Research Papers in Education

ISSN: 0267-1522 (Print) 1470-1146 (Online) Journal homepage: http://www.tandfonline.com/loi/rred20

Do the teacher and school factors of the dynamic model affect high- and low-achieving student groups to the same extent? a cross-country study

Gudrun Vanlaar, Leonidas Kyriakides, Anastasia Panayiotou, Machteld Vandecandelaere, Léan McMahon, Bieke De Fraine & Jan Van Damme

Research Papers in Education, 2015
http://dx.doi.org/10.1080/02671522.2015.1027724

Do the teacher and school factors of the dynamic model affect high- and low-achieving student groups to the same extent? a cross-country study

Gudrun Vanlaar[a,b]*, Leonidas Kyriakides[c], Anastasia Panayiotou[c], Machteld Vandecandelaere[a], Léan McMahon[d], Bieke De Fraine[a] and Jan Van Damme[a]

[a]*Centre for Educational Effectiveness and Evaluation, KU Leuven, Dekenstraat 2 box 3773, 3000 Leuven, Belgium;* [b]*Center for Education Policy Analysis, Stanford University, Stanford, CA, USA;* [c]*Department of Education, University of Cyprus, Nicosia, Cyprus;* [d]*The Economic and Social Research Institute, Dublin, Ireland*

(*Received 28 April 2014; accepted 6 March 2015*)

Background. The dynamic model of educational effectiveness (DMEE) is a comprehensive theoretical framework including factors that are important for school learning, based on consistent findings within educational effectiveness research. *Purpose.* This study investigates the impact of teacher and school factors of DMEE on mathematics and science achievement, and identifies factors with equalising qualities in terms of helping low-achieving student groups to catch up with their better-achieving peers. *Sample.* Data were retrieved from a large-scale, longitudinal project conducted in 571 classes in 334 schools in 6 European countries (Belgium/Flanders, Cyprus, Germany, Greece, Ireland and Slovenia). In each country, a sample of about 50 schools was drawn, and tests in mathematics and science were administered to all grade 4 students (N = 10,742) at the beginning and end of school-year 2010–2011. *Design and methods*. Data on teacher factors were collected through student questionnaires, and data on school factors were collected through teacher questionnaires. Two-level regression models were applied. Interaction effects between the factor and the group composition were estimated, while controlling for prior achievement. *Results.* Our results confirm the importance of most tested teacher factors (except for modelling and qualitative structuring) and all tested school factors of DMEE for effective math and science education. The majority of these factors appear to make an even greater difference for low-achieving student groups. *Conclusions*. Our results provide further validity to DMEE at classroom and school level, and indicate that most factors make a greater difference for low-achieving student groups. Thus, effective teaching is an important aspect for reducing the achievement gap. This illustrates the importance of placing the most effective teachers in schools with the highest percentage of underachieving students. Policy should encourage good teachers to teach in low-achieving schools by making these jobs more attractive, both financially and in terms of work conditions.

Keywords: dynamic model of educational effectiveness; mathematics and science; multilevel regression analysis; teacher effectiveness, school effectiveness, primary education

*Corresponding author. Email: gudrun.vanlaar@ppw.kuleuven.be

Introduction

Despite major advances in model development and theory-driven research during the latest decades, the field of educational effectiveness research still suffers a lack of well-developed theoretical models (Creemers and Kyriakides 2006; Scheerens 2014a, 2014b). Due to this lack of theory development, not many studies in educational effectiveness aim at validating components of effective education for their generalisability across countries and cultures. The dynamic model of educational effectiveness (DMEE) is probably the best attempt till date to describe the factors and processes of effective education in a comprehensive and general way.

The dynamic model (DMEE) is a comprehensive theoretical framework including factors that are of importance when it comes to school learning, based on general research findings within educational effectiveness research. DMEE distinguishes four hierarchical levels, i.e. the student, the class, the school and the system. Each level is associated with factors referring to characteristics or practices that have proven to be effective elements of education (Creemers and Kyriakides 2006). DMEE is tested by its developers in several studies in different countries (Creemers and Kyriakides 2009; De Jong, Westerhof, and Kruiter 2004; Kyriakides 2005, 2007, 2008; Kyriakides and Creemers 2008, 2009). However, whether the factors of this model are effective to the same extent for all types of student groups, and in different educational systems, has not yet sufficiently been investigated.

Many countries have developed policies to promote equal educational opportunities. Educational research has proven that effective education and good teachers *can* help underachieving students to catch up with their peers. Despite this knowledge, many attempts from policy-makers to narrow the gap between low- and high-achieving student groups have no – or sometimes even an opposite – effect. Identifying effective teacher factors that are associated with learning outcomes may help researchers, policy-makers, educational leaders and teachers to develop ways to improve effectiveness in education. Further testing of DMEE is therefore important as it might help to establish an evidence-based and theory-driven approach in policy-making; hence the need to test its validity in a cross-national context.

In this study, we aim at testing DMEE in a cross-national setting, with a special focus on the *differential effects of teacher and school factors for low- and high-achieving student groups*. To that end, we use a rich and longitudinal database, collected in 571 classes in 334 schools across six European countries.

The paper unfolds as follows: the first section provides an overview of the theoretical framework upon which the European study has been based; we start with a brief description of DMEE at teacher and school level; and next, we give a short overview of the literature on differential teacher and school effects. In the second section, we provide a more detailed description on the sample and data, and we elaborate on the analysis processes. The third section presents the results; and in the fourth section, we interpret the results and discuss practical implications.

Theoretical framework

The DMEE

A full description of DMEE goes beyond the scope of this paper. In what follows, we give a short overview of the teacher and school factors of DMEE that we tried to test in the current study. For an extended overview, we refer to the book *The Dynamics of Educational Effectiveness* (Creemers and Kyriakides 2008).

Teacher factors of DMEE

DMEE refers to eight *teacher factors* that are related to student outcomes and describe teachers' instructional role: *orientation, structuring, questioning, teaching-modelling, application, time management, teacher role in making classroom a learning environment* and *classroom assessment*. These eight factors do not merely refer to one approach of teaching, such as the direct and active teaching approach (Joyce, Weil, and Calhoun 2000) or the constructivist approach (Brekelmans, Sleegers, and Fraser 2000; Schoenfeld 1998), but combine factors that are derived from different approaches. That way, an integrated approach is adopted to define effective teaching: DMEE refers not only to teacher-guided instruction models (e.g. direct teaching and mastery learning), but also to more student-guided teaching and learning models (Muijs and Reynolds 2001). A short description of each teacher factor follows.

(A) *Orientation* refers to providing the reason(s) for which specific tasks or lessons take(s) place and/or challenging students to identify the reason(s) why the lesson involves a particular activity. It is expected that when orientation tasks take place during a lesson, students will be motivated to participate in the activities of the lesson, and, thus, student engagement rates will be increased (e.g. De Corte 2000; Paris and Paris 2001).

(B) *Structuring* refers to reviewing the objectives of a lesson or series of lessons, calling attention to important aspects, pointing out transitions between different parts of a lesson and summing up the main objectives of the lesson at the end (Rosenshine and Stevens 1986). Structuring assists students to comprehend information as an integrated whole with recognition of the association between parts and, that way, facilitates the memorisation of the information.

(C) *Questioning* refers to stimulating student interaction by raising different types of questions (i.e. product and process questions) at appropriate difficulty level, and appropriate dealing with student responses. Previous research has shown that effective teachers tend to more frequently use process questions, which expect students to provide explanations of their thinking methods (Askew and William 1995; Evertson et al. 1980). Additionally, questioning techniques refer to the clarity of a question posed by the teacher based on the students' individual abilities as well as the provision of sufficient time for the students to respond.

(D) *Teaching-modelling* is associated with teaching higher-order thinking skills, especially problem-solving. Teachers are expected to encourage students to use and develop their own strategies to help them effectively deal with different problematic situations (Grieve 2010; Kyriakides, Campbell, and Christofidou 2002; Vanlaar et al. 2014).

(E) *Application* refers to the opportunities provided to students to practice and apply new knowledge. Effective teachers are expected to encourage application tasks by using seatwork or small-group tasks (Borich 1992). Application activities are expected to serve the purpose of practice and also to act as a linkage of the content covered in a lesson with the next steps of learning. Thus, application tasks should not merely constitute a repetition of the content covered during a lesson, but also require students to discover and employ more complex ways of thinking.

(F) *The classroom as a learning environment* comprises five elements, i.e. teacher–student interaction, student–student interaction, students' treatment by the teacher, competition between students and classroom disorder (Creemers and Kyriakides 2008). Generally, this factor refers to the extent to which teachers are able to establish on-task behaviour through promotion of interactions and to teachers' attempts to create an efficient and supportive environment for learning in the classroom (Vandecandelaere et al. 2012; Walberg 1986).

(G) *Time management:* according to DMEE, effective teachers are capable of organising and managing the classroom environment so that time used for organisational purposes (for giving instructions, organising the materials to be used in a lesson, etc.) does not result in significant loss of teaching time (Creemers and Reezigt 1996). Time management is considered an important indicator of a teacher's ability to manage the classroom effectively (Evertson and Weinstein 2006).

(H) *Assessment* is seen as an integral part of teaching (Stenmark 1992), and special attention is given to formative assessment. Research findings indicate that teachers who emphasise formative assessment are more effective. Information gathered from assessment is expected to enable teachers to identify their students' needs, as well as to evaluate their own practice (e.g. Black 2001; Black and Wiliam 1998, 2010; Kyriakides 2005; Shepard 1989).

DMEE is based on the assumption that each of these factors can be defined and measured by using five dimensions: frequency, focus, stage, quality and differentiation (Creemers and Kyriakides 2008). In this study, however, even though it was possible to measure most factors, we were unable to measure all five dimensions of each factor.

School factors of DMEE

The school policy aspects that are emphasised in DMEE are concerned with teaching, creating and improving the learning environment and school evaluation. The factors at the school level in DMEE are expected to have direct and indirect effects on student achievement. School factors can influence classroom-level factors, especially the teaching practice. DMEE refers to factors at the *school level* that are related to key concepts of quantity of teaching, provision of learning opportunities and quality of teaching. Within the current study, the following aspects of these policies were tested: evaluation of teaching, evaluation of school learning environment, collaboration and interaction between teachers, provision of learning resources, partnership policy, relation with community and school policy in relation to quantity of teaching, learning opportunities and quality of teaching (Creemers and Kyriakides 2008; Kyriakides et al. 2010).

Differential effects for low- and high-achieving student groups

Earlier research generally indicated that effective education matters most for underachieving students, i.e. students with less advantaged background characteristics such as low socioeconomic status (SES) or not speaking the language of instruction at home (Campbell et al. 2004b; Scheerens and Bosker 1997). As early as 1966, the famous Coleman report revealed that school effects in the US were almost twice as large for black students as for white students (Coleman et al. 1966). More recently, Nye, Konstantopoulos and Hedges (2004) found larger teacher effect variance in low-SES schools compared with high-SES schools.

Despite this knowledge, however, few studies have examined which practices are most effective for underachieving students. The studies that did investigate differential teacher or school effects for different student groups usually found no, little or mixed differential effects (Konstantopoulos 2009; Konstantopoulos and Chung 2011; Muijs et al. 2005; Strand 1999). For instance, Raudenbush and Bryk (1992) found that the differences between public and private schools were twice as large for low-SES students as for high-SES students, suggesting that attending private schools is of more benefit to at-risk students. Sammons, Nuttall and Cuttance (1993) studied differential effects in reading in primary education and discovered that differential effects could only be demonstrated for the *prior achievement* of students, but not with respect to their SES, or ethnicity. Campbell et al. (2004a) found no, or little, differential effects with regard to different socioeconomic and achievement groups. These inconsistent findings are partly due to the use of different variables to investigate differential effects: ethnicity, SES and prior achievement. Indeed, identifying the sources of educational achievement gaps is not straightforward and presents a confusing and uncertain picture – particularly about the impact of schools on different students (Hanushek and Rivkin 2009).

Even so, a handful of studies reveal some teacher factors that can influence the association between students' at-risk status and their educational achievement, either directly or indirectly. For example, research investigating the effects of teacher relationships (e.g. Maulana et al. 2011) suggests that a good teacher-student relationship matters even more for minority students (e.g. den Brok et al. 2010). Hamre and Pianta found that it is extra important for students with high functional risk to have close relations with their teachers (2005). High-risk students tend to benefit even more from high-quality instruction than their low-risk peers do (Curby, Rimm-Kaufman, and Ponitz 2009). Furthermore, connecting learning to real-life experiences and stressing practical applications, putting emphasis on basic skills and offering external rewards have been found particularly important to low-SES students (Campbell et al. 2004a). Vanlaar et al. (2014) found negative effects of constructivist class practices on reading comprehension for high-risk students, while the more traditional class practices had no differential effects.

Thus, although different types of students have different educational needs, the existing literature still does not give sound answers on how to more effectively implement and monitor education strategies for vulnerable, low-achieving students (Kyriakides 2007). If teachers and schools have greater effects for underachieving students, it is important for governments and practitioners to know which practices can help at-risk students to catch up with their more advantaged peers. More research is needed to clarify from which practices underachieving student groups can benefit the most.

The present study addresses this need by answering the following *research question:* 'Do the teacher and school factors of DMEE have the same impact for high- and low-achieving student groups on science and math learning in grade 4?'

Method

This part is divided in two sections: (1) description of the data, including the setting in which the data were gathered, the sample characteristics and measurement procedures, and (2) description of the multilevel models used to analyse the differential effects of the teacher and school factors.

Data

Setting

We used data from the European collaborative research project 'Establishing a knowledge-base for quality in education: Testing a dynamic theory of educational effectiveness'. The project was conducted in six countries, i.e. Belgium/Flanders, Cyprus, Germany, Greece, Ireland and Slovenia. Math and science tests were administered to 10,742 grade 4 students at the beginning and end of the school-year 2010–2011. The central aim of the project was to further develop and test the validity of DMEE in different educational settings, i.e. in relation to the diversity of student intake, processes and prospective outcomes. Thus, the project aimed to improve the knowledge base on effectiveness of education.

By testing students from different countries, diversity is likely to be assured with this project. Although all participating countries are OECD countries (except Cyprus) and members of the European Union, the learning environment and school organisation in these countries differ in certain ways, giving us a broad spectrum to test and validate the teacher and school factors of the DMEE.

Another advantage of our data is that prior knowledge was measured, which makes the data suitable for investigation of the differential effects of teacher and school factors (Creemers, Kyriakides, et al. 2013). Only with correction for prior achievement, it is possible to isolate teacher and school factors (Campbell et al. 2004a; Kyriakides 2004; Strand 2010).

Sample

In each participating country, a stratified sampling procedure (Cohen, Manion, and Morrison 2000) was used to select a sample of at least 50 primary schools. Tests in mathematics and science were administered to all grade 4 students at the beginning and at the end of school-year 2010–2011. In total, 10,742 fourth grade students participated, spread over 571 classes and 334 schools. Not all students completed the two mathematics and science tests at the beginning and end of the school-year. The

Table 1. Sample description.

Country		Belgium	Cyprus	Germany	Greece	Ireland	Slovenia	All
N students		1954	1990	1228	1196	2423	2155	10,946
Class size[a]	Mean	19.74	17.17	19.49	17.41	21.44	19.41	19.18
	SD	4.72	4.10	4.83	5.30	7.30	4.43	5.44
	N	99	116	63	69	113	111	571
School size[b]	Mean	1.94	1.97	1.17	1.41	1.85	1.88	1.71
	SD	0.70	0.74	0.38	0.61	0.96	0.67	0.76
	N	51	59	54	49	61	59	333

[a]*N* students per class in our sample.
[b]*n* classes per school in our sample.

data informing this paper include the test results for students who completed tests at the two waves of data collection. Hence, the analyses cover 9321 students (86.8% of total sample) who completed both mathematics tests and 9148 students (85.2%) who completed both science tests, spread over 561 classes and 329 schools. In each school, all teachers ($N = 3010$) were requested to fill in a questionnaire about their school policy. Table 1 gives for each country the number of participating students, the average class size per country and the average number of classes per school in our sample.

Measures

Achievement tests. The construction of the mathematics and science tests was mainly based upon the released items of TIMSS 2007. TIMSS (Trends in International Mathematics and Science Study) constitutes an international comparative study, organised by the International Association for the Evaluation of Educational Achievement (IEA), aiming to provide insights into student achievement in mathematics and science.[1] The tests measure three cognitive domains: knowing, reasoning and applying. The math test covers the following content categories: whole numbers; fractions and proportionality; measurement, estimation and number sense; data representation, analysis and probability; and geometry, patterns, relations and functions. The science questions cover: earth science; life science; physical science; environmental issues; and nature of science. Thus, using items from TIMSS 2007, two parallel tests were developed for both subjects.

For both subjects, the test scores were vertically equated using Item Response Theory (IRT). IRT is a widely used statistical technique to compute standardised estimates. It enables to vertically equate achievement scores of different levels on one single measurement scale, so that, for instance, the math scores from several sequential grades can be compared. In this study, all test scores were calibrated on the same scale (with mean = 300 and SD = 60 for the first test) to make them comparable across test versions, countries and measurement occasions (Kyriakides et al. 2014). The test scores measured at the start of grade 4 are an indicator of students' prior achievement. These individual prior achievement scores were aggregated to class and school averages to create indicators for the class and school composition. This resulted in four group composition variables: class mean prior achievement math class mean prior achievement science, school mean prior achievement math, and school mean prior achievement science. The test scores measured at the end of grade 4 form the dependent variables. Figures 1 and 2 visualise the distribution of the test scores and the group composition variables by country.

Factors of the dynamic model. Data on the factors were derived from questionnaires, based on instruments developed in earlier studies to test the DMEE (Creemers and Kyriakides 2010a, 2010b; Kyriakides and Creemers 2008). The instruments were tailored for our project following discussion and agreement among the collaborating researchers of the six participating countries. For each item, the researchers indicated the applicability and relevance to the educational context of their country and whether the students and teachers could reasonably be expected to answer the questions. Based on this information, a student questionnaire that measures the teacher factors and a teacher questionnaire that measures the school factors were developed. Initially, an English version of the questionnaires was developed; next, procedures of translation and back translation were used to produce the questionnaires in the

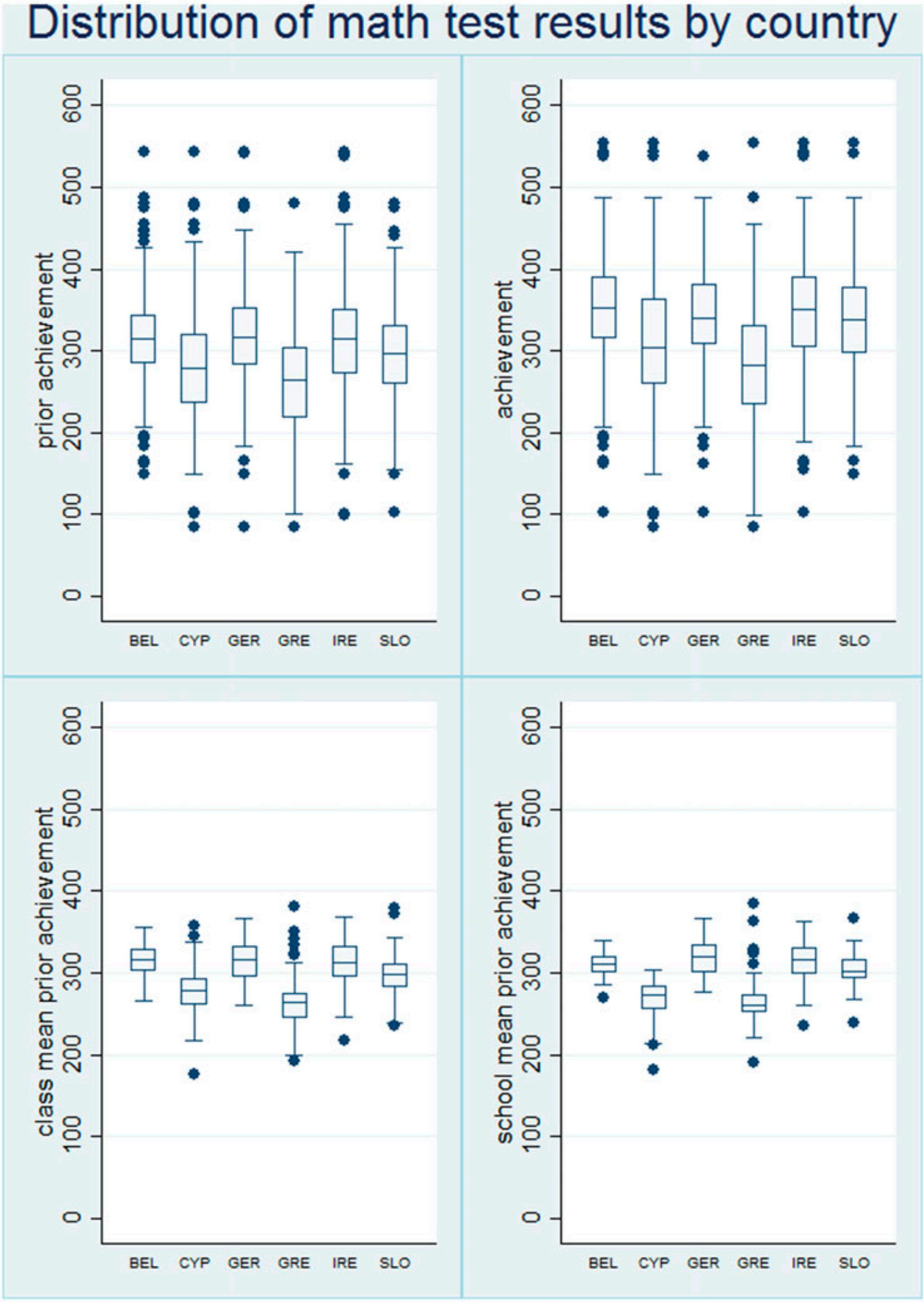

Figure 1. Mathematics achievement.

four other European languages, i.e. Dutch, German, Greek and Slovenian (Kyriakides et al. 2014).

The factors are latent constructs (Fisher, Waldrip, and den Brok 2005); more specifically, generalised variables are based on items measured at a lower level, and

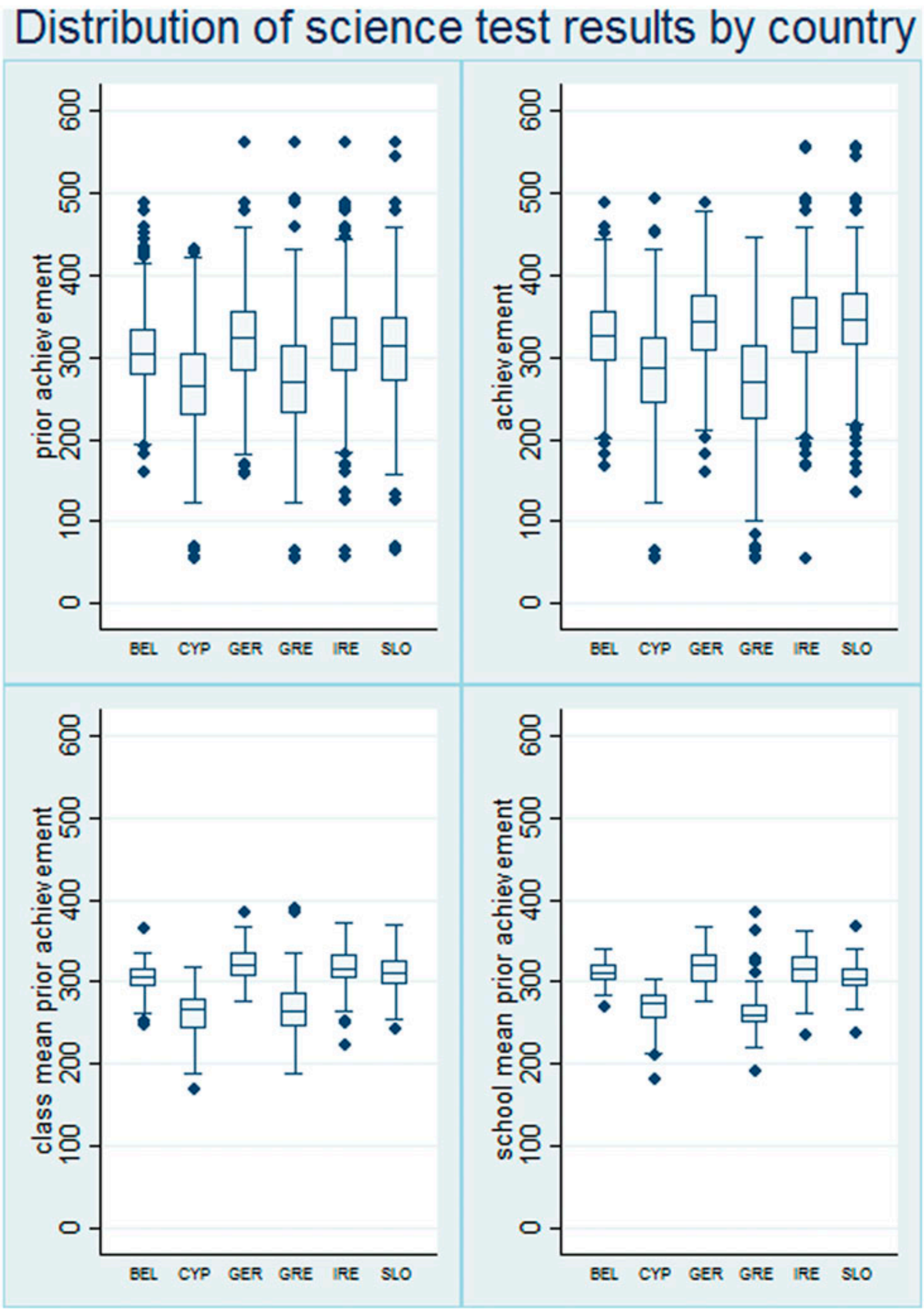

Figure 2. Science achievement.

aggregated to indicate a group-level characteristic. The data analytic technique to construct latent variables is *structural equation modelling* (SEM). SEM tests the relationships among observed items by merging multivariate regression and factor analysis. Thus, factor analysis is used to compress the information of several related items, which measure the same latent construct, into one factor (Savalei and Bentler 2010; Schreiber et al. 2006).

Teacher factors. In addition to the achievement tests, all students completed a questionnaire at the end of grade 4. The survey addressed teacher and classroom practices; students were asked to indicate on a five-point Likert scale the extent to which their teacher behaves in a certain way during their lessons. Although there is some controversy about the use of *student ratings*, we found enough evidence in the literature that student ratings provide reliable information on their teacher's behaviour and the class environment (Kyriakides et al. 2014). For instance, De Jong and Westerhof (2001) compared sixth graders' perception of teacher behaviour to the opinion of external observers and found that the quality of aggregated student perceptions is as good as – or even better than – the quality of data from external observers. Students are in a key position to rapport about their teacher's behaviour in the classroom. Student ratings constitute a main source of information regarding opportunity for learning, the degree of rapport and communication between teacher and students, the way and frequency of assessment, application, classroom equity, etc. (Carle, 2009; Creemers, Kyriakides, and Sammons 2010; Kyriakides 2005; Kyriakides et al. 2014).

Moreover, during preliminary analyses, we tested whether there was enough consistency between the ratings of students within one class to ensure that the items can be generalised to the class level. To measure this homogeneity, ANOVA was used. This is called a *generalisability study,* which tests whether the students' ratings are a reliable source of information on their teacher's behaviour and the usage of practices in their class (Brennan 2003, 2010). ANOVA[2] results confirmed that the items can be aggregated to the class level as, for all the questionnaire items, the between-group variance was higher than the within-group variance ($p < 0.05$) (Kyriakides et al. 2014; Panayiotou et al. 2014).

In the current study, the teacher factors were computed based on aggregated student questionnaire items. During the SEM analysis, some questionnaire items had to be removed as it was shown that their contribution to the factor was not significantly different from zero. Based on the results of SEM analysis, the teacher factors were computed for each class in our sample. The teacher factor scores that were used in the multilevel analysis are given in Appendix 1. For each factor, the items that were included to predict the respective factor for each class, with their factor loadings, are given.

School factors. School factors were measured by surveying all teachers in the participating schools. The teachers completed the questionnaire during the last term of the school-year. A four-point Likert scale was used to collect data on teachers' perceptions of the school policy. As before, preliminary ANOVA showed that the teacher ratings are generalisable, since the between-school variance was higher than the within-school variance ($p < 0.05$). Also, to compute the school factors, SEM analysis was used (Creemers, Panayiotou, et al. 2013). Appendix 2 gives an overview of the items that were used to calculate the school factor scores, together with their factor loadings.

Missing values. Missing values of the teacher and school factors were imputed using SPSS software (Schafer and Graham 2002; SPSS Inc 2011). The proportion of missing data was limited (maximum 8% per variable), the sample size was large enough and we built a comprehensive imputation model including all available variables that were relevant and not causing multicollinearity as auxiliary variables, that way minimising potential bias (Graham 2009).

Multilevel analyses

Our data have a *hierarchical structure* (students nested in classes, nested in schools, nested in countries). Multilevel modelling techniques have been recognised as a prevailing tool in analysing this kind of hierarchically structured data-sets (Goldstein 2003; Hox 2010; Raudenbush and Bryk 1992). Multilevel models portray significant advantages such as accounting for the dependency of observations within a group and identifying the variance at different levels (Luyten and Sammons 2010). It is recommended to have as many units as possible at the top level of the multilevel hierarchy (minimum 25), and to have enough observations within each group (Maxwell, Kelley, and Rausch 2007; Roy et al. 2007).

For our study, the data fitted best in a two-level model. We could not include a country level because there are only six countries. And the current study mainly comprises primary schools with only one class per grade; thus, the school and class level comprise often the same observations (average number of classes per school in our sample is shown in Table 1). Therefore, we estimated two-level models – (i) students within classes for the models testing the teacher factors, and (ii) students within schools for the models testing the school factors – using MLwiN software (Rasbash et al. 2012). We chose not to include 'country' as a covariate in the analysis, because this might take away some of the variety in teacher and school factors. Before running the models, we centred all predictors around the grand mean, in order to facilitate the interpretation of the results (Hox 2010). Next, we estimated four series of multilevel models:

(a) differential effects of teacher factors on math achievement,
(b) differential effects of teacher factors on science achievement,
(c) differential effects of school factors on math achievement and
(d) differential effects of school factors on science achievement.

Each of these series was executed in the following order:

(1) We ran a two-level null model (i.e. without any explanatory variables) to investigate the distribution of the variance over the two levels (student and class level for series a and b, and student and school level for series c and d).
(2) We added student prior achievement for math (series a and c) or science (series b and d) to the model.
(3) The group composition (class or school mean of prior math or science achievement, respectively) was added; this third model was named the 'basic model'. This basic model can be seen as a type B value-added model since a correction is made for student prior achievement and for group composition (Timmermans, Doolaard, and de Wolf 2011).
(4) We added one factor plus its interaction with the group composition to the basic model to test whether the effect of this factor differs for high- and low-achieving classes/schools. This fourth step was repeated for every factor

separately, resulting in one model for each factor. Each model underwent the following tests:

(a) We tested how well this model fits to the data by comparing the log-likelihood with that of the basic model. The likelihood is the probability of the data set given the model with its parameter estimates. We use the log likelihood as an indicator of the model fit, where higher values indicate a better fit. Statistical significance was tested using a chi-squared test, with two degrees of freedom (since two parameters were added to the basic model that is compared). The log-likelihood ratio test can only be used when two models are based on the same data, with an equal number of observations, which was a reason for imputing the missing values of the factors.

(b) To test whether an effect is statistically significant from zero, it is necessary to take into account the standard error of the estimates. To that end, a two-sided Wald-test was done (estimate divided by standard error), which is the most commonly used method to test the significance of a regression estimate.

(c) To have a measure of the size of this effect, we calculated the explanatory power of the factors, by comparing the second-level variance of the model with this variance of the basic model. The proportion of variance that is reduced after adding the extra variables to the basic model tells us how much of the differences between classes/schools can be explained by these added variables.

Results

Math and science achievement

For *mathematics*, 24% of the variance lies at school/class level, and 76% of the variance lies at student level. After adding prior achievement to the model, 71/69% of the school/class-level variance, and 38% of the student-level variance is explained. Adding the group composition (mean of prior achievement) explains another 8/7% of the variance at school/class level. Comparing the null model with the basic

Table 2. Multilevel analyses of teacher factors: summary of results.

	Mathematics			Science		
Teacher factor	Main effect	Interaction effect	%	Main effect	Interaction effect	%
Struct QuanT	8.12**	−0.38***	5.7	9.55**	−0.09	2.2
Struct QuaL	3.93	−0.30**	1.7	2.55	−0.32**	2.1
Questioning	10.42**	−0.40***	5.1	5.93	−0.24*	1.8
Modelling	5.20	−0.34**	3.3	6.36	−0.25*	1.5
Application	9.86**	−0.49***	6.1	14.18***	−0.16	3.5
TS interaction	16.95***	−0.56***	9.8	22.51***	−0.23†	6.8
Time Manag	19.91***	−0.05	10.9	12.48***	0.02	2.8
Misbehaviour	11.26***	0.08	5.3%	5.69*	0.10	1.0
Assessment	11.26***	−0.47***	8.6	12.16***	−0.14†	3.4

Notes: % = per cent of class variance explained by factor + factor *x* group composition.
Significance level (based on Wald-test).
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

model, there is a large improvement of the model fit (school: $\chi^2(2) = 4699.21$; $p < 0.000$/ class: $\chi^2(2) = 4819.75$; $p < 0.000$).

For *science*, 31/32% of the variance lies at school/class level and 69/68% lies at student level. Prior achievement explains 62/60% of the school/class-level variance and 29% of the student variance. Group composition explains another 13/26% of the school/class variance. Thus, these two prior-achievement covariates together cause a large model-fit improvement (school: $\chi^2(2) = 3432.76$, $p < 0.000$; class: $\chi^2(2) = 3581.16$, $p < 0.000$).

Differential effects of teacher factors

The results of the two-level models showing at least one significant teacher-factor effect are shown in Appendix 3 and compared with the basic model. Table 2 gives a summary of the findings.

Quantitative structuring is associated with both math and science achievement, but is particularly strongly associated with math achievement among low-achieving classes. Quantitative structuring and the interaction with group composition explain 5.7% of the class variance in math achievement and 2.2% of the class variance in science achievement.

Qualitative structuring has no significant main effects, but the interaction with group composition is significant. Qualitative structuring appears to be more effective for low-achieving classes than for high-achieving classes. These findings hold for both math and science achievement. The math model explains 1.7% more of the class variance compared to the basic model, while the science model explains an extra 2.1% of the class variance.

The more *adequate questions* a teacher poses during the lessons, the higher the math achievement of the students, especially in low-achieving classes. The effect on science achievement is not statistically significant, but the interaction effect with group composition is negative and significant. The quality of the questions explains 5.1% of the variance in math and 1.8% of the variance in science.

Modelling appears to have no significant main effects, neither on math nor on science. There is a significant negative interaction effect with class mean, indicating that the effect is larger for low-achieving classes, and explaining 3.3% of the class variance in math and 1.5% of the class variance in science.

Application is an effective class practice for both math and science, and for math, this effect is stronger among low-achieving classes. The explanatory power for the class variance in math achievement is 6.1% and for science achievement is 3.5%.

The quality of the *interaction between the teacher and their students* is associated with both math and science achievement. Concerning math achievement, this is more important in low-achieving classes. Almost 10% of the class variance in math, and 6.8% of the variance in science, is explained by this factor and its interaction.

Time management appears to be important for both math and science achievement, and for all types of classes. Time management explains 10.9% of the class variance in math achievement, and 2.8% of the class variance in science achievement.

Effective *dealing with misbehaviour* is associated with math achievement, explaining 5.3% of the variance. The relation with science achievement is less strong, but nevertheless significant, and explains 1% of the class variance. Dealing well with misbehaviour is equally important in high- and low-achieving classes.

Table 3. Multilevel analyses of school factors: summary of results.

	Mathematics			Science		
School factor	Main effect	Interaction effect	%	Main effect	Interaction effect	%
Evaluation Teaching	15.58***	−0.49***	11.0	20.34***	−0.20	7.1
Evaluation SLE	13.59**	−0.31†	3.7	16.61**	−0.31†	0.1
Collaboration	20.22***	−0.48**	9.3	25.64***	−0.48**	7.6
Resources	17.54***	−0.30*	6.3	18.37***	−0.46***	−0.3
Community	17.97***	−0.27*	10.7	19.74***	−0.18	6.9
Partnership	21.90***	−0.34*	6.7	23.16***	−0.42**	3.1
PT quantity	12.21***	−0.13	5.2	10.12**	−0.46***	2.1
PT learning opp	11.90***	−0.29*	2.9	12.09**	−0.42**	0.4
PT quality	16.13***	−0.31**	8.8	13.38***	−0.30**	4.4

Notes: % = per cent of school variance explained; PT: policy on teaching; SLE: school learning environment.
Significance (based on Wald-test).
†$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Finally, *assessment* appears to be an indicator of effective education. For math achievement, assessment is more effective in low-achieving classes. For science achievement, assessment has the same effect for all student groups. Assessment explains 8.6% of the variance in math achievement and 3.4% of the variance in science achievement.

Differential effects of school factors

Appendix 4 shows the results for the multilevel models analysing the school factors and their interactions with group composition, i.e. the school mean of prior achievement. Table 3 gives a summary of the findings. All tested school factors are positively associated with both math and science achievement at the end of grade 4. Almost all school factors have a significant negative interaction effect with group composition, meaning that these school factors are more important in schools with a low average achievement level. The exceptions, which appear to be equally important for high- and for low-achieving student groups, are: evaluation of teaching (for science), relationship with community (for science) and policy on the quantity of teaching (for math).

As we found for the teacher factors, the school factors consistently explain a higher percentage of the second-level variance in math achievement than in science achievement. The percentage of school-level variance in *math* achievement explained by the school factors and their interaction with the group composition varies between 11% (evaluation of teaching) and 2.9% (policy on learning opportunities), with an average of 7.2%. For *science* achievement, the explanatory power of the school factors and their interactions with group composition varies between 7.6% (collaboration) and 0% (resources), with an average of 3.5% (see Table 3).

All models, both the ones analysing teacher factors and the ones analysing school factors, resulted in a significant improvement of the model fit ($p < 0.01$).

Discussion

Aims of this paper were to investigate whether the teacher and school factors of the DMEE are valid elements of effective primary education across six countries, and to identify potential factors that can help low-achieving student groups to progress and keep pace with their high-achieving peers.

Theoretical implications

This European study created important data for testing a model of educational effectiveness in different contexts. It is the first study that allows to test and validate the teacher and school factors of the DMEE in six European countries. The results show that almost all tested teacher factors and all tested school factors are important elements of effective math and science education, which suggests that these factors of the DMEE can be generalised to a broader context, at least within Europe.

However, we were not able to validate the teacher factor *orientation*. Also, the factor *learning environment* seems to fall apart in two factors, which might indicate that this factor is too broadly defined or has a higher cultural loading and gives, therefore, inconsistent results in cross-cultural research. Moreover, the study was not in a position to measure the five dimensions proposed by the dynamic model, and therefore this assumption could not be tested. In fact, this study raises some doubts on whether the five dimensions can be used to measure the functioning of teacher and school factors in different countries. There is still need for studies investigating the validity of the dynamic model. One implication of the findings of this study is that we need to more carefully consider the context specificity of the functioning of factors, especially the use of dimensions to measure them through better instruments. Our study demonstrates the importance of validation and replication in research (Duncan et al. 2013; Makel and Plucker 2014).

Concerning differential effectiveness, the results suggest that most factors have a greater effect on low-achieving student groups, especially for mathematics achievement. This finding might indicate that implementing effective practices is not only important for student learning in general, but also for promoting equity. These findings support most of the literature on differential effectiveness described in the theoretical framework; this study also demonstrates that teacher and school factors make a greater difference for low-achieving classes and schools (e.g. Kyriakides and Creemers 2011). No factor was found to be more beneficial for high-achieving student groups. Our findings do not allow to formulate conclusive answers on how these interactions between factors and the achievement level of the student group exist, or which underlying processes cause them.

Where most effectiveness studies focus on interaction effects of student-level characteristics, a notable contribution of this study is the examination of the interaction between effectiveness factors and the achievement level of the student group (i.e. aggregated achievement score). Since many schools have more or less homogenous student populations, the interaction with the group is an interesting approach for practitioners.

The significance of this study also lies in its investigation of the effects of teacher and school factors on science achievement, which is an understudied subject in educational effectiveness research (most studies focus solely on either mathematics or language learning).

Importance of cross-national research

Teacher effectiveness research has been criticised for showing ethnocentric tendencies. Due to a lack of theory development, not many studies in educational effectiveness aim at validating components of effective education for their generalisability across countries and cultures. A lack of cross-national perspectives and inter-cultural collaboration between educational effectiveness researchers is seen as intellectually detrimental (Panayiotou et al. 2014; Reynolds 2000). Over the last two decades, several education policy-makers attempted to raise standards based on the simplistic application or transplantation of knowledge and ideas from one country, or one culture, to another (Panayiotou et al. 2014). Several educational effectiveness researchers voiced concerns about the potential hazards of improvident transplantation of educational policies from one culture to another (e.g. Reynolds 2006; Scheerens 2012). Examples of this practice include the proposed lengthening of the school-day and shortening of school-holidays discussed in some American states, following Japan, where students spend the most hours per year in school. Another example is the trend in British primary schools towards whole-class lecturing, which is clearly inspired by the popularity of this approach in the East Asian countries that have the highest rankings in international assessment studies (Reynolds 2000).

Obviously, in an era when educational policy follows the international route, it is unfortunate that EER appears to mainly conduct within-country studies rather than cross-national research. Cross-national studies on educational effectiveness are required to gain understanding about the complex structures of education policy across different countries and cultures, and to explain how policies affect student outcomes in different settings. This cross-national study contributes to the development of the international dimension of educational effectiveness research.

Limitations

This transnational nature of this study might also have implications for the interpretation of our results. Although the mathematics and science tests were carefully developed for the TIMSS-study (Kyriakides et al. 2014), the associations between factors and student achievement might somehow reflect associations between certain cultural aspects within the different countries and the extent to which the test materials reflect their math and science curricula. We chose not to include 'country' as a covariate in the analysis, because it was felt this might take away some of the variety in teacher and school factors. This was based on the arguable assumption that the six countries are relatively similar culturally, due to the fact that they are all members of the European Union and the OECD. It should be acknowledged that our findings might partially be caused by the fact that some factors are more frequently found within high-achieving countries, or vice versa.

Although this cross-national study is a step in the right direction, it should be noted that it is also important to examine the generic nature of the teacher and school factors in countries outside the European Union with different educational contexts in which even greater variation can be found.

Cross-national research also has its drawbacks. When using measurement instruments in different countries, the researcher needs to assure that the items measure the same thing in every country. Only items that show high factor loadings for both the across and the within-country analyses can be assumed to be valid across countries. However, 9 out of 49 student questionnaire items had to be removed because they acted differently in different countries. As a result, some factors are based on two or three items, while it is recommended to construct a latent variable on more items. This is an important limitation of this study. Due to the omission of some items in order to obtain reliable factor scores for all six countries, the remaining items do not always cover all aspects of the teacher and school factors. Our data did not allow to investigate the precise implementation of the factors. A score on a few items does not fully reflect the daily practices in a class/school, and is, arguably, a poor representation of very complex and rich interactions between human beings. Unfortunately, there were no means for observations in classes and schools in our project, which would have yielded extra information on teacher behaviour in class and/or the application of certain education policies within the daily school context.

It should be noted that the teacher factors are not subject-specific and that there are also other possible ways of modelling the data (for instance, we could have used a repeated measures design to focus more on learning gains). Also, the data on the teacher and school factors are based on questionnaires; therefore, limitations of using perceptual methods to measure factors should be acknowledged. Student ratings of teacher performance were used for the measurement of the teacher factors. Although student ratings have been questioned in terms of reliability, recent studies demonstrate their capability of providing valid data on the classroom behaviour of their teachers. Our data seem to support these studies, since the student ratings showed enough consistency within classes. This finding can be seen as a confirmation that young students can be considered as good sources of information on class characteristics, as they 'taste the dinner that is served' (Kyriakides et al. 2014).

Conclusion

In conclusion, despite the limitations, this study contributes to the literature by giving more information on how the teacher and school factors of the DMEE interact with different types of student groups. The findings suggest that the factors of the DMEE are effective for most student groups, and especially for underachieving student groups. A better comprehension of this interaction is required in order to help inform policy decisions when aiming at creating equal educational opportunities for all students, especially for underachieving student groups. A comprehensive, general theoretical framework in educational effectiveness might help policy-makers to make evidence-based decisions, on school reformations for instance. The DMEE is an attempt in the right direction. Although more studies are needed, this study suggests that cross-national research can reveal both universal and differential components of effectiveness, and might actually help to get more insights into differential teaching effects for low-achieving student groups. Professional development programmes need an evidence-based theory to rely on and it is the task of educational researchers to further develop, replicate and validate their findings (Makel and Plucker 2014).

Although more sound research is needed to understand the interaction between teacher factors and different student types, our findings suggest that there are differential effects and that differences in teacher factors might partly explain the differences in achievement levels in different classes and countries. These findings suggest that it is important that the most effective teachers are placed in schools with the highest percentage of underachieving students. Policy-makers should encourage 'good' teachers to work in schools with high percentages of underachieving students by enhancing these jobs financially and in terms of work conditions.

Acknowledgements

This study uses data that was collected in the framework of a three-year project (2009–2012) entitled 'Establishing a knowledge-base for quality in education: Testing a dynamic theory of educational effectiveness', funded by the ESF (European Science Foundation 08-ECRP-012). Additional funding was received from the FWO (Fonds voor Wetenschappelijk Onderzoek) for the data collection and analyses in Flanders. The authors wish to thank the Policy Research Centre on Educational and School Careers (http://steunpuntssl.be) and the Centre for Educational Effectiveness and Evaluation (KU Leuven) for the data collection in Flanders.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was funded by the ESF (European Science Foundation 08-ECRP-012). Additional funding was received from the FWO (Fonds voor Wetenschappelijk Onderzoek) for the data collection and analyses in Flanders.

Notes

1. The TIMSS 2007 released items can be downloaded from http://timssandpirls.bc.edu/TIMSS2007/items.html. More information on IEA can be found at http://www.iea.nl.
2. ANOVA = Analysis of variance.

Notes on contributors

Gudrun Vanlaar is a doctoral researcher investigating which educational practices are the most effective for high-risk students, i.e. students with a low socioeconomic or immigrant background, and students with special educational needs. She is interested in multilevel modelling and causal inference techniques.

Leonidas Kyriakides is a professor of Educational Research and Evaluation. His main research interests are in the area of educational effectiveness and especially in modelling the dynamic nature of educational effectiveness and in using research to promote quality and equity in education. Leonidas is the chair of the EARLI SIG on Educational Effectiveness and acted as chair of the AERA SIG on School Effectiveness and Improvement.

Anastasia Panayiotou participated in several international projects as a PhD student. Her main research interests are in the area of educational effectiveness research, teacher professional development and teacher evaluation.

Machteld Vandecandelaere is a doctoral researcher interested in educational effectiveness, the effects of kindergarten retention, causal inference techniques and longitudinal studies.

Léan McMahon is a social science researcher, specialised in research on evaluation of effectiveness in education and health environments, mixed methods research and policy decision-making.

Bieke De Fraine is an assistant professor and head of the Centre for Educational Effectiveness and Evaluation. Her main research interests include longitudinal educational effectiveness studies and growth curve modelling.

Jan Van Damme is an emeritus professor. His research interests include educational effectiveness, evaluation and the longitudinal study of students' pathways through education.

References

Askew, M., and D. William. 1995. *Recent Research in Mathematics Education*, 5–16. London: Office for Standards in Education.

Black, P. 2001. "Formative Assessment and Curriculum Consequences." In *Curriculum and Assessment*, edited by D. Scott, 7–23. Westport, CT: Ablex.

Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice* 5 (1): 7–74. doi:10.1080/0969595980050102.

Black, P., and D. Wiliam. 2010. "Inside the Black Box: Raising Standards Through Classroom Assessment." *Phi Delta Kappan* 92 (1): 81–90. doi:10.1177/003172171009200119.

Borich, G. D. 1992. *Effective Teaching Methods*. New York: Macmillan.

Brekelmans, M., P. Sleegers, and B. Fraser. 2000. "Teaching for Active Learning." In *New Learning*, edited by R. J. Simons, J. van der Linden, and T. M. Duffy, 227–242. Dordrecht: Kluwer Academic.

Brennan, R. L. 2003. "Generalizability Theory." *Journal of Educational Measurement* 40 (1): 105–107. doi:10.1111/j.1745-3984.2003.tb01098.x.

Brennan, R. L. 2010. "Generalizability Theory and Classical Test Theory." *Applied Measurement in Education* 24 (1): 1–21. doi:10.1080/08957347.2011.532417.

Campbell, J., L. Kyriakides, D. Muijs, and W. Robinson. 2004a. *Assessing Teacher Effectiveness. Developing a Differentiated Model*. London: Routledge Falmer.

Campbell, J., L. Kyriakides, D. Muijs, and W. Robinson. 2004b. "Effective Teaching and Values: Some Implications for Research and Teacher Appraisal." *Oxford Review of Education* 30 (4): 451–465. doi:10.1080/0305498042000303955.

Carle, A. C. 2009. "Evaluating College Students' Evaluations of a Professor's Teaching Effectiveness Across Time and Instruction Mode (Online Vs. Face-to-Face) Using a Multilevel Growth Modeling Approach." *Computers & Education* 53 (2): 429–435. doi:10.1016/j.compedu.2009.03.001.

Cohen, L., L. Manion, and K. Morrison. 2000. *Research Methods in Education*. London: Routledge/Falmer.

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. Hood, F. D. Weinfeld, and L. York. 1966. *Equality of Educational Opportunity.* Washington, DC: US Office of Education.

Creemers, B., and L. Kyriakides. 2006. "Critical Analysis of the Current Approaches to Modelling Educational Effectiveness: The Importance of Establishing a Dynamic Model." *School Effectiveness and School Improvement* 17 (3): 347–366. doi:10.1080/09243450600697242.

Creemers, B., and L. Kyriakides. 2008. *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge.

Creemers, B., and L. Kyriakides. 2009. "Situational Effects of the School Factors Included in the Dynamic Model of Educational Effectiveness." *South African Journal of Education* 29 (3): 293–315. http://www.ajol.info/index.php/saje/article/viewFile/45171/28661.

Creemers, B., and L. Kyriakides. 2010a. "Explaining Stability and Changes in School Effectiveness by Looking at Changes in the Functioning of School Factors." *School Effectiveness and School Improvement* 21 (4): 409–427. doi:10.1080/09243453.2010.512795.

Creemers, B., and L. Kyriakides. 2010b. "Using the Dynamic Model to Develop an Evidence-based and Theory-driven Approach to School Improvement." *Irish Educational Studies* 29 (1): 5–23. doi:10.1080/03323310903522669.

Creemers, B. P. M., L. Kyriakides, and P. Sammons. 2010. *Methodological Advances in Educational Effectiveness Research*. London: Taylor & Francis.

Creemers, B., L. Kyriakides, A. Panayioutou, W. Bos, H. Holtappels, M. Pfeifer, M. Venneman et al. 2013. *Establishing a Knowledge Base for Quality in Education: Testing a Dynamic Theory for Education. Handbook on designing evidence-based strategies and actions to promote quality in education*. Münster: Waxmann. ISBN 978-3-8309-2933-8.

Creemers, B., A. Panayiotou, L. Kyriakides, J. Van Damme, E. Smyth, H. Wendt, D. Zupanc, and E. Penderi. 2013. *Impact on Student Achievement of the School Learning Environment and the Policy on Teaching*. Paper presented at the annual meeting of AERA, San Francisco, CA.

Creemers, B., and G. J. Reezigt. 1996. "School Level Conditions Affecting the Effectiveness of Instruction." *School Effectiveness and School Improvement* 7 (3): 197–228. doi:10.1080/0924345960070301.
Curby, T. W., S. E. Rimm-Kaufman, and C. Ponitz. 2009. "Teacher–Child Interactions and Children's Achievement Trajectories Across Kindergarten and First Grade." *Journal of Educational Psychology* 101 (4): 912–925. doi:10.1037/a0016647.
De Corte, E. 2000. "Marrying Theory Building and the Improvement of School Practice: A Permanent Challenge for Instructional Psychology." *Learning and Instruction* 10 (3): 249–266. doi:10.1016/S0959-4752(99)00029-8.
De Jong, R., and K. J. Westerhof. 2001. "The Quality of Student Ratings of Teacher Behaviour." *Learning Environments Research* 4 (1): 51–85. doi:10.1023/A:1011402608575.
De Jong, R., K. J. Westerhof, and J. H. Kruiter. 2004. "Empirical Evidence of a Comprehensive Model of School Effectiveness: A Multilevel Study in Mathematics in the 1st Year of Junior General Education in the Netherlands." *School Effectiveness and School Improvement* 15 (1): 3–31. doi:10.1076/sesi.15.1.3.27490.
den Brok, P., J. van Tartwijk, T. Wubbels, and I. Veldman. 2010. "The Differential Effect of the Teacher-student Interpersonal Relationship on Student Outcomes for Students with Different Ethnic Backgrounds." *British Journal of Educational Psychology* 80 (2): 199–221. doi:10.1348/000709909X465632.
Duncan, G. J., M. Engel, A. Claessens, and C. J. Dowsett. 2013. *The Value of Replication for Developmental Science*. Working Paper. http://sites.uci.edu/gduncan/files/2013/06/Replication-paper-single-spaced.pdf.
Evertson, C. M., C. W. Anderson, L. M. Anderson, and J. E. Brophy. 1980. "Relationships Between Classroom Behaviors and Student Outcomes in Junior High Mathematics and English Classes." *American Educational Research Journal* 17 (1): 43–60. doi:10.3102/00028312017001043.
Evertson, C. M., and C. S. Weinstein. 2006. *Handbook of Classroom Management. Research, Practice, and Contemporary Issues*. Mahwah, NJ: Lawrence Erlbaum.
Fisher, D., B. Waldrip, and P. den Brok. 2005. "Students' Perceptions of Primary Teachers' Interpersonal Behavior and of Cultural Dimensions in the Classroom Environment." *International Journal of Educational Research* 43 (1–2): 25–38.
Goldstein, H. 2003. *Multilevel Statistical Models*. Oxford: Oxford University Press.
Graham, J. W. 2009. "Missing Data Analysis: Making it Work in the Real World." *Annual Review of Psychology* 60: 549–576. doi:10.1146/annurev.psych.58.110405.085530.
Grieve, A. M. 2010. "Exploring the Characteristics of 'Teachers for Excellence': Teachers' Own Perceptions." *European Journal of Teacher Education* 33 (3): 265–277. doi:10.1080/02619768.2010.492854.
Hamre, B. K., and R. C. Pianta. 2005. "Can Instructional and Emotional Support in the First-grade Classroom Make a Difference for Children at Risk of School Failure?" *Child Development* 76 (5): 949–967. doi:10.1111/j.1467-8624.2005.00889.x.
Hanushek, E. A., and S. G. Rivkin. 2009. "Harming the Best: How Schools Affect the Black–White Achievement Gap." *Journal of Policy Analysis and Management* 28 (3): 366–393. doi:10.1002/pam.20437.
Hox, J. J. 2010. *Multilevel Analysis, Techniques and Applications*. New York/Hove: Routledge.
Joyce, B., M. Weil, and E. Calhoun. 2000. *Models of Teaching*. Boston, MA: Allyn & Bacon.
Konstantopoulos, S. 2009. "Effects of Teachers on Minority and Disadvantaged Students' Achievement in the Early Grades." *The Elementary School Journal* 110 (1): 92–113. doi:10.1086/598845.
Konstantopoulos, S., and V. Chung. 2011. "Teacher Effects on Minority and Disadvantaged Students' Grade 4 Achievement." *The Journal of Educational Research* 104 (2): 73–86. doi:10.1080/00220670903567349.
Kyriakides, L. 2004. "Differential School Effectiveness in Relation to Sex and Social Class: Some Implications for Policy Evaluation." *Educational Research and Evaluation* 10 (2): 141–161. doi:10.1076/edre.10.2.141.27907.
Kyriakides, L. 2005. "Extending the Comprehensive Model of Educational Effectiveness by an Empirical Investigation." *School Effectiveness and School Improvement* 16 (2): 103–152. doi:10.1080/09243450500113936.

Kyriakides, L. 2007. "Generic and Differentiated Models of Educational Effectiveness: Implications for the Improvement of Educational Practice." In *International Handbook of School Effectiveness and Improvement*, edited by T. Townsend, 41–56. Dordrecht: Springer.

Kyriakides, L. 2008. "Testing the Validity of the Comprehensive Model of Educational Effectiveness: A Step towards the Development of a Dynamic Model of Effectiveness." *School Effectiveness and School Improvement* 19 (4): 429–446. doi:10.1080/09243450802535208.

Kyriakides, L., J. Campbell, and E. Christofidou. 2002. "Generating Criteria for Measuring Teacher Effectiveness Through a Self-evaluation Approach: A Complementary Way of Measuring Teacher Effectiveness." *School Effectiveness and School Improvement* 13 (3): 291–325. doi:10.1076/sesi.13.3.291.3426.

Kyriakides, L., and B. Creemers. 2008. "Using a Multidimensional Approach to Measure the Impact of Classroom-level Factors Upon Student Achievement: A Study Testing the Validity of the Dynamic Model." *School Effectiveness and School Improvement* 19 (2): 183–205. doi:10.1080/09243450802047873.

Kyriakides, L., and B. Creemers. 2009. "The Effects of Teacher Factors on Different Outcomes: Two Studies Testing the Validity of the Dynamic Model." *Effective Education* 1 (1): 61–85. doi:10.1080/19415530903043680.

Kyriakides, L., and B. Creemers. 2011. "Can Schools Achieve Both Quality and Equity? Investigating the Two Dimensions of Educational Effectiveness." *Journal of Education for Students Placed at Risk (JESPAR)* 16 (4): 237–254. doi:10.1080/10824669.2011.610269.

Kyriakides, L., B. Creemers, P. Antoniou, and D. Demetriou. 2010. "A Synthesis of Studies Searching for School Factors: Implications for Theory and Research." *British Educational Research Journal* 36 (5): 807–830. doi:10.1080/01411920903165603.

Kyriakides, L., B. Creemers, A. Panayiotou, G. Vanlaar, M. Pfeifer, C. Gašper, and L. McMahon. 2014. "Using Student Ratings to Measure Quality of Teaching in Six European Countries." *European Journal of Teacher Education* 37 (2): 125–143. doi:10.1080/02619768.2014.882311.

Luyten, H., and P. Sammons. 2010. "Multilevel Modelling." In *Methodological Advances in Educational Effectiveness Research*, edited by B. P. M. Creemers, L. Kyriakides, and P. Sammons, 246–276. London: Routledge.

Makel, M. C., and J. A. Plucker. 2014. "Facts Are More Important than Novelty: Replication in the Education Sciences." *Educational Researcher* 43 (6): 304–316. doi:10.3102/0013189X14545513.

Maxwell, S. E., K. Kelley, and J. R. Rausch. 2007. "Sample Size Planning for Statistical Power and Accuracy in Parameter Estimation." *Annual Review of Psychology* 59 (1): 537–563. doi:10.1146/annurev.psych.59.103006.093735.

Maulana, R., M. C. Opdenakker, P. den Brok, and R. Bosker. 2011. "Teacher-student Interpersonal Relationships in Indonesia: Profiles and Importance to Student Motivation." *Asia Pacific Journal of Education* 31 (1): 33–49. doi:10.1080/02188791.2011.544061.

Muijs, D., J. Campbell, L. Kyriakides, and W. Robinson. 2005. "Making the Case for Differentiated Teacher Effectiveness: An Overview of Research in Four Key Areas." *School Effectiveness and School Improvement* 16 (1): 51–70. doi:10.1080/09243450500113985.

Muijs, D., and D. Reynolds. 2001. *Effective Teaching. Evidence and Practice*. London: Sage.

Nye, B., S. Konstantopoulos, and L. V. Hedges. 2004. "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26 (3): 237–257. doi:10.3102/01623737026003237.

Panayiotou, A., L. Kyriakides, B. Creemers, L. McMahon, G. Vanlaar, M. Pfeifer, B. Matevz, and G. Rekalidou. 2014. "Teacher Behavior and Student Outcomes: Results of a European Study." *Educational Assessment, Evaluation and Accountability* 26: 73–93. doi:10.1007/s11092-013-9182-x.

Paris, S. G., and A. H. Paris. 2001. "Classroom Applications of Research on Self-regulated Learning." *Educational Psychologist* 36 (2): 89–101. doi:10.1207/S15326985EP3602_4.

Rasbash, J., C. Charlton, W. J. Browne, M. Healy, and B. Cameron. 2012. *MLwiN (Version 2.27) [Computer Software]*. Bristol: University of Bristol, Centre for Multilevel Modelling.

Raudenbush, S. W., and A. S. Bryk. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Reynolds, D. 2000. "School Effectiveness: The International Dimension." In *The International Handbook of School Effectiveness Research*, edited by C. Teddlie, 232–256. London: Routledge.
Reynolds, D. 2006. "World Class Schools: Some Methodological and Substantive Findings and Implications of the International School Effectiveness Research Project (ISERP)." *Educational Research and Evaluation* 12 (6): 535–560. doi:10.1080/13803610600874026.
Rosenshine, B., and R. J. Stevens. 1986. "Teaching Functions." In *Handbook of Research on Teaching*. 3rd ed, edited by M. C. Wittrock, 376–391. New York: Macmillan.
Roy, A., D. K. Bhaumik, S. Aryal, and R. D. Gibbons. 2007. "Sample Size Determination for Hierarchical Longitudinal Designs with Differential Attrition Rates." *Biometrics* 63 (3): 699–707. doi:10.1111/j.1541-0420.2007.00769.x.
Sammons, P., D. Nuttall, and P. Cuttance. 1993. "Differential School Effectiveness: Results from a Reanalysis of the Inner London Education Authority's Junior School Project Data." *British Educational Research Journal* 19 (4): 381–405. doi:10.1080/0141192930190407.
Savalei, V., and P. M. Bentler. 2010. *Structural Equation Modeling*. In *The Corsini Encyclopedia of Psychology*, edited by I. B. Weiner and W. E. Craighead, 1714–1716. Hoboken, NJ: John Wiley.
Schafer, J. L., and J. W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7 (2): 147–177. doi:10.1037//1082-989X.7.2.147.
Scheerens, J. 2012. *School Leadership Effects Revisited. Briefs in Education*. Dordrecht: Springer. ISBN:978-94-007-2767-0.
Scheerens, J. 2014a. "Evidence Based Educational Policy and Practice: The Case of Applying the Educational Effectiveness Knowledge Base." *Journal of Educational, Cultural and Psychological Studies* 9: 83–99. doi:10.7358/ecps-2014-009-sche.
Scheerens, J. 2014b. "Relevance of the Findings for Educational Policy and Future Research." In *Effectiveness of Time Investments in Education.* Briefs in Education, edited by J. Scheerens, 143–157. Dordrecht: Springer.
Scheerens, J., and R. Bosker. 1997. *The Foundations of Educational Effectiveness*. Oxford: Elsevier Science.
Schoenfeld, A. H. 1998. "Toward a Theory of Teaching-in-context." *Issues in Education* 4 (1): 1–94. doi:10.1016/S1080-9724(99)80076-7.
Schreiber, J. B., A. Nora, F. K. Stage, E. A. Barlow, and J. King. 2006. "Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review." *The Journal of Educational Research* 99 (6): 323–338. doi:10.3200/JOER.99.6.323-338.
Shepard, L. A. 1989. "Why We Need Better Assessment." *Educational Leadership* 46 (2): 4–8.
SPSS 20.0. 2011. *SPSS Inc. [Computer Software]*. Chicago, IL: IBM.
Stenmark, J. K. 1992. *Mathematics Assessment: Myths, Models, Good Questions and Practical Suggestions*. Reston, VA: NCTM.
Strand, S. 1999. "Ethnic Group, Sex and Economic Disadvantage: Associations with Pupils' Educational Progress from Baseline to the End of Key Stage 1." *British Educational Research Journal* 25 (2): 179–202. doi:10.1080/0141192990250204.
Strand, S. 2010. "Do Some Schools Narrow the Gap? Differential School Effectiveness by Ethnicity, Gender, Poverty, and Prior Achievement." *School Effectiveness and School Improvement* 21 (3): 289–314. doi:10.1080/09243451003732651.
Timmermans, A. C., S. Doolaard, and I. de Wolf. 2011. "Conceptual and Empirical Differences among Various Value-added Models for Accountability." *School Effectiveness and School Improvement* 22 (4): 393–413. doi:10.1080/09243453.2011.590704.
Vandecandelaere, M., S. Speybroeck, G. Vanlaar, B. De Fraine, and J. Van Damme. 2012. "Learning Environment and Students' Mathematics Attitude." *Studies in Educational Evaluation* 38 (3–4): 107–120. doi:10.1016/j.stueduc.2012.09.001.
Vanlaar, G., K. Denies, M. Vandecandelaere, J. Van Damme, J. Verhaeghe, M. Pinxten, and B. De Fraine. 2014. "How to Improve Reading Comprehension in High-risk Students: Effects of Class Practices in Grade 5." *School Effectiveness and School Improvement* 25 (3): 408–432. doi:10.1080/09243453.2013.811088.
Walberg, H. J. 1986. "What Works in a Nation Still at Risk." *Educational Leadership* 44 (1): 7–10. http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ343725.

Appendix 1. Items of the student questionnaire per teacher factor, with their factorloadings

Items per teacher factor	Factor loadings
Structuring quantitative	
At the beginning of the lesson, the teacher starts with what we covered in the previous lessons.	0.57
We spend time at the end of the lesson to go over what we have just learned.	0.71
Structuring qualitative	
In mathematics and science, we start the lesson with things that are easy to understand. As the lesson goes on, what we cover is more difficult.	0.54
My teacher helps us to understand how different activities (such as exercises, subject matter) during a lesson are related to each other.	0.72
Our teacher has good ways of explaining how the new things we are learning are related to things we already know.	0.84
Questioning	
When a pupil gives a wrong answer, the teacher helps her/him to understand her/his mistake and find the correct answer.	0.65
When I give a wrong answer to a question, the teacher helps me to understand my mistake and find the correct answer.	0.65
When the teacher asks us a question about the lesson, he/she asks us for the answer, but does not ask us to explain how we worked out the answer.	0.42
Our teacher uses words that are hard to understand when he/she asks us a question.	0.74
Modelling	
When we have problem-solving exercises and tasks in Mathematics and Science lessons, our teacher helps us by showing us easy ways or tricks to solve the exercises or tasks.	0.66
Our teacher lets us use our own easy ways or tricks to solve the exercises or tasks we have in Mathematics and Science.	0.48
Application	
At the end of each lesson, the teacher gives us exercises on what we have just learned.	0.56
During the lesson, our teacher often covers the same things that we have already learned or done exercises in.	0.60
When one of the pupils in the class is having difficulties with the lesson, our teacher goes to help him/her straight away.	0.70
Teacher–student interaction (part of the factor learning environment)	
Our teacher encourages us to ask questions if there is something that we don’t understand during the lesson.	0.67
Our teacher makes us feel that we can ask him/her for help or advice if we need it.	0.67
The teacher gives all pupils the chance to take part in the lesson.	0.62
Our teacher encourages us to work together with our classmates during mathematics and science lessons.	0.57
Time management	
Our teacher keeps on teaching us even though it is break-time or the lesson is supposed to be over.	0.65

(*Continued*)

Appendix 1. (*Continued*)

Items per teacher factor	Factor loadings
There are times we don't have the necessary materials for the lesson to take place (e.g. test tubes, thermometers, calculators, rulers)	0.62
There are times when I do not have anything to do during a lesson.	0.69
Dealing with misbehaviour (part of the factor learning environment)	
When a pupil gives a wrong answer in mathematics and science class, some of the other children in the class make fun of her/him.	0.82
There are some pupils in the classroom who tease some of their classmates during mathematics and science lessons.	0.80
When the teacher talks to a pupil after they have been naughty, sometimes after a while, that pupil will be naughty again.	0.65
The teacher has to stop teaching the class because one of the pupils is being naughty	0.45
Assessment	
A few days before the test, my teacher gives us similar exercises to those that will be in the test.	0.52
When we go over our homework, our teacher finds what we had problems with and helps us to overcome these difficulties.	0.75

Appendix 2. Items of the teacher questionnaire per school factor, with their factorloadings

Items per school factor	Factor loadings
Evaluation of school policy on teaching	
My school keeps systematic records concerned with: Student absenteeism, Teacher absenteeism; Special educational needs of students	0.57
The principal and/or other members of the school staff observe the way the teaching policy is put into practice and presents the results of their observations to staff	0.78
To evaluate the implementation of the school policy on teaching, we collect information from: teachers, students, parents	0.81
Teachers' capacity to implement the school policy on teaching (e.g. quantity of education, quality of education, provision of learning opportunities for students) is evaluated within the school	0.88
Information collected during evaluation of the school policy on teaching is used for re-designing the policy or for taking new decisions	0.89
Aspects of my school's policy on teaching which are considered problematic are evaluated further and/or in more detail	0.79
Evaluation of school learning environment	
My school keeps systematic records concerned with: Problems that arise among students during break time; The use of educational tools for teaching supplied by the school (e.g. maps, software etc.)	0.50
We evaluate the extent to which student discipline problems during break time are reduced as a result of the school policy	0.51
The principal and/or school staff observe the implementation of the learning environment policy and present the results of their observations to staff	0.94
Our school identifies the professional development/further education needs of its teachers	0.61
The evaluation of the school policy on the broader learning environment (e.g. further school climate, students' behaviour outside the classroom, the cooperation and interaction between teachers, the support of teachers and students, the learning objectives ...) is carried out in a way that refers to a single aspect of the policy each time (i.e. evaluation focuses on student behaviour, relations with parents etc. separately)	0.94
Information collected during the evaluation of the policy on the broader learning environment is used for re-designing the policy or for taking new decisions	0.93

(*Continued*)

Appendix 2. (*Continued*)

Items per school factor	Factor loadings
School policy evaluation results are useful to pinpoint areas in teaching for which we need support and/or further training	0.95
Collaboration and interaction between teachers	
At staff meetings in our school, we discuss and take decisions on issues concerned with: developing trust between teachers and children	0.67
We take into account research findings when we form a school policy concerned with: teacher collaboration	0.80
The teachers in my school cooperate with each other by exchanging ideas and material when teaching specific units or series of lessons.	0.71
Provision of sufficient learning resources	
I feel that I am positively influenced by the staff meetings in relation to use of visual aids and technological equipment in teaching (e.g. overhead projector, computer)	0.82
We take into account research findings when we form a school policy concerned with: use of resources for teaching provided by the school	0.90
Discussions at staff meetings help me to improve my practice in: Using different educational tools for teaching provided by the school	0.81
There is material on notice-boards in the school relevant to: the use of different educational tools for teaching provided by the school	0.75
Partnership policy	
The teachers in my school participate in educational school-based seminars (e.g. workshops) which deal with specific issues that the school faces	0.83
My school invites specialists in to conduct in-service training for teachers (e.g. an expert that works on developing students' creativity or other types of in-service)	0.78
The management team (principal and deputy heads) organises in-service seminars for a specific group of teachers when they think it is needed (e.g. newly appointed teachers)	0.82
Relation with community	
In parent-teacher meetings organised by the school, the way in which parents can help in dealing with the following issues are discussed:	0.89
Student absenteeism	
Homework	0.95
Addressing children's educational needs (e.g. gifted children, children with learning difficulties, children with special interests)	0.96
Parents providing learning opportunities in the school through activities organised on their own initiative (e.g. educational visits, educational games)	0.75

(*Continued*)

Appendix 2. (*Continued*)

Items per school factor	Factor loadings
At staff meetings, we usually make decisions on the ways in which parents can be involved in the learning process	0.59
My school has a clear policy for parental involvement in the learning process	0.56
In my school, there is an opportunity for different groups/people outside the school to become involved with and cooperate in the learning process of (for example, a basketball player of a local team together with teachers teaches different basketball techniques)	0.54
Policy on quantity of teaching	
I feel that I am positively influenced by the staff meetings in relation to the following:	0.52
Dealing with student absenteeism	
Homework	0.94
Policy on creating learning opportunities	
I feel that I am positively influenced by the staff meetings in relation to	
long-term planning of teaching	
interaction with students during break time	0.96
Discussions at staff meetings help me to improve my practice in: providing learning opportunities to students beyond the ones offered by the formal curriculum	0.92
There is material on notice-boards in the school relevant to: provision of learning opportunities beyond the ones provided by the formal curriculum	0.85
Policy on Quality of teaching	
I feel that I am positively influenced by staff meetings in relation to the following:	
Student evaluation	0.97
Structuring of lessons	0.97
Student orientation (i.e. helping students to understand why a unit is taught)	0.97
Using exercises to help students apply their learning (i.e. giving them tasks which apply the concepts taught to a situation in everyday life)	0.97
Asking questions and making good use of them	0.97
Strategies for learning	0.97
The learning environment of the classroom (e.g. promoting interaction among students)	0.97
We take into account research findings (e.g. recently published articles in scientific journals, results of research studies) in developing the school policy on teaching	0.94

Appendix 3. Multilevel analyses of the teacher factors and their interaction with the prior achievement level of the class

Mathematics

Fixed Part	BasicModel Est.	BasicModel S.E.	StructQuanT Est.	StructQuanT S.E.		StructQuaL Est.	StructQuaL S.E.		QuestQuaL Est.	QuestQuaL S.E.		Modelling Est.	Modelling S.E.	
Intercept	332.29	0.86	332.40	0.84***		332.90	0.87***		332.59	0.84***		332.33	0.85	***
PriorAch	0.68	0.01***	0.68	0.01***		0.68	0.01***		0.68	0.01***		0.68	0.01	***
CMPA	0.31	0.03***	0.30	0.03***		0.30	0.03***		0.29	0.03***		0.30	0.03	***
StructQuanT			8.12	2.85**										
CMPA.StructQuanT			−0.38	0.09***										
StructQuaL						3.93	2.97							
CMPA.StructQuaL						−0.30	0.10**							
QuestQuaL									10.42	3.76**				
CMPA.QuestQuaL									−0.40	0.12***				
Modelling												5.20	3.87	
CMPA.Modelling												−0.34	0.12**	
Applic														
CMPA.Applic														
TSinteract														
CMPA.TSinteract														
TimeManag														
CMPA.TimeManag														
Misbehaviour														
CMPA.Misbehaviour														
Assessment														
CMPA.Assessment														
Random Part	Est.	S.E.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
Class variance	268.23	24.31	252.87	23.40	5.7%	263.71	24.03	1.7%	254.67	23.51	5.1%	259.43	23.74	3.3%
Student variance	2154.29	32.53	2152.72	32.50	0.1%	2152.92	32.51	0.1%	2153.77	32.52	0.0%	2154.44	32.53	0.0%
−2*loglikelihood (diff.)	98606.3		98578.4	28.0***		98594.4	11.9**		98585.3	21.1***		98594.8	11.6**	

Fixed Part	Application Est.	Application S.E.		TSinteraction Est.	TSinteraction S.E.		TimeManag Est.	TimeManag S.E.		Misbehaviour Est.	Misbehaviour S.E.		Assessment Est.	Assessment S.E.	
Intercept	332.52	0.84	***	332.94	0.84	***	332.64	0.87	***	332.00	0.90	***	332.56	0.83	***
PriorAch	0.68	0.01	***	0.68	0.01	***	0.68	0.01	***	0.68	0.01	***	0.68	0.01	***
CMPA	0.29	0.03	***	0.27	0.03	***	0.24	0.03	***	0.26	0.03	***	0.29	0.03	***
StructQuanT															
CMPA.StructQuanT															
StructQuaL															
CMPA.StructQuaL															
QuestQuaL															
CMPA.QuestQuaL															
Modelling															
CMPA.Modelling															
Applic	9.86	3.69**													
CMPA.Applic	−0.49	0.12***													
TSinteract				16.95	4.06***										
CMPA.TSinteract				−0.56	0.13***										
TimeManag							19.91	3.03***							
CMPA.TimeManag							−0.05	0.09							
Misbehaviour										11.26	2.40***				
CMPA.Misbehaviour										0.08	0.07				
Assessment													11.26	3.22***	
CMPA.Assessment													−0.47	0.10***	
Random Part	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
Class variance	251.95	23.35	6.1%	242.04	22.79	9.8%	238.90	22.61	10.9%	253.91	23.47	5.3%	245.07	22.96	8.6%
Student variance	2153.43	32.51	0.0%	2153.34	32.50	0.0%	2153.78	32.51	0.0%	2153.78	32.52	0.0%	2153.22	32.50	0.0%
−2*loglikelihood (diff.)	98580.0	26.3***		98565.3	41.0***		98562.5	43.8***		98584.2	22.1***		98569.2	37.1***	

SCIENCE

Fixed Part	BasicModel Est.	BasicModel S.E.	StructQuanT Est.	StructQuanT S.E.		StructQuaL Est.	StructQuaL S.E.		QuestQuaL Est.	QuestQuaL S.E.		Modelling Est.	Modelling S.E.	
Cons	319.73	0.96***	319.86	0.95***		320.31	0.97***		319.90	0.95***		319.74	0.95***	
PriorAch	0.54	0.01***	0.54	0.01***		0.54	0.01***		0.54	0.01***		0.54	0.01***	
CMPA	0.36	0.03***	0.35	0.03***		0.35	0.03***		0.35	0.03***		0.36	0.03***	
StructQuanT			9.55	3.22**										
CMPA.StructQuanT			−0.09	0.09										
StructQuaL						2.55	3.29							
CMPA.StructQuaL						−0.32	0.10**							
QuestQuaL									5.93	4.21				
CMPA.QuestQuaL									−0.24	0.11*				
Modelling												6.36	4.32	
CMPA.Modelling												−0.25	0.12*	
Applic														
CMPA.Applic														
TSinteract														
CMPA.TSinteract														
TimeManag														
CMPA.TimeManag														
Misbehaviour														
CMPA.Misbehaviour														
Assessment														
CMPA.Assessment														
Random Part	Est.	S.E.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
Class variance	379.75	30.54	371.21	29.97	2.2%	371.71	29.99	2.1%	372.97	30.06	1.8%	374.03	30.19	1.5%
student variance	1930.64	29.45	1930.60	29.45	0.0%	1930.47	29.44	0.0%	1930.81	29.45	0.0%	1930.60	29.44	0.0%
−2*loglikelihood (diff.)	95954.9		95945.3	9.6**		95945.3	9.6**		95948.2	6.7*		95948.4	6.5*	

Fixed Part	Application Est.	Application S.E.		TSinteraction Est.	TSinteraction S.E.		TimeManag Est.	TimeManag S.E.		Misbehaviour Est.	Misbehaviour S.E.		Assessment Est.	Assessment S.E.	
Cons	319.87	0.95***		320.21	0.96***		319.78	1.01***		319.30	1.01***		319.89	0.95***	
PriorAch	0.54	0.01***		0.54	0.01***		0.54	0.01***		0.54	0.01***		0.54	0.01***	
CMPA	0.35	0.03***		0.31	0.03***		0.32	0.03***		0.34	0.03***		0.34	0.03***	
StructQuanT															
CMPA.StructQuanT															
StructQuaL															
CMPA.StructQuaL															
QuestQuaL															
CMPA.QuestQuaL															
Modelling															
CMPA.Modelling															
Applic	14.18	4.10***													
CMPA.Applic	−0.16	0.13													
TSinteract				22.51	4.62***										
CMPA.TSinteract				−0.23	0.12†										
TimeManag							12.48	3.55***							
CMPA.TimeManag							0.02	0.09							
Misbehaviour										5.69	2.67*				
CMPA.Misbehaviour										0.10	0.08				
Assessment													12.16	3.65***	
CMPA.Assessment													−0.14	0.10	
Random Part	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
Class variance	366.44	29.70	3.5%	353.88	29.02	6.8%	369.18	29.85	2.8%	376.14	30.41	1.0%	366.85	29.73	3.4%
student variance	1930.86	29.45	0.0%	1931.08	29.45	0.0%	1930.48	29.44	0.0%	1930.19	29.46	0.0%	1930.85	29.45	0.0%
−2*loglikelihood (diff.)	95941.2	13.7**		95927.9	27.0***		95942.5	12.4**		95948.9	6.0*		95941.6	13.3**	

Notes: $p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; CMPA: class mean prior achievement.; %Expl: percent of the variance explained by the addition to the basic model.

Appendix 4. Multilevel analyses of the school factors and their interaction with the prior achievement level of the school

Mathematics	Basic Model			EvaluationTeaching			EvaluationSLE			Collaboration			Resources		
Fixed Part	Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.	
Intercept	332.02	1.00***		332.52	0.97***		332.14	0.99***		332.59	0.97***		332.29	0.98***	
PriorAch	0.68	0.01***		0.68	0.01***		0.68	0.01***		0.68	0.01***		0.68	0.01***	
SMPA	0.32	0.03***		0.29	0.03***		0.32	0.03***		0.29	0.03***		0.32	0.03***	
EvaluationTeaching				15.58	3.76***										
SMPA.EvaluationTeaching				−0.49	0.14***										
EvaluationSLE							13.59	4.79**							
SMPA.EvaluationSLE							−0.31	0.18†							
Collaboration										20.22	4.50***				
SMPA.collaboration										−0.48	0.16**				
Resources													17.54	3.73***	
SMPA.Resources													−0.30	0.13*	
Community															
SMPA.Community															
Partnership															
SMPA.Partnership															
PTquantity															
SMPA.PTquantity															
PTlearningOpp															
SMPA.PTlearningOpp															
PTquality															
SMPA.PTquality															
Random Part	Est.	S.E.		Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
School variance	227.50	25.00		202.57	23.09	11.0%	219.17	24.48	3.7%	206.27	23.38	9.3%	213.07	24.04	6.3%
Student variance	2199.89	32.79		2198.11	32.75	0.1%	2198.89	32.77	0.0%	2197.72	32.75	0.1%	2197.28	32.74	0.1%
−2*loglikelihood (diff.)	98608.0			98574.6	33.4***		98595.4	12.6**		98576.9	31.0***		98582.3	25.7***	

Mathematics	Community			Partnership			PTquantity			PTlearningOpp			PTquality		
Fixed Part	Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.	
Intercept	332.69	0.97***		332.78	0.99***		332.25	0.98***		332.22	0.99***		332.44	0.97***	
PriorAch	0.68	0.01***		0.67	0.01***		0.68	0.01***		0.68	0.01***		0.68	0.01***	
SMPA	0.27	0.04***		0.28	0.04***		0.32	0.03***		0.31	0.03***		0.30	0.03***	
EvaluationTeaching															
SMPA.EvaluationTeaching															
EvaluationSLE															
SMPA.EvaluationSLE															
Collaboration															
SMPA.collaboration															
Resources															
SMPA.Resources															
Community	17.97	3.26***													
SMPA.Community	−0.27	0.11*													
Partnership				21.90	4.34***										
SMPA.Partnership				−0.34	0.15*										
PTquantity							12.21	3.37***							
SMPA.PTquantity							−0.13	0.12							
PTlearningOpp										11.90	3.49***				
SMPA.PTlearningOpp										−0.29	0.13*				
PTquality													16.13	3.18***	
SMPA.PTquality													−0.31	0.11**	
Random Part	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
School variance	203.13	23.12	10.7%	212.20	23.99	6.7%	215.76	24.25	5.2%	220.96	24.61	2.9%	207.52	23.48	8.8%
Student variance	2196.42	32.73	0.2%	2196.77	32.73	0.1%	2199.40	32.77	0.0%	2197.92	32.75	0.1%	2197.01	32.74	0.1%
−2*loglikelihood (diff.)	98568.2	39.8***		98579.3	28.7***		98593.9	14.1***		98593.2	14.8***		98575.4	32.6***	

SCIENCE	Basic Model			EvaluationTeaching			EvaluationSLE			Collaboration			Resources		
Fixed Part	Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.	
Cons	319.21	1.12***		319.52	1.10***		319.34	1.12***		320.02	1.10***		319.50	1.12	
PriorAch	0.54	0.01***		0.54	0.01***		0.54	0.01***		0.54	0.01***		0.54	0.01***	
SMPA	0.39	0.04***		0.36	0.04***		0.38	0.04***		0.35	0.04***		0.38	0.04***	
EvaluationTeaching				20.34	4.24***										
SMPA.EvaluationTeaching				−0.20	0.14										
EvaluationSLE							16.61	5.39**							
SMPA.EvaluationSLE							−0.31	0.17†							
Collaboration										25.64	5.04***				
SMPA.collaboration										−0.48	0.15**				
Resources													18.37	4.24***	
SMPA.Resources													−0.46	0.14***	
Community															
SMPA.Community															
Partnership															
SMPA.Partnership															
PTquantity															
SMPA.PTquantity															
PTlearningOpp															
SMPA.PTlearningOpp															
PTquality															
SMPA.PTquality															
Random Part	Est.	S.E.		Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
School variance	316.33	31.78		293.95	30.15	7.1%	316.10	31.69	0.1%	292.19	30.01	7.6%	317.30	31.74	−0.3%
Student variance	1989.05	29.94		1987.37	29.91	0.1%	1985.92	29.94	0.2%	1985.45	29.88	0.2%	1982.83	29.95	0.3%
−2*loglikelihood (diff.)	95962.4			95936.5	25.9***		95948.2	14.2***		95926.4	36.0***		95935.3	27.1***	

SCIENCE	Community			Partnership			PTquantity			PTlearningOpp			Ptquality		
Fixed Part	Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.		Est.	S.E.	
Cons	320.05	1.12***		320.20	1.13***		319.37	1.11***		319.56	1.12***		319.81	1.11***	
PriorAch	0.54	0.01***		0.54	0.01***		0.54	0.01***		0.54	0.01***		0.54	0.01***	
SMPA	0.33	0.04***		0.34	0.04***		0.39	0.04***		0.38	0.04***		0.36	0.04***	
EvaluationTeaching															
SMPA.EvaluationTeaching															
EvaluationSLE															
SMPA.EvaluationSLE															
Collaboration															
SMPA.collaboration															
Resources															
SMPA.Resources															
Community	19.74	3.71***													
SMPA.Community	−0.18	0.12													
Partnership				23.16	4.94***										
SMPA.Partnership				−0.42	0.16**										
PTquantity							10.12	3.77**							
SMPA.PTquantity							−0.46	0.12***							
PTlearningOpp										12.09	3.94**				
SMPA.PTlearningOpp										−0.42	0.14**				
PTquality													13.38	3.61***	
SMPA.PTquality													−0.30	0.12**	
Random Part	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.	Est.	S.E.	%Expl.
School variance	294.46	30.18	6.9%	306.62	31.08	3.1%	309.74	31.31	2.1%	315.14	31.73	0.4%	302.57	30.77	4.4%
Student variance	1986.06	29.89	0.2%	1985.06	29.88	0.2%	1985.33	29.88	0.2%	1985.65	29.94	0.2%	1987.27	29.91	0.1%
−2*loglikelihood (diff.)	95931.1	31.3***		95936.7	25.7***		95940.4	22.0***		95946.2	16.2***		95943.3	19.2***	

Notes: †$p < 0.1$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; SMPA: school mean prior achievement.