena in a simplified or reduced manner. It is generally made of two parts: the conceptual and the formal one.

The conceptual part usually involves abstract terms. These terms are usually units like facts, concepts, and variables. The formal part is a system of relationships among these units. Looking at the background of educational effectiveness, it is possible to distinguish three disciplines and thus the models required for each (Scheerens & Bosker, 1997). The economic approach, which focuses on the “Education Production function” (Monk, 1992), the educational-psychological approach with main focus on “Effective Instruction and Learning Conditions” (Creemers, 1994) and the generalist-educationalist approach, that focuses on integrated, multilevel school effectiveness modelling (Bosker & Scheerens, 1994). The third category will be our main concern in this dissertation because it is in a sense a combination of the first two approaches. Details of these approaches can be read in the research contributions of Creemers (1994) and Scheerens (1990).

2. Statistical methodology

The most popular statistical analysis methods in educational effectiveness research are concerned with modelling a single response variable, measured for a sample of subjects, coming from a certain population. This concept can also be extended to measuring several aspects, several response variables, of a sample of subjects from a certain population and the data analysed as multivariate data. However, a design where, for each subject, several measurements of the same variable are recorded under different experimental conditions is often termed a repeated measures design. If the variable is measured repeatedly over time for each subject, we describe this as longitudinal data. Longitudinal techniques are therefore those applied to studies in which subjects are measured repeatedly through time. This is in contrast to cross-sectional studies in which a single outcome is measured for each subject. Let’s consider an example presented in the book of Verbeke and Molenberghs (2000) to make this clearer. They describe a repeated measurement study design where for each subject, diastolic blood pressure is recorded under several experimental conditions (sitting, standing, lying etc.). If the time at which each measurement is taken, is also recorded, then the diastolic blood pressure is measured repeatedly over time for each subject, representing a longitudinal study.

However, it will not be longitudinal if emphasis is laid only on the position of the patient and not on the time the measurement is made. In this case the order of taking the repeated measurements can be different for the different patients. Some could start with lying while others start with standing. It is worth noting that repeated measures can also be done over spatial dimensions resulting in a spatial data. The latter of the design above can be seen as spatio-temporal data if the distance of the centre of gravity (standing > sitting > lying) of the patient is considered from the floor together with the time of the repeated measurements. Though a medical field example, similar settings can be obtained in educational effectiveness studies.

Although longitudinal data can also be viewed as a multivariate design in terms of its data structure, there are many fundamental differences, which affect the mode of analysis of such data. The analyses of such complex data structure therefore require special statistical techniques to properly model, carefully program and judiciously run the relevant software.

During the modelling of repeated measurements, it happens that correlation comes in to play a very pertinent role. This type of correlation called a serial correlation which is almost always present in longitudinal data only and not in other forms of clustered data needs special attention. It is therefore worthwhile that this serial correlation is taken into account when considering growth modelling as the statistical model choice. The importance of accounting for serial correlation cannot be undermined if school effects are to be properly estimated. Correcting for serial correlation can greatly affect the estimation of model variance and covariance parameters which are used to estimate the school effect. To know how to deal with this correlation, it is important to distinguish between several kinds of correlation (Fahrmeir & Tutz, 2001). In general, there are four major types of correlations that can be distinguished. These are clustered data (baby rats from the same mother), longitudinal data (repeated measurements over time), spatial data (repeated measurement of soil fertility from the centre of an experimental farm), and data of measurements of different variables, taken on the same subject. Analogous studies can be designed in educational research, think for example of correlation in clustered data as children from the same family attending the same school. Or correlations in longitudinal data as students' repeated measurements through their entire primary or secondary school. Collateral for spatial data can be looking at children's repeated measurements with respect to their home distance from school. Finally, several student outcomes can be measured on the same student. This dissertation will focus only on two of these correlations, the second category (clustered data) and a combination of the second and last categories (clustered data and repeated measurements).

There are several methods proposed in the literature (Fahrmeir & Tutz, 2001) on the types of statistical techniques that can be applied to handle longitudinal data depending on the type of response. Amongst which is the general linear model for continuous and normally distributed responses. These models can be extended to linear mixed models, by including random effects in the model, which allow for the modelling of subject-specific characteristics of individuals. It will be possible to cover most of the heterogeneity of the data by taking subject-specific differences into account, which is not possible by allowing only fixed effects that only yield population-averaged models.

3. Data description

3.1. Longitudinal research in secondary education (LOSO)

The first data that is considered in the research project leading to this dissertation is the LOSO which is the acronym for the Dutch form "Longitudinaal Onderzoek in het Secundair Onderwijs" (Anumendem et al., in press; De Fraine et al., 2007; Van de gaer et al., 2009). This is a large scale research project in which 6411 students from 90 secondary schools in Flanders are followed from September 1990 till date even after the completion of secondary school studies. In order to follow up students changing schools, the research was limited to two regions in

Flanders taking into account their representativeness in the educational setting in Flanders.

Data collection on the student characteristics was made possible via a number of questionnaires while the student outcomes (cognitive or non-cognitive), Dutch and mathematics achievement were assessed a number of times. These assessments were done via curriculum based tests administered at five different measurement occasions: at the beginning and end of first grade, and at the end of the second, fourth and sixth grades. A common scale for the Dutch language was obtained for the five measurement occasions using Item Response Theory. For the non-cognitive student outcomes, the same questionnaire was administered four times during secondary school. This questionnaire includes items on well-being, academic self-concept, integration in the class, degree of effort for studying and teacher-student relationship. There are also background variables which include gender, age, socioeconomic status of family, language spoken at home and initial cognitive abilities.

3.2. Longitudinal research in primary education (SiBO)

The second data set considered for this project is the SiBO-data (Schoolloopbanen in het BasisOnderwijs). The SiBO data are from a large scale project on school careers in primary education which started in September 2002 in the last class of the kindergarten (Maes, Ghesquière, Onghena, & Van Damme, 2002). This project involves 210 schools with over 8500 pupils, who were followed through their primary school career. The data obtained come from a reference sample which is a representation of the Flemish primary school population. After getting this sample, three additional samples were taken: that is the GOK-sample (Gelijke-OnderwijsKansenbeleid [Equal Educational Opportunities Policy]), a method-school-sample and a Ghent-sample. Data collection at the school, teacher and pupil level was by means of questionnaires and tests. There are also cognitive and non-cognitive effectiveness criteria similar to that discussed for the LOSO-project. Similar background variables as for the data from the LOSO-project are also available in this SiBO-project data.

4. Structure of the dissertation

This dissertation consists of four manuscripts that have been submitted to peer-reviewed journals in the domains of educational effectiveness research and statistical methods.

Manuscript I deals with the hypothesis that school effects will be larger when the intercept refers to a later point in time (May et al., 2004). Though there are a number of publications on growth curve models today, the relationship between the initial status and the time at which it is defined was however first indicated by Rogosa, Brandt, and Zimowski (1982) and Rogosa and Willett (1985). The initial status is not the only concern in growth curve models but also its relationship with the growth rate and covariates. The first manuscript intends to add more flesh to the vivid discussions already found in the literature on the implications of poorly defining the time variable (Duncan et al.,1999; McArdle, 1988; Willett & Sayer, 1994). Many others have build on these ideas and thus brought about research in the invariance of growth curve parameters due to changes in the definition of the initial status. In educational effectiveness research as well as other research fields, the initial status is most often not the natural origin, but another one defined by factors which are most of the time different from the origin of the process being investigated. The first manuscript gives a simple, clear and straightforward illustration of the statistical methodology of a three level quadratic growth curve model in a compact form. It goes further to apply this on the LOSO data to analyse two different students' outcomes: Dutch language achievement and student well-being.

Manuscript II presents a growth curve model with serial correlation at one level. In any educational process where time is required to acquire knowledge, current knowledge is built on the previous. Such an influence of previous knowledge on the current, will certainly present itself in the structure of the repeated measurement data and consequently into the growth model errors. The second manuscript intends to expand the knowledge on how to conscientiously construct a statistical model that can handle such data. Goldstein, Healy, and Rasbash (1994) have indicated the possibility of correlated level 1 residuals for multilevel models by considering autocorrelated models for both discrete and continuous time. In their work, they illustrate this fact only for level 1 residuals though in their discussion they indicate the possibility of further research into models with higher level serial correlation using the example of repeated measurements nested within students who are in turn nested within schools. Maas and Snijders (2003) have indicated the need for other families of covariance matrices with structures more complex than the common compound symmetry model but less than the complete unstructured model. The use of autoregressive models (order 1) for equally time spaced data has also been mentioned by a number of these researchers (Goldstein et al., 1994; Mass & Snijders, 2003; Snijders & Bosker, 1999). The main purpose of Manuscript II is firstly to introduce a serial correlation method to multilevel (three and more levels) growth curve models for school effectiveness research and to apply it to real data to show how it impacts the estimates of school effects. Secondly, this method will be extended to handle level 2 serial correlation functions in the special case of three-level data to show the implications of extending from the common two level models discussed in the literature (Little et al., 2006; Mass & Snijders, 2003; Snijders & Bosker, 199). This method will then be used to demonstrate for the first time the complex structure of serial correlation at student and school levels and how this affects the estimates of the school effects.

While longitudinal studies are universally accepted by researchers as a key design to study the changes in a student performance over time, it comes with a number of challenges in terms of data structure and statistical analysis. Children, for example, change with time by virtue of their experience in school, and also because the structures, functions, and compositions of the schools they attend also change with time. There are possibly other sources of this change that are neither of the children's nor the schools' making (purely stochastic) which cannot be captured by researchers but are in any case present in a way as to influence the outcome of interest. One can say that the sources and consequences of changes in student outcomes often constitute the object of study in school effectiveness research (Rowan & Denk, 1982). Scheerens and Bosker (1997) have shown how school effects are underestimated when the school variance structure is misspecified or not taken into account.

Manuscript III extends further the methodology introduced in the second manuscript to look at two levels of change simultaneously in a three level model. The method of the third manuscript in addition to breaking the level-1 error as in manuscript II, also includes a complex serial correlation at the second level, resulting in what is called in this dissertation "double serial correlation" (DSC). The main interest of this manuscript is then to compare school effects estimates from multilevel nonlinear growth curve models with and without such a complex serial correlation correction.

Manuscript IV introduces a bivariate transition multilevel growth curve model for longitudinal data with more than one student outcome (reading comprehension and mathematics), with unequal number of measurement occasions. The measurement and explanation of the school effects on students' reading comprehension and mathematics achievement in primary school children has so far not been studied in depth. Generally, in school effectiveness studies that have addressed the relationship between reading and mathematics achievement, the focus has been more on the effects of schools at some point in time. Other studies have focused on mathematics achievement as a predictor of reading comprehension (Lerkkanen et al., 2005) or reading as a predictor of mathematics performance (Kirsch et al., 2002). These studies make the key assumption of deciding in advance which of the outcomes is dependent on the other. A few multivariate studies have nonetheless been done to investigate the possibility of a relationship between overall reading ability and mathematics, by looking at the common characteristics responsible for high performance in mathematics and reading (De Maeyer et al., 2010). However, research investigating the relationship between parameters of pupils' growth in reading comprehension and mathematics without any assumption on their functional dependence is still wanting. Manuscript IV focuses on reading comprehension which requires more advanced cognitive and linguistic skills. While studies have shown that individual differences in reading comprehension in particular during primary school are stable (de Jong & van der Leij, 2002), differences between schools however have not been addressed in depth.

MANUSCRIPT 1: The impact of coding time on the estimation of school effects[1]

Anumendem, D. N., De Fraine, B., Onghena, P., & Van Damme, J.

Abstract

Multilevel growth curve models are becoming invaluable in educational research because they model changes in student outcomes efficiently. The coding of the time variable in these models plays a crucial role as illustrated in this study for the case of a three-level quadratic growth curve model. This paper shows clearly how the choice of a time coding affects school effects estimates and their interpretation. A new definition for school effects for growth curve models with random intercepts and slopes is proposed. This study recommends that the choice of a time coding should not only be based on the ease of interpretation and model convergence but also on its consequences on the student status and growth parameter estimates. The current application illustrates that in general the school effects for student growth in well-

[1] Manuscript has been accepted for publication in Quality and Quantity International Journal of Methodology.

being and language achievement in secondary school, are greater for student growth than for student status.

1. Introduction

Growth curve models have become very popular in educational research and other fields of research like psychology and medicine. These models are a flexible method used to model change over time (Singer & Willett, 2003; Willett, 1997). In educational research for example, the importance of growth curve models is stressed by many authors (May, Supovitz, & Perda, 2004; Raudenbush, 1989, 1995; Teddlie & Reynolds, 2000). The theoretical analysis of change over time as in growth curves can even be traced as far back as Wishart (1938) and Tucker (1958).

Growth curve models are related to multilevel models where measurement occasions are nested within persons. They are also related to the structural equation framework because latent growth trajectories are assumed (Curran & Hussong, 2002). The shape of these underlying trajectories varies across studies. In the social science field and especially educational effectiveness research, linear growth curves are applied regularly, which is not a problem when the study is limited to data with two or three time points. In a linear growth curve model, the change over time is modelled by two growth parameters: the intercept (status) and the slope (growth). The (rather common) limitation of growth curve models to linear growth curves when more complex models would fit better, might result in unreliable conclusions. This paper shows that more complex growth curves can bring much more contributions in the knowledge required for understanding change in longitudinal studies with more than three measurement occasions. In the current paper, a more complex model will be applied: a quadratic growth curve model where the change over time is modelled by three growth parameters: the intercept, the slope and the acceleration (quadratic growth).

It is also rather common that growth curve models are limited to two-level data even when more than two would be more plausible. Typically, these take the form of measurements nested within persons. While the analysis of two-level data has been well documented (Verbeke & Molenberghs, 2000), considerably less attention has been given to the analysis of three-level data. Three level data usually take the form of measurement occasions nested within persons and persons nested within schools, hospitals or firms. The addition of the third level makes it possible to see whether change patterns differ over schools, hospitals or firms as the case may be. Three-level growth curve models thus enable the researcher to investigate whether students in certain schools make more progress or why patients in some hospitals show a faster recovery process. Three-level growth curve models in educational effectiveness research also allow for the calculation of the percentage of the variance in the growth parameter that lies between schools. This percentage is also know as the intraclass correlation coefficient (ICC) and is interpreted as the school effect on a certain student outcome. However, most textbooks on multilevel models or growth curve models (such as Singer & Willett, 2003; Snijders & Bosker, 1999) do not discuss the calculation of this school effect (% of the variance between schools) on the student growth. The current paper will therefore discuss this topic and propose a clear way of calculating the school effect.

In educational effectiveness research, these three-level models have given rise to a lively debate whether schools have a larger effect on student status (intercept) than on student growth (slope) (De Fraine, Van Landeghem, Van Damme, & Onghena, 2005; De Fraine, Van Damme, & Onghena, 2007; May, Supovitz and Perda, 2004; Raudenbush 1989, 1995; Teddlie & Reynolds, 2000). Some researchers believe that schools have a larger impact on their students' growth than on their students' outcomes at a certain point in time. This conviction stems from the study of Raudenbush (1989, 1995) in which the school effect on student initial status for mathematics was 14% whereas the school effect on the learning rates was over 80%. May et al., (2004) indicate that time coding might affect the size of the school effect. They hypothesize that the school effect on student status will be larger when the intercept refers to a later point in time. The current study will investigate whether the time coding affects this debate in terms of producing larger effects on student status than on student growth.

The current paper goes into one of the topics which is unavoidable for all researchers applying growth curve models: the choice of the time coding. With the fast growing use of longitudinal statistical analysis methods in several fields of research and in educational effectiveness research in particular, it is invaluable that the questions regarding the effects of time coding and scaling be addressed properly. The importance of this time coding is highlighted by several researchers who call for in-depth studies on the problem (Raudenbush, 2001a, 2001b; Raudenbush & Bryk, 2002). One important aspect is the definition of the status (intercept) through the time coding (Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1985). A number of articles and handbooks have been written that discuss the effects of changing the meaning of the intercept in growth curve models (Duncan, Duncan, & Strycker, 2006; McArdle, 1988; Rogosa, 1995; Willett & Sayer, 1994). A few others have investigated the effect of the time coding on the correlation between intercept and linear slope (Mehta & West, 2000; Rogosa, 1995; Rogosa & Willett, 1985). These authors have shown that this correlation can change radically by simply changing the time point to which the intercept refers. Mehta and West (2000) also studied how the choice of a specific time coding affects the variance of the intercept. It has also been shown that shifting the intercept does not change the slope parameter in a linear growth curve model. However, this shift of intercept has not been discussed in relation to the impact on school effects.

The coding of time in growth curve models also has important implications for the interpretation of the model (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004; Kreft, de Leeuw, & Aiken, 1995; Mehta & West, 2000). In other words, changing the time coding comes down to changing the research question of interest. Mehta and West (2000) indicate that time coding effects in growth curve models parallel the effects of centring predictor variables in multilevel models.

However, most of the studies on time coding effects mentioned above operate within the framework of two-level linear growth curve models. They are only of limited interest for researchers that investigate non-linear shaped growth curves and for researchers looking at three-level models. This paper therefore opens up theoretical and practical implications of a time coding dependence of parameter estimates for the more complex three-level quadratic growth curve model. These types of models can enable the researcher to investigate the school effects on the student linear and quadratic growth in any outcome of interest.

There are several plausible reasons for considering an intercept different from the initial status by coding the time variable. First, coding time could be used especially in growth curve models to address particular substantive research questions (Raudenbush, 2001a, 2001b; Raudenbush & Bryk, 2002; May, Supovitz & Perda, 2004). A second reason to code time could be when there is a critical need for interpretable parameters estimates (Kreft, de Leeuw, & Aiken, 1995). Third, coding of time might also be used to circumvent software convergence problems (De Fraine et al., 2005; Verbeke & Molenberghs, 2000). And lastly coding of time might also be used to solve boundary value problems in parameter estimation (Verbeke & Molenberghs, 2000). For example changing the unit of time from years to months will have a non-linear effect on the variance of the outcome but implicitly affect the school effect estimate and possibly its interpretation.

It is the intention of this study to obtain a common framework on which one can base the coding of time to gain general approbation of the interpretation, in terms of the parameter estimates, and most importantly the school effects. To investigate the consequences of different time coding schemes on school effects estimates, we will study two important types of estimates: (1) the variances and covariances and (2) the intraclass correlation as will be defined, based on these estimates. Also the formulas for the means (fixed effects) will be provided.

The consequences of changing the time coding are investigated in two ways. First, the exact expressions of the effects of changing the time coding on the estimation of the covariance parameters are derived mathematically. Second, these insights are applied on a dataset of students' well-being in secondary schools to estimate corresponding school effects. In this application, a three-level quadratic growth curve model is fitted with a complex modelling of between and within school covariance structures.

2. Research questions

This paper addresses three main research questions (a-c) after unveiling some of the ramifications around the definition of ICC in multilevel growth curve models. This study also opens a way forward with a new definition of the ICC for multilevel growth curve models with random intercepts only and those with random intercepts and slopes. Before answering these four research questions, a matrix representation of a three-level quadratic growth curve model in the style of Laird and Ware (1982), is formulated. This is because as the number of levels in a multilevel quadratic growth curve model increases, the mathematical representation become very cumbersome.

a) How do different time coding schemes affect the estimates of the fixed effects, school effects and the interpretation of the intercept and intraclass correlation coefficient in a three-level quadratic growth curve model?
b) How true is the hypothesis that school effects on student status are larger when the intercept refers to a later point in time?
c) How large are the school effects on their students' growth compared to their students' status under different time coding schemes?

3. Method

In the current study, the impact of time coding is studied for the case of a three-level quadratic growth curve model. This paper extends the two-level models, often seen in the literature (Hoffman, 2007; Kwok, West, & Green, 2007; Laird & Ware, 1982; Lange & Laird, 1989; Mehta & West, 2000; Verbeke & Molenberghs, 2000), to three-level linear growth models (Spyros, 2008) and further extends the three-level linear growth curve model to a three-level quadratic growth curve model. The three-level quadratic growth curve model considered in this paper is more complex than most three-level random intercept models (Biesanz et al., 2004; Fitzmaurice, Laird, & Ware, 2004; Yang, Goldstein, & Heath, 2000). The complexity arises in the number of random effects specified at the different levels.

The quadratic growth curve is chosen here as the most parsimonious of all non-straight line polynomials. The first step is to derive mathematical expressions (in matrix notation) for three-level quadratic growth curve models. The next step is the discussion of unconditional growth curve models which is followed by an extension to conditional growth curve models. Lastly an elaboration on the effects of altering time coding schemes is made.

3.1. Unconditional Growth Curve Model

Consider for a start, a standard two-level quadratic growth curve model as the one presented in equation 1 below

$$Y_{jk} = \beta_{j0} + \beta_{j1} t_{jk} + \beta_{j2} t_{jk}^2 + \varepsilon_{jk}, \tag{1}$$

Equation 1 represents an unconditional growth curve model for a person j. Y_{jk} is the response variable for the j^{th} person at the k^{th} measurement occasion. β_{j0} represents the intercept which in this equation is the expected value of Y_{jk} for the j^{th} person when the time t_{jk} is zero. β_{j1} is the regression coefficient corresponding to the expected change in Y_{jk} for a 1-unit change in time (t_{jk}) for the j^{th} person. β_{j2} represents the acceleration or deceleration in growth.

Equation 1 can be repeated for all the individuals in the study and once again be arranged as matrices (following Laird & Ware, 1982) and shown in equation 2a below.

$$\mathbf{Y}_j = X_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \tag{2a}$$

This model assumes that $\boldsymbol{\varepsilon}_j$ and $\boldsymbol{\beta}_j$ (which contains student-specific regression coefficients) are independent and X_j is the design matrix for level-1 covariates (1, time and time squared). Allowing the $\boldsymbol{\beta}_j$ to be modelled with a multivariate regression model, equation 2a can be modified to get 2b

$$\mathbf{Y}_j = Z_j\boldsymbol{\beta} + X_j\boldsymbol{\upsilon}_j + \boldsymbol{\varepsilon}_j \qquad \text{(2b)},$$

where $Z_j = X_j K_j$. Details of this model can be found in Verbeke and Molenberghs (2000).

Extending the model now to a three-level unconditional quadratic growth curve in matrix form and introducing a new subscript, yields the following result with i representing the school and j the student. The vectors of repeated measurements are stacked into the matrices:

$$\mathbf{Y}_i = X_i\boldsymbol{\delta} + X_i\boldsymbol{\nu}_i + X_i\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_i \qquad (3)$$

Where Y_i is a response or outcome vector for school i. X_i is a matrix of time covariates. $\boldsymbol{\delta}$ is a vector containing the fixed effects. $\boldsymbol{\nu}_i$ is a vector containing school random effects, $\boldsymbol{\upsilon}_i$ is a vector containing student random effects and $\boldsymbol{\varepsilon}_i$ is the vector of residual components. The dimensions of the vectors $\boldsymbol{\nu}_i$ and $\boldsymbol{\upsilon}_i$ depend on the number of random effects allowed in the model.

3.2. Conditional growth curve model

The model described by equation 4a below, is an extension of equation 3 to include covariates additively. It is a conditional three-level quadratic growth curve model with fixed effects parameters $\boldsymbol{\delta}$, student-specific effects $\boldsymbol{\upsilon}_i$ and school-specific effects $\boldsymbol{\nu}_i$. Using the notation similar to that in Laird and Ware (1982), equation 4b is derived. Care must be taken with these notations because though simple in visualisation, they are more complex in concept. In the process of stacking, vectors will have also vector components in order to avoid representing them as very high dimensional matrices.

$$\mathbf{Y}_i = X_i\boldsymbol{\delta} + T_i\boldsymbol{\nu}_i + Z_i\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_i \qquad \text{(4a)}$$

Where $\mathbf{Y}_i$ is the response or outcome vector, X_i , T_i and Z_i are matrices of known covariates at measurement occasion level, school level and student level respectively.

$$\begin{cases} \mathbf{Y}_i = X_i\boldsymbol{\delta} + T_i\boldsymbol{\nu}_i + Z_i\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\nu}_i \sim N(\mathbf{0}, \Psi_v), \\ \boldsymbol{\upsilon}_j \sim N(\mathbf{0}, \Psi_u), \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma\ ) \\ \boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_N, \boldsymbol{\upsilon}_1, ..., \boldsymbol{\upsilon}_N, \boldsymbol{\varepsilon}_1, ..., \boldsymbol{\varepsilon}_N \ \textit{independent} \end{cases} \tag{4b}$$

Where $\mathbf{Y}_i$ is the n_i -dimensional response or outcome vector for school i , and $1 \le i \le N$, with N being the total number of schools. Here n_i refers to the number of students in school i , while N is fixed for a particular study. This means n_i will take values like $n_1, \ldots, n_N$ with n_1 and n_N being the dimensions of the response vectors for school 1 and school N respectively. The number of measurement occasions per student in each school is also allowed to vary and denoted by m_{ij}. Moreover, X_i , T_i and Z_i are $(n_i \times p), (n_i \times q)$ and $(n_i \times q')$ dimensional matrices of known covariates respectively. $\boldsymbol{\delta}$ is the p-dimensional vector containing the fixed effects. $\boldsymbol{\nu}_i$ is the q-dimensional vector containing school random effects, $\boldsymbol{\upsilon}_i$ is the q'-dimensional vector containing student random effects and $\boldsymbol{\varepsilon}_i$ is the n_i -dimensional vector of residual components. For covariance matrices, Ψ_v is a general $(q \times q)$ covariance symmetric matrix, Ψ_u is a general $(q' \times q')$ covariance symmetric matrix and Σ is a $(n_i \times n_i)$ covariance matrix which may be assumed (though not necessarily) to depend on i and j only through their dimensions n_i and m_{ij}. This means the linear mixed model of Laird and Ware becomes a multilevel linear mixed model. This matrix notation is very handy when looking at models with several levels. It is used later on in this paper, to ease demonstration and make the necessary changes tractable.

3.3. Coding time and parameter estimates for a three-level quadratic growth curve model

The effects of changes in time coding will be considered as a linear transformation. Equation 1 can be written in matrix form as follows:

$$\underbrace{\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{114} \end{pmatrix}}_{Y_1} = \underbrace{\begin{pmatrix} 1 & t_{111} & t^2_{111} \\ 1 & t_{112} & t^2_{112} \\ 1 & t_{113} & t^2_{113} \\ 1 & t_{114} & t^2_{114} \end{pmatrix}}_{X_1} \underset{\beta_1}{\begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{12} \end{pmatrix}} + \underset{\varepsilon_1}{\begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{113} \\ \varepsilon_{114} \end{pmatrix}}$$

If a new coding is given to the time variable, the design matrix X_i (factor loading matrix) changes in the above matrix equation. This change affects the solution of this system of equations as will be shown shortly. Let t_{ijk} be the time variable defined by the scale $[1,...,k]$ meaning k measurement occasions (in this application there are four measurement occasions just to simplify the mathematical steps). Let's assume that it is always possible to code this time variable to a new time $t*_{ijk}$ such that the relationship between the new time and the former is linear as $t*_{ijk} = a + bt_{ijk}$ where $b \neq 0$. This gives a similar system of equations as the one above with the new design matrix being $X*_i$

$$X*_i = \begin{pmatrix} 1 & t*_{ij1} & t*^2_{ij1} \\ 1 & t*_{ij2} & t*^2_{ij2} \\ 1 & t*_{ij3} & t*^2_{ij3} \\ 1 & t*_{ij4} & t*^2_{ij4} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & t_{ij1} & t^2_{ij1} \\ 1 & t_{ij2} & t^2_{ij2} \\ 1 & t_{ij3} & t^2_{ij3} \\ 1 & t_{ij4} & t^2_{ij4} \end{pmatrix}}_{X_i} \underbrace{\begin{pmatrix} 1 & a & a^2 \\ 0 & b & 2ab \\ 0 & 0 & b^2 \end{pmatrix}}_{P}$$

$$X*_i = X_i P \qquad (5)$$

The relationship between the original design matrix and the transformed design matrix is represented in equation 5 and by simply making choices for the constants a and b in the linear relationship, the transformation matrix P is defined. The inverse transformation is represented by equation 6 below.

$$X_i = X*_i P^{-1} \qquad (6)$$

P^{-1} is the inverse of the matrix P and is the matrix of the inverse transformation shown in equation 6. This inverse matrix is always possible to get because the matrix P is non-singular.

$$\text{where } P^{-1} = \frac{1}{|P|} adjP = \frac{1}{b^2}\begin{pmatrix} b^2 & -ab & a^2 \\ 0 & b & -2a \\ 0 & 0 & 1 \end{pmatrix}$$

$adjP$ is called the adjoint matrix of the transformation matrix P and is defined to be the transpose of the cofactor matrix of P and $|P|$ is the determinant of P.

The expectation of $\mathbf{Y}_i$ from equation 3 is given by $\boldsymbol{\mu} = X_i\boldsymbol{\delta}$ and becomes $\boldsymbol{\mu}^* = X^*_i P^{-1}\boldsymbol{\delta}$ under the linear transformation of equation 6. It is possible to relate the fixed effects parameters estimated under the original time variable ($\boldsymbol{\delta}$) to those under the transformed time variable ($\boldsymbol{\delta}^*$) by $\boldsymbol{\delta}^* = P^{-1}\boldsymbol{\delta}$ with parameters for a quadratic growth curve being:

$$\beta_0^* = \beta_0 - \frac{a}{b}\beta_1 + \frac{a^2}{b^2}\beta_2 \tag{7}$$

$$\beta_1^* = \frac{1}{b}\beta_1 + \frac{-2a}{b^2}\beta_2 \tag{8}$$

$$\beta_2^* = \frac{1}{b^2}\beta_2 \tag{9}$$

Equations 7-9 are different from those obtained in the case of a linear growth model (see Stoel & Van den Wittenboer, 2003). β_0, β_1, β_2 are the intercept, linear slope and quadratic slope for the fixed effects parameters under the original time variable and β_0^*, β_1^*, β_2^* are those under the transformed time variable. The transformation obtained when $b=1$ in simple terms represents a shift in the origin from one measurement occasion to another. This special case appears frequently in educational research and this specific situation will be discussed later in detail.

The fixed effect estimates are not the only parameters affected by a linear transformation of the time variable but even substantially are the random effects. The covariance matrix of the observations across schools, students and measurement occasions for the unconditional three-level growth curve model shown in equation 3 can be expressed as $V_{yy} = X_i(\Psi_v + \Psi_u)X_i^T + \Theta_\varepsilon$ in which there is no T_i or Z_i matrix. After the transformation, this is obtained

$$\begin{aligned}
V*_{yy} &= X*_i P^{-1}(\Psi_v + \Psi_u)(X*_i P^{-1})^T + \Theta_\varepsilon \\
&= X*_i \left[P^{-1}\Psi_v (P^{-1})^T + P^{-1}\Psi_u (P^{-1})^T \right](X*_i)^T + \Theta_\varepsilon \\
&= X*_i \left[\Psi_v^* + \Psi_u^* \right](X*_i)^T + \Theta_\varepsilon
\end{aligned}$$

where $\Psi_v^* = P^{-1}\Psi_v(P^{-1})^T$ and $\Psi_u^* = P^{-1}\Psi_u(P^{-1})^T$

It is then possible to calculate the components of the transformed covariance matrix for the school level and the student level in terms of the components of the original covariance matrix. An elaboration is shown with the school level covariance matrix (Ψ_v). The student-level for this model will follow in a similar fashion.

$$\Psi_v^* = P^{-1}\Psi_v(P^{-1})^T$$

$$where \quad P^{-1} = \frac{1}{b^2}\begin{pmatrix} b^2 & -ab & a^2 \\ 0 & b & -2a \\ 0 & 0 & 1 \end{pmatrix} \; and \quad \Psi_v = \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix},$$

v_{11} is the variance of the student status, v_{22} is the variance of the student growth, v_{33} is the variance of the general curvature of the growth curve or the acceleration. v_{21}, v_{31} and v_{32} are the covariances of the student status and the student growth parameters, student status and acceleration, and student growth and acceleration respectively at the school level.

The above matrix equations can then be solved as follows:

$$\Psi_v^* = P^{-1}\Psi_v(P^{-1})^T = \frac{1}{b^4}\begin{pmatrix} b^2 & -ab & a^2 \\ 0 & b & -2a \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}\begin{pmatrix} b^2 & 0 & 0 \\ -ab & b & 0 \\ a^2 & -2a & 1 \end{pmatrix}$$

By multiplying these three matrices out gives the following as the solutions for the transformed components of Ψ_v^* in terms of the original components of Ψ_v :

$$v*_{11} = v_{11} + \frac{a^2}{b^2} v_{22} + \frac{a^4}{b^4} v_{33} - \frac{2a}{b} v_{12} + \frac{2a^2}{b^2} v_{13} - \frac{2a^3}{b^3} v_{23},$$

$$v*_{12} = -\frac{a}{b^2} v_{22} - \frac{2a^3}{b^4} v_{33} + \frac{1}{b} v_{12} - \frac{2a}{b^2} v_{13} + \frac{3a^2}{b^3} v_{23},$$

$$v*_{13} = \frac{a^2}{b^4} v_{33} + \frac{1}{b^2} v_{13} - \frac{a}{b^3} v_{23},$$

$$v*_{22} = \frac{1}{b^2} v_{22} + \frac{4a^2}{b^4} v_{33} - \frac{4a}{b^3} v_{23},$$

$v*_{23} = -\frac{2a}{b^4} v_{33} + \frac{1}{b^3} v_{23}$ and $v*_{33} = \frac{1}{b^4} v_{33}$

The effects of changes in the time coding on the random effects in a quadratic growth curve model relationships are more complex than the ones shown in Garst (2000), for a linear growth curve. The process of coding time and the above relationships do not seem to be a simple generalization as some authors put it (Mehta & West, 2000).

The relationships shown indicate that all the components of the transformed covariance matrix change. The variance of the quadratic slope (v_{33}) and its associated covariances with the intercept and linear slope variances play a very important role in the estimation and interpretation of the student status and student growth.

3.4. *Shifting the intercept*

Consider a typical transformation in the educational and social sciences of the form $t*_{ijk} = a + t_{ijk}$ meaning $b = 1$(Biesanz et al., 2004; De Fraine et al., 2005, 2007; Wilkins & Ma 2002, 2003). When considering a linear growth curve model, a shift in the origin results in a change in the intercept and not the slope parameter with $b = 1$. On the contrary a quadratic growth curve model with a similar shift in the origin does not only affect the intercept but also the linear slope parameter. The quadratic slope parameter is invariant with $b = 1$. The expressions for the mean, variance and covariance parameters under this transformation now simplify to the following:

$$\beta_0{*} = \beta_0 - a\beta_1 + a^2\beta_2,$$
$$\beta_1{*} = \beta_1 - 2a\beta_2,$$
$$and \quad \beta_2{*} = \beta_2 \quad respectively.$$

$$v{*}_{11} = v_{11} + a^2 v_{22} + a^4 v_{33} - 2av_{12} + 2a^2 v_{13} - 2a^3 v_{23} \tag{10}$$

$$v{*}_{12} = -av_{22} - 2a^3 v_{33} + v_{12} - 2av_{13} + 3a^2 v_{23} \tag{11}$$

$$v{*}_{13} = a^2 v_{33} + v_{13} - av_{23} \tag{12}$$

$$v{*}_{22} = v_{22} + 4a^2 v_{33} - 4av_{23} \tag{13}$$

$$v{*}_{23} = -2av_{33} + v_{23} \tag{14}$$

$$v{*}_{33} = v_{33} \tag{15}$$

Also note that the mean and variance of the acceleration parameter is unchanged when only the intercept is shifted. Based on the mathematical relationships of equations 10-15, it is not possible to confirm the hypothesis that school effects will be larger when the intercept refers to a later point in time. For the case of a linear model, the mathematical relationships show a clear trend with few conditions. For example in the linear case $v{*}_{11} = v_{11} + a^2 v_{22} - 2av_{12}$ and implies simply that with a negative covariance between the random intercept and linear slope, $v{*}_{11} \geq v_{11}$ for all values of a . Under this condition of negative covariance (meaning students with low status grow faster than students with high status), the transformed variance will increase as the origin shifts to later time points. The three-level quadratic growth curve model results in tables 1A, 2A, and 3A of the appendix A can be used to verify the mathematical relationships presented in equations 10-15.

Despite the various reasons for coding the time variable discussed earlier on in this section, the bone of contention still remains in terms of the varying interpretation of the results after a time variable transformation. School effects estimation and interpretation change drastically when different time coding schemes and interpretations of student status are used. The possibilities of obtaining valid intercepts (estimate and interpretation), and yet meaningless school effect estimates is enormous.

Consider a regression model given as $Y_{jk} = \beta_{j0} + \beta_{j1} X_{jk} + \beta_{j2} X_{jk}^2 + \varepsilon_{jk}$, β_{j0} is called the Y_{jk} intercept of this regression equation when $X_{jk} = 0$ is in the scope of the model. When $X_{jk} = 0$ is not in the scope of the model then β_{j0} does not have any particular meaning as a separate term in the regression equation (Kutner, Nachtsheim, Neter, & Li, 2005). Consider for example that X_{jk} is the age of student k in secondary school j, then β_{j0} as an intercept will be nonsense considering that X_{jk} cannot take zero for students in secondary school. In any case, if the ages of the students in secondary school range between 12 and 18 years, one may be interested in a particular age or mean age and so modify the regression equation correspondingly. For example, the age could be re-coded as "age minus 12". In that case, $X_{jk} = 0$ refers to the age at the beginning of secondary school and therefore the intercept is meaningful. While the latter example remains reasonable and sometimes even very convenient in terms of answering specific research questions, it complicates the use of the intercept variance in the estimation of school effects. Indeed, Raudenbush and Bryk (2002) emphasize that researchers should be cautious with the choice of location for level-1 covariates because of its implications in the interpretation of the variances and covariances of the model.

While it is possible to estimate the intraclass correlation coefficient (ICC) for a particular student outcome at the start of grade 7 of secondary school for example, it is not possible in this case to use the ICC estimate as a school effect and get a meaningful interpretation. The school effect obtained using the ICC calculated at the start of grade 7 represents the percentage of variance in the particular student outcome due to the differences in the schools they attend. However, this school effect estimate is not very correct because the students have just barely started school and the schools cannot solely be responsible for the variability. In fact if it were ever possible in practice, to account for all the recruitment characteristics responsible for this difference, then the school effect at the start of the year should be zero. School effects estimates using ICC will make more sense only when considered at a later time than the start of the study because the impact of the school would have been felt by the students and consequently their performance.

The use of the ICC for the estimation of school effects on student growth is handicapped because it is limited to random intercept models. This paper proposes a new definition which is general enough to take care of quadratic and higher order polynomials growth curves with more than three levels of hierarchy. The choice of a later time also becomes one of great interest because of the many possibilities that would be available in any repeated measurement study when the meaningful intercept occurs at a point different from the initial status.

3.5. Small variance components estimation problems

The problem of estimating small variance components is very frequent when considering quadratic three-level growth curve models. However, this problem is also very common with multilevel models of many levels when the point of convergence is close to the boundary parameter space. When such a problem is encountered during data analysis, the most common temptation is for researchers to set these variance components to zero and move on with the analysis. The time transformation illustrated above can be used to circumvent this problem. Consider the transformation $t*_{ijk} = a + bt_{ijk}$ and let $a = 0$ and $b = \frac{1}{c}$ then the covariance parameters under this transformation now simplify to the following

$v*_{11} = v_{11}$, $v*_{12} = cv_{12}$, $v*_{13} = c^2 v_{13}$, $v*_{22} = c^2 v_{22}$, $v*_{23} = c^3 v_{23}$, and $v*_{33} = c^4 v_{33}$

This transformation is applied only to the design matrix of the random part of the growth model and results in enlarged covariance parameters. Therefore the fixed effects estimates remain unchanged while the multiplicative factor used cancels out in the school effect calculation. This implies that the peak of the loglikelihood is well away from the boundary. This transformation helps to make the system of normal equations to be more stable and the maximization algorithm used to solve the system of equation, runs without any convergence problems. It is worth noting that this technique does not apply to convergence problems due to model misspecifications. An example of model misspecification can be the convergence problem resulting from the estimation of negative variance components.

4. The intraclass correlation coefficient in growth curve models

The intraclass correlation coefficient (ICC) is considered as a measure of school effects in educational effectiveness research. In this domain the use of growth models is increasing with much interest centred on the school effects on student status and even recently on student growth. It is important to unveil some of the ramifications around the definition of ICC in growth curve models. The most common definition of ICC for growth curve models is described by the proportion of the total variance that lies “between” people (Singer & Willett, 2003). Although this definition is commonly used in the social science fields like behavioural measurement, psychometrics, and educational effectiveness, a definition of the ICC is lacking when models deviate from being a random intercept only model. This section will throw more light into this issue and open a way forward for new definitions of the ICC for growth curve models that contain random slopes.

4.1. The intraclass correlation in two-level random intercept models

For a 2-level model with only a random intercept, the definition of the ICC can be written in a straightforward way mathematically as $\frac{v_{11}}{v_{11}+\sigma^2}$ (Snijders & Bosker, 1999; Verbeke, & Molenberghs 2000). This ICC represents the proportion of the variance at the second level. It can also be interpreted as the expected correlation between two randomly chosen subjects within the same group. In educational research, one can talk of two students within a school or two measurement occasions within a student. This definition for the case of a two-level random intercept model posses no problem in terms of consistency with many other authors (Fritzmaurice, Laird, & Ware, 2004; Goldstein, 1995; Raudenbush & Bryk, 2002).

4.2. The intraclass correlation in three-level random intercept models

In a three-level model (as in repeated measurements within students and students within schools) with only a random intercept, there are two ways to calculate the ICCs (Siddiqui, Hedeker, Flay, & Hu, 1996). Once the total variance is estimated as a sum of all variances at the three levels, then the first definition of ICCs for the second and third levels can be defined for the student and school levels as

$ICC_{school} = \frac{v_{11}}{v_{11} + u_{11} + \sigma^2}$ and $ICC_{student} = \frac{u_{11}}{v_{11} + u_{11} + \sigma^2}$

Where v_{11} and u_{11} are the variances for the intercepts at the school and student level respectively. σ^2 is the level-1 variance (Davis & Scott, 1995). A second method to define these ICCs is:

$ICC_{school} = \frac{v_{11}}{v_{11} + u_{11} + \sigma^2}$ and $ICC_{student} = \frac{u_{11} + v_{11}}{v_{11} + u_{11} + \sigma^2}$. Here the student ICC indicates the correlation between any two measurement occasions for the same student in the same school (Siddiqui et al., 1996). The two definitions are valid as they represent different truths about the same model. The first method represents the proportion of variance at the student and school level while the second represents an estimate of the expected correlation between two randomly selected subjects in the same school. In any case, in educational research, the main interest is on the ICC at the school level which is used as a measure of the school effect. The ICC at the student-level is not of prior interest here, thus the difference between the two methods for this case is not so important.

4.3. *The intraclass correlation in three-level random intercept and slope model*

A multilevel growth curve model with more than just a random intercept is much more complex and the definition of the ICC becomes trickier resulting in inconsistencies between different authors (Singer & Willett, 2003). Unlike for the random-intercept only model where the intraclass correlation coincides with the proportion of variance in the response or outcome variable between groups (Raudenbush & Bryk, 2002), models with a random intercept and slopes do not have such a coincidence. For a three-level growth curve model with a random intercept and random slopes, this coincidence does not occur because the intercept and slopes components are allowed to covary; as a result the total variance is no longer the sum of the variances of the different components.

A proposal for calculating the school effects in a linear growth curve model with a random intercept and slope has been made by Raudenbush and Bryk (2002) using "the percentage of variation that lies between schools for both the initial status and growth" to measure the school effect. They give a formal definition (without referring to it as an ICC) of the percentage of variance between schools on the student initial status and student growth as:

% variance between schools on student's initial status = $\frac{v_{11}}{v_{11}+u_{11}}$

% variance between schools on student's linear growth = $\frac{v_{22}}{v_{22}+u_{22}}$,

Here v_{22} and u_{22} are the variances for the linear slopes at the school and student levels respectively.

In a three-level quadratic growth curve model, the percentage of variance between schools for the student status, linear and quadratic growth components as defined above can be written as: $\frac{v_{11}}{v_{11}+u_{11}}$, $\frac{v_{22}}{v_{22}+u_{22}}$ and $\frac{v_{33}}{v_{33}+u_{33}}$ respectively. Here, v_{33}

and u_{33} are the variances for the quadratic slopes for the school and student levels respectively.

A generalisation of this definition of proportions is proposed in this paper for a three level quadratic growth curve model with random intercept, linear slope and quadratic slope, to reconcile the two definitions. That is the definition for a random intercept only model (ICC) and that for the random intercept and slopes model (% variance between schools). This will enable the estimation of school effects on the students' linear and quadratic growth in a particular outcome variable.

Consider the three-level quadratic growth curve model described by equation 4b and which allows for three random regression parameters which are the intercept, linear slope and quadratic slope. Let the variance of the error term ε_{ijk} be σ^2, the variances of the student and school level intercepts are u_{11} and v_{11} respectively. The variances for the corresponding student and school level linear and quadratic slope residuals are u_{22}, v_{22}, u_{33} and v_{33}. It is then possible to define the percentage of variance between schools on each of these regression parameters (intercept, linear slope and quadratic slope) as follows:

% variance between schools on $\beta_{ijp} = \rho_{p+1}$

$$\rho_{p+1} = \begin{cases} \dfrac{v_{11}}{v_{11} + u_{11} + \sigma^2} & \textit{when } P = 1 \\ \dfrac{v_{(p+1)(p+1)}}{v_{(p+1)(p+1)} + u_{(p+1)(p+1)}} & \textit{when } P > 1 \end{cases} \qquad (16),$$

$\textit{where } p = 0,\ldots,P-1 \textit{ and } P \textit{ is the number of random effects allowed}.$

For example with a random intercept only model, P=1 and $p = 0$, and the percentage of variance between schools on the intercept as defined by equation 16, corresponds to the definition of the traditional intraclass correlation coefficient.

A second example can be a random intercept, linear slope and quadratic slope model with P=3 and $p = 0,1,2$. The second example will result in the estimation of three percentages of variance between schools that is on the intercept (ρ_1), linear slope (ρ_2) and quadratic slope (ρ_3).

4.4. Definition of the percentage of variance between schools for models with more than three levels

This paper puts forward another proposed definition which can be used to generalize the above definition for models with more than three levels of hierarchy. An example in educational research could be repeated measurements within students within classes within schools within educational systems or even within countries. This is done with an additional index which replaces the different letters used to denote the different levels, to avoid running out of letters of the alphabet. Let $v_{q(p+1)(p+1)}$ denote the variance of the qth level random regression coefficient. q takes values 1,…, Q with Q not less than 2. It is then possible to define the percentage of variance between units of the q level on each of these regression parameters as follows:

% variance between units of level q on $\beta_{ijp} = \rho_{q(p+1)}$,

$$\text{where } \rho_{q(p+1)} = \begin{cases} \dfrac{v_{q11}}{\sum_{q=1}^{Q} v_{q11}} & \textit{when } P = 1 \textit{ and } q = 1,\dots,Q \\ \dfrac{v_{q(p+1)(p+1)}}{\sum_{q=2}^{Q} v_{q(p+1)(p+1)}} & \textit{when } P > 1 \textit{ and } q = 2,\dots,Q \end{cases} ,$$

where $p = 0,\dots,P-1$. Where P is the number of random coefficients allowed and Q is the number of levels of hierarchy in the data. In this new definition, the percentage of variance between level q units for the case of a random intercept model, that is with $P = 1$ will corresponds to the definition of the intraclass correlation at the qth level (ρ_{q1}).

5. Application to educational data

5.1. Data

The data that will be considered in this application are the LOSO-data which is the acronym for the Dutch form “Longitudinaal Onderzoek in het Secundair Onderwijs”. This is a large-scale research project on secondary education (Van Damme et al., 2006). In this project, over 6000 students from about 57 schools in Flanders (Belgium) are followed during secondary education. The student outcome studied is the student well-being at school. The outcome is obtained by the use of the questionnaire on student well-being and was administered to the students four times during secondary school: at the end of Grade 7, Grade 8, Grade 10 and Grade 12. Student well-being at school is measured on a 32-item scale. The sample with 3788 students from 53 schools is studied for the application of the time transformations and school effect estimations described earlier. The sample includes only students who did not change schools and students who did not repeat a grade. It is also restricted to those students who have at least three measurements, in order to be able to estimate individual student growth curves.

The main software programme used in this is SAS 9.1 (SAS Institute Inc. 2003). The PROC MIXED procedure of SAS is very flexible and suitable for fitting multilevel models, hierarchical linear models, and growth curve models (Singer, 1998). Two other software programmes were used for comparability with the outputs of some of the models done in SAS: MLwiN 2.02 (Rasbash et al., 2000) and Mplus 4 (Muthén & Muthén, 2006).

5.2. Time coding schemes

The operationalization of time as described earlier is $t*_{ijk} = a + t_{ijk}$ and $t*_{ijk} = \frac{1}{c} t_{ijk}$, where the new time variable is $t*_{ijk}$. In the original time coding of the data, the origin refers to the start of Grade 7. In what follows, four alternative time codings will be applied in which the origin refers to end of Grade 7, Grade 8, Grade 10 and Grade 12. In this study, six different time codings will be compared: original time, (time minus 1), (time minus 2), (time minus 4), and (time minus 6). In other words, the constant a takes values 0 (original time), -1 (coding A), -2 (coding B), -4 (coding C), -6 (coding D) and the transformation of time by scaling (coding E) which considers the time in decades by letting $c = 10$. Three-level quadratic growth curve models are then fitted for the original and the other five different transformed time variables. The three columns of the design matrices below represent time^0 (1), time^1 (time) and time^2 (time squared).

Time code O $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \end{pmatrix}$ (original time), Time code A $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \end{pmatrix}$ (time - 1), Time code B $\begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{pmatrix}$ (time - 2), Time code C $\begin{pmatrix} 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \end{pmatrix}$ (time - 4),

Time code D $\begin{pmatrix} 1 & -5 & 25 \\ 1 & -4 & 16 \\ 1 & -2 & 4 \\ 1 & 0 & 0 \end{pmatrix}$ (time - 6), Time code E $\begin{pmatrix} 1 & 1/10 & 1/10^2 \\ 1 & 2/10 & 4/10^2 \\ 1 & 4/10 & 16/10^2 \\ 1 & 6/10 & 36/10^2 \end{pmatrix}$ (time/10)

The design matrices shown above have a great impact on the schools and students fixed effects and random effects parameters. The next paragraph looks into the impact of time coding on the intraclass correlation or proportion of variance between schools for three-level growth curves.

6. Results

Three different specifications of a three level growth curve model are considered as follows: Model A is a three-level linear growth curve model with random intercept and linear slopes for both the school and student levels. Model B is a three-level quadratic growth curve model with random intercept and linear slopes for the school and student levels. Model C is a three-level quadratic growth curve model with random intercept, linear and quadratic slopes for both the school and student levels. The ordinary deviance statistics in combination with individual and school growth profiles is used to show that a quadratic growth fits better than a linear one. However, in order to determine how many random effects to specify in the quadratic growth curve model, a mixture of chi-square asymptotic null distribution as the null distribution of the deviance statistic is used. The degrees of freedom are calculated as the number of random effects in the models. Model C is shown to have the best fit compared to any other reduced model in terms of random effects. However, the results of model A and B are still considered to show the effect of choosing less fitting models instead of the more appropriate quadratic growth curve with all possible random effects. The result of this model fit is presented on table A1. The results of the parameter estimates of the fixed effects for model C with the different time codings are in table 1. A summary of the results for the three growth curve models described for the student well-being is given in table 2. Detailed results of all the parameter estimates can be found in tables A2, and A3 of the appendix A.

Table 1

The estimates of the fixed effects of the students' well-being with a quadratic growth curve.

	Estimate(model based Std errors)					
Effects	Original	Code A	Code B	Code C	Code D	Code E
Intercept	3.96(0.030)*	3.74(0.012)*	3.55(0.016)*	3.30(0.020)*	3.21(0.022)*	3.96(0.030)*
Linear slope	-0.24(0.018)*	-0.20(0.014)*	-0.16(0.009)*	-0.09(0.005)*	**-0.01(0.012)**	-0.24(0.018)*
Quadratic Slope	0.02(0.002)*	0.02(0.002)*	0.02(0.002)*	0.02(0.002)*	0.02(0.002)*	0.02(0.002)*

The results of table 1 show obviously that changing the time coding by shifting the intercept affects the intercept and the linear slope of the growth curve but not the quadratic slope which in all these transformations is significantly different from zero ($p<0.0001$). The fixed effects results of the scaling transformation (Code E) are the same as those of the original time as expected because the transformation is done only on the random effects design matrix. This is a good transformation if the reason of transforming the time variable is to avoid small variance estimation problems. Care must be taken in any case when deciding to shift the intercept to later time points because some parameters may even become insignificant. This can be seen with the time coding D at the end of grade 12 which shows that the linear slope is not significantly different from zero ($p=0,5818$) at a 5% level.

Table 2

The percentage of the variance of the well-being at the school level for the three models

School level	Original time	Time code A	Time code B	Time code C	Time code D	Time code E
<u>Model A</u>						
% var Intercept	5.9	5.2	5.2	7.4	11.5	5.9
% var Slope	18.0	18.0	18.0	18.0	18.0	18.0
<u>Model B</u>						
% var Intercept	5.4	5.1	5.17	7.17	10.60	5.40
% var Slope	13.7	13.7	13.7	13.7	13.7	13.7
<u>Model C</u>						
% var Intercept	9.7	4.3	3.3	4.9	7.6	9.7
% var Lin. Slope	29.7	26.4	21.1	15.2	/	29.7
% var Quad Slope	40.0	40.0	40.0	40.0	40.0	40.0

Looking at the results of table 2 for model A and model B, the percentage of variance at the school level on the student status for student well-being increases in general as the origin is shifted to later time points. At the start of Grade 7 (original time coding), about 5% of the differences in student well-being on the student status are situated at school level but by the end of Grade 12 (Time code D), this school effect has grown to about 11% for model A and 10% for model B. This result ties with the findings of May et al. (2004) even though a more complex model and data structure has been considered. The linear (model A) and quadratic (model B) models indicate a constant variance in student linear growth though a little higher for the linear model than for the quadratic model (18%; 13.73%).

Model C shows that the school effect on the student status decreases from 9.7% to 3.3% by the end of grade 8 and then increases to 7.6% at the end of grade 10. On the other hand, the school effect on the students' linear growth in well-being decreases from 29.7% to 15.2% at the end of grade 10. Since the linear growth parameter is not significant at the 5% level of significance, no school effect parameter on linear growth was estimated. No school effect was estimated for linear growth in well-being because one of the corresponding variance parameter estimated at the end of grade 12 was not significant. The school effect on the student growth in well-being from model C seems consistently higher than the school effect on the student status thus confirming the findings of Raudenbush (1995) that schools have a higher effect on the students' growth than status at a certain point in time. The quadratic parameter representing the rate at which the students' grow over time is also estimated and its corresponding school effect is estimated as 40%. This means that schools have an even greater effect on the rate of growth with time than the growth in the student's well-being. These data also support that schools have a larger effect on student growth than on student status.

7. Discussion and Conclusion

Growth curve models are growing in terms of their application in several research domains. The current paper was intended to shed some light on the issue of time coding in these growth curve models. Some studies have addressed the effects of changes in the time scale, but most of these studies were restricted to the common linear two-level growth curve model. In the current study, a broader framework was chosen and the effects of changes in time coding were illustrated for the quadratic three-level growth curve model.

The paper employed two methods to study the effects of time coding in growth curve models: mathematical formulas and an elaborated application. First, formulas were derived that indicated that different time coding schemes affect both the estimates of the fixed effects (growth parameters) and the variances and covariances (random effects). Formulas were derived for a linear transformation of the time variable and for the special cases where the intercept is shifted and also when some variance parameters are very small. These formulas are easy to use in reverse order to calculate the parameters that would have been found when another time scale would have been used. Second, the effect of shifting the intercept was illustrated through an application on educational data. For these analyses, the school effects were compared for three types of models (A, B and C) and six different time codings. The five of these time coding schemes differed according to what measurement occasion was chosen as the origin and the sixth was a time scale transformation for very small variance parameters. This application illustrated that, when the intercept refers to another point in time, this affects the estimation of the size of the school effects.

The current article also gave an overview on the meaning and calculation of the ICC, which is interpreted in educational effectiveness research as 'the school effect'. Because of the lack of a clear definition on how to calculate school effects in random slope models, a clear formula was proposed. This formula makes it possible to compare school effects on different growth parameters (such as the intercept and the slope). This formula was given for models with three levels and it was elaborated to models with more levels in the hierarchy and more random coefficients.

The main conclusion of this study is that the estimation of the size of the school effect (intraclass correlation) can be strongly determined by the choices for the time scale. The hypothesis that school effects are larger when the intercept refers to a later point in time (May et al., 2004) was only partially confirmed. The analyses for the well-being subscribed a general increase of school effects at a later time point. The current study also contributes to the discussion in the field of educational effectiveness whether school effects are larger for student growth (slope) or for student status (intercept). The school effect on the student growth in well-being seems consistently higher than the school effect on the student status thus confirming the findings of Raudenbush (1995).

In the study, only raw school effects were calculated. This means that no covariates (such as student prior achievement, intelligence, sex or socio-economic status) were added to the model in order to make a more fair comparison between schools (Teddlie & Reynolds, 2003). One would expect such net school effects to be smaller than the raw school effects though research on this decrease for conditional quadratic growth curve models are rare.

When model C is used instead of model B, the school effect on student linear growth in well-being increases from about 14% to 30% with the untransformed time. Though the results are not as strong as those obtained by Raudenbush (1995), they are in agreement with the fact that schools affect students' growth more than their status in well-being. Our three-level quadratic growth curve model as specified in model C produces the largest school effect on both the students' linear and quadratic growth in well-being.

And only for model C, the estimate of the school effect on students' linear slope was affected by the time coding (as was expected from the formulas that were derived). The effects of the time coding in model C on the estimate of school effect for the slope was quite impressive. However, no conclusion can be made in terms of how the school effects will behave (increase or decrease) when the time coding is changed without considering what type of model is used. Further research with other response variables may help to generalise our findings.

Then, how should researchers decide on what time coding they should use? The coding of time can be chosen to solve substantive problems, stretching from easily and readily interpretable parameter estimates, to interests in the understanding of school effects at particular points in time. Based on the results of this study, we want to advise researchers using growth curve models to be very explicit on the time coding used. Studies should report very clearly to what point in time the intercept is referring. This is important, because different time coding schemes alter the estimations. In fact, changing the time coding comes down to changing the research question. The current paper has shown that school effects estimates might change drastically when the time coding is changed.

MANUSCRIPT 2: Multilevel serial correlation correction of school effect estimation in growth curve models[2]

Anumendem, D. N., De Fraine, B., Onghena, P., & Van Damme, J.

Abstract

The analysis of repeated measurement data can be carried out very efficiently using multilevel methods albeit the standard assumption of serially uncorrelated level 1 residuals usually fails. The need for appropriate modelling of this serial correlation is invaluable in statistical analysis. This study examines educational effectiveness research data with unequally spaced three-level repeated measurements. We introduce the semi-variogram for multilevel data, which determine the presence and the form of serial correlation in repeated measurements of students' response variables. We propose modelling of serially correlated residuals at level 1 with appropriate functions and go further to model serial correlation at level 2. This proposed method is used finally to show the huge impact of serial correlation correction on school effects estimates.

[2] Manuscript submitted for publication.

1. Introduction

The general belief in school effectiveness research is that schools have a larger impact on their students' growth than on their students' outcomes at a certain point in time. This belief emanates mainly from the research of Raudenbush (1989, 1995) in which he showed that the proportion of the school effect on student initial status for mathematics was 14% whereas that for learning rates or students' progress over time was over 80%. This study and many others have accreted the use of growth in student outcomes over time to the point of becoming the incumbent method among some researchers who see it as the most appropriate criterion for assessing school effectiveness. The investigation of such changes in students' outcomes has necessitated the call for more studies with repeated measurements over time.

The assessment of school effectiveness using student growth in outcomes has resulted in the general call for longitudinal studies of school effects (Teddlie & Reynolds, 2000). They see longitudinal school effectiveness studies as the most sophisticated in nature and note at the same time the increasing need of such studies in the field of school effectiveness. They propose two main reasons for this increase. The first is that cross-sectional designs may underestimate the impact of schools and the second is that cross-sectional designs do not provide a proper framework for studies in school effectiveness processes.

While longitudinal studies are universally accepted by researchers as a key design to study change over time, it comes with a number of challenges in terms of data structure and statistical analysis. Children change with time by virtue of their experience in school, and also because the structures, functions, and compositions of the schools they attend change with time. There are possibly other sources of this change that are neither of the children's nor the schools' making, which cannot be captured by the researcher. Accordingly, the inebriety caused by these sources is present in a way as to influence the outcome of interest. One can say that the sources and consequences of this genre of stochastic change often constitute the object of study in school effectiveness research (Rowan & Denk, 1982).

It is very possible that the small school effects reported by many studies result from the statistical methods used (Scheerens & Bosker, 1997). Scheerens and Bosker (1997) also show how school effects are underestimated when the school variance structure is misspecified or not taken into account at all. For a three-level situation as shown later in the current study, within student variance can also be misspecified and result to another underestimate of the school effects. However, the method proposed in this paper is not only geared towards obtaining larger school effects but also to achieving better and more accurate estimates of variance components.

While Ferron, Dailey, and Yi (2002) have looked at two-level models of change with misspecification of level 1 error structure, relatively little has been done for three-level models. Our study breaks the level 1 error into two components, a pure measurement error which can be corrected for misspecification (by the use of covariance structures like; simple, compound symmetry, banded, autoregressive (1), toeplitz etc.) and a serial correlation component which must be accounted for with a proper structure (by the use of gaussian, power and exponential functions). The model is extended further to include serial correlations at the second level. The principal aim of this paper is then to study the effects of serial correlation correction on school effects estimates in a multilevel growth curve model setting.

Although multilevel growth curve models are very popular among educational researchers, and time series most prominent among econometricians, some researchers have considered the multilevel growth curve models as a genre of the time series (Goldstein, Healy & Rasbash, 1994; Verbeke & Molenberghs, 2000; Willms & Raudenbush, 1989). Time series and their applications have been used in several fields in which repeated measurements or observations are used. (Diggle, 1990; Diggle, Liang & Zeger, 1994; Goldstein, Healy & Rasbash, 1994; Huitema & McKean, 2007; Verbeke, Lessaffre, & Brant, 1998). Repeated measures with a continuous time setting can thus be viewed as short time series especially when applied to growth models. Fundamentally the use of the term time series is just to emphasize what is observed in practice during repeated measurement investigations (Diggle, 1990). Short time series are often present in medicine where systolic and diastolic blood pressure repeated observations are used for example to assess the effectiveness of a hypertension drug. However, time series and their applications have been very rarely used in the field of educational effectiveness research.

Most authors use multilevel growth curve models because of their elegance and flexibility to model observations from individuals taken at different sets of time points (Van den Noortgate & Onghena, 2006; Cools, Van den Noortgate & Onghena, 2008). For example in growth studies for student's academic achievements, observations may be taken more often in the earlier classes than in the later classes. Say at the beginning of year 1, end of year 1, end of year 2, end of year 4 and end of year 6. However, researchers have failed very often to model level 1 variance due to serially correlation. The standard assumption frequently used by these researchers using growth curve models for repeated measurement data is that the within-subject residuals are serially independent or uncorrelated. This assumption fails very often especially in cases where such repeated measurements are made close together and even more when unequally spaced. There is even a high chance of always finding serial dependence between repeated measurements. Such a dependency can be expressed in terms of serial correlation functions. Growth curve models with serial correlation have been used often in time series analysis of longitudinal data in the field of economics, medicine and geostatistics. However, very little is known about such serial correlations in educational effectiveness research even for the simple two-level models. Therefore to further extend the two-level models to three-level growth curve models stand out as an even greater challenge. This challenge comes in because of the possibilities of serial correlation at higher levels other than level 1 which until now have never been investigated.

In any educational process where time is required to acquire knowledge, current knowledge is built on the previous. Such an influence of previous knowledge on the current, will certainly present itself in the structure of the repeated measurement data and consequently into the growth model errors. Goldstein, Healy, and Rasbash (1994) have indicated the possibility of correlated level 1 residuals for multilevel models by considering autocorrelated models for both discrete and continuous time. In their work, they illustrate this fact only for level 1 residuals though in their discussion they indicate the possibility of further research into models with higher level serial correlation using the example of repeated measurements nested within students who are in turn nested within schools. Their discussion on such an autocorrelation will mean correlated level 2 structures with the standard independence assumption at level 1. Maas and Snijders (2003) have indicated the need for other families of covariance matrices with structures more complex than the common compound symmetry model but less than the complete unstructured model. The use of autoregressive model or order 1 for equally time spaced data has also been mentioned by a number of these researchers (Goldstein et al., 1994; Mass & Snijders, 2003; Snijders & Bosker, 1999).

The main purpose of this dissertation is firstly to introduce a serial correlation method to multilevel (three and more levels) growth curve models for school effectiveness research and apply it to real data to show how it impacts the estimates of school effects. Secondly, this method will be extended to handle level 2 serial correlation functions in the special case of three-level data to show the implications of extending from the common two level models discussed in the literature (Little et al., 2006; Mass & Snijders, 2003; Snijders & Bosker, 199). This method will then be used to demonstrate for the first time the complex structure of serial correlation at student and school levels and how this affects the estimates of the school effects. This gap in the current literature can be a good starting point, considering that it is very common to find some serial dependency among level 1 observations (measurement occasions) in multilevel longitudinal data. The term "level 1 serial correlation" is used when referring to the correlation of the repeated measurements of the students. Similarly, the "level 2 level serial correlation" refers to the serial correlation of the average school repeated measurements over time.

2. Research questions

This paper introduces a method of detecting the presence and the type of serial correlation in multilevel growth curve models by drawing inspiration from authors like Diggle (1990), Diggle, Liang, and Zeger (1994), and Verbeke and Molenberghs (2000). In order to inculcate this method in educational effectiveness research, the article shows applications of this statistical method to school effectiveness data and finally goes on to give answers to the following three main research questions.

1) Is it possible to find serial correlation in repeated measurements of students' outcomes in educational effectiveness research (at level 1 or level 2)?
2) What is the best functional form that can be used to describe the serial correlation process when it exists?
3) How are the school effects on students' status and students' growth affected in the presence of serial correlation?

3. Method

3.1. Three sources of random variability

School effect estimates are obtained from the estimates of the maximum likelihood or restricted maximum likelihood estimates of the variances at the different levels. This implies errors in the estimation of these variances will be carried over to the school effect estimation. There are at least three possible types of random variation that can be of great interest in any longitudinal study (Diggle, Liang, & Zeger, 1994; Verbeke & Molenberghs, 2000).

The first type of random effect is stochastic variation between units (schools or students), resulting from the fact that the sample units are drawn at random from a population of such units. The second type of variation is due to serial correlation, which is the part of any unit's (school or student) observed measurement profile as a result of time-varying stochastic processes operating within that unit. For example students well-being measured sequentially will reflect to some extent biological, psychological, and emotional processes operating within them (Diggle, Liang & Zeger, 1994; Verbeke & Molenberghs, 2000). This type of stochastic variation results in correlation between pairs of measurements on the same unit which in turn depends on the time separation between the pairs of measurements. Typically, this correlation is a decreasing function of the time separation between these measurements (horizontal axis) and the variance attributed to this correlation is represented by the height of the curve in Figure 1. The third and last source of random variation is the measurement error, which is due to the measurement process itself. A summary of these three sources of random variation showing how the total variance is partitioned is given in Figure 1.

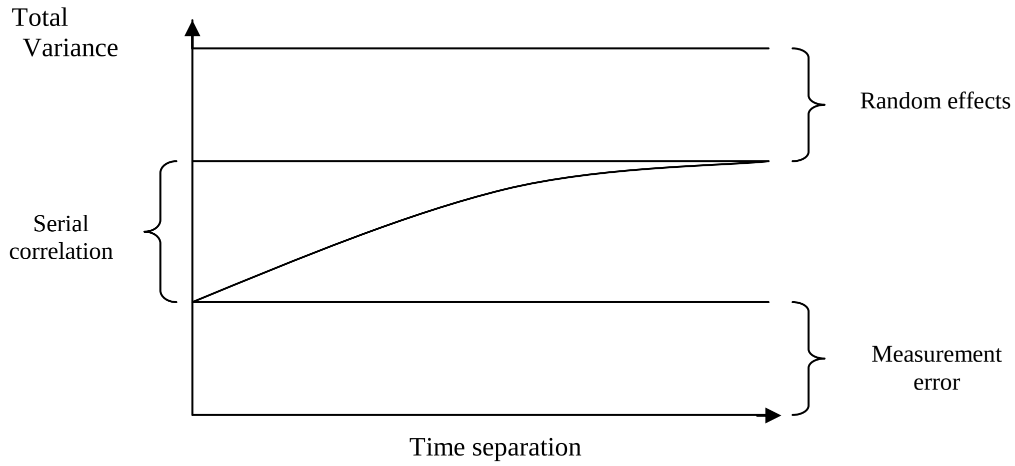


Figure 1. A hypothetical variogram showing the partitioning of total variance into variances due to random effects, serial correlation and measurement error

Adequate modelling of the serial correlation is very essential not only to obtain valid inferences for parameters in the mean model (which for many studies in the fields of economics, medicine and biostatistics are of primary interest (Verbeke, Lessaffre, & Brant, 1998)), but also for the variance/covariance parameters which are of primary interests in school effectiveness research. The estimates of these variance components are used to obtain school effects estimates. Diggle (1990) and Diggle, Liang, and Zeger (1994), have worked on this problem of serial correlation for the case of a two-level model, with random intercept only and under stationarity conditions and they have shown that taking serial correlation into account could result in huge improvements of the model predictions.

3.2. Detecting serial correlation in multilevel growth curve models

This section proposes a method of detecting serial correlation for two-level models and then generalizes the two-level growth curve models to models with three levels or more. It begins with some background information on semi-variogram for two-level random intercepts models. Considerable attention is then devoted to the extension of these two-level random intercepts and slopes growth curve models to three-level models (Anumendem, De Fraine, Onghena, & Van Damme, in press), and finally the application of the proposed extended models to real data.

3.3. Accounting for serial correlation

A standard assumption frequently used for repeated measurement data as indicated previously, is that the within-subject residuals are serially independent or uncorrelated serially. This assumption fails very often especially in cases where such repeated measurements are made very close to each other (hourly or daily) for fast changing responses like body temperature, well-being and blood pressure. But such correlation can also be very frequent in repeated measurement made in months and years for slowly changing responses, like children's growth in height and children's academic growth in language or mathematics achievement. Once the presence of serial correlation is acknowledged, any statistical method used has to take this into account. The model proposed in this light is given by equation 1 representing a three-level growth curve model with a serial correlation at level 1 or level 2.

3.4. Three-level serially correlated growth curve model

A proposed multilevel linear mixed model (multilevel growth curve model) extended to include level 1 serial correlation is given below

$$\begin{cases} \mathbf{Y}_i = X_i\boldsymbol{\delta} + T_i\boldsymbol{\nu}_i + Z_i\boldsymbol{\upsilon}_i + \boldsymbol{\varepsilon}_{(s)i} + \boldsymbol{\varepsilon}_{(e)i} \\ \boldsymbol{\nu}_i \sim N(\mathbf{0}, \Psi_1), \\ \boldsymbol{\upsilon}_i \sim N(\mathbf{0}, \Psi_2), \\ \boldsymbol{\varepsilon}_{(s)i} \sim N(\mathbf{0}, \tau_1^2 H_i) \\ \boldsymbol{\varepsilon}_{(e)i} \sim N(\mathbf{0}, \sigma^2 I_{n_i}) \\ \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_N, \boldsymbol{\upsilon}_1, \ldots, \boldsymbol{\upsilon}_N, \boldsymbol{\varepsilon}_{(s)1}, \ldots, \boldsymbol{\varepsilon}_{(s)N}, \boldsymbol{\varepsilon}_{(e)1}, \ldots, \boldsymbol{\varepsilon}_{(e)N} \ \textit{independent} \end{cases} \tag{1}$$

Where $\mathbf{Y}_i$ is the n_i-dimensional response or outcome vector for school i, and $1 \le i \le N$, with N being the total number of schools and n_i taking values like $n_1, \ldots, n_N$ with n_1 and n_N being the dimensions of the response vectors for school 1 and school N respectively. Moreover, X_i, T_i and Z_i are $(n_i \times p), (n_i \times q)$ and $(n_i \times q')$ dimensional matrices of known covariates respectively. $\boldsymbol{\delta}$ is a p-dimensional vector containing fixed effects. $\boldsymbol{\nu}_i$ is the q-dimensional vector containing school random effects, $\boldsymbol{\upsilon}_i$ is the q'-dimensional vector containing student random effects and $\boldsymbol{\varepsilon}_{(e)i}$ is the n_i-dimensional vector of residual components. For covariance matrices, Ψ_1 is a general $(q \times q)$ covariance symmetric matrix, Ψ_2 is a general $(q' \times q')$ covariance symmetric matrix. The time series assumption is usually that the serial effect of $\boldsymbol{\varepsilon}_{(s)i}$ (or $\boldsymbol{\upsilon}_{(s)i}$ for level 2) is a population phenomenon, independent of the students (or schools). The serial correlation matrix H_i is assumed to depend only on i through its dimension n_i and through the measurement occasions t_{ijk}. It is assumed further that the (k, k') element $h_{ijkk'}$ of H_i is modelled as $h_{ijkk'} = g(|t_{ijk} - t_{ijk'}|)$ for some decreasing function $g(\bullet)$ with $g(0) = 1$. Examples of such decreasing functions include $g(u) = \phi^{-u}$, $g(u) = \exp(-\phi u)$, and $g(u) = \exp(-\phi u^2)$ which are the power, exponential and Gaussian serial correlation functions respectively. This implies that the correlation between $\varepsilon_{(s)ijk}$ and $\varepsilon_{(s)ijk'}$ only depends on the time interval between response y_{ijk} and $y_{ijk'}$ and decreases with increases in this interval. The serial correlation function depends on a pair of times and only under the assumption of stationarity does this pair of times simplify to the time lag (Box, Jenkins, & Reinsel, 2008).

Serial correlations can be modelled by using flexible parametric models as proposed by Lessaffre, Asefa, and Verbeke (1999), in which they assume a parametric form for the correlation function by using fractional polynomials. This paper considers the empirical semi-variogram (nonparametric technique) described later because it does not require fitting of the linear mixed model itself.

4. Semi-variogram theoretical framework

4.1. Random intercepts multilevel growth curve models

Historically, the semi-variogram (also frequently called variogram in some literature (Diggle, 1990; Diggle, Liang & Zeger, 1994)) has been widely used in spatial statistics to identify the covariance structure in geostatistical data (Little et al., 2006). Diggle (1988) first introduced it for the case of a random intercept linear mixed model. This was later extended by Verbeke, Lesaffre, and Brant (1998) to situations containing not only random intercepts (models with nonconstant variance) but also random slopes. The semi-variogram applied to longitudinal data, can be seen as one-half the expected squared difference between residuals obtained on the same subject. A semi-variogram can be used as a diagnostic tool and can also suggest appropriate models for the covariance (Fritzmaurice, Laird, & Ware, 2004).

A semi-variogram is especially used to describe the association among repeated values and easily estimated with irregular observation times (Diggle, 1990). In this, an estimate of the semi-variogram also called sample variogram (Diggle, Liang, & Zeger, 1994) is obtained from smoothing the scatter plot of the $\sum_{i=1}^{N} \frac{n_i(n_i - 1)}{2}$ half-square differences $V_{ijk} = \frac{(r_{ij} - r_{ik})^2}{2}$ between pairs of residuals within subjects ($r_{ij} - r_{ik}$) versus the corresponding lags $U_{ijk} = |t_{ij} - t_{ik}|$ at measurement occasions j and k for subject i. The n_i are the responses or observations of subject i with N being the total number of subjects. The semi-variogram is estimated from the data as the plot of the half-squared differences (V_{ijk}) and the time lags

(U_{ijk}) by fitting a non-parametric curve and denoted by $\hat{V}_{ijk}(U_{ijk})$. Usually LOESS (locally weighted scatterplot smoothing) is used because of its great combination of classical methods like linear and nonlinear regression. Loess smoothing combines both the simplicity of linear least squares regression and the flexibility of nonlinear regression. This is done by fitting simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data (Cleveland & Devlin, 1988).

In a random intercepts model (in which random intercepts are the only random effects), the marginal covariance matrix for a three-level multilevel model is given for two separate model fit by Equations 2 and 3 for level 1 and level 2 serial correlations respectively.

$$v^2 J_{n_i} + u^2 J_{n_i} + \tau_1^2 H_i + \sigma^2 I_{n_i} \qquad (2)$$

$$v^2 J_{n_i} + u^2 J_{n_i} + \tau_2^2 H_i + \sigma^2 I_{n_i} \qquad (3)$$

The notations are described earlier on in detail except for J_{n_i} which is $(n_i \times n_i)$ and containing only ones while v^2 and u^2 denote the variances of the school level random intercepts and student level random intercepts respectively. The variances due to the level 1 and level 2 serial correlations are represented by τ_1^2 and τ_2^2 respectively. This implies that the residuals r_{ijk} of the jth student of the ith school at the kth measurement occasion have constant variances, $v^2 + u^2 + \tau_1^2 + \sigma^2$ and $v^2 + u^2 + \tau_2^2 + \sigma^2$ with serial correlation variances at the student level (τ_1^2), and school level (τ_2^2).

Level 1 Serial Correlation

The correlation between two residuals r_{ijk} and $r_{ijk'}$ from the same student j from school i at measurement occasions k and k', is given by

$$\rho(|r_{ijk} - r_{ijk'}|) = \frac{u^2 + \tau_1^2 g(|r_{ijk} - r_{ijk'}|)}{v^2 + u^2 + \tau_1^2 + \sigma^2} \qquad (4)$$

Level 2 Serial Correlation

A similar correlation can be obtained between two residuals r_{ik} and $r_{ik'}$ from the same school i at measurement occasions k and k' is

$$\rho(|r_{ik} - r_{ik'}|) = \frac{v^2 + \tau_2^2 g(|r_{ik} - r_{ik'}|)}{v^2 + u^2 + \tau_2^2 + \sigma^2} \tag{5}$$

4.2. Random intercepts and slopes multilevel growth curve models

When growth curve models with random slopes in addition to the random intercepts are used to model the growth process, introducing serial correlation to the level 1 residuals requires complex statistical methods and software. This paper discusses and elaborates on some of the methods already mentioned in the literature for random intercepts only, while showing how they can be extended and applied to school effectiveness research under a number of conditions. It has often been shown in other fields of research that in growth curve models with random intercepts and slopes, the covariance structure of Equations 2 and 3 is usually dominated by its first two components. This implies that to efficiently study the random effects, it is necessary to first remove the variability in the data due to the random effects at both the school and student level. The residuals obtained can then be explored using the empirical semi-variogram as illustrated previously. There are two possibilities described in this paper on how to go about this exploration of residuals: examining subject-specific residuals and examining transformed residuals.

The first, is looking at subject specific residuals $\mathbf{Y}_i - X_i\hat{\boldsymbol{\delta}} - T_i\hat{\mathbf{v}}_i - Z_i\hat{\boldsymbol{\upsilon}}_i$, in which $\hat{\mathbf{v}}_i = E(\hat{\mathbf{v}}_i | \hat{\boldsymbol{\upsilon}}_i, \mathbf{Y}_i)$ and $\hat{\boldsymbol{\upsilon}}_i = E(\hat{\boldsymbol{\upsilon}}_i | \mathbf{Y}_i)$ are the empirical Bayes estimates for the $\mathbf{v}_i$ and $\boldsymbol{\upsilon}_i$ obtained by fitting a specific multilevel growth curve model. The main drawbacks of this approach are the strong dependence of the $\mathbf{v}_i$ and $\boldsymbol{\upsilon}_i$ on the normality assumption specified by the model on the random effects and also the form of the variance covariance structure imposed on Equations 2 and 3. This simply means wrong normality assumptions will have great repercussions on the preceding analyses. Caution must therefore be taken when exploring the functional form of the subject-specific residuals if an assumed covariance function had been used previously because this could indirectly contaminate the serial correlation if present. Despite the above shortcomings, in educational effectiveness research one usually encounters huge amount of data, reducing the chances of non-normality of the random effects. If the covariance structure is in addition left to be more general (unstructured) then one can proceed using these subject specific residuals. Morrell, Pearson, Ballentine, and Brant (1995) have used such residuals for the case of a two level nonlinear growth curve model.

The second is to look instead at transformed residuals. This was proposed by Verbeke, Lesaffre, and Brant (1998) and which also serves as a simple informal check for the need of a serial component in a linear mixed model. This technique requires the transformation of ordinary least squares residuals $\mathrm{r}_i = \mathbf{Y}_i - X_i\hat{\boldsymbol{\delta}}_{\mathrm{OLS}}$ based on some presumed mean structure and ignoring any dependence among the repeated measurements. The transformation process entails projecting the ordinary least squares residuals orthogonally to the columns of T_i for the school level random effects and Z_i for student level random effects. In this case the transformed residuals are left only with variability not explained by the random effects. This second method is very complex and thus only really makes a big gain when applied to small or very moderate datasets and in which restrictions have been put on the covariance structure of the random effects.

5. Application

5.1. Data

The data that will be considered in this application stem from the LOSO research conducted on secondary school students from the academic year 1990-1991 to the academic year 1996-1997 (Van Damme et al., 2006). In this project, over 6000 students from about 57 schools in Flanders (Belgium) are followed during secondary education. Two student outcomes are studied in this paper: Dutch achievement (the mother tongue of most students) and student well-being at school. Achievement in Dutch was measured via curriculum-based tests administered at five different measurement occasions: at the start of first year (Grade 7), and at end of first year (Grade 7), second year (Grade 8), fourth year (Grade 10) and sixth year (Grade 12). A common scale for the Dutch language scores for the five measurement occasions was obtained using Item Response Theory. The questionnaire on student well-being was also administered four times during secondary school: at the end of Grade 7, Grade 8, Grade 10 and Grade 12. However, for comparison purposes, only four of the measurement occasions of the language achievement were used just as for well-being. Student well-being at school is measured on a 32-item scale. Two samples (3788 students from 53 schools are used for well-being and 3311 students from 52 schools are used for Dutch language achievement) are studied for the application. The student samples include only students who did not change schools and students who did not repeat a grade. The sample is also restricted to those students who have measurements at least three time points in other to be able to estimate individual student growth curves. School effects in all the models will be estimated at the end of year 1(grade 7).

SAS 9.1 (SAS Institute Inc., 2003) and its PROC MIXED is used because it is very flexible and suitable for fitting multilevel models or hierarchical linear models, and growth curve models (Singer, 1998). MLwiN 2.02 (Rasbash et al., 2000) was also used because it provides a wide range of multilevel models together with plotting, diagnostic and data manipulation facilities.

5.2. Comparing models with and without serial correlation

The first research question on whether it is possible to find serial correlation in repeated measurements of students' outcomes in educational effectiveness research is dealt with by considering three growth models. A three-level quadratic growth curve model is applied to data in which measurements are nested within students and students are in turn nested within schools. The three models considered are: a model without serial correlation (model A), a model with serial correlation at level 1 (model B) and a model with serial correlation at level 2 (model C). The first model (model

A) is the most frequently used and makes a very strong assumption about the nature of the residuals. The standard assumption being that within subject residuals are serially uncorrelated. In model B, this assumption will be tested by modelling a serial correlation at level 1, thus modelling how the within-student residuals are serially correlated. In this paper the verification of a serial correlation at level 2 (model C) will be a step forward in the field of educational effectiveness in general and the analysis of repeated measurement multilevel data in particular. In model C, the serial correlation of the within-school residuals will be modelled.

The three models described earlier with a random intercept only are fitted to determine whether or not serial correlation exists in the data. Using the results obtained, the verification of the presence of serial correlation is done by taking the difference between -2 times the log of restricted maximum likelihood estimates of model A and model B which is 414.4 (with a variance due to serial correlation of 0.084). The difference is then compared to a chi-square distribution with 2 degrees of freedom and seen to be clearly significant at a 5% significance level. This is an indication that the amount of serial correlation between the repeated measurements of the students cannot be neglected and this paper admonishes researchers to consider correcting for this in their models. A similar conclusion is drawn when comparing model A to model C. The difference in their -2 times log of restricted maximum likelihood values is 200.8 with 2 additional parameters (with a variance due to serial correlation of 0.01). This is still an insurmountable difference when considered with 2 degrees of freedom though less than the serial correlation at level 1. This indicates once again that average repeated measurements of the school are also strongly serially correlated. It is also worth noting that the fixed effects parameter estimates for both model A and B are almost identical but for model C there are some slight differences in parameters as well as their standard errors. Thus taking serial correlation into account will not only be useful for a true interpretation of the random variability in the data but also to obtain valid inferences for the growth profiles (student growth as well as school growth). The same three models A, B and C this time with a random intercept and a random slope, are once again fitted to the data and the results compared. Comparing the fits of models A to B and A to C indicates deviance values of 127.6 (2 degrees of freedom) and 102.1 (2 degrees of freedom) respectively.

The results of the students' language achievement not presented, indicate that the difference in -2loglikelihood for models A and B with a random intercept only is 62.8 (2 degrees of freedom). This is significant at a 5% level of significance though with a smaller difference in comparison with the same calculations for the students' well-being outcome. Comparing model A and C for the students' language achievement shows a difference of 319.9 (2 degrees of freedom). For the case of a model with random intercepts and random slopes, the differences between models A and B and models A and C are 41.1 (2 degrees of freedom) and 159.1 (2 degrees of freedom) respectively, and are both significant at a 5% level of significance.

5.3. Serial correlation functions

As indicated earlier, the repeated measurement framework of time series is usually made of many more subjects in comparison to the number of measurement occasions. In this case the amount of information available to describe the shape of the serial correlation is limited thus giving ascendancy over the search of serial correlation functions to a set of plausible functions instead of looking at an infinite set of these functions. The consideration of a plausible set of functions in a sense is much more practical and easier to implement with available software. In SAS Proc Mixed, such functions would have to be positive definite to ensure that matrices created with entries of this functional forms are also positive definite (Littel et al., 2006). Three of these types of functions whose correlation asymptotically dies to zero will be considered in this application: the spatial power, exponential and Gaussian models of serial correlations. However, some researchers have settled down to only the exponential and Gaussian serial correlations given that the behaviour of spatial power and exponential are in general similar (Verbeke, Lesaffre, & Brant, 1998; Verbeke & Molenberghs, 2000).

Power function *Exponential function*

$$\sigma^2\begin{pmatrix} 1 & \rho^{l_{12}} & \rho^{l_{14}} & \rho^{l_{16}} \\ \rho^{l_{12}} & 1 & \rho^{l_{24}} & \rho^{l_{26}} \\ \rho^{l_{14}} & \rho^{l_{24}} & 1 & \rho^{l_{46}} \\ \rho^{l_{16}} & \rho^{l_{26}} & \rho^{l_{46}} & 1 \end{pmatrix} \qquad \sigma^2\begin{pmatrix} 1 & \exp(-l_{12}/\rho) & \exp(-l_{14}/\rho) & \exp(-l_{16}/\rho) \\ \exp(-l_{12}/\rho) & 1 & \exp(-l_{24}/\rho) & \exp(-l_{26}/\rho) \\ \exp(-l_{14}/\rho) & \exp(-l_{24}/\rho) & 1 & \exp(-l_{46}/\rho) \\ \exp(-l_{16}/\rho) & \exp(-l_{26}/\rho) & \exp(-l_{46}/\rho) & 1 \end{pmatrix}$$

Gaussian function

$$\sigma^2\begin{pmatrix} 1 & \exp(-l_{12}^2/\rho^2) & \exp(-l_{14}^2/\rho^2) & \exp(-l_{16}^2/\rho^2) \\ \exp(-l_{12}^2/\rho^2) & 1 & \exp(-l_{24}^2/\rho^2) & \exp(-l_{26}^2/\rho^2) \\ \exp(-l_{14}^2/\rho^2) & \exp(-l_{24}^2/\rho^2) & 1 & \exp(-l_{46}^2/\rho^2) \\ \exp(-l_{16}^2/\rho^2) & \exp(-l_{26}^2/\rho^2) & \exp(-l_{46}^2/\rho^2) & 1 \end{pmatrix}$$

where σ^2 is the variance attributed to the particular serial correlation function considered and ρ represents how fast the correlations decrease as a function of the time lags l_{ij}. The time lags ($l_{12}, l_{24} = l_{46}, l_{14}, l_{26}, l_{16}$) are estimated from the data easily to be (1, 2, 3, 4, 5) thus reducing the parameters to be estimated to σ^2 and ρ.

The need for improving statistical acumen in the modelling of growth curves models in educational effectiveness research is inevitable. Therefore the semi-variogram proposed for checking serial correlation in multilevel longitudinal data could become a very essential tool available to researchers in the future. The modified version of the semi-variogram for three-level data will be applied in order to investigate the possibility of distinguishing which of these functions best represent the serial correlation structure in the data. Figures 2a and 2b showing two simulated representative samples of the student well-being, for three different serial correlation functions (g(u)) for random intercepts only with serial correlations of with rates of decrease of 0.5 and 2 respectively. It is clear from Figure 2a that the semi-variogram will find it very difficult to make a clear cut demarcation between the Gaussian and exponential functions. In any case Figure 2b shows that possibility of making the difference between the Gaussian and exponential function when the rate of decrease is 2. The figures will guide our choice of serial correlation function when the semi-variogram is obtained. With as few as four measurement occasions, the number of time lags greatly depends on the time spacing between the measurement occasions. For example four measurements at equally spaced times say, year 1, 2, 3, and 4 will produce only three time lags (1, 2, and 3) whereas four measurements at unequally spaced times 1, 2, 4, and 6 produces five time lags (1, 2, 3, 4, and 5). It is thus better to get four unequally spaced times measurements than get six equally spaced times when researching on serial correlation.

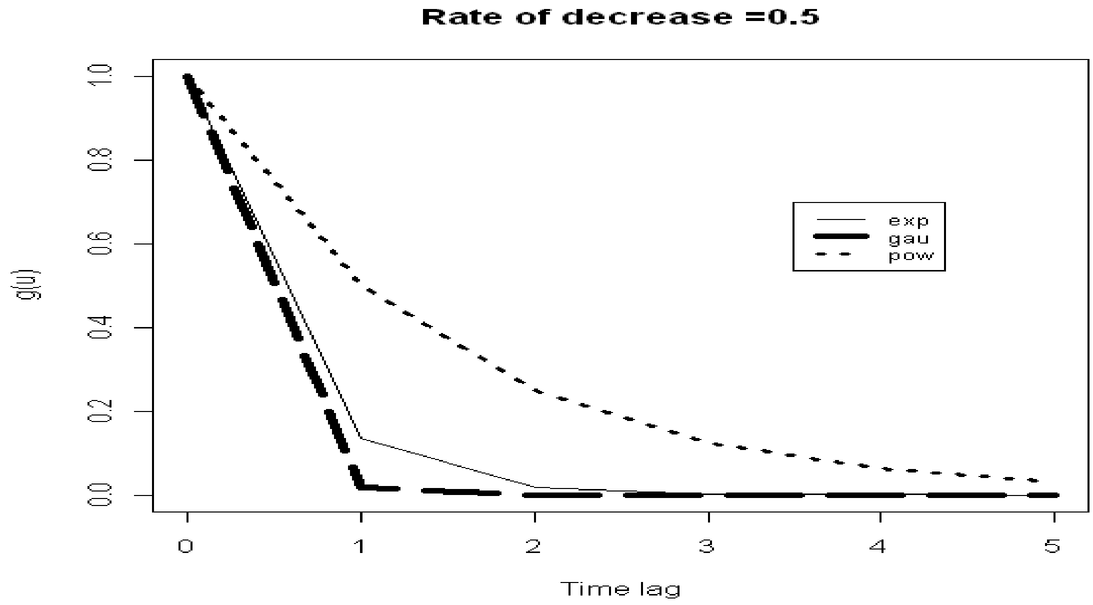


Figure 2a. Serial correlation with rate decrease of 0.5

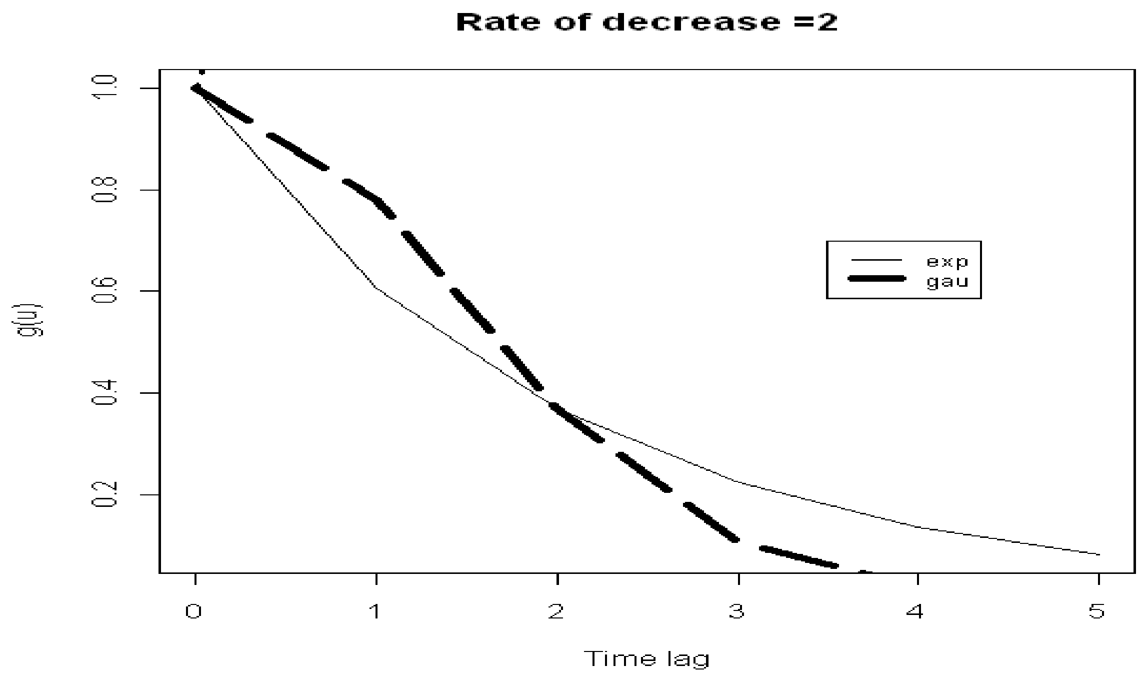


Figure 2b. Serial correlation with rate of decrease of 2

The graphs of three different serial correlation functions corresponding to the three functional forms are shown in Figures 2a and 2b, simulated with properties of the students' well-being outcome. The log likelihood values point to the indistinguishable fact of these three functions at level 1. The three serial correlation functions show a difference of 0.008% which is no doubt an inane difference. Looking at the three serial correlation functions for level 2, the results indicate a difference of 0.1% which is still very small but higher than at level 1. However, the distinction between the exponential and power serial correlations still remains practically impossible using deviance statistics because they all show the same model fit statistics for both the random intercepts model and the random intercepts and slopes model.

Looking only at the results of the deviance statistics, will indicate that the three-level growth curve models with serial correction at level 1 do not depend very much on the type of serial correlation function used. The differences between the -2loglikelihood values are almost negligible possibly because of the few number of measurement occasions present in the data. The -2loglikelihood values for the spatial power and exponential functions remain identical confirming why other authors have restricted their search to exponential and Gaussian serial correlations only. When serial correlation is included at level 2, the difference between the loglikelihood values for the Gaussian serial correlation and spatial power (or exponential) increases but still not as much as to call for too much attention.

While the difference between the plausible serial correlation functions is relatively small, the level 1 serial correlation correction for the random intercepts model, using the Gaussian serial correlation function shows the largest value of the -2loglikelihood. For level 2 corrections, the exponential and spatial power functions larger values of the -2loglikelihood. Similar arguments can be put forward when looking at the results obtained for the students' language achievement outcome. The Gaussian function seems better for level 1 serial correlation while exponential and spatial power result in a better fit for level 2 serial correlation corrections.

5.4. *Application of semi-variogram*

Figures 3 and 4 show semi-variograms obtained by using subject-specific residuals (students and schools) with student random effects only and with school random effects only respectively. It was earlier explained that such subject-specific residuals will only be valid for large datasets and unstructured variance covariance matrices. Note that these plots in Figures 3 and 4 represent the variance plotted against time lag because they are easily estimated from data which are unequally spaced in time instead of correlation plots. Correlation plots in any case will be quite similar only that they will instead be decaying from left to right as the lags increase as seen in Figures 2a and 2b. The two plots both indicate an indubitable presence of a serial correlation either at level 1 or level 2 as was illustrated earlier on in Figure 1 (total variability partition). The semi-variograms also point out the fact that a decaying function could as well be used to represent the serial correlation. It was also indicated earlier on that with few measurement occasions, the shape of the serial correlation cannot be captured completely. Hence, the semi-variogram estimated for this particular example cannot make a good distinction between a Gaussian serial correlation and an exponential serial correlation functions. However, it does distinguish clearly the power function from the other two functions.

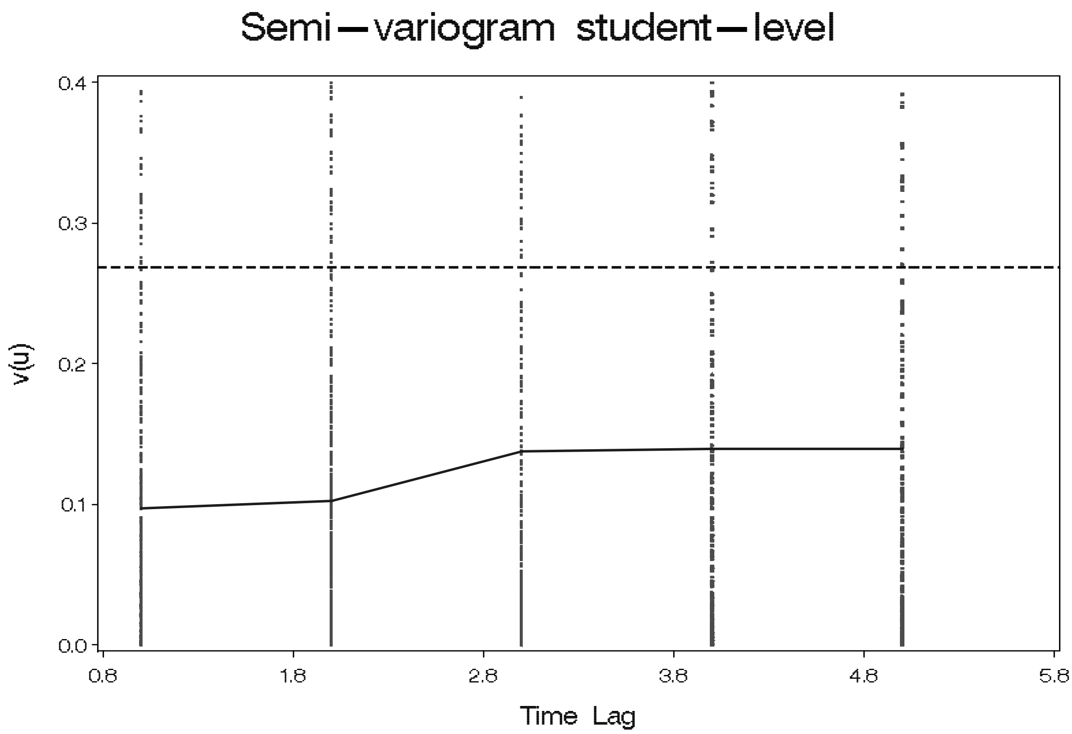


Figure 3. Semi-variogram of Level 1 serial correlation

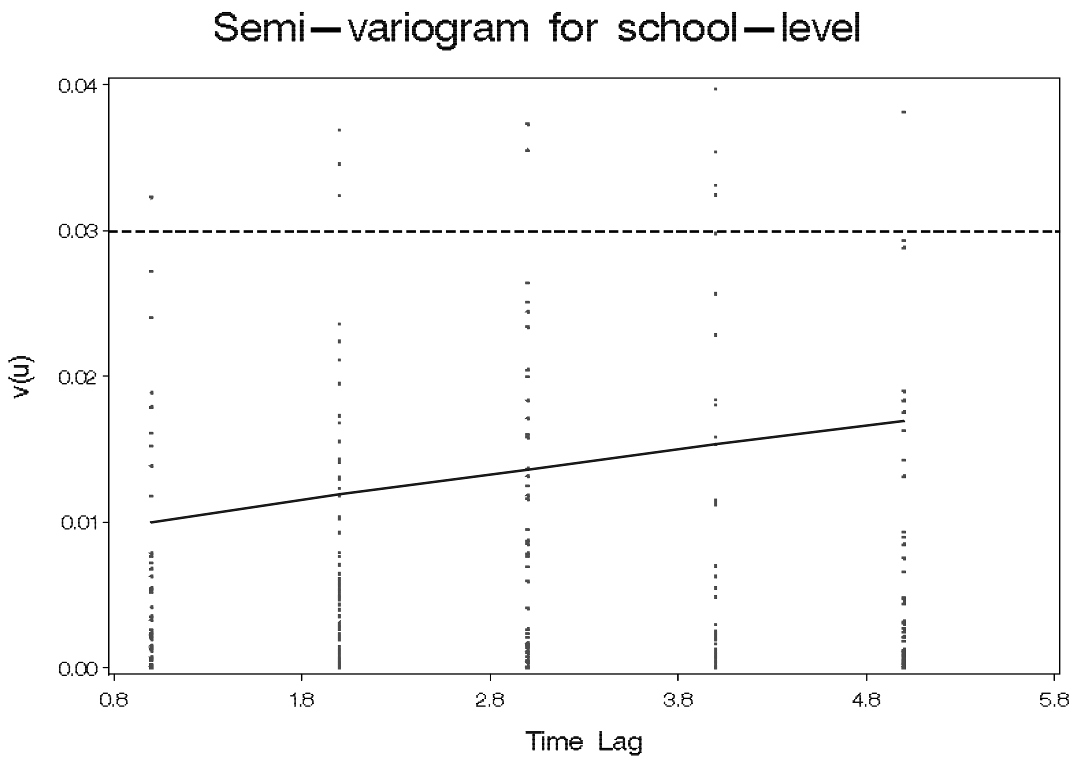


Figure 4. Semi-variogram of Level 2 serial correlation

5.5. *School effects estimates with serial correlation correction*

After finding that the data are serially correlated, the next step is to estimate the school effects taking this correlation into account. The Gaussian serial correlation was considered for the measurement occasions correlation though the results indicated by the semi-variogram of Figure 3. The plot of Figure 4 shows that the power function will be the best functional form for the serial correlation at level 2. The results presented in Tables 1 and 2 for the students' well-being and language achievement reveal that serial correlation correction at level 1 can result in a larger school effect estimate for the student well-being (3.3% to 4.8%) and language achievement (41% to 48%). Correction of serial correlation at level 2 does show a different pattern for the well-being outcome as for the language achievement outcome. The school effect for the status of the students' well-being when level 2 serial correlation is taken into account, reduces from 3.3% to 2.3% while for the language achievement, it increases from 41.5% to about 43%.

Table 1

Random effects estimates of students' well-being for the model A, B and C with random intercepts only and serial correlation

parameters	Without serial correlation	Level 1 serial Correlation(gau)	Level 2 serial Correlation(gau)
Variance			
School level intercept	0.009(0.0026)*	0.009(0.0025)*	0.006(0.0028)*
Student level intercept	0.146(0.0044)*	0.095(0.0077)*	0.147(0.0044)*
Residual	0.119(0.0018)*	0.082(0.0024)*	0.114(0.0018)*
Serial correlation			
School serial	/	/	0.007(0.0018)*
Student serial	/	0.084(0.0065)*	/
School effect			
Student status	3.3%	4.8%	2.3%

* Significant at 5% level of significance

Table 2

Random effects estimates of students' language achievement for the model A, B and C with random intercepts only and serial correlation

Parameters	Without serial correlation	Level 1 serial Correlation(pow)	Level 2 serial Correlation(gau)
Variance			
School level intercept	0.306(0.0636)*	0.304(0.0630)*	0.316(0.0676)*
Student level intercept	0.241(0.0075)*	0.234(0.0088)*	0.243(0.0074)*
Residual	0.191(0.0030)*	0.096(0.0089)*	0.179(0.0028)*
Serial correlation			
School serial	/	/	0.0225(0.0050)*
Student serial	/	0.112(0.033)*	/
School effect			
Student status	41.5%	47.9%	42.8%

* Significant at 5% level of significance

As earlier indicated, school effectiveness research also seeks estimates of school effects on students' growth. In this light, quadratic growth curve models with random intercepts and slopes are considered to enable the estimation of school effects on students' growth. The same models are then reconsidered twice, once with level 1 serial correlation correction and once with level 2 serial correlation. The results of these models for the students' well-being and students' language achievement are presented in Tables 3 and 4 respectively.

Table 3

Random effects estimates of students' well-being for the model A, B and C with random intercepts & slopes and serial correlation

	Without serial correlation	Level 1 serial correlation(gau)	Level 2 serial Correlation(gau)
Variances			
School level intercept	0.012(0.0035)*	0.01(0.0029)*	0.002(0.0057)*
School level slope	0.001(0.0002)*	0.0007(0.0002)*	0.0001(0.0003)*
Student level intercept	0.217(0.0079)*	0.110(0.0079)*	0.221(0.0079)*
Student level slope	0.004(0.0004)*	0.00001(--)	0.0047(0.0003)*
Residual	0.0970(0.0019)*	0.0820(0.0025)*	0.0940(0.0018)*
School effect			
Student status	5.2%	8.3%	0.9%
Student growth	20%	98.5%	2.1%

* Significant at 5% level of significance

Table 4.

Random effects estimates on students' language achievement for the model A, B and C with random intercepts & slopes and serial correlation

	Without serial correlation	Level 1 serial correlation(gau)	Level 2 serial correlation(pow)
Variances			
School level intercept	0.263(0.0551)*	0.265(0.0556)*	0.232(0.0508)*
School level slope	0.002(0.0005)*	0.002(0.0005)*	0.002(0.0006)*
Student level intercept	0.196(0.0080)*	0.168(0.0084)*	0.189(0.0098)*
Student level slope	0.002(0.0005)*	0.00001(--)	0.003(0.0005)*
Residual	0.175(0.0034)*	0.042(0.086)*	0.166(0.0032)*
School effect			
Student status	57.3%	61.2%	55.1%
Student growth	50.0%	99.5%	40.0%

* Significant at 5% level of significance

Results in Table 3 show that with a level 1 serial correlation, the school effect on both the students' status and growth increase from 5.2% to 8.3% and 20% to 98.5% respectively. The interesting fact about these results seems to be the greater increase on the students' growth than status. It is worth noting that with level 1 serial correlation correction, the variance of the student level linear slope is very small and just approximated to 0.00001 without any estimates for the standard errors. With serial correlation at level 2, the school effects on the students' status and growth both decrease drastically with little or no school effects (0.9%) on the students' status at the end of year 1 (Grade 7) but with a much higher though still small school effect on the students' growth in well-being (2.1%). Table 4 shows that the school effect on students' status and growth in language achievement increase with serial correlation considered at level 1, from 57.3% to 61.2% and 50% to 99.5% respectively. The school effect on students' status and growth in language achievement both decrease with level 2 serial correlation though far less than for the students' well-being.

6. Discussion and Conclusion

Researchers in educational effectiveness research generally split up total variance into two components that is the measurement error and random effects of the different levels present in the data. The level 1 variance is thus considered to be the error variance and the rest considered to be random effects due to random samples (groups) present. This is often done under the assumption of independent level 1 residuals. The results of this study show that this assumption is in general not realistic when working with longitudinal data. It illustrates that there is a very important serial correlation component left out and how incorporating this component can greatly improve the accuracy of the models and the estimated school effects. Analysis on repeated measurements with more than three times will very often show some amount of serial correlation and it is thus necessary to check the data and take the required measures when need be.

The results show that more often than not, repeated measurement data will be serially correlated which is in line with the findings by early researches in longitudinal data. Diggle et al. (1994) have indicated that the presence of random effects other than random intercepts (e.g. random linear slopes) in addition to measurement error will inundate the effect of serial correlation. This idea made them to limit their study to level 1 serially correlated growth curve models with random intercepts only. The findings of this article nonetheless indicate contrasting results to that of Diggle et al. (1994) in the sense that when level 1 or level 2 serial correlation corrections for random intercepts and slopes quadratic growth curves are considered, they show a great reduction in the variance of the random effects. This means the random effects do not overshadow the presence of the serial correlation always and therefore modelling only the random intercept effects cannot correct for serial correlation in growth curve models with random intercepts and slopes. The finding of this study showed in addition that all forms of serial correlation corrections considered for a three-level growth curve model with random intercept and slopes fitted better either at level 1 or level 2.

Another key finding of this article is that the use of the semi-variogram in repeated measurement data with as few as four time points (as well as time lags), is most useful for the detection of the presence of serial correlation, but less splendid for the type of correlation function present when the serial correlation is small. The results from the semi-variogram, however, still indicate that the presence of serial correlation should be considered much more seriously than the type of serial correlation function when significant but small serial correlation is present.

The three-level quadratic growth curve model considered in this article with serial correlation correction at level 1 or level 2 shows that repeated measurement data in educational effectiveness research can have serially correlated residuals. The results of this paper indicate that school effects on the student status and student growth are in general underestimated without level 1 serial correlation correction. This may be due to the overestimation of the level 1 variance. It seems that models without serial correlation correction ascribe too much variability to the level 1 random effect. Correction of level 2 serial correction seems not to show a general direction. When level 2 serial correlation is considered for a random intercept model, the school effects for the students' status in well-being reduced while that for the students' language achievement increased. This is simply an indication that the two outcomes are different stochastically in terms of variability around the students' and schools' growth profiles.

The advice to researchers of multilevel growth curves will be to include a serial correlation component in their models for educational data no matter the type of outcome. This paper focused on the correction of serial correlation at level 1 or level 2, however, further research is required to combine both levels of serial correlation in a single three-level quadratic growth curve model. This paper also advices researchers to consider unequal time spacing between measurements when they intend to do as few as four repeated measurements.

MANUSCRIPT 3: Double serial correlation for multilevel growth curve models[3]

Anumendem, D. N., Verbeke, G., De Fraine, B., Onghena, P., & Van Damme, J.

Abstract

Multilevel growth curve models for repeated measures data have become increasingly popular and stand as a flexible tool for investigating longitudinal change in students' outcome variables. In addition, these models allow the estimation of school effects on students' outcomes though making strong assumptions about the serial independence of level-1 residuals. This paper introduces a method which takes into account the serial correlation of level-1 residuals and also introduces such serial correlation at level- 2 in a complex double serial correlation multilevel growth curve model. The results of this study from both real and simulated data show a great improvement in school effects estimates compared to those that have previously been found using multilevel growth curve models without correcting for double serial correlation for both the students' status and growth criteria.

[3] Manuscript submitted for publication.

1. Introduction

Educational effectiveness research is a domain that tries to furnish answers to questions like: Do schools really differ with respect to their impact on pupils? How big is this impact? How can it be assessed? What are the factors that cause the effect? Can this effect be found in all types of educational contexts, such as in primary schools, secondary schools, and in the different countries around the world? What are the theoretical explanations behind research findings in this area (Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000)? It is of course invaluable that a criterion (if possible two or more criteria) be chosen to evaluate school effectiveness. Our study will look into two criteria for educational effectiveness: student status and student growth, with a focus on correctly estimating these effects.

The assessment of school effectiveness using student growth in outcomes has resulted in the general call for longitudinal studies of school effects (Teddlie & Reynolds, 2000). While longitudinal studies are universally accepted by researchers as a key design to study the changes in a student performance over time, it comes with a number of challenges in terms of data structure and statistical analysis. Children, for example, change with time by virtue of their experience in school, and also because the structures, functions, and compositions of the schools they attend also change with time. There are possibly other sources of this change that are neither of the children's nor the schools' making (purely stochastic) which cannot be captured by researchers but are in any case present in a way as to influence the outcome of interest. One can say that the sources and consequences of changes in student outcomes often constitute the object of study in school effectiveness research (Rowan & Denk, 1982).

School effectiveness researchers, with the exception of Raudenbush and Willms (1995) have almost always found only small school effects estimates for the students' status in different student outcomes. It is possible that the small school effects reported by many studies are a result of the use of poor statistical methods (Scheerens

& Bosker, 1997). Scheerens and Bosker have shown how school effects are underestimated when the school variance structure is misspecified or not taken into account at all.

While Ferron, Dailey, and Yi (2002) have looked at two-level models of change with misspecification of level-1 error structure, this paper looks at two levels of change in a three level model. However, it breaks the level-1 error or unexplained variability, into two components; a pure measurement error which can be corrected for misspecification and a serial correlation component which must be adequately accounted for with a proper structure. This model is then extended to include a complex serial correlation at the second level, resulting in what is called in this paper "double serial correlation" (DSC). The main interest of this paper is then to compare school effects estimates from multilevel nonlinear growth curve models with and without serial correlation correction.

Multilevel growth curve models have been considered as a class of time series by many researchers (Goldstein, Healy, & Rasbash, 1994; Verbeke & Molenberghs, 2000; Willms & Raudenbush, 1989). These models are a subset of time series models because they can be viewed as windows in a time series. In multilevel longitudinal studies, it is very common to find that there is some dependency among level-1 units with time. This dependency can be expressed in terms of serial correlation functions, which are generally positive definite. Models with dependency of level-1 units are very popular in the analysis of longitudinal data in the fields of economics, medicine and geostatistics to name a few. In any case very little is known about serial correlations in educational effectiveness research and three-level quadratic growth curve models stand as a greater challenge with the possibility of a DSC. One of the main aims of this paper is to introduce this method for three-level longitudinal data and apply it to show its impact on the estimates of school effects. Starting with level-1 and later extend the method to handle two levels of serial correlation functions in the special case of three-level growth curve models which is innovative to all fields in general. A DSC is proposed for a three-level growth curve model in the next section.

2. DSC multilevel growth curve model

There are at least three possible types of random variation in general (Diggle, Liang, & Zeger, 1994; Verbeke & Molenberghs, 2000) that can be of interest in a longitudinal study. A summary of these three will help in the understanding of later concepts. The first is random effects which are stochastic variations between units resulting from the fact that the sample units are drawn from a population of such units. In this area of research one can think of units, such as countries, educational systems, schools and the students. Assuming a three -level sampling where a sample of schools is drawn from a population of schools and within the schools, a sample of students is drawn from a population of students and these students are observed repeatedly. This means we have a school random effect and a student random effect.

The second type of random variation is serial correlation which is the part of any unit's (country, educational system, school or student) observed measurement profile which may be the result of time-varying stochastic processes operating within that unit. For example, a student's well-being measured sequentially will reflect to some extent biological, psychological and emotional processes operating within them. This type of stochastic variation results in correlation between pairs of measurements on the same unit which in turn depends on the time separation between the pairs of measurements. Typically, this correlation is a decreasing function of the time separation between these measurements. Finally the third type of random variation is the measurement error which is random variation due to the measurement process itself.

This paper proposes a multilevel growth curve model (three levels) with DSC which is an extension to the linear mixed model proposed by Verbeke and Molenberghs (2000). The extension is with respect to the number of levels in the nested data (from two to three) and the serial correlation (from level-1 only to levels -1 and -2). The resulting model can be written as

$$\begin{cases} \mathbf{Y}_i = X_i\boldsymbol{\delta} + T_i\mathbf{v}_i + Z_i\boldsymbol{\upsilon}_{ij} + \boldsymbol{\upsilon}_{(s)ij} + \boldsymbol{\varepsilon}_{(s)i} + \boldsymbol{\varepsilon}_{(e)i} \\ \mathbf{v}_i \sim N(\mathbf{0}, \Psi_1), \\ \boldsymbol{\upsilon}_{ij} \sim N(\mathbf{0}, \Psi_2), \\ \boldsymbol{\varepsilon}_{(s)i} \sim N(\mathbf{0}, \tau_1^2 H_i) \\ \boldsymbol{\upsilon}_{(s)ij} \sim N(\mathbf{0}, \tau_2^2 H'_{ij}) \\ \boldsymbol{\varepsilon}_{(e)i} \sim N(\mathbf{0}, \sigma^2 I_{n_i}) \\ \mathbf{v}_1, \ldots, \mathbf{v}_N, \boldsymbol{\upsilon}_{11}, \ldots, \boldsymbol{\upsilon}_{NM}, \boldsymbol{\upsilon}_{(s)11}, \ldots, \boldsymbol{\upsilon}_{(s)NM}, \boldsymbol{\varepsilon}_{(s)1}, \ldots, \boldsymbol{\varepsilon}_{(s)N}, \boldsymbol{\varepsilon}_{(e)1}, \ldots, \boldsymbol{\varepsilon}_{(e)N} \textit{ independent} \end{cases} \tag{1}$$

Where $\mathbf{Y}_i$ is the n_i-dimensional response or outcome vector for school i, and $1 \le i \le N$, with N being the total number of schools. n_i depends on the particular school while N is fixed for a particular study. This means n_i will take values like $n_1, \ldots, n_N$ with n_1 and n_N being the dimensions of the response vectors for school 1 and school N respectively. The dimension of the response vector per student within each school is also allowed to vary and denoted by m_{ij}. Moreover, X_i , T_i and Z_i are $(n_i \times p), (n_i \times q)$ and $(n_i \times m_{ij})$ dimensional matrices of known covariates respectively. Vector $\boldsymbol{\delta}$ is p-dimensional containing fixed effects, while $\mathbf{v}_i$ is the q-dimensional vector containing school random effects. Vector $\boldsymbol{\upsilon}_{ij}$ is the m_{ij}-dimensional vector containing student random effects and $\boldsymbol{\varepsilon}_{(e)i}$ is the n_i-dimensional vector of residual components. For covariance matrices, Ψ_1, of level-3 random effects is a general $(q \times q)$ covariance symmetric matrix, Ψ_2 is a general $(m_{ij} \times m_{ij})$ covariance symmetric matrix of level-2 random effects. The usual assumption of such a model is that the serial effects $\boldsymbol{\upsilon}_{(s)ij}$ and $\boldsymbol{\varepsilon}_{(s)i}$ are population phenomena, independent of the school and students respectively. The serial correlation matrices H_i and H'_{ij} are assumed to depend only on i and j through their dimensions n_i and m_{ij} respectively, and the measurement occasions t_{ijk}. It is assumed further that the (k, k') element $h_{ijkk'}$ of H'_{ij} is modelled as $h_{ijkk'} = g_1(|t_{ijk} - t_{ijk'}|)$ and $h_{ikk'}$ of H_i is modelled as $h_{ikk'} = g_2(|t_{ik} - t_{ik'}|)$ for some decreasing function $g(\bullet)$ with $g(0) = 1$. Examples of such decreasing functions include $g(u) = \phi^{-u}$, $g(u) = \exp(-\phi u)$, and $g(u) = \exp(-\phi u^2)$ which are the power, exponential and Gaussian serial correlation functions respectively and these functions are allowed to be specified differently at the different levels of serial correlation. The correlation between $\varepsilon_{(s)ijk}$ and $\varepsilon_{(s)ijk'}$ at level-1 only depends on the time interval between response y_{ijk} and $y_{ijk'}$ and decreases as this interval increases. Similarly the correlation between $\upsilon_{(s)ik}$ and $\upsilon_{(s)ik'}$ at level-2 only depends on the time interval between response y_{ik} and $y_{ik'}$ and decreases as this interval increases.

The serial correlation function can be specified to depend on a pair of times and only under the assumption of stationarity does this pair of times simplify to the time lag (Box & Jenkins, 1970). This kind of correlation can be modelled by using flexible parametric models as proposed by Lesaffre, Asefa, and Verbeke (1999). They assume a parametric form for the correlation function by using fractional polynomials. However it is also possible to explore residual covariance structure and consequently the serial correlation function, using the empirical semi-variogram (nonparametric technique).

3. Detecting DSC in random intercept and slope multilevel growth curve models

Historically the semi-variogram also frequently called variogram (Diggle, 1990; Diggle, Liang, & Zeger, 1994) has been widely used in spatial statistics to represent the covariance structure in geostatistical data. Diggle (1988) first introduced it for the case of a random intercept linear mixed model. The use of the term variogram or semi-variogram depends on the way the mathematical expression is perceived. This was later extended by Verbeke, Lesaffre, and Brant (1998) to situations containing not only random intercepts (constant variance) but slopes too. The semi-variogram can be used as a diagnostic tool and can also suggest appropriate models for the covariance (Fitzmaurice, Laird, & Ware, 2004).

We propose in this paper an extension of the applicability of the semi-variogram to models with more than two levels. Starting with some background knowledge of semi-variogram for random intercepts models, random slopes are then incorporated into the three-level models. Finally this proposed model is applied to real data. An advantage of the semi-variogram is that it can be used to describe the association among repeated values and easily estimated with irregular observation times (Diggle, 1990). In this case, an estimate of the semi-variogram $\hat{V}_{ijk}(U_{ijk})$ also called empirical or sample variogram (Diggle, Liang, & Zeger, 1994) is easily obtained from smoothing the scatter plot of the $\sum_{i=1}^{N}\frac{n_i(n_i-1)}{2}$ half-square differences $V_{ikk'}=\frac{(r_{ik}-r_{ik'})^2}{2}$ between pairs of residuals within subjects $(r_{ik}-r_{ik'})$ versus the corresponding lags $U_{ikk'}=|t_{ik}-t_{ik'}|$ at measurement occasions k and k' for subject i. The technique uses information on lags rather than the measurement occasions themselves. The half-squared differences are then plotted against the time lags by fitting a non-parametric curve using a loess curve. The loess function combines the simplicity of linear least squares regression with the flexibility of nonlinear regression. The loess curve is obtained by fitting simple models to localized subsets of the data and then using them to build up a function that describes the deterministic part of the variation in the data (Cleveland & Devlin, 1988).

In a random intercept model (in which only random intercepts are the random effects), the marginal covariance matrix for a three-level multilevel model is given by

$$v^2 J_{n_i} + u^2 J'_{n_i} + \tau_1^2 H_i + \tau_2^2 H'_{ij} + \sigma^2 I_{n_i} \qquad (2)$$

The notations are defined as in equation 1 above except for J_{n_i} is $(n_i \times n_i)$ matrix containing only ones and J'_{n_i} which is a blocked diagonal matrix with J_{n_i} blocks. The school level random intercepts variance and student level random intercepts variance are denoted by v^2 and u^2 respectively. This implies that the residuals r_{ijk} have a constant variance $v^2 + u^2 + \tau_1^2 + \tau_2^2 + \sigma^2$ with serial correlation at the level-1 and level-2.

3.1. Double serial correlation (DSC)

$$\rho(|t_{ijk}-t_{ijk'}|,|t_{ik}-t_{ik'}|)=\frac{u^2+\tau_1^2 g_1(|t_{ijk}-t_{ijk'}|)+\tau_2^2 g_2(|t_{ik}-t_{ik'}|)}{v^2+u^2+\tau_1^2+\tau_2^2+\sigma^2} \qquad (3)$$

If observations for level-1 units within the same level-2 units are made at the same time points with equal or unequal time spacing, then the time lags will be the same for both levels-1 and-2. These types of data are for example present in studies where a student cohort in a specific sample of schools is followed by taking test at fixed measurement occasions. And if the serial correlation function is the same for both levels then equation 3 can be simplified as follows

$$\rho(|t_{ijk}-t_{ijk'}|)=\frac{u^2+\left[\tau_1^2+\tau_2^2\right]g(|t_{ijk}-t_{ijk'}|)}{v^2+u^2+\tau_1^2+\tau_2^2+\sigma^2} \qquad (4)$$

Considering a model with double serial correlation for a three-level growth curve model with random intercepts and slopes, the marginal covariance matrix will be of the form

$$T_i\,\Psi_1 T_i^T+Z_i\,\Psi_2 Z_i^T+\tau_1^2 H_i+\tau_2^2 H'_{ij}+\sigma^2 I_{n_i} \qquad (5)$$

Where T_i and Z_i are matrices of known level 3 and level 2 covariates respectively. Ψ_1 and Ψ_2 are covariance matrices for level 3 and level 2 random effects, H_i and H'_{ij} are level 1 and level 2 serial correlation matrices. The serial correlation variances for level 1 and level 2 are represented by τ_1^2 and τ_2^2. σ^2 is the level 1 error variance and I_{n_i} is an identity matrix.

The first approach that can be used to investigate the presence of serial correlation in subject specific residuals given by the use of $\mathbf{Y}_i-X_i\hat{\boldsymbol{\delta}}-T_i\hat{\mathbf{v}}_i-Z_i\hat{\boldsymbol{\upsilon}}_{ij}$, in which $\hat{\mathbf{v}}_i=E(\hat{\mathbf{v}}_i\,|\hat{\boldsymbol{\upsilon}}_{ij},\mathbf{Y}_i)$ and $\hat{\boldsymbol{\upsilon}}_{ij}=E(\hat{\boldsymbol{\upsilon}}_{ij}\,|\mathbf{Y}_i)$ are the empirical Bayes estimates for the $\mathbf{v}_i$ and $\boldsymbol{\upsilon}_{ij}$ obtained by fitting a specific multilevel growth curve model. The main drawbacks of this approach are the strong dependence of the $\hat{\mathbf{v}}_i$ and $\hat{\boldsymbol{\upsilon}}_{ij}$ on the normality assumption on the random effects and also the form of the variance covariance structure imposed on equations 1 and 2. This simply means wrong normality assumptions will have repercussions on the analyses thereafter. Caution must therefore be taken when exploring the functional form of the subject-specific residuals if an assumed covariance function had been used previously because this could indirectly contaminate the serial correlation if present. Despite the above shortcomings, in longitudinal educational effectiveness research one usually encounters huge amount of data and in addition maximum likelihood estimators for the fixed effects and variance components are still consistent even with non-normality of random effects distribution (Verbeke & Lesaffre, 1997). If the covariance structure is allowed to be more general (unstructured) then one can proceed using these subject specific residuals. Morrell, Pearson, Ballentine, and Brant (1995) have used such residuals for the case of a two level nonlinear growth curve model.

The second approach is to look instead at transformed residuals proposed by Verbeke, Lesaffre, and Brant (1998). This approach serves as a simple informal check for the need of a serial component in a linear mixed model. This technique requires the transformation of ordinary least squares residuals $\mathrm{r}_i = \mathbf{Y}_i - X_i\hat{\boldsymbol{\delta}}_{\mathrm{OLS}}$ based on some presumed mean structure and ignoring any dependence among the repeated measurements. The transformation process entails projecting the ordinary least squares residuals orthogonally to the columns of T_i for the school level random effects and Z_i for student level random effects. In this case the transformed residuals are left only with variability not explained by the random effects. This second method

3.2. School effects

In educational effectiveness research, the school effect is measured by percentage of variance in the outcome that is between schools (ICC for cross-sectional data). However, for growth curve models with random intercepts and slopes, there exists no general agreement on what a 'school effects' is. We therefore propose a definition of school effect as the percentage of variance between schools for growth curve models with random intercepts and slopes. This proposed definition can be extended to three or more levels of hierarchy. The semi-variogram described above can be applied in a straightforward manner with the addition of random slopes at the second and third levels. This results in the inclusion of two new terms to equation 2. This means the covariance between the random intercepts and the random slopes is taken into account when estimating the variances of the different levels. However, this covariance parameter is less important when defining the percentage of variance at a particular level as shown in the next paragraph.

An example of a model with more than two levels in educational research could be repeated measurements within students within classes within schools within educational systems or even within countries. To make the presentation simple, an additional index is used in place of the different letters used previously. This index is used to denote the different levels in order to avoid running out of letters of the alphabet. Let $v_{q(p+1)(p+1)}$ denote the variance of the qth level random regression coefficient, and that q takes values 1,…, Q with Q not less than 2. With this notation, it is then possible to define the percentage of variance between units of the q level on each of these growth model parameters as follows:

% variance between units of level q on $\beta_{ijp} = \rho_{q(p+1)}$,

$$\text{where } \rho_{q(p+1)} = \begin{cases} \dfrac{v_{q11}}{\sum_{q=1}^{Q} v_{q11}} & \textit{when } P=1 \textit{ and } q=1,\dots,Q \\ \dfrac{v_{q(p+1)(p+1)}}{\sum_{q=2}^{Q} v_{q(p+1)(p+1)}} & \textit{when } P>1 \textit{ and } q=2,\dots,Q \end{cases},$$

where $p = 0,\dots,P-1$. Where P is the number of random coefficients allowed and Q is the number of levels of hierarchy in the data. In this new definition, the percentage of variance between level q units for the case of a random intercept model (that is with $P=1$) corresponds to the definition of the intraclass correlation (ICC) at the qth level (ρ_{q1}) for a random intercept model. For a random intercept and slope model (P=2), the definition indicates the estimation of two school effects: one effect for the intercept, and another effect for the slope.

4. Application to educational data

4.1. Data

The data considered for the application results from a large scale research project on secondary education (Van Damme et al., 2006). In this project, over 6000 students from about 57 schools in Flanders (Belgium) are followed. Several student outcomes are studied in this research but the application of the method proposed in this article has been limited to two student outcomes being the student's Dutch achievement (the mother tongue of most students) and well-being at school. Achievement in Dutch was measured via curriculum based tests administered at five different measurement occasions: at the start of first year (Grade 7), end of first year (Grade 7), second year (Grade 8), fourth year (Grade 10) and sixth year (Grade 12). A common scale for the Dutch language scores for the five measurement occasions was obtained using Item Response Theory (IRT). The questionnaire on students' well-being was administered four times during secondary school and measured on a 32-item scale. The chosen student samples for the application include only students who did not change schools and those who did not repeat a grade.

4.2. Software

Two software programmes are used in this study because of their different merits and demerits. Firstly, PROC MIXED of SAS 9.1 (SAS Institute Inc., 2003) is very flexible and suitable for fitting multilevel models, hierarchical linear models, and growth curve models (Singer, 1998). Secondly, MLwiN 2.02 (Rasbash et al., 2000) which provides a wide range of multilevel models together with plotting, diagnostic and data manipulation facilities. And lastly, it enabled us to compare results of multilevel models fitted in SAS to those of same multilevel model fitted in

MlwiN. However, SAS has the upper hand because it allows the programming of the more complex DSC multilevel growth curve model.

4.3. Results of random intercept only models

A quadratic growth curve model fitted the data best, but with different random effect specifications at the student and the school levels. Several authors stress the importance of nonlinear modelling (Bauer & Cai, 2008; Harring, 2009) when it represents the growth profile better. The first set of models used to obtain the results of Table 1, considered only random intercepts for both the students' and schools' growth profiles. The fixed part of the model is assumed to be quadratic and considered reasonable from the variance exploration and deviance statistics. A mixture of chi-square distribution is used as null distribution for the likelihood ratio test statistic (Verbeke & Molenberghs, 2000) for the inclusion of random linear and quadratic growth effects.

There are four multilevel random intercept quadratic growth curve models fitted to the data. However more emphasis is put on the different serial correlation structures specified. The results of the first column of Table 1 come from the model without any serial correlation. Under this model specification, the school effect on the students' status in well-being is estimated to be about 3%. Such a small school effect on the students' status for a noncognitive outcome is in agreement with the current literature in school effectiveness research. The second and third models in Table 1 are improvements on model 1 and are obtained by specifying a Gaussian serial correlation at level-2 and level-1 respectively. When the level-2 serial correlation variability (the stochastic variation of schools from the school mean profile) is modelled, the overall school effect estimation drops to 2.2%. Considering the student level serial correlation correction (model 3), the school effect on the students' status in well-being increases to 4.3%. The school effects estimates from model 2 and 3 are far larger than that from model 1. It is worth noting that model 2 fits better than model 1 when looking at the deviance statistic while model 3 fits even better than models 1 and 2.

In order to take into account the serial correlation at levels 1 and 2 simultaneously, model 4 with a DSC is fitted to the data. The school effect is again estimated without all the stochastic variability at the student and school levels. The school effect estimate from model 4 is then 3.4%, which is less than that of model 3 but greater than that of model 2. In any case one may tend to favour this model because it fits better than the first three, and corrects for both levels of serial correlation. In comparison to the model without serial correlation correction, the results of model 4 show an increase of about 17% in the school effect on the students' status in well-being.

Other spatial serial correlation functions considered were spatial power and exponential. However, test of model fit using deviance statistics indicate that the choice of a serial correlation function did not matter much at level-1 (see Table 2a of the appendix B). But at level 2, the results show a slight difference in favour of the Gaussian serial correlation function over the other two functional forms.

Table 1

Well-being outcome modelled using four different three level quadratic growth curve models

	Model1	Model 2	Model 3	Model 4
Parameters				
Fixed effects				
Intercept	3.9708(0.0192)	3.935 (0.0285)	3.9668(0.0191)	3.9417(0.0282)
Linear slope	-0.2473(0.0089)	-0.2285 (0.0172)	-0.2465(0.0084)	-0.2328(0.0170)
Quadratic slope	0.0198(0.0012)	0.0178 (0.0024)	0.0200(0.0012)	0.0185(0.0024)
Variances				
School intercept	0.0080 (0.0024)	0.0055(0.0027)	0.0085(0.0025)	0.0063(0.0027)
Student intercept	0.1443(0.0044)	0.1472(0.0044)	0.0953(0.0077)	0.0985(0.0077)
Residual	0.1228(0.0019)	0.1142(0.0018)	0.0818(0.0024)	0.0794(0.0023)
-2loglikelihood	14751.3	14550.5	14336.9	14159.1
Degree ff	6	8	8	10
Serial correlation				
Level 2 serial	/	0.0073(0.0018)	/	0.0063(0.0018)
Level 1 serial	/	/	0.0844(0.0065)	0.0801(0.0065)
School effect				
Initial status	2.9%	2.2%	4.3%	3.4%

Model 1: without serial correlation; Model 2: with level 2 Gaussian serial correlation correction only Model 3: with level 1 Gaussian serial correlation correction only; Model 4: with both levels 1 and 2 Gaussian serial correlation corrections.

However, basing the choice of the serial correlation functional form on the results of Table 2a will be wrong. The reason for this being the absence of a formal test for two multilevel growth curve models which differ only in the type of serial correlation functions used in each. Two such models will have the same number of parameters as such even the famous deviance statistics cannot be applied with zero degree of freedom. More appropriate results are therefore obtained using the semi-variogram technique shown in Figures 1a and 1b. Figure 1a shows that the level 1 serial correlation is most likely a Gaussian form while Figure 1b indicates that level 2 has an exponential form. In fact the results of Table 2a only help to indicate that serial correlation cannot be neglected in the model specification since useful comparisons can only be made when compared to the same models without DSC.

When the students' Dutch achievement outcome is considered, the results on Table 2 below are obtained. Looking at the results for model 1 (no serial correlation) in the first column of Table 2, the school effect on the student status estimate is 40%. The models with serial correlation correction all result in larger school effects. Unlike the school effects estimates of the students' well-being outcome, the double serial correlation correction for the Dutch achievement produced the highest school effect estimate in comparison to level 1 and level 2 corrections each alone. The DSC correction results in an increase of about 11.3% in the school effect estimate compared to the model 1 with no serial correction. Model 4 once again shows to be the best fit for the data when compared to the other three using once again deviance statistic.

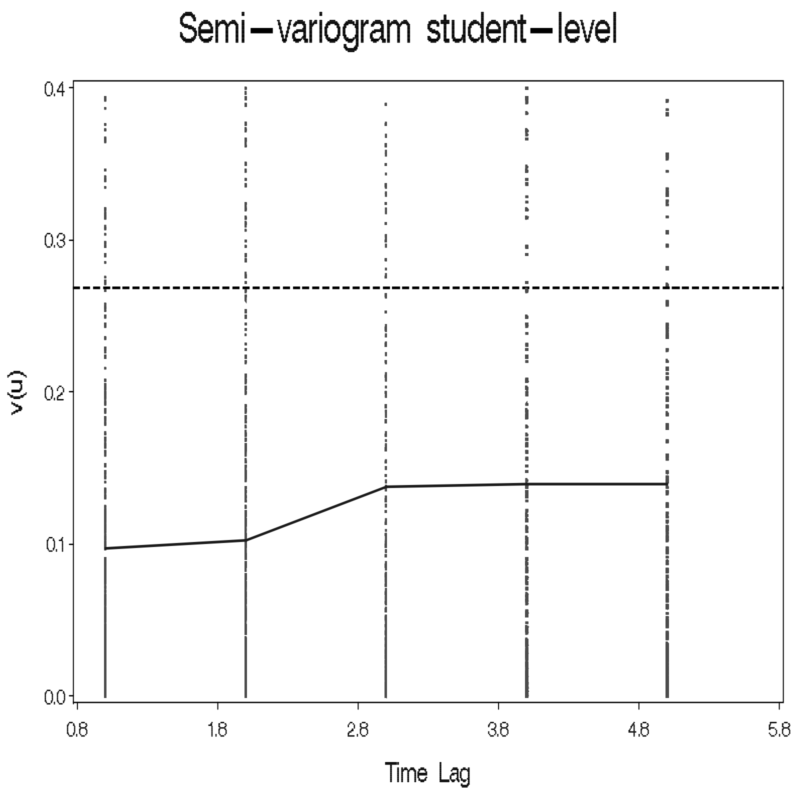


Figure 1a: Semi-variogram of Level 1 serial correlation

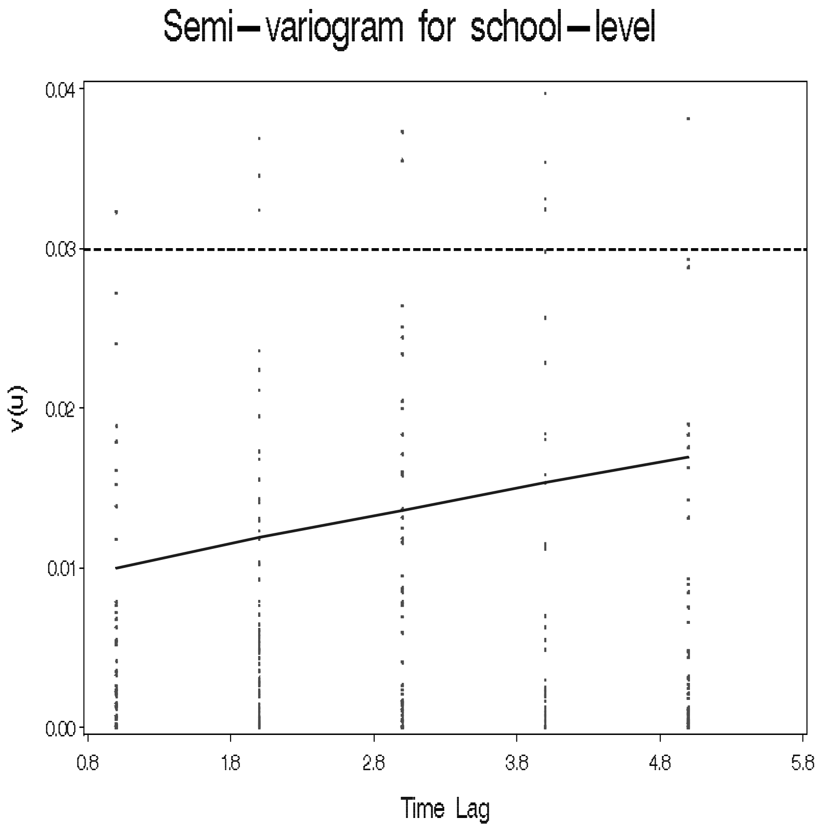


Figure 1b: Semi-variogram of Level 2 serial correlation

Table 2

Dutch achievement outcome modelled using four different three level quadratic growth curve models

parameters	Model1	Model 2	Model 3	Model 4
Fixed effects				
intercept	4.4815(0.0782)	4.4577(0.0878)	4.4802(0.0787)	4.4599(0.0878)
Linear slope	-0.0883(0.0112)	-0.0642(0.0273)	-0.0886(0.0111)	-0.0654(0.0277)
Quadratic slope	0.0214(0.0016)	0.0171(0.0038)	0.0214(0.0016)	0.0173(0.0039)
Variances				
School intercept	0.2938(0.0601)	0.3028(0.0636)	0.2977(0.0609)	0.3028(0.0633)
Student intercept	0.2460(0.0075)	0.2475(0.0075)	0.2354(0.0083)	0.2409(0.0081)
Residual	0.1929(0.0030)	0.1811(0.0028)	0.1519(0.0089)	0.1375(0.0129)
-2loglikelihood	19775.3	19459.4	19712.2	19408.8
Degree ff	6	8	8	10
Serial correlation				
Level 2 serial	/	0.0221(0.0049)	/	0.0195(0.0056)
Level 1 serial	/	/	0.0546(0.0083)	0.0530(0.0118)
School effect				
Initial status	40%	41.4%	43.5%	44.5%

Model 1: without serial correlation; Model 2: with level 2 serial correlation correction only Model 3: with level 1 serial correlation correction only; Model 4: with both levels 1and 2 serial correlation corrections.

4.4. Results of random intercept and slope models

In order to see the impact of modelling serial correlation on students' growth in well-being, three multilevel quadratic growth curve models with random intercepts, linear and quadratic slopes at the school level and random intercepts and linear slopes for the student level were fitted. Though the results of the model random effects selection and fit for the student well-being in Table 1a of the appendix B show the possibility of a random slope at the student level, the time needed for convergence was far longer.

Three models are used: model 1 with no serial correlation, model 2 with level-1 serial correlation and model 3 with the DSC. The results of these three models are presented in Table 3a in the appendix B. The school effect estimate on the initial status with serial correlation correction at level-1 seems to be better than for the double serial correlation in the presence of random slopes. In any case, the complex DSC model results in a higher school effects estimate compared to the model without any serial correlation. It also has the best fit compared to the other two models. However, the increase in the school effect estimate on the students' linear growth in well-being increases enormously with either the level-1 or DSC correction.

5. Simulation study

A small simulation study was conducted to evaluate the impact of the number of level-3 and level-2 sample units, and number of measurement occasions on the multilevel growth curve model with DSC, to detect the presence of level-1 and most especially level-2 serial correlations. It also allows an evaluation of the problem when school effects are estimated by using growth models without serial correlation correction on data with such a property. We also looked at the impact on the growth criteria used (student status and student growth). Using the well-being outcome characteristics (intercept =3.941, linear slope =-0.2328 and quadratic slope = 0.0185), data was simulated with level-2 and level-3 random effects variances of 0.5 and 0.62 respectively. We also included a level-1 error variance of 0.51. Level-1 and level-2 Gaussian serial correlations were specified as 0.4*ranexp (0)*$((1/\mu)^2)$. Where μ (the rate of decrease of the serial correlation function) was fixed at 0.8 and ranexp is a random value generator function for variables with an exponential distribution. Six datasets were generated with six different level 3 units (30, 40, 50, 60, 70, & 90), 50 level-2 units per level-3 unit and 5 measurement occasions for each level-2 unit. These generated samples ranged between 7500 and 22550 observations. A multilevel growth curve model without and with DSC is fitted to the data using PROC MIXED of SAS 9.2.

The school effect estimates for the student status and linear growth is then obtained from the estimates of the level-2 and level-3 variance components. The results indicated that except for the case of 30 schools (level-3); the DSC model was able to detect the presence of the two level serial correlations present in the data. The school effect estimates after correcting for DSC, indicated that school effect on the student status decreases (83% to 62%) with increase in the number of schools from 30 to 90. However, the school effect on the student growth was relatively stable varying between 42% and 45% inclusively as shown in Table 3.

Table 3

Multilevel growth curve model with DSC results on six simulated datasets (with 50 students per school and 5 measurement occasions per student).

	Number of schools					
	30	40	50	60	70	90
Covariance parameter	Estimate					
School level						
Intercept	0.4526	0.5547	0.5807	0.5777	0.5573	0.4967
Intercept, slope	0.1697	0.2454	0.2447	0.2286	0.2329	0.2402
Slope	0.3222	0.3254	0.3053	0.3014	0.2973	0.3221
Student level						
Intercept	0.1866	0.1116	0.2570	0.3344	0.3404	0.3463
Intercept, slope	0.4340	0.4344	0.3945	0.3780	0.3758	0.3674
Slope	0.4046	0.3991	0.4001	0.4006	0.4089	0.4025
Serial correlation						
School level	1.36E-21	3.67E-21	0.0018	0.0017	0.0010	0.0009
Student level	0.1073	0.1527	0.0617	0.0083	0.0030	0.0012
Residual	0.6500	0.6423	0.6474	0.6393	0.6425	0.6516
School effect						
Status	71%	83%	69%	63%	62%	59%
Growth	31%	45%	43%	43%	42%	44%

In order to see how the DSC model performs compared to a model without this specification, the two types of models were fitted on each set of simulated data and the results summarised with bar charts as in Figures 2a and 2b. The results indicate that correcting for serial correlation at both levels generally gave larger school effect estimates on both the student status and growth. The gain in school effect estimation by correcting for DSC seems more pronounced on the student status when the number of schools is less than 50. For the impact of the DSC corrected model on student growth, the number of schools has to be less than 40.

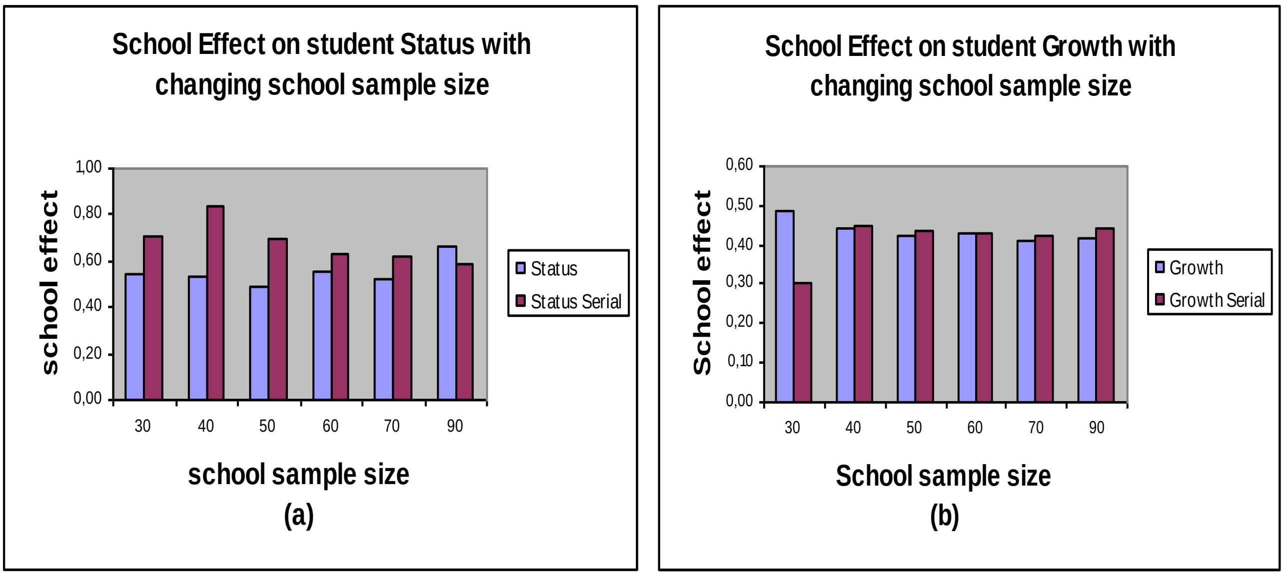


Figure 2: School effect estimates on student status (a) and student growth (b) with & without double serial correlation correction using multilevel data with different school sample sizes.

The number of schools was then fixed at 50 and the number of students per school varied (30, 40, 50, 60, and 80) and new samples obtained on which the models with and without serial correlation were fitted. The school effect for the student status and growth were estimated and the results summarised using the bar charts as shown in Figures 3a and 3b. The results revealed that when the number of students per school is less than 60 (except 30 students), the school effect on the student status are underestimated by the model without DSC.

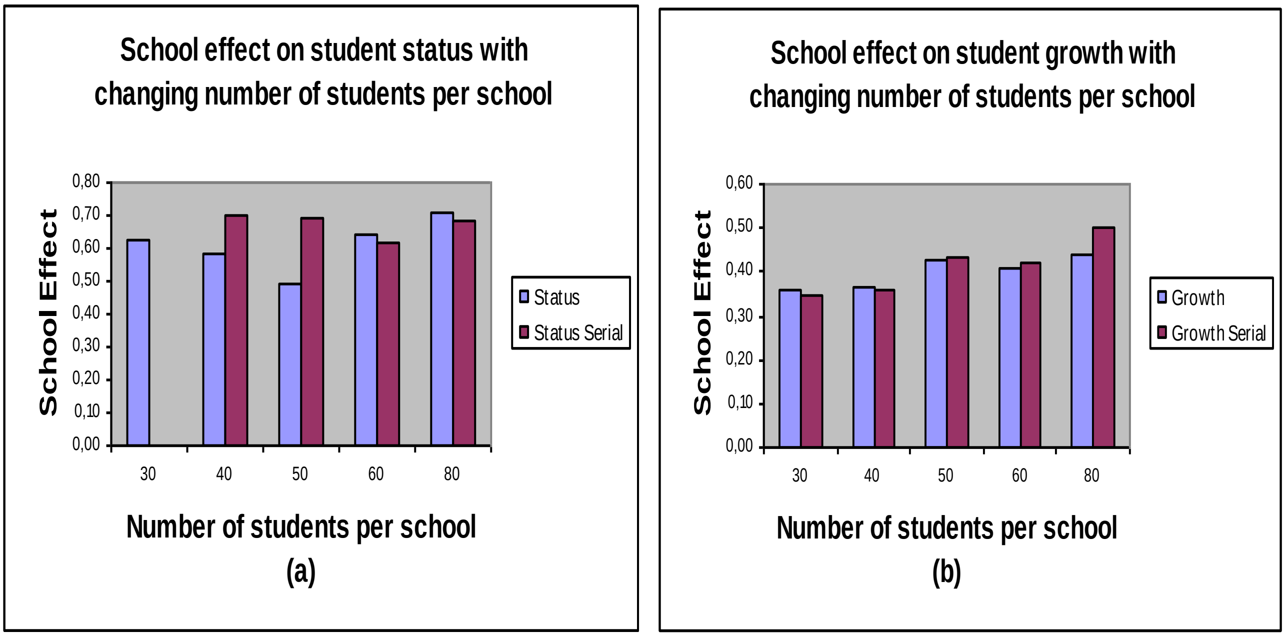


Figure 3: School effect estimates on student status (a) and student growth (b) with & without double serial correlation correction using multilevel data with different number of students per school.

A similar procedure was used to investigate when the number of schools was fixed at 50, the number of students per school at 50 and the number of measurement occasions varied from 4 to 8. The results indicate that 5 to 6 repeated measures are enough to be able to detect level-1 and -2 serial correlations when they exist in the data.

In school effectiveness studies, policy makers, administrator, parents and researchers often want to know the ranking of schools. Usually schools are ranked using the estimates of school level residuals which are of course conditioned on the model used and its assumptions. The study reveals that when DSC correction is omitted in the models when in fact it does exists, 21 out of the 52 schools are incorrectly ranked using the students Dutch language achievement outcome. Some schools even gained up to 4 places on the ranking when the appropriate multilevel growth curve model with DSC correction is used compare to when neglected. These school ranking results are very crucial to parents and school administrators and consequently school residuals must be estimated correctly.

6. Discussion and Conclusion

The importance of nonlinear modelling, and the need of modelling level-1 residuals properly, have been stressed by a few publications in the area of educational effectiveness research. However the modelling of level-1 residuals serially in a three level model is rather new and as shown in this article, can greatly affect the estimates of school effects on students' status and growth in academic achievement outcomes as well a non-cognitive outcome like well-being.

According to the results of this study, the modelling of serial correlation for such longitudinal data, improves the estimates of school effects on both the students' academic achievement in Dutch and the students' well-being at school. This article further illustrates that for longitudinal data with more than two levels, residuals at the higher levels could equally be serially correlated and thus needing consideration if school effects and other important model parameters must be estimated properly.

The simulation study showed that such longitudinal data with level-1 and -2 serial correlations can exist in many situations under different conditions. However, not all possibilities have been explored in this paper but with a dataset of about 50 schools, about 50 students per school and at least 5 equally spaced or 4 unequally spaced measurement occasions, DSC can be detected. The simulation results also seem to hint that the estimation of the school effect on growth is less affected by changes in number of schools and DSC correction as compared to the estimation of the school effect on student status.

The impact of the DSC model on the school effects on the students' status and growth shows the importance of the model in growth curve applications in general. From these findings, it is recommended that checks for the presence of serial correlation should always be made whenever longitudinal data (with at least four measurement lags for unequally spaced measurement occasions, or five equally spaced measurement occasions) are used to model growth. It also emphasises the gain in model fit and validity of school effects estimation with the use of the complex DSC for growth curve models with three levels.

MANUSCRIPT 4: Growth in reading comprehension and mathematics achievement in primary school: A bivariate transition multilevel growth curve model approach[4]

Anumendem, D. N., De Fraine, B., Onghena, P. & Van Damme, J.

Abstract

There is an increasing interest in the field of educational effectiveness research for studying changes in pupils' outcomes over time. However, most studies focus on the growth in only one effectiveness criterion, which is problematic given that school effects are only moderately consistent over different criteria. Moreover, the consistency issue has seldom been studied through multivariate growth curve models. The current study investigates school effects on pupils' growth in both mathematics and reading comprehension (and their relation) in primary schools taking previous changes in mathematics into account through a bivariate transition multilevel growth curve model. Using the proposed two-stage effectiveness criteria, the results indicate some level of consistency between school effects. It revealed that schools in which pupils show a stronger growth in mathematics tend to also show a stronger growth in reading comprehension. Earlier growth in mathematics was found to predict subsequent growth in reading comprehension.

[4] Manuscript submitted for publication.

1. Introduction

The measurement and explanation of school effects on both students' reading comprehension and mathematics achievement in primary school children has seldom been studied. Some studies have focused on mathematics achievement as a predictor of reading comprehension (Lerkkanen et al., 2005) or reading as a predictor of mathematics performance (Kirsch et al., 2002). These studies make the key assumption of deciding in advance which of the outcomes is dependent on the other. A few multivariate studies have nonetheless been done to investigate the possibility of a relationship between overall reading ability and mathematics, by looking at the common characteristics responsible for high performance in mathematics and reading (e.g. De Maeyer et al., 2010). The current study goes further, because it investigates growth in reading comprehension and mathematics without any assumption on their functional dependence.

Reading achievement in primary school contains two distinguishable components: word reading and reading comprehension. Reading comprehension is a complex process and requires not only the fluent decoding of words but also understanding vocabulary, making inferences and relating the ideas to prior knowledge. Reading comprehension in primary schools stretches from the understanding of the meaning of words to the meaning of a short text and this is highly dependent on age. Many researchers have argued that true measurement of reading comprehension skills can only be obtained at later stages of primary school say from the third grade on (7 or 8 years old) (Adams, Treiman, & Pressley, 1998; Topping & Fisher, 2003; Van Keer, 2004). This paper focuses on reading comprehension which requires more advanced cognitive and linguistic skills. For this reason pupils were only tested from the end of grade 3 on. While studies have shown that individual differences in reading comprehension in particular during primary school are stable (de Jong & van der Leij, 2002), differences between schools however have not been addressed in depth. It might also be interesting to find out if this stability mentioned by de Jong and van der Leij (2002) remains when considered in conjunction with mathematics.

Mathematics achievement in primary schools usually embodies different components like: arithmetic, basic number knowledge, memorisation of arithmetic operations like addition, subtraction, and conceptual understanding and procedural knowledge (Dowker, 1998; Donlan & Gourlay, 1999). Studies have shown that pupils’ skill in mathematics develops rapidly during formal instruction (Geary, 1990; Jordan, Hanich, & Kaplan, 2003).

Most studies in the field of educational effectiveness are limited to only one outcome variable as the effectiveness criterion. Or these studies report on several effectiveness criteria, but they model each criterion separately. Many researchers have called for the development of better statistical methods capable of handling more than one effectiveness criterion in the same model (Teddlie, Reynolds, & Sammons, 2000; Van de gaer et al., 2009). When effectiveness criteria are modelled separately, the underlying assumption is that these different effectiveness criteria are independent of each other. As far as the school effects estimation is concerned, we hypothesize that this assumption is most often not tenable with reading comprehension and mathematics in primary school. The implication of such an assumption is for example, that a pupil’s or school’s score in mathematics is not related to the pupil’s or school’s performance in reading comprehension. On the contrary, some studies have suggested that mathematics achievement and reading as a whole may depend on similar predictors (Bull & Johnston, 1997; Tymms, 1999), strengthening the need for combining these effectiveness criteria. Other studies have even found that similar linguistic abilities are needed for both mathematics problem solving and reading comprehension tasks (Jordan et al., 2002). Further studies have identified four components of mathematical problem solving as; translation, integration, solution planning and execution, with the first two of these components highly correlated with reading comprehension skills (Mayer, 1987).

In addition to the plea for multiple criteria, educational effectiveness researchers advocate studying student outcomes over time (De Fraine, Van Damme, & Onghena, 2007; Raudenbush, 1995; Singer & Willett, 2003). It is argued that growth in student outcomes over time is a very essential criterion since learning means changing (Teddlie, Reynolds, & Sammons, 2000). In order to investigate changes in student outcomes over time, longitudinal data are invaluable. This paper will address both the plea for multiple effectiveness criteria and the plea for studying growth. The use of growth curves as a statistical method introduces another level of choice of criterion beyond the choice of pupil outcome variable. This paper tries in the next section to delineate these two stages of school effectiveness criteria. Firstly, the choice of the dependent or outcome variables with a correlation strong enough to lend credence to a multivariate model instead of separate univariate models. Secondly, one has to make a choice of which growth curve parameters to use to model the student outcomes and to estimate the school effect. These parameters could be obtained at a certain point in time (intercept) or as a growth parameter (linear or quadratic).

2. Two-Stage Effectiveness Criteria

Generalisation of results of different studies with different criteria has most often been problematic given that school effects are only moderately consistent over different criteria. Moreover, the consistency issue has seldom been studied through growth curve models. This paper introduces what is called “two-stage effectiveness criteria” to study school effect consistency in multivariate multilevel growth curve models (MMGCM). In stage 1, the researcher chooses one or more student outcomes (reading comprehension, mathematics achievement, well-being, etc.) and in stage 2, the growth parameters are chosen (initial status, linear change, quadratic slope, etc.). This process can result in two or more effectiveness criteria depending on the number of outcome variables and the nature of the growth. For example, two outcome variables and a random intercept and linear growth model will have four possible effectiveness criteria: an intercept (student status) and a linear slope (student growth) each for both outcomes.

The current study has two first stage criteria which are mathematics and reading comprehension and two second stage criteria (intercept and linear slope). The four effectiveness criteria in this study are therefore: mathematics intercept, mathematics slope, reading comprehension intercept and reading comprehension slope. These four effectiveness criteria (see Figure 1- illustrate pupils' growth in both mathematics and reading comprehension (and their relation) in primary schools through a bivariate multilevel growth curve model (BMGCM).

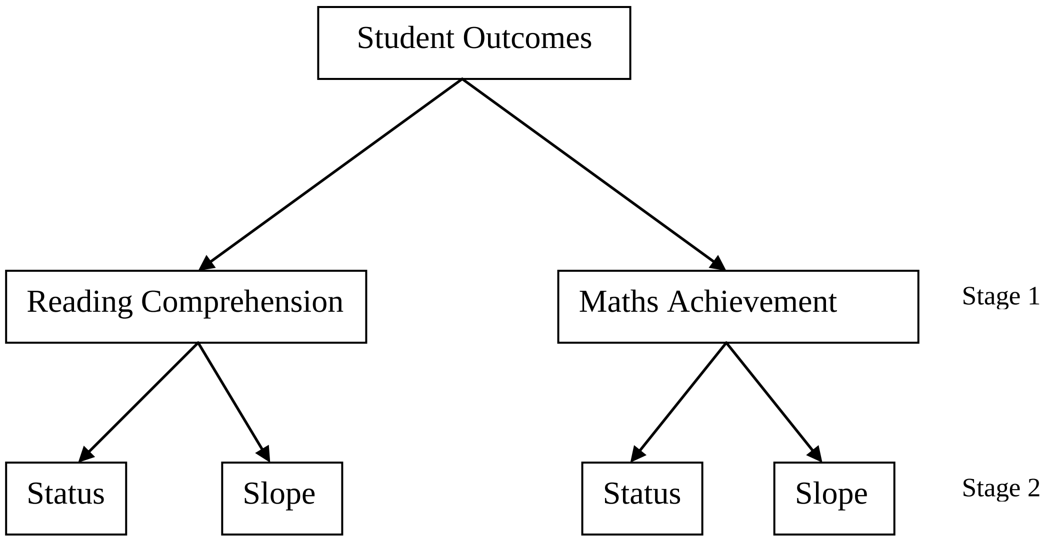


Figure 1 A proposed schema of the two-stage school effectiveness criteria for a bivariate linear growth curve model.

The advantages of MMGCM are enormous. They are not only statistically powerful, but are also capable of answering a wider range of research questions more efficiently. These questions could stretch from stability to consistency of school effects. The choice of characteristics taking into account their dependence, can greatly reduce the chance of making a type 1 error (Hox, 2002) and as a consequence, improves generalisability of the study findings.

3. Objectives

The main objectives of the current study are summarised by the following research questions:

1. What is the nature of the evolution of mathematics achievement and reading comprehension from the end of Grade 3 to the end of Grade 6 in primary school? Are there differences in the growth trajectories for these two outcomes?
2. What is the correlation between the mathematics and reading comprehension growth profiles at the pupil level and at the school level?
3. How large is the school effect on the pupil status and pupil growth? Is this effect similar for the two outcomes?
4. What is the impact on the estimates of the school effects when the dependence of mathematics and reading comprehension is taken into account? In other words, we will compare school effects for two separate univariate models and one bivariate model.
5. Can the prior growth in mathematics explain the subsequent differences in pupils' status and growth in reading comprehension and mathematics achievement?

4. Methodology

In this section, we will discuss three models: (1) the univariate multilevel growth curve model (UMGCM), (2) the bivariate multilevel growth curve model (BMGCM) and (3) the extension to a transition model (the bivariate transition multilevel growth curve model, BTMGCM).

4.1. Univariate Multilevel Growth Curve Model (UMGCM)

The univariate multilevel growth curve models (UMGCM) will be applied to the mathematics scores on the one hand and the reading comprehension scores on the other hand. These two univariate multilevel growth curve models are each of the form

$$\mathbf{y}_i \mid \mathbf{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \Sigma_i) \qquad (1)$$

In equation 1 above, the $\mathbf{y}_i$'s are vectors representing all the measurements for the i th school.

Each outcome or response measurement y_{ijk} denotes the kth measurement for the jth student from the i th school. This means the vector of responses $\mathbf{y}_i = (y_{i11}, y_{i12}, \ldots, y_{i1m_j}, \ldots, y_{in_im_j})^T$.

4.2. *Bivariate Multilevel Growth Curve Model (BMGCM)*

The research questions will be answered by applying two univariate multilevel growth curve models (UMGCM) and two bivariate multilevel growth curve models (BMGCM). The two univariate multilevel growth curve models are each of the form In the bivariate model, the two outcomes are combined through the proper specifications of a bivariate distribution for all the random effects taking into account the dependence of the growth processes. In this combined model, a bivariate normally distributed response is considered for the new response Y. Where Y= (Y_1, Y_2) $\sim N((X_1\boldsymbol{\beta}_1, X_2\boldsymbol{\beta}_2),(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2))$ and the mean structures and variance covariance matrices are allowed to be different.

A multivariate response can be incorporated into a multilevel growth curve model by creating an extra lowest level, which is called level zero in this paper. In the growth curve model setting, the two responses are nested within the measurement occasions which are in turn nested within the students and finally within the schools. The main purpose of the level 0 is to define the double response per pupil. Our interest is then to use this model to assess the relationship between the growth parameters of the two response variables (reading comprehension and mathematics achievement).

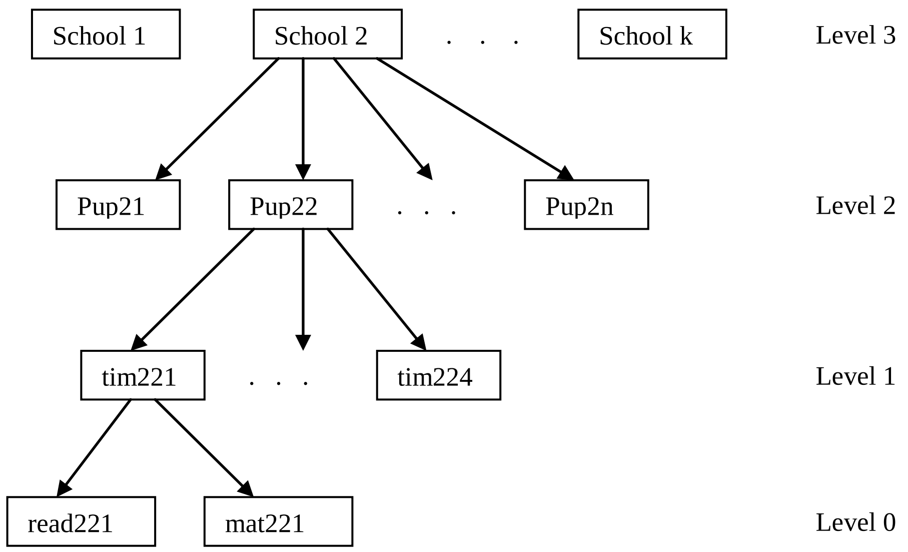


Figure 2 Data collection structure for reading comprehension and mathematics outcome variables.

Modelling the two outcome variables simultaneously, accounts for the dependence between the outcomes and thus improves the parameter estimates of the model. This is usually of great importance when association structures change with time (Fieuws & Verbeke, 2004). In this study, we will fit a model, which has a structure of a four-level model but with the lowest level called level 0 because its variability is not of interest. The reason being that the level 0 index is used only to differentiate between the response variables. In this case the structure of the data fits into a multilevel growth curve model.

$$\left.\begin{aligned} Y_{ijk} = {} & \beta_{01} z_{1ijk} + \beta_{02} z_{2ijk} + \beta_{11} t_{ijk} z_{1ijk} + \beta_{12} t_{ijk} z_{2ijk} + \beta_{21} t_{ijk}^{2} z_{1ijk} + \beta_{22} t_{ijk}^{2} z_{2ijk} + \\ & (v_{00k} + v_{10k} t_{ijk} + v_{20k} t_{ijk}^{2} + u_{0ik} + u_{1ik} t_{ijk} + u_{2ik} t_{ijk}^{2} + \varepsilon_{ijk}) z_{1ijk} + \\ & (v'_{00k} + v'_{10k} t_{ijk} + v'_{20k} t_{ijk}^{2} + u'_{0ik} + u'_{1ik} t_{ijk} + u'_{2ik} t_{ijk}^{2} + \varepsilon'_{ijk}) z_{2ijk} \end{aligned}\right\} \qquad (2)$$

where

$$z_{1ijk} = \begin{cases} 1 & if\ Reading\ comprehension \\ 0 & if\ Mathematics\ achievement \end{cases}$$

$$z_{2ijk} = \begin{cases} 0 & if\ Reading\ comprehension \\ 1 & if\ Mathematics\ achievement \end{cases}$$

This means our model can be written as

$$Y_{ijk} = \begin{cases} \beta_{01} + \beta_{11} t_{ijk} + \beta_{21} t_{ijk}^2 + v_{00k} + v_{10k} t_{ijk} + v_{20k} t_{ijk}^2 + u_{0ik} + u_{1ik} t_{ijk} + u_{2ik} t_{ijk}^2 + \varepsilon_{ijk} & if\ z_{1ijk} = 1 \\ \beta_{02} + \beta_{12} t_{ijk} + \beta_{22} t_{ijk}^2 + v'_{00k} + v'_{10k} t_{ijk} + v'_{20k} t_{ijk}^2 + u'_{0ik} + u'_{1ik} t_{ijk} + u'_{2ik} t_{ijk}^2 + \varepsilon'_{ijk} & if\ z_{2ijk} = 1 \end{cases}$$

The school level variance (level 3) is given by:

$$\begin{pmatrix} v_{00k} \\ v_{10k} \\ v_{20k} \\ v'_{00k} \\ v'_{10k} \\ v'_{20k} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{v_0}^2 & & & & & \\ \sigma_{v_1 v_0} & \sigma_{v_1}^2 & & & & \\ \sigma_{v_2 v_0} & \sigma_{v_2 v_1} & \sigma_{v_2}^2 & & & \\ \sigma_{v'_0 v_0} & \sigma_{v'_0 v_1} & \sigma_{v'_0 v_2} & \sigma_{v'_0}^2 & & \\ \sigma_{v'_1 v_0} & \sigma_{v'_1 v_1} & \sigma_{v'_1 v_2} & \sigma_{v'_1 v'_0} & \sigma_{v'_1}^2 & \\ \sigma_{v'_2 v_0} & \sigma_{v'_2 v_1} & \sigma_{v'_2 v_2} & \sigma_{v'_2 v'_0} & \sigma_{v'_2 v'_1} & \sigma_{v'_2}^2 \end{pmatrix} \right),$$

and for the student level variance (level 2):

$$\begin{pmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u'_{0jk} \\ u'_{1jk} \\ u'_{2jk} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & & & & & \\ \sigma_{u_1 u_0} & \sigma_{u_1}^2 & & & & \\ \sigma_{u_2 u_0} & \sigma_{u_2 u_1} & \sigma_{u_2}^2 & & & \\ \sigma_{u'_0 u_0} & \sigma_{u'_0 u_1} & \sigma_{u'_0 u_2} & \sigma_{u'_0}^2 & & \\ \sigma_{u'_1 u_0} & \sigma_{u'_1 u_1} & \sigma_{u'_1 u_2} & \sigma_{u'_1 u'_0} & \sigma_{u'_1}^2 & \\ \sigma_{u'_2 u_0} & \sigma_{u'_2 u_1} & \sigma_{u'_2 u_2} & \sigma_{u'_2 u'_0} & \sigma_{u'_2 u'_1} & \sigma_{u'_2}^2 \end{pmatrix} \right).$$

The level 1 matrix components represent parameters associated with the error terms of the two growth processes

$$\begin{pmatrix} \varepsilon_{ijk} \\ \varepsilon_{ijk'} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right)$$

In vector notation we can simple write

$$\mathbf{v}_k \sim MVN(\mathbf{0}, \Omega_v),\ \boldsymbol{\upsilon}_{jk} \sim MVN(\mathbf{0}, \Omega_u)\ and\ \ \boldsymbol{\varepsilon}_{ijk} \sim MVN(\mathbf{0}, \Omega_\varepsilon)$$

Where $\mathbf{0}$ is a zero mean vector and Ω_v and Ω_u are respectively the covariance matrices for the school and student levels. An extension of this unconditional growth curve model to a conditional model is possible. Conditional versions of the BMGCM can enable the estimation of general and specific effects for the combined responses or for each response in the model respectively.

4.3. Bivariate Transition Multilevel Growth Curve Model (BTMGCM)

A common problem with multivariate outcome data is the possibility of incomplete observations in the outcome vector. There are a number of reasons why some observations might be absent in a study. When incomplete observations are missing at random or even completely at random, maximum likelihood estimates obtained from multilevel growth curve models (Dempster, Laird, & Rubin, 1977) or the full maximum likelihood estimates for latent growth models (Muthén, 2004), are still valid. However sometimes because of the design of the study, the statistical method used or the type of pupil outcomes to be considered, attritions occur in one outcome variable and not in the other. The situation in this study is summarized in Table 1 with the (X) indicating that a test was administered at that primary school grade. Students took a mathematics test at 7 occasions, while the reading comprehension test was administered at 4 occasions.

Table 1
Overview of the measurement occasions of the mathematics and reading comprehension tests.

Outcome	Begin grade1	End grade1	End grade2	End grade3	End grade4	End grade5	End grade6
Mathematics	X	X	X	X	X	X	X
Reading Comp				X	X	X	X

A bivariate transition multilevel growth curve model (BTMGCM) is introduced in this section as a way of circumventing the problem of missing reading comprehension scores at the beginning of grade 1, end of grades 1 and 2. This is considered as a better alternative to deleting the available mathematics scores obtained at those measurement occasions. The purpose of this model is to account for any possible dependence of the pupils reading comprehension and mathematics growth curves on these prior mathematics achievement scores.

Transition models are a specific class of conditional models. In a transition model, an outcome (Y_{ijk}) in a longitudinal sequence is described as a function of previous outcomes or history $\mathbf{h}_{ijk} = (Y_{ij1},\ldots,Y_{ijk-1})$ (Diggle et al., 2002; Fahrmeir & Tutz, 2001). The order of a transition model is the number of previous measurements that is still considered to influence the current outcome. This is a model which is simple to fit and understand yet strong enough to enable the investigation of the complex relationship that current processes have with their history. These models have been discussed in detail in textbooks such as Diggle et al. (2002), Molenberghs and Verbeke (2006) and Fahrmeir and Tutz (2001). However, extensions to handle more than one student outcome and in a multilevel growth curve model setting have never been done. It is in this context that the following BTMGCM is introduced, firstly, to solve the problem of unequal number of measurement occasions for the two pupil outcomes. And secondly, the model provides a powerful framework that can throw more light on the question of dependence of growth in one outcome on previous growth in a different outcome. The formulation of a bivariate transition model is given as follows:

$$Y_{ijk} = \begin{cases} \beta_{01} + \beta_{11}t_{ijk} + \beta_{21}t_{ijk}^2 + \kappa_1(h_{ijk},\beta) + v_{00k} + v_{10k}t_{ijk} + v_{20k}t_{ijk}^2 + u_{0ik} + u_{1ik}t_{ijk} + u_{2ik}t_{ijk}^2 + \varepsilon_{ijk} & if\ z_{1ijk} = 1 \\ \beta_{02} + \beta_{12}t_{ijk} + \beta_{22}t_{ijk}^2 + \kappa_2(h_{ijk},\beta) + v'_{00k} + v'_{10k}t_{ijk} + v'_{20k}t_{ijk}^2 + u'_{0ik} + u'_{1ik}t_{ijk} + u'_{2ik}t_{ijk}^2 + \varepsilon'_{ijk} & if\ z_{2ijk} = 1 \end{cases}$$

where κ_1, κ_2 are functions (most often linear) of the history (h_{ijk}). In the special case of this study $\kappa_1 = \kappa_2 = \kappa$. The β's indicate the possibility of separate models for the independent variables of the growth curve model. In compact form, the bivariate transitional growth curve model can be written as

$$\mathbf{y}_i \left| (\mathbf{b}_i, \kappa(\mathrm{h}_i, \boldsymbol{\beta})) \sim N(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \kappa(\mathrm{h}_i, \boldsymbol{\beta}), \Sigma_i) \right. .$$

The next section proceeds with the application of the models described so far. First of all the paper examines if BMGCMs are more realistic and statistically backed to use instead of two separate UMGCMs. Next it compares the BMGCM with the bivariate growth model controlling for previous changes in mathematics using BTMGCM. The results from the three models are then investigated for any fundamental changes in the conclusions.

5. Application

The data used for this study were collected as part of a longitudinal research project to describe and explain pathways through primary education, SiBO (Schoolloopbanen in het BasisOnderwijs). There were about 200 Flemish schools that participated in this study, which started in September 2003 and followed one cohort of pupils throughout their career in primary school (Grade 1 to Grade 6). All the pupils took mathematics achievement tests at 7 occasions and reading comprehension tests (Dutch language) at four occasions (see previous section). Grade-appropriate tests with common scales for the reading comprehension and mathematics scores were obtained separately for four measurement occasions and seven measurement occasions respectively, using Item Response Theory. A number of background variables were also collected including: socio-economic status of the family, gender, language spoken at home, age and ethnic-cultural background. The sample used for this paper had 194 schools with 6250 pupils.

The average growth profile was explored for both mathematics and reading comprehension and looking at the deviance statistics and parsimony, we settled down to a linear growth model for both outcomes. The time variable (linear slope) is coded 0 for end of Grade 3, 1 for end of Grade 4, 2 for end of Grade 5 and 3 for end of Grade 6. Two main software programmes are used in this study because of their different merits. First, SAS 9.1 (SAS Institute Inc., 2003) is used because of its PROC MIXED which is very flexible and suitable for fitting hierarchical linear models and growth curve models (Singer, 1998). Secondly, MLwiN 2.02 (Rasbash et al., 2000) provides a wide range of multilevel models together with plotting diagnostics.

6. Results

The results of the BMGCM reveal interesting improvements in the estimates of school effects and correlations in comparison with UMGCMs. First and foremost, the level-1 correlation for the joint growth processes for mathematics and reading comprehension is 0.17 and significant at a 5% level ($p < 0.0001$) indicating the need of fitting a bivariate model instead of two separate univariate models to the data.

Looking at the results of the first column of Table 2a for the student level of the UMGCM, it is clear that all the variance-covariance parameters are significant except for the covariance between the pupil's status and growth in reading comprehension. This seems to indicate that the pupils' level for reading comprehension at the end of grade 3 has no significant relationship with the pupils' growth thereafter. However, the negative correlation between the pupils' status in mathematics and growth in mathematics (-0.182) is significant at a 5% level. Pupils with a high mathematics score at the end of grade 3 generally grow less between the end of grade 3 and end of grade 6.

The second column of Table 2a shows the results of the BMGCM with four extra parameters rendering the possibility to answer many more research questions. The non-significant correlation between pupils' status and growth in reading comprehension for the UMGCM is now significant under the BMGCM with value -0.085. This correlation is in the same direction as that between pupils' mathematics status and growth though weaker. Also the effect of pupils' status on growth in mathematics is significant and seems stronger in the BMGCM (-0.222) than in the UMGCM (-0.182). The cross covariances were all significant with corresponding correlations of 0.691 between pupils' mathematics intercept and reading comprehension intercept. Pupils with high scores in mathematics also tend to have high scores in reading comprehension at the end of grade 3.

Table 2a

Random Effects of the univariate multilevel growth curve models (UMGCM) compared with bivariate multilevel growth curve models (BMGCM) with 4 measurements for Mathematics and Reading Comprehension (***Student Level Variance Parameters)***

	UMGCM			BMGCM		
Variance parameter	Estimate	Std error	Correlation	Estimate	Std error	Correlation
Math status	57.783	1.254	1	67.629	1.435	1
Math slope	1.058	0.081	1	0.594	0.087	1
Read status	36.926	0.879	1	47.184	1.038	1
Read slope	0.861	0.081	1	0.888	0.083	1
Covariance parameter						
Math status - slope	-1.425	0.251	-0.182	-1.408	0.293	-0.222
Read status - slope	**-0.325**	**0.205**	**0.058**	-0.555	0.243	-0.086
Math status - Read status	/	/		39.025	1.011	0.691
Math status - Read slope	/	/		1.255	0.279	0.162
Math slope - Read status	/	/		-0.501	0.263	-0.095
Math slope - Read slope	/	/		0.284	0.062	0.391

Estimates in bold are not significant at a 5% level. Mat=mathematics, Read=reading comprehension, status is the student intercept at the end of grade 3 and slope= linear growth.

Table 2b

Random Effects of the univariate multilevel growth curve models (UMGCM) compared with bivariate multilevel growth curve models (BMGCM) with 4 measurements for Mathematics and Reading Comprehension (***School Level Variance Parameters)***

	UMGCM			BMGCM		
Variance parameter	Estimate	Std error	Correlation	Estimate	Std error	Correlation
Math status	17.853	2.121	1	15.440	1.917	1
Math slope	0.961	0.121	1	1.079	0.137	1
Read status	14.044	1.655	1	12.373	1.508	1
Read slope	0.456	0.068	1	0.676	0.093	1
Covariance parameter						
Math status - slope	-1.665	0.390	-0.402	-1.231	0.383	-0.302
Read status - slope	-0.681	0.249	-0.269	-0.712	0.274	-0.246
Math status - Read status	/	/		9.277	1.439	0.671
Math status - Read slope	/	/		**-0.050**	**0.299**	**-0.015**
Math slope - Read status	/	/		**-0.474**	**0.325**	**-0.130**
Math slope - Read slope	/	/		0.497	0.090	0.582

Estimates in bold are not significant at a 5% level. Mat=mathematics, Read=reading comprehension, status is the student intercept at the end of grade 3 and slope= linear growth.

Another positive correlation of 0.162 was observed between the pupils' mathematics intercept and reading comprehension slope. This means the pupils with a high score in mathematics at the end of grade 3 tend to grow faster in reading comprehension subsequently. A correlation of 0.391 is estimated between pupils' mathematics slope and reading comprehension slope. This indicates that fast growing pupils in mathematics also grew fast for reading comprehension. There is a significant and negative correlation for the pupils' reading comprehension intercept and slope in mathematics (-0.095). The negative correlation means that high achievers in reading comprehension at the end of grade 3 generally had a slower growth in mathematics.

The results for the school level presented in Table 2b show some positive and significant correlations between the average mathematics and reading comprehension intercepts (0.672) and between their slopes (0.581). Schools with high end of year 3 scores in mathematics also have high scores in reading comprehension. Similarly schools with a steeper average slope in mathematics turn to have a steeper slope in reading comprehension too. The Table 2b results also show negative and significant correlations between the average intercept and average growth of schools for both mathematics and reading comprehension of -0.301 and -0.246 respectively. This means that schools with a high average mathematics score at the end of year 3 tend to have a slower average growth in mathematics during the subsequent grades, and so do the schools with high average reading comprehension. There were two correlations not significant at the 5% level for the relationship between schools' average intercept in mathematics and average growth in reading comprehension and between the schools' average intercept in reading comprehension and average growth in mathematics. These results may suggest that the school average growth in mathematics is not influenced by the average reading comprehension at the end of grade 3 and that the growth in reading comprehension is not influenced by the mathematics status too.

After considering that a bivariate growth model was the better model compared with two separate univariate growth models, the bivariate transition growth model was fitted to handle the difference in number of measurement occasions for reading comprehension and mathematics. The bivariate transition multilevel growth curve model (BTMGCM) is suggested in this study not only to solve inequality in the number of measurement occasions between the two outcome variables but also as a means of answering the fifth research question of the current study. In this special design of the transition model, the previous measurement covariates are constructed as changes in the mathematics achievement of the pupils between the beginning and end of first grade and between the end of grade 1 and the end of grade 2. Two such second order transition growth models are fitted. The BTMGCM (I) includes the two covariates (math2_1 and math3_2) as main effects only and the BTMGCM (II) adds the interactions between the two covariates and the time variable.

$$E(Y_{ijk}) = \begin{cases} \beta_{01} + \beta_{11} t_{ijk} + \alpha_{10} math2\_1 + \alpha_{11} math3 - 2 + \alpha_{12} math2\_1 * t_{ijk} + \alpha_{13} math3\_2 * t_{ijk} & if\ z_{1ijk} = 1 \\ \beta_{02} + \beta_{12} t_{ijk} + \alpha_{20} math2\_1 + \alpha_{21} math3\_2 + \alpha_{22} math2\_1 * t_{ijk} + \alpha_{23} math3\_2 * t_{ijk} & if\ z_{2ijk} = 1 \end{cases}$$

Table 3

Comparing the fixed effects estimates and model fits of a bivariate multilevel growth curve model (BMGCM) and two bivariate transitional multilevel growth curve models (BTMGCM (I) and BTMGCM (II)).

	BMGCM	**BTMGCM (I)**	**BTMGCM (II)**
Fixed effects parameter	Estimate (std error)	Estimate (std error)	Estimate (std error)
Intercept Reading Comp	43.224 (0.245)	46.051(0.362)	46.633 (0.395)
Intercept Mathematics	90.571 (0.312)	88.809 (0.479)	88.884 (0.497)
Time*Reading Comp	5.249 (0.068)	5.305 (0.071)	4.518 (0.135)
Time*Mathematics	5.563 (0.082)	5.584 (0.087)	5.248 (0.144)
Math2_1*Reading Comp	/	-0.145 (0.017)	-0.168 (0.017)
Math2_1*Mathematics	/	0.066 (0.019)	0.055 (0.018)
Math3_2*Reading Comp	/	**-0.014 (0.019)**	-0.039 (0.018)
Math3_2*Mathematics	/	0.179 (0.021)	0.185 (0.022)
Math2_1*Time*Read Comp	/	/	0.030 (0.005)
Math2_1*Time*Mathematics	/	/	0.018 (0.005)
Math3_2*Time* Read Comp	/	/	0.033 (0.005)
Math3_2*Time*Mathematics	/	/	**0.008 (0.006)**
-2loglikelihood	238524.3	189018.9	189005.5
Degrees of freedom	28	32	36

Estimates in bold are not significant at a 5% level. Math2_1=change in mathematics between start of grade 1 and end of grade 1, Math3_2= change in mathematics between end of grade 1 and end of grade 2. Reading Comp=reading comprehension, status is the student intercept at the end of grade 3.

The results of Table 3 show quite some differences between the fixed effects estimates of the BMGCM and the two versions of the BTMGCM. The BTMGCM (I) considers the dependence on the main effects of the two previous growths in mathematics (Math2-1 and Math3-2) and BTMGCM (II) also includes the interaction effect of these previous changes with time. The deviance statistics can be used to compare the fits of the models using the difference in the loglikelihood values and difference in degrees of freedom and a chi-square distribution as the null distribution for the likelihood ratio test statistic. Comparing BMGCM and BTMGCM (I) indicate a deviance value of 49506.3 with only 4 degrees of freedom is very significant ($p <$ 0.0001) indicating that there will be a significant loss in information by trying to reduce the BTMGCM (I) to BMGCM. A similar conclusion is established between

BTMGCM (I) and BTMGCM (II) in favour of BTMGCM (II) (p-value =0.017). The estimates of the earlier prior change in mathematics (start and end of grade 1) affect the pupils’ growth in both reading comprehension and mathematics from the end of grade 3 to the end of grade 6. On the other hand the later prior change in mathematics (between end of grade 1 and end of grade 2) seems to impact only the pupils’ subsequent growth in reading comprehension not their growth in mathematics.

A summary of the fixed effects result is presented as growth profiles for the UMGCM, BMGCM and the final bivariate transition growth curve model in Figure 3. The graphs of Figure 3(a) reveal that the average pupil score in reading comprehension at the end of grade 3 is underestimated by the univariate model (ReadUMGCM) and improved by the bivariate model (ReadBMGCM). In any case, the BTMGCM (for pupils with 1SD difference in prior math achievement (ReadBTMGCM+1SD) for both Math2_1 and Math3_2), gives the largest estimates for the reading comprehension score at the end of grade 3. A possible explanation for the underestimation by the UMGCM and BMGCM is that these models assume falsely that the growth processes only started at the end of grade 3. The BTMGCM (II) seems to indicate that growth in mathematics of the previous grades has a role to play in the average reading comprehension performance of pupils at the end of the third grade.

The graphs of Figure 3(b) show no big difference between the BMGCM and BTMGCM for their estimates of the average score in mathematics at the end of grade 3. This means that the change in prior mathematics achievement has a bigger influence on pupils’ subsequent development in reading comprehension than in mathematics.

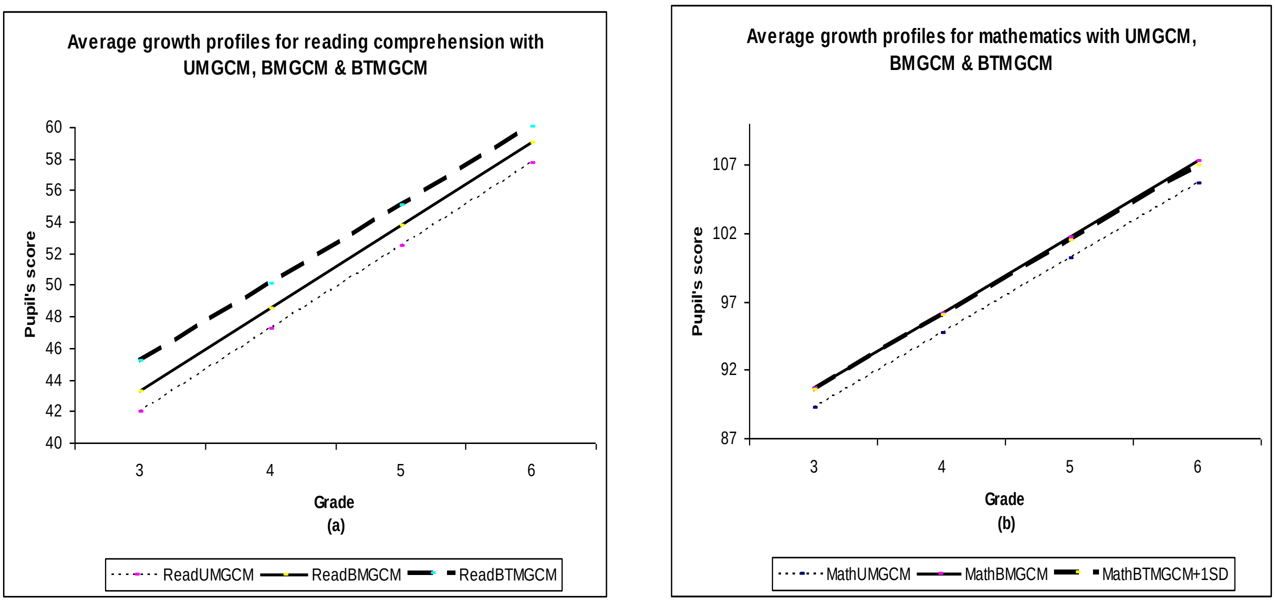
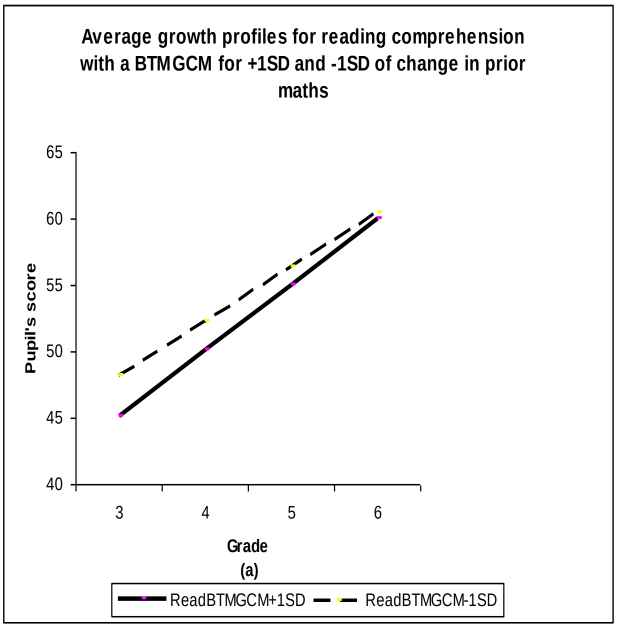


Figure 3. Average growth profiles for the UMGCM, BMGCM and BTMGCM (for +1SD difference in prior change) a) reading comprehension and b) mathematics.

Also the growth in pupils' reading comprehension is larger in the BMGCM probably because this model assumes that this growth is only due to the instruction for reading comprehension. There seems to be quite a bit of contribution of skills gained due to the first two years of mathematics instruction to subsequent performances in reading comprehension.

Figure 4(a) shows that on the one hand , pupils with a plus one standard deviation difference (+1SD) in prior mathematics (Math2_1) achievement (ReadBTMGCM+1SD) tend to have a lower end of grade 3 reading comprehension score but a steeper growth subsequently. On the other hand it indicates that, pupils with minus one standard deviation difference (ReadBTMGCM-1SD), show a less steep growth in reading comprehension. Not surprising though, pupils with a +1SD difference (MathBTMGCM+1SD) in prior mathematics achievement showed a steeper growth than pupils with a -1SD difference (MathBTMGCM-1SD) from the end of grade3 to the end of grade 6.



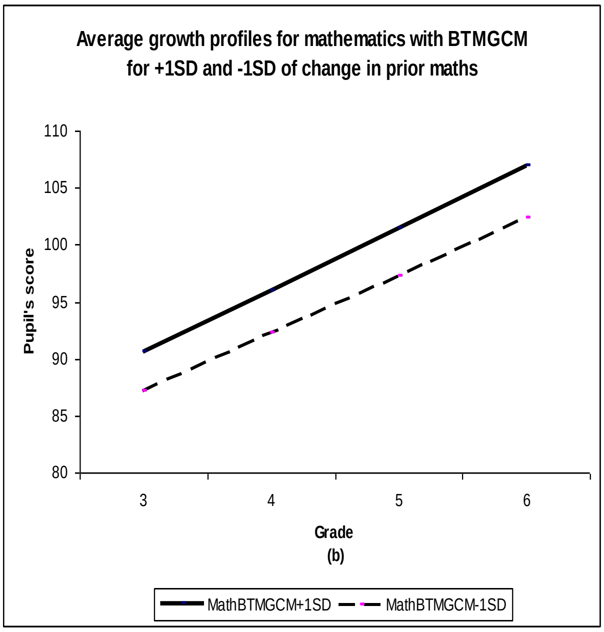


Figure 4. BTMGCM profiles for pupils with +1SD and -1SD of difference in prior mathematics achievement a) Reading comprehension and b) Mathematics.

After considering the BTMGCM with earlier and later prior change in mathematics achievement as covariates, the results of variance-covariance parameters of the BMGCM are then compared with those of the BTMGCM and presented in Table 4.

Table 4a

Estimates of variance and covariance components for the bivariate multilevel growth curve model (BMGCM) compared with the bivariate transition multilevel growth curve model (BTMGCM (I)) (***Student Level Variance Parameters)***

	BMGCM			BTMGCM (I)		
Variance parameter	Estimate	Std error	Correlation	Estimate	Std error	Correlation
Math status	67.629	1.435	1	62.776	1.520	1
Math slope	0.594	0.087	1	0.599	0.094	1
Read status	47.184	1.038	1	46.355	1.165	1
Read slope	0.888	0.083	1	0.882	0.091	1
Covariance parameter						
Math status – Math slope	-1.408	0.293	-0.222	-1.884	0.318	-0.307
Read status – Read slope	-0.555	0.243	-0.086	-0.552	0.268	-0.086
Math status - Read status	39.025	1.011	0.691	37.922	1.107	0.703
Math status - Read slope	1.255	0.279	0.162	0.984	0.307	0.132
Math slope - Read status	-0.501	0.263	-0.095	-0.667	0.254	-0.127
Math slope - Read slope	0.284	0.062	0.391	0.245	0.067	0.337

Math=mathematics, Read=reading comprehension, status is the student intercept at the end of grade 3 and slope= linear growth.

Table 4b

Estimates of variance and covariance components for the bivariate multilevel growth curve models (BMGCM) compared with the bivariate transition multilevel growth curve model (BTMGCM (I)) ***(School Level or Between Variance Parameters)***

	BMGCM			BTMGCM (I)		
Variance parameter	Estimate	Std error	Correlation	Estimate	Std error	Correlation
Math status	15.440	1.917	1	13.746	1.839	1
Math slope	1.079	0.137	1	1.177	0.158	1
Read status	12.373	1.508	1	9.391	1.300	1
Read slope	0.676	0.093	1	0.676	0.099	1
Covariance parameter						
Math status – Math slope	-1.231	0.383	-0.302	-1.606	0.408	-0.399
Read status – Read slope	-0.712	0.274	-0.246	-0.766	0.266	-0.304
Math status - Read status	9.277	1.439	0.671	8.023	1.325	0.706
Math status - Read slope	**-0.050**	**0.299**	**-0.015**	**-0.303**	**0.301**	**-0.099**
Math slope - Read status	**-0.474**	**0.325**	**-0.130**	-0.893	0.329	-0.268
Math slope - Read slope	0.497	0.090	0.582	0.579	0.102	0.649

Estimates in bold are not significant at a 5% level using the Wald test. Math=mathematics, Read=reading comprehension, status is the student intercept at the end of grade 3 and slope= linear growth.

The results are quite similar in terms of the direction of the covariance though the BMGCM seem to yield higher estimates for most of the parameters as compared to the BTMGCM (I). However, one conspicuous difference is the significant correlation (-0.268) between the average school intercept in reading comprehension and average slope in mathematics for the BTMGCM, which is not significant in the BMGCM. This means schools with a higher average score in reading comprehension at the end of grade 3 do not grow as fast in mathematics in comparison to schools with a lower average score. In other words, the higher the school's average reading comprehension score at the end of grade 3, the lower the growth of the school in mathematics from the end of grade 3 to the end of grade 6.

The school effect estimates for mathematics and reading comprehension on the pupils' status and growth were obtained for the different growth models described previously in this paper. The school effects are estimated as the proportion of variance accounted by the school level compared to that at the pupil level (Anumendem, De Fraine, Onghena, & Van Damme, in press). The results of the school effect estimates under the three different models are summarised in Table 5 below. The results indicate that univariate growth curve models seem to overestimate the effect of schools on the pupils at the end of grade 3 for both reading comprehension (27.5%) and mathematics (23.6%). The improved BMGCM estimates the same effects as 22.8% for reading comprehension and 18.6% for mathematics. However, the UMGCM seem to underestimate the school effect on growth with estimates of 34.6% and 47.6% for reading comprehension and mathematics respectively. Looking at the same estimates using the BMGCM, they increase to 43.2% and 64.5% for reading comprehension and mathematics respectively. This means that assuming a joint bivariate growth in the pupils reading comprehension and mathematics achievement can result in larger estimates of the effects of schools on the pupils' growth than treating the outcomes as if they were independent of each other.

The BTMGCM is fitted to remove the false assumption that the bivariate growth process started at the end of grade 3 for both pupil outcomes ignoring the first three measurements of mathematics. It is possible that knowledge acquired during the first two grades of primary school, might be responsible not only in the better fit of the model as shown earlier but also help avoid the fallacy of missing at random assumption for this data structure. With the BTMGCM, changes were observed in the school effect estimates on the pupils' status in reading comprehension (16.8%) and the growth in mathematics (66.5%).

Table 5

The school effect estimates for the pupils' status and growth in reading comprehension and mathematics at the end of grade 3, for the three growth curve models described.

At the end of Grade 3			
Reading comprehension			
School effect on student status	27.5%	22.8%	16.8%
School effect on student linear growth	34.6%	43.2%	43.4%
Mathematics achievement			
School effect on student status	23.6%	18.6%	18.0%
School effect on student linear growth	47.6%	64.5%	66.5%

This table also indicates that schools have a larger effect on their pupils' mathematics growth than on their growth in reading comprehension.

7. Discussion and Conclusion

Considering the correlation results, the UMGCM shows no significant relationship between the pupils' status and growth in reading comprehension. However, this effect became significant in the BMGCM indicating better power in the latter model. The BMGCM results also illustrate that pupils' who are higher achievers in mathematics at the end of grade 3 are also higher achievers in reading comprehension at the end of grade 3. Pupils with a stronger growth in mathematics also show a stronger growth in reading comprehension. Pupils who score high in mathematics at the end of grade 3 grow more in reading comprehension subsequently. However, pupils who are higher achievers in reading comprehension at the end of grade 3 experience a slower growth in mathematics.

At the school level, the correlation between the average reading comprehension at the end of grade 3 and the average mathematics at the end of grade 3 was significant. This implies schools with high achieving pupils in mathematics also have high achievers in reading comprehension. Also schools with a high average growth in reading comprehension have a high average growth in mathematics.

The significant correlation between the overall mathematics achievement and reading comprehension growth profiles is an indication that statistically a BMGCM approach is more appropriate. The BMGCM also resulted in about 35% increase in the school effect estimate on pupils' growth in mathematics and about 25% increase for pupils' growth in reading comprehension. However, the school effect on the pupils' status dropped by 21% and 17% for mathematics and reading comprehension respectively. The univariate model seems to overestimate the differences between schools by the end of grade 3. The UMGCM also underestimates the effect of schools on pupils' linear change in mathematics as well as reading comprehension. This is far enough evidence of the need for more than one criterion to better estimate the effects of schools on primary school children. The results of the BMGCM also show more clearly that the effect of schools is more pronounced on the pupils' growth criterion than on the status. The school effects on the pupils' growth is about 3.7 times larger than on the pupils' status at the end of grade 3 for mathematics achievement and about 2.6 times more for reading comprehension. The same comparison at the beginning of grade 3 indicates that the school effect on pupils' growth is about 1.9 times for mathematics and about 1.5 times for reading comprehension. Analysis at the end of grade 6 (not presented) corroborates these findings and also results of previous studies on longitudinal data (Raudenbush, 1989; 1995). It is nonetheless advisable to fit models with many more first stage criteria than just two as in this study (mathematics and reading comprehension), in order to generalise these findings. This can be considered a relative advantage of the choice of the growth criterion over the status criterion at the second stage for both first stage criteria because schools seem to have more impact on it.

The results of the transition model showed that changes in mathematics achievement in the first and second grade could predict the change in reading comprehension in the later grades. When prior growth in mathematics is taken into account, the schools seem to help the low achievers at the end of grade 3 to catch up with their higher achieving peers by the end of grade 6. In other words pupils with a larger gain in earlier mathematics achievement grow faster in reading comprehension subsequently though they are low achievers at the end of grade 3.

The BMGCM showed no significant correlation between the schools' average mathematics score at the end of grade 3 and the growth in reading comprehension. A similar non-significant result is obtained for the correlation between the school's average score in reading comprehension at the end of grade 3 and the growth in mathematics. However, with the introduction of the BTMGCM, the correlation between the schools' average score in reading comprehension and the schools' average growth in mathematics became significant. The school level correlations indicate that schools that are effective in the pupils' average mathematics achievement are also effective in their average reading comprehension attainment on both the status and growth criteria.

From the findings of this study, we recommend that researchers in the field of school effectiveness should consider multiple criteria to enable this field of research come up with improved school effect estimates. We encourage researchers to make more use of longitudinal data and the two-stage criteria proposed, to enable researchers to answer a wider range of scientifically relevant questions to school effectiveness research. We acknowledge the computational difficulties that such complex models with multiple stage 1 and 2 criteria will bring while recognising the invaluable contribution it will make to the field of educational effectiveness. The multivariate transition model proposed in this paper can be used by researchers to avoid false missingness assumptions or even the loss of data and in addition can answer other very relevant research questions.

We were not able in our study to investigate the more appropriate joint causal change relationship. It would have been interesting for example to investigate whether the change in mathematics influences the change in reading comprehension and not the other way around. Nevertheless, this study serves as a strong foundation on which more complex educational research methods can be developed.

GENERAL DISCUSSION

The last part of this dissertation summarises the main findings of the four manuscripts presented and looks at alternative approaches. There are three sections in this general discussion. The first section summarises and discusses the main results. The second section focuses on the contributions of this research to educational effectiveness research in particular and to other research fields. Finally, section three investigates possible alternative approaches and directions for future research.

1. Summary and discussion of the main results

Manuscript 1 investigated growth curve models as a statistical method which is fast gaining grounds in various research domains in general and educational effectiveness research in particular. In this dissertation, a broader framework was chosen and the effects of changes in time coding were illustrated for a quadratic three-level growth curve model. Though, the formula for estimating school effect for random slopes GCMs was derived based on models with three levels, an elaboration is made to cover models with more than three levels in the hierarchy and more random coefficients. The main conclusion of this manuscript is that the estimation of the size of the school effect is strongly determined by the choices of the time scale. Also, the hypothesis that school effects are larger when the intercept refers to a later point in time was only partially confirmed. The current study also contributes to the discussion in the field of educational effectiveness whether school effects are larger for student growth (slope) or for student status (intercept). However, no conclusion can be made in terms of how the school effects will behave (increase or decrease) when the time coding is changed without considering what type of model and outcomes are used. Though, the coding of time can be chosen to solve substantive problems, researchers should be very explicit on the time coding used in order to make replications or extensions of their studies possible. Altering the initial status, however without clearly stating so can make research results and interpretation to lose their transparency. It is therefore important to get a common framework on which educational effectiveness researchers can base the coding of time in order to gain general acceptable results.

Manuscript 2 looked at the traditional way of splitting total variance into two components that is the measurement error and random effects of the data and compared it with a more proper version including a serial component. The results showed that more often than not, repeated measurement data will be serially correlated which confirmed findings of other researches in longitudinal data (Diggle et al., 1994). The results also indicated that the random effects do not always overshadow the presence of the serial correlation. Another key finding of this manuscript is that the use of the semi-variogram in repeated measurement data with as few as four time points is still a very useful tool for the detection of the presence of serial correlation. However, more focus should be on modelling the presence of serial correlation than the exact functional form of serial correlation present.

The results of manuscript two showed for the first time that serial correlation is also present at level 2. The results indicated that school effects on the student status and student growth are in general underestimated without level 1 serial correlation correction. This may be due to the overestimation of the level 1 variance. The impact of level 1 and 2 serial correlation correction was different for the two students' outcomes. This indicated that the two outcomes were different stochastically in terms of variability around the students' and schools' growth profiles. The advice is therefore to include a serial correlation component in GCMs for educational data no matter the type of outcome considered. And also to consider unequal time spacing between measurement occasions when few repeated measurements are foreseen.

Manuscript 3 looked further into the research work of manuscript two and introduced a GCM with a serial correlation at each level of the hierarchy. A double serial correlation multilevel GCM was established and applied to the LOSO data for the case of a three-level data. According to the results of this manuscript, the modelling of serial correlation greatly improved the estimates of school effects on both the students' academic achievement in Dutch and the students' well-being at school. The simulation study results showed that such longitudinal data with level-1 and -2 serial correlations can exist in many situations under different conditions. The simulation results also hinted that the estimation of the school effect on growth is less affected by changes in number of schools. From these findings, it is recommended that checks for the presence of serial correlation should always be made whenever longitudinal data are used to model growth. It also emphasized the gain in model fit and validity of school effects estimation with the use of the complex double serial correlation for growth curve models with three levels.

Manuscript 4 investigated multilevel growth curve models with more than one school effectiveness criterion. In order to investigate changes in student outcome over time, longitudinal data are invaluable. The use of growth curve as a statistical methodology introduced another level of choice of criterion beyond the choice of pupils' outcome variables. This manuscript delineated clearly between these two stages of school effectiveness criteria: Firstly, the choice of the dependent or outcome variables with a correlation strong enough to lend credence to a multivariate model instead of separate univariate models and secondly, the choice of growth parameters used to estimate the school effect.

2. Usefulness of the research to educational effectiveness

Firstly, based on the results of this study, researchers are advised to be very explicit on the time coding used. Studies should report very clearly at what point in time the intercept is referring to. In fact, it is argued that changing the time coding comes down to changing the research question. Also, researchers can choose the coding of time to solve substantive problems, stretching from easily and readily interpretable parameter estimates, to interests in the understanding of school effects at particular points in time.

Secondly, the research results of this dissertation showed that random effects do not overshadow the presence of the serial correlation always. Therefore, modelling only the random intercept effects cannot correct for serial correlation in growth curve models with random intercepts and slopes. Another key point that researchers have to bear in mind, is that the use of the semi-variogram in repeated measurement data with as few as four time points is possible and practical. In order to properly correct for serial correlation in longitudinal data, researchers are advised to consider unequal time spacing between measurements when they intend to do as few as four repeated measurements. This can enormously reduce cost and yet properly model the growth in the outcome of interest.

Thirdly, from the findings of this study, it is recommended that researchers in the field of school effectiveness should consider multiple criteria to enable this field of research come up with improved school effect estimates. Researchers are encouraged to make greater use of longitudinal data and the two-stage criteria proposed. This model, though a little more complex, can be employed by researchers to answer a wider range of scientifically relevant questions to school effectiveness research. The multivariate transition growth curve model proposed in this dissertation can be used to avoid false missing data assumptions or even the loss of data and in addition enable researchers answer other very relevant research questions.

3. Limitations and directions for future research

3.1. *Gain in extending unconditional growth curve models to conditional models*

In the dissertation, only raw school effects were calculated. This means that no covariates[5] (such as student prior achievement, intelligence, sex or socio-economic status) were added to the model in order to make a more fair comparison between schools (Teddlie & Reynolds, 2003). It is expected that such net school effects will be smaller than the raw school effects though research on this decrease for conditional quadratic growth curve models are rare. Unconditional multilevel growth curve models (UMGCM) usually have only time as the independent variable and so time is not really considered as a covariate. Of course, UMGCM are flexible enough for different independent variables at the different levels to be taken into account. However, UMGCM, despite the presence of only the time variable and higher orders of time are already complex enough to model the growth processes involved. The more complex the model, the better it will seem to represent the growth process but caution is always required not to exaggerate the number of parameters to estimate in the model (parsimony).

Unconditional quadratic multilevel growth curve models already have two covariates which are the linear time and quadratic time. Such a model for a three-level data is already very complex when the variance-covariance matrices for the school (3x3) and student (3x3) levels are considered to be unstructured. On the other hand if the research interest is to estimate differential school effect then one may be allowed to include few of these covariates but this creates the risk of running out of degrees of freedom because the number of parameters to estimate quickly skyrockets. A trade-off is therefore needed to weight parsimony of the growth curve model being used. Since the prior achievement seems always to explain most of the school level variance and this is already take into account when UMGCM are used. CMGCM do

[5] Except for the transitional model in manuscript 4, where the prior growth in mathematics was introduced by two covariates.

not necessarily need to correct for prior achievement or aptitude any more but time-varying covariates.

3.2. Multilevel growth curve modelling versus latent growth modelling

Due to the increasing interest in studying changes in pupils' outcomes by educational effectiveness researchers, two main statistical methodologies have been proposed to answer research questions involving multiple student outcomes. This dissertation has investigated multiple effectiveness criteria through multivariate multilevel growth curve models. Nonetheless, the use of a bi-parallel process multilevel latent growth is also possible. This model comes down to modelling two dependent outcome variables as a multivariate variable with the repeated measurements of the two variables the repeated measurements becoming variables in themselves and assumed independent of each other. This is unlike the bivariate multilevel growth curve model (BMGCM) which considers independence only for the repeated measurements of the same outcome variable but allows the dependence of the two growth processes. The BMGCM enables a statistical test for the need of a multivariate growth model over separate univariate growth models. Arguments have been made by users of latent growth model which claim that constraints can always be used on latent growth models to make both approaches similar in terms of their outputs. While this is true under a very restricted class of models, deviations from the fundamental theory behind them are inevitable due to too many constraints. There is therefore a great need for more advanced statistical research to enable a proper comparison of the two approaches in terms of their similarities and dissimilarities.

REFERENCES

Adams, M. J., Treiman, R., & Pressley, M. (1998). Reading, writing, and literacy. In I. E. Sigel & K. A. Renninger (Eds.), *Handbook of child psychology, vol.4: Child psychology in practice* (pp. 275-355). New York: Wiley.

Anumendem, D. N., De Fraine, B., Onghena, P., & Van Damme, J. (in press). The impact of coding time on the estimation of school effects. *Quality and Quantity*.

Bauer, D. J. & Cai, L. (2009). Consequences of unmodeled nonlinear effects in multilevel models. *Journal of Educational and Behavioral Statistics, 34,* 97-114.

Biesanz, J. C., Deeb-Sossa, N., Aubrecht, A. M., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9,* 30-52.

Bosker, R. J., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. In R.J. Bosker, B.P.M. Creemers & J. Scheerens (Eds.), Conceptual and methodological advances in educational effectiveness research [Special issue]. *International Journal of Educational Research*, *21*, 159-180.

Box, G. E., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis – forecasting and control* (4th Ed.). New Jersy: Wiley.

Bull, R., & Johnston, R. S. (1997). Children’s arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, *65*, 1-24.

Cleveland, W. S. & Devlin, S.J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of American Statistics Association, 83*, 596–610.

Cools, W., Van den Noortgate, W. & Onghena, P. (2008). ML-Des: A program for designing efficient multilevel studies. *Behavior Research Methods, 40,* 236-249.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Curran, P. J., & Hussong, A. M. (2002). Structural equation modeling of repeated measures data: latent curve analysis. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data* (pp. 59-85). Mahwah, NJ: Lawrence Erlbaum.

Davis, P., & Scott, A.: The effect of interviewer variance on domain comparisons. *Survey Methodology, 21,* 99-106 (1995).

De Fraine, B., Van Damme, J., & Onghena, P. (2007). A Longitudinal analysis of gender differences in academic self-concept and language achievement: A multivariate latent growth curve approach. *Contemporary Educational Psychology, 32,* 132-150.

De Fraine, B., Van Landeghem, G., Van Damme, J., & Onghena, P. (2005). An analysis of well-being in secondary school with multilevel growth curve models and multilevel multivariate models. *Quality and Quantity, 39,* 297-316.

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading, 6,* 51-77.

De Maeyer, S., Van den Bergh, H., Rymenans, R., Van Petegem, P., & Rijlaarsdam, G. (2010). Effectiveness criteria in school effectiveness studies: Further research on the choice for a multivariate model. *Educational Research Review, 5,* 81-96.

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

Diggle, P. J. (1988). An approach to the analysis of repeated measures. *Biometrics, 44*, 959-971.

Diggle, P. J. (1990). *Time Series: A biostatistical introduction.* Oxford: Oxford University Press.

Diggle, P. J., Heagerty, P. J., Liang, K. –Y., & Zeger, S. L.(2002). *Analysis of longitudinal data (2nd* ed.). Oxford Science Publications, Oxford: Clarendon Press.

Diggle, P. J., Liang, K. –Y., & Zeger, S. L.(1994). *Analysis of longitudinal data*. Oxford: Clarendon Press.

Donlan, C., & Gourlay, S. (1999). The importance of non-verbal skills in the acquisition of place-value knowledge: Evidence from normally-developing and language-impaired children. *British Journal of Developmental Psychology, 17,* 1-19.

Dowker, A. (1998). Individual differences in normal arithmetical development. In C. Donlan (Ed.), *The development of mathematical skills* (pp 275-302). East Sussex, U.K.: Psychology Press.

Duncan, T. E., Duncan, S. C., & Strycker, L. A.(2006). *An introduction to latent variable growth curve modeling: Concepts, issues and applications*. Mahwah, NJ: Laurence Erlbaum.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues and applications*. Mahwah, NJ: Erlbaum.

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

Ferron, J., Dailey, R., Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research, 37,* 379–403.

Fieuws, S., & Verbeke, G. (2004). Joint modeling of multivariate longitudinal profiles: pitfalls of the random effects approach. *Statistics in Medicine, 23,* 3093-3104.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis*. New Jersey: Wiley.

Garst, H. (2000). Longitudinal research using structural equation modeling applied to studies of determinants of psychological well-being and personal initiative in East Germany after the unification. Doctoral dissertation, University of Amsterdam.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49*, 363-383.

Goldstein, H. (1995). Multilevel models in educational and social research. New York: Oxford University Press.

Goldstein, H., Healy, M. J. R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine, 13*, 1643-1655.

Harring, J. R. (2009). A nonlinear mixed effects model for latent variables. *Journal of Educational and Behavioral Statistics, 34,* 293-318.

Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavior Research, 42,* 609-629.

Hox, J. (2002). *Multilevel analysis. Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Huitema, B. E. & McKean, J. W. (2007). An approved portmanteau test for autocorrelated errors in interrupted time-series regression models. *Behavior Research Methods, 39*, 343-349.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74*, 834-850.

Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: Findings of a two-year longitudinal study. *Journal of Educational Psychology, 94,* 586-597.

Kirsch, I., De Jong, J., LaFontaine, D., McQueen, J., Mendelovits, J., & Monseur, C. (2002). *Reading for change. Performance and engagement across countries. Results from PISA 2000.* Paris: OECD.

Kreft, I. G. G., de Leeuw, J., & Aiken, L. S. (1995). The effects of different forms of centering in hierarchical linear models. *Multivariate Behavior Research, 30,* 1-22.

Kwok, O., West, S. G. & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavior Research, 42,* 557-592.

Laird, N., & Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963-974.

Lange, N., & Laird, N. (1989). The effect of covariance structure on variance estimation in balance growth-curve models with random parameters. *Journal of American Statistical Association, 84*, 241-247.

Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2005). Mathematical performance predicts progress in reading comprehension among 7-years olds. *European Journal of Psychology of Education, 2,* 121-137.

Lesaffre, E., Asefa, M., & Verbeke, G. (1999). Assessing the goodness-of-fit of the Laird and Ware model: an example: the Jimma infant survival differential study. *Statistics in Medicine, 18,* 835-854.

Littel, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. & Schabenberger, O. (2006). *SAS system for mixed models* (2nd Ed.). Cary, NC: SAS Institute Inc.

Maas, C. J. M. & Snijders, T. A. B. (2003). The multilevel approach to repeated measures for complete and incomplete data. *Quality and Quantity*, *37,* 71-89.

Maes, F., Ghesquière, P., Onghena, P., & Van Damme, J. (2002). *Longitudinaal onderzoek in het basisonderwijs. Van doelstellingen tot onderzoeksopzet* [Longitudinal research in primary education. From objectives to research design]. LOA-rapport nr. 1. Leuven: Steunpunt LOA.

May, H., Supovitz, J. A., & Perda, D. (2004). A longitudinal study of the impact of America's choice on student performance in Rochester, New York, 1998-2003. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

Mayer, R. E. (1987*). Educational psychology: A cognitive approach. Boston*: Little, Brown.

McArdle, J. J. (1988). *Dynamic but structural equation modeling of repeated measures data*. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (2nd Ed.). New York: Plenum.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5,* 23-43.

Molenberghs, G., & Verbeke, G., (2006). *Models for discrete longitudinal data*. Springer Series in Statistics, New-York: Springer-Verlag.

Monk, D. H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis*, 14, 307-332.

Morrell, C. H., Pearson, J. D., Ballentine C. H. & Brant, L. J. (1995). Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association*, *90*, 45-53.

Muthén, B. (2004). Latent variable analysis. Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology in the social sciences* (pp. 345-368). London: Sage.

Muthén, L. K., & Muthén, B. O. (2006). *Statistical analysis with latent variables: Mplus User's Guide*. (4th Ed.). Los Angeles, CA: Authors.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., et al., (2000). *A user's guide to MlwiN*. London: Institute of Education.

Raudenbush, S. W. (1989). The analysis of longitudinal, multilevel data. *International Journal of Educational Research 13,* 721–740.

Raudenbush, S. W. (1995). Statistical models for studying the effects of social context on individual development. In Gottman, J. (Ed.). *The Analysis of Change* (pg. 165-201). Hillsdale, NJ: Lawrence Erlbaum.

Raudenbush, S. W. (2001a). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 50,* 501–525.

Raudenbush, S. W. (2001b). Towards a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New* *methods for the analysis of change* (pp. 35–64). Washington, DC: American Psychology Association.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods.* (2nd Ed.). London: Sage.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20,* 307-335.

Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research," plus supplemental questions. In: J.M. Gottman (Ed.), *The analysis of change* (pp. 3-65). Mahwah, NJ: Lawrence Erlbaum.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203-228.

Rogosa, D. R., Brandt, D., & Zimowski, M. A. (1982). Growth curve approach to the measurement of change. *Psychology Bulletin, 90,* 726-748.

Rowan, B., & Denk, C. E. (1982). *Modelling the academic performance of schools using longitudinal data: An analysis of school effectiveness measures and school and principal effects on school-level achievement.* San Francisco, Calif.: Far West Laboratory for Educational Research and Development.

SAS Institute Inc. (2003). *SAS User's Guide: Statistics* (9th Ed.). Cary, NC: SAS Institute Inc.

Scheerens, J. (1990). School effectiveness and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1,* 61-80.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness.* Oxford, UK: Pergamon.

Scheerens, J., & Creemers, B. P. M. (1989a). Conceptualising school effectiveness'. *International Journal of Educational Research, 13,* 689-706.

Siddiqui, O., Hedeker, D., Flay, B. R., & Hu, F. B. (1996). Intraclass correlation estimates in a school-based smooking prevention study: outcome and mediating variables, by gender and ethnicity. *American Journal of Epidemiology, 144,* 425-433.

Singer, J. D. (1998). Using SAS proc mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24,* 323-355.

Singer, J. D., & Willett, J. B. (2003) *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

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

Spyros, K. (2008). Computing power of tests of the variance of treatment effects in designs with two levels of nesting. *Multivariate Behavior Research, 43,* 327-352.

Stoel, R. D., & Van den Wittenboer, G. (2003). Time dependence of growth parameters in latent growth curve models with time invariant covariates. *Methods Psychology Research, 8,* 21-41.

Teddlie, C., Reynolds, D., & Sammons, P. (2000). The methodology and scientific properties of school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 55-133). London: Falmer Press.

Topping, K. J., & Fisher, A. M. (2003). Computerised formative assessment of reading comprehension: field trials in the UK. *Journal of Research in Reading, 26,* 267-279.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23,* 19-23.

Tymms, P. (1999). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading, 22*, 27-36.

Van Damme, J., & De Fraine, B. (2005, April). *The school effect on student status and growth: A longitudinal multilevel modelling study*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Van Damme, J., Opdenakker, M.-C., Van Landeghem, G., De Fraine, B., Pustjens, H., & Van de gaer, E. (2006). *Educational effectiveness: An introduction to* *international and Flemish research on schools, teachers, and classes*. Leuven: Acco.

Van de gaer, E., De Fraine, B., Pustjens, H., Van Damme, J., De Munter, A., & Onghena, P. (2009). School effects on the development of motivation toward learning tasks and the development of academic self-concept in secondary education: a multivariate latent growth curve approach. *School Effectiveness and School Improvement, 20,* 235-253.

Van den Noortgate, W., Onghena, P. (2006); Analysing repeated measures data in cognitive research: A comment on regression coefficient analyses. *European Journal of Cognitive Psychology, 18,* 937-952.

Van Keer, H. (2004). Fostering reading comprehension in fifth grade by explicit instruction in reading strategies and peer tutoring. *British Journal of Educational Psychology, 74,* 37-70.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer Series in Statistics, New-York: Springer-verlag.

Verbeke, G., Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & data Analysis , 23,* 541-556.

Verbeke, G., Lesaffre, E., & Brant L. J. (1998). The detection of residual serial correlation in linear mixed models. *Statistics in Medicine, 17,* 1391-1402.

Wilkins, J. L. M., & Ma, X. (2002). Predicting student growth in mathematical content knowledge. *Journal of Educational Research, 95,* 288-298.

Wilkins, J. L. M., & Ma, X. (2003). Modeling change in student attitude toward and beliefs about mathematics. *Journal of Educational Research, 97,* 52-63.

Willett, J. B. (1997). Measuring change: What individual growth modeling buys you. In E. Amsel & K. A Renninger (Eds), *Change and development: Issues of theory, method, and application* (pp. 213-243). Mahwah, NJ: Lawrence Erlbaum.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychology Bulletin, 116,* 363-381.

Willms, J. D. (1992). *Monitoring school performance: A guide for educators*. London: The Falmer Press.

Willms, J. D., & Raudenbush, S.W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, *26*, 209-232.

Wishart, J. (1938). Growth rate determination in nutrition studies with bacon pig and their analysis. *Biometrika*, *30*, 16-28.

Yang M., Goldstein H., & Heath A. (2000). Multilevel models for repeated binary outcomes: attitudes and voting over the electoral cycle. *Journal of the Royal Statistical Society: Series A, 163,* 49-62.

APPENDICES

Appendix A

Table A1

Results model for the student well-being outcome

	Random effects		-2loglikelihood(reml)
Model	Student	School	
Mod 1	Int lin_slope quad_slope	Int lin_slope quad_slope	5094.3 (6 rand effects)
Mod 2	Int lin_slope	Int lin_slope quad_slope	10846.8(5 rand effects)
Mod 3	Int lin_slope quad_slope	Int lin_slope	10907.1(5 rand effects)
Mod 4	Int lin_slope	Int lin_slope	10916.3(4 rand effects)

	G^2	Mixture of Chi-square	P-value
Mod 2 vs Mod 1	5752.5	$\chi^2_{5:6}$	p($\chi^2_{5:6}$ >5752.5) <0.0001
Mod 3 vs Mod 1	5812.8	$\chi^2_{5:6}$	p($\chi^2_{5:6}$ >5752.5) <0.0001
Mod 4 vs Mod 2	69.2	$\chi^2_{4:5}$	p($\chi^2_{4:5}$ >5752.5) <0.0001
Mod 4 vs Mod 3	9.2	$\chi^2_{4:5}$	p($\chi^2_{4:5}$ >5752.5) =0.0333

Table A2

Results for random effects estimates of model B for well-being

		Estimate(model based Std errors)				
Level	Cov	Original time	Time code A	Time code B	Time code C	Time code D
School	v_{11}	0.0124*	0.0099*	0.0089*	0.0112*	0.0192*
	v_{21}	-0.0016*	-0.0009*	-0.0001*	0.0013*	0.0027*
	v_{22}	0.0007*	0.0007*	0.0007*	0.0007*	0.0007*
Student	υ_{11}	0.2173*	0.1859*	0.1634*	0.1449*	0.1620*
	υ_{21}	-0.0179*	-0.0135*	-0.0091*	-0.0002*	0.0087*
	υ_{22}	0.0044*	0.0044*	0.0044*	0.0044*	0.0044*
Residual	σ^2	0.09694*	0.09694*	0.09694*	0.09694*	0.09694*
-2loglikelihood		14410.2	14410.2	14410.2	14410.2	14410.2

cov = covariance * significant at the 5% level

Table A3

Results for random effects estimates of model C for well-being

		Estimate(model based Std errors)					
Level	Cov	Original	Code A	Code B	Code C	Code D	Code E
School	v_{11}	0.0314*	0.0121*	0.0091*	0.0144*	0.0184*	0.0314*
	v_{21}	-0.0160*	-0.0045*	**0.0006**	**0.0009**	**0.0029**	-0.1599*
	v_{22}	0.0115*	0.0066*	0.0031*	0.0007*	0.0043*	11.529*
	v_{31}	0.0019*	0.0007*	**-0.0002**	-0.0008*	**0.0001**	0.1924*
	v_{32}	-0.0014*	-0.0011*	-0.0007*	**0.0001**	0.0008*	-14.324*
	v_{33}	0.0002*	0.0002*	0.0002*	0.0002*	0.0002*	18.834*
Student	υ_{11}	0.2017*	0.1765*	0.1782*	0.1862*	0.1344*	0.2017*
	υ_{21}	-0.0235*	**-0.0039**	0.0039*	-0.0037*	-0.0220*	-0.2349*
	υ_{22}	0.0272*	0.0184*	0.0116*	0.0039*	**0.0044**	27.222*
	v_{31}	**-0.0004**	-0.0026*	-0.0043*	-0.0062*	-0.0061*	-0.0390
	v_{32}	-0.0025*	-0.002*	-0.0015*	-0.0005*	**0.0006**	-24.612*
	v_{33}	0.0003*	0.0003*	0.0003*	0.0003*	0.0003*	25.147*
Residual	σ^2	0.091*	0.0904*	0.0905*	0.0905*	0.0905*	0.0905*
-2loglikelihood		14203.4	14203.4	14203.4	14203.4	14203.4	14203.4

Cov = covariance * significant at the 5% level

Appendix B

Table 1a

Results of model fit for the student well-being outcome

	Random effects		-2loglikelihood(reml)
Model	Student	School	
Mod 1	Int linslop quadslop	Int linslop quadslop	5094.3 (6 rand effects)
Mod 2	Int linslop	Int linslop quadslop	10846.8(5 rand effects)
Mod 3	Int linslop quadslop	Int linslop	10907.1(5 rand effects)
Mod 4	Int linslop	Int linslop	10916.3(4 rand effects)

	G^2	Mixture of Chi-square	P-value
Mod 2 vs Mod 1	5752.5	$\chi^2_{5:6}$	p($\chi^2_{5:6}$ >5752.5) <0.0001
Mod 3 vs Mod 1	5812.8	$\chi^2_{5:6}$	p($\chi^2_{5:6}$ >5812.8) <0.0001
Mod 4 vs Mod 2	69.2	$\chi^2_{4:5}$	p($\chi^2_{4:5}$ >69.2) <0.0001
Mod 4 vs Mod 3	9.2	$\chi^2_{4:5}$	p($\chi^2_{4:5}$ >9.2) =0.0333

Int=intercept ; linslop= linear slope; quadslop=quadratic slope; rand effects=random effects

Table 1b

Results of model fit for the student Dutch achievement outcome

	Random effects		-2loglikelihood(reml)
	Student	School	
Model 1:	Int linslop quadslop	Int linslop quadslop	19321.8
Model 2:	Int linslop	Int linslop quadslop	19325.5
Model 3:	Int linslop quadslop	Int linslop	19442.3

	G^2	Mixture Chi-square	P-value
Mod 2 vs Mod 1	3.7	$\chi^2_{5:6}$	p($\chi^2_{5:6}$ >3.7)=0.655
Mod 3 vs Mod 1	120.5	$\chi^2_{5:6}$	p($\chi^2_{5:6}$ >120.5) <0.0001

Int=intercept ; linslop= linear slope; quadslop=quadratic slope; rand effects=random effects

Table 2a

Comparing the -2loglikelihoods of three serial correlation functions for the students' well-being outcome

Models	Serial correlation	-2loglikelihood
Random intercept		
	Level 1	
Model 1	Gaussian	14550.5
Model 2	Exponential	14549.2
Model 3	Power	14549.2
	Level 2	
Model 4	Gaussian	14336.9
Model 5	Exponential	14353.1
Model 6	Power	14353.1
Random intercept & slope		
	Level 1	
Model 7	Gaussian	14281.0
Model 8	Exponential	14280.4
Model 9	Power	14280.4
	Level 2	
Model 10	Gaussian	14255.5
Model 11	Exponential	14271.3
Model 12	Power	14271.3

Table 3a

Three multilevel quadratic growth curve models with random intercepts and slopes

	Model1	Model 2	Model 3
Parameters			
Fixed effects			
Intercept	3.960 (0.0295)	3.955(0.0297)	3.945(0.0293)
Linear slope	-0.2400(0.0184)	-0.2398(0.0187)	-0.2346(0.0184)
Quadratic slope	0.0192(0.0024)	0.0194(0.0024)	0.0187(0.0024)
Variances			
School level			
Intercept	0.0302(0.0088)	0.0311(0.0091)	0.0139(0.0132)
Linear slope	0.0119(0.0037)	0.0122(0.0039)	0.0044(0.0062)
Student level			
Intercept	0.2196(0.0079)	0.1119(0.0095)	0.1108(0.0093)
Linear slope	0.0046(0.0004)	0.00001(-)	0.00001(-)
Residual	0.0948(0.0018)	0.0820(0.0024)	0.0799(0.0023)
-2loglikehood	14300.1	14221.1	14129.5
Degree ff	13	15	17
Level 2 serial	/	/	0.0045(0.0036)
Level 1 serial	/	0.0777(0.0069)	0.0798(0.0067)
Correlation			
(intercept vs slope)			
School level	-0.0158(0.0053)	-0.0163(0.0054)	-0.0061(0.0067)
Student level	-0.0185(0.0015)	-0.0022(0.0009)	-0.0021(0.0009)
School effect			
Initial status	12%	21%	11.1%
Student growth	72%	99%	99%

Model 1: without serial correlation; Model 2: with student level serial correlation correction only; Model 3: with both school and student level serial correlation corrections.