School Effectiveness and School Improvement
Vol. 18, No. 2, June 2007, pp. 125–145

Routledge
Taylor & Francis Group

Educational Leadership and Pupil Achievement: The choice of a valid conceptual model to test effects in school effectiveness research

Sven De Maeyer[a]*, Rita Rymenans[a], Peter Van Petegem[a], Huub van den Bergh[b], and Gert Rijlaarsdam[c]
[a]*Institute for Educational and Informational Sciences, University of Antwerp, Belgium,* [b]*University Utrecht, The Netherlands,* [c]*University of Amsterdam, The Netherlands*

(Received 13 January 2006; accepted 29 May 2006)

School effectiveness research often uses multilevel models in which only direct effects of characteristics of schools on pupil achievement are modelled. Recently, more attention is given to conceptual models that assume indirect and antecedent effects. In this paper, we elaborate on these models and show that the findings from school effectiveness research are influenced by the conceptual model underlying the analyses. We do this by focusing on the effect of ''integrated leadership'' on 2 outcome measures: mathematics achievement and mother tongue achievement. Four different conceptual models will be tested by means of multilevel structural equation modelling. Therefore we use data from a school effectiveness research in Flanders (850 fourth graders and 847 sixth graders within 47 schools). The results of these analyses show that the conclusion whether integrated leadership has an effect or not on pupil achievement, is dependent on the choice of the conceptual model.

Introduction

A characteristic of schools that appears in many lists of effectiveness-enhancing factors is educational leadership. Nevertheless, there are no consistent results for the relationship between educational leadership and pupil achievement (Scheerens &

*Corresponding author. Institute for Educational and Informational Sciences, EduBROn, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium.
Email: sven.demaeyer@ua.ac.be

ISSN 0924-3453 (print)/ISSN 1744-5124 (online)/07/020125–21

DOI: 10.1080/09243450600853415

Bosker, 1997). According to Hallinger and Heck (1996, 1998) and Witziers, Bosker, and Krüger (2003), this can be due to conceptual and methodological choices.

Most of the studies in school effectiveness research start from a direct effect model. Typically, a multilevel regression model is set up and different school and/or class characteristics are added to the model. The question arises as to whether this is the most appropriate way to analyse school effectiveness, and, more specifically, the effect of educational leadership. Is it valid to assume that school leadership has an autonomous and direct effect on pupil achievement? It seems more plausible that school leadership produces an effect through other processes taking place in the school and the class (such as school climate or the teachers' proficiency). Nevertheless, most school effectiveness research models the autonomous effects of school and class characteristics. A possible reason for this approach is that, until recently, it was very difficult to set up more complex models and to test them statistically. Recent developments in statistical analysis techniques, like the possibility to test multilevel path-models with a multilevel structural equation analysis, have made it possible to do so.

As such, these direct effect models are not problematic, although they give less insight into the complex educational reality. The use of the direct effect model only becomes problematic when it leads to substantially different findings than a more internal valid model. This validity question is addressed in this article. The main purpose of this study is to demonstrate the importance of the choice of a sound conceptual model for testing the effect of educational leadership. The findings in this article can be generalised for school effectiveness research. Based on these findings, we show the importance of the conceptual and methodological choices in school effectiveness research in general.

First, we give a brief review of the relevant literature on the influence of educational leadership on pupil achievement and focus on key elements in the debate on the effect of this characteristic of schools. Based on this review, we deduce four possible conceptual models. Thereafter, we test these models based on data from a school effectiveness study carried out in Flemish technical secondary education (De Maeyer & Rymenans, 2004). We show that the findings strongly depend on the chosen conceptual model. In the discussion, we will reflect on the consequences of these findings for future school effectiveness research.

The Effect of Educational Leadership on Achievement

The Far West Model

Since the earliest studies in the field of school effectiveness research, scholars have paid attention to the effects of school leadership on pupil achievement (Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979; Edmonds, 1979; Rutter, Maughan, Mortimore, Ouston, & Smith, 1979). These researchers assumed that there is a direct relationship between school leaders' behaviour and the learning outcomes of the

pupils in their school. These studies were criticised by Bossert, Dwyer, Rowan, and Lee (1982), because they did not explain how the behaviour of school leaders could affect the learning outcomes (Hallinger & Heck, 1998; Krüger, Witziers, Sleegers, & Imants, 1998; Witziers et al., 2003).

Based on the results of a longitudinal case study in secondary education, Bossert and his colleagues developed a descriptive model of educational leadership in their Far West Laboratory in San Francisco (Bossert et al., 1982). The so-called Far West model was innovative because it depicted school leadership in an integrated way. The actions of a school leader occupy a central place in the model. They are influenced by the context in which a school leader has to work: educational policy at the macrolevel, the school environment, and the leader's own vision and experience. The school leader's actions influence the learning results in an indirect way, through actions directed at the organisation of the school and at the pedagogical and educational climate (Krüger et al., 1998).

The Far West model has been empirically validated in some American studies. An example is the study of Heck, Marcoulides, and Lang (1991) that examined these relationships in two different cultural contexts, California and the Marshall Islands.

> The association between principal instructional leadership and school achievement tested in this research study provided empirical support for the Bossert et al. (1982) model, indicating that school achievement can be readily predicted through an examination of principals' leadership practices. (Heck et al., 1991, p. 132)

Based on this Far West model, Hallinger (1983) developed the ''Principal Instructional Management Rating Scale'' (PIMRS). In this instrument the concept of ''instructional management'' occupies a central place. This concept focuses on a school leader's mental and physical actions that enable others in an organisation to carry out their work in such a way that the educational goals of the organisation are achieved (Krüger et al., 1998). In the PIMRS instrument 10 leadership functions or practices are measured that can be reduced to three dimensions: (1) defining the school mission, (2) directing the educational programme, and (3) promoting a climate focused on learning. This PIMRS instrument is used in different American research studies to examine the influence of school leadership on achievement, but it has not yet led to similar results.

Within the research on the effects of school leadership, different approaches to this concept have led to different results (Marks & Printy, 2003). Two conceptual models have dominated the discussion on educational leadership: ''instructional leadership'' and ''transformational leadership'' (Hallinger, 2003; Marks & Printy, 2003). Recently, Marks and Printy developed an alternative conceptualisation, ''integrated leadership'', based on a case study in American primary and secondary education:

> Transformational and shared instructional leadership are complementary, in our view, but neither conceptualization embraces the other. When they operate in tandem, however, the leadership approaches are integrated. (Marks & Printy, 2003, p. 373)

Conceptual Models for the Relation Between School Leadership and Achievement

Beside the conceptualisation of educational leadership, the way in which the effect of it on achievement is conceptualised has an autonomous influence on the (absence of) results. Most of the research studies examine (in an implicit way) a direct effect of school leadership. According to Bossert et al. (1982), an indirect effect, through the way in which the school is organised and the pedagogical and educational climate, seems more plausible. Following Pitner (1988), Hallinger and Heck (1996, 1998) described five conceptual models for the relationship between leadership and achievement:

1. The direct effect model assumes a direct relationship between school leaders' practice and pupil achievement. Moreover, this model assumes that these effects can be measured in a valid way, keeping all other possible variables constant.
2. The direct effect model with antecedents is a variant of the previous model in which effects of the context on leadership and achievement are assumed. A context factor often used is the mean socioeconomic status of the school population. Its effect is not seen as an interaction effect between leadership and achievement.
3. The indirect effect model assumes that the effect of school leadership goes through intermediate variables, such as the organisation of the school and the school climate.
4. The indirect effect model with antecedents extends the previous model with context factors that have an influence on leadership by a school leader. In this model, leadership is seen both as a function of the context and as a factor influencing achievement through the organisation and climate of the school.
5. The recursive effect model assumes that the relations between the school leadership, intermediate variables, and achievement are interactive. This model departs from the assumption that leaders adapt their behaviour to the organisation in which they work, and that their ideas and practices can change during time.

Hallinger and Heck (1996, 1998) present an extensive review of 41 international effectiveness studies on the effect of school leadership. From this review, they conclude that only two of these studies assumed a recursive model. This can be ascribed to the fact that studying this model needs a longitudinal design to test the dynamics between these different aspects. That is essential given the fact that this model assumes that leadership at moment 1 can influence achievement at moment 2, which can in its turn influence leadership at another moment, and so on.

Furthermore, their review showed that the direct effect model is limited in its usability for examining the effects of school leadership on achievement (see also Witziers et al., 2003). Even with the most advanced statistical models (such as structural equation modelling) there is no convincing evidence for these direct effects (Hallinger & Heck, 1996, 1998). Given its conceptual limits, this model cannot make

a substantial contribution to the understanding of the relationship between educational leadership and achievement (Hallinger & Heck, 1996, 1998; Witziers et al., 2003).

From their review, Hallinger and Heck conclude that educational leadership has an indirect effect on pupil achievement. Although the effect is relatively small, it seems to be significant, which leads to the conclusion that educational leadership contributes to the effectiveness of schools. Another important conclusion from this review is that it was not until the introduction of more advanced methods of analysis like multilevel structural equation modelling that more consistent results from these effects were found.

> It is interesting to note that greater consistency in findings of principal effects only emerged after these methods began to be used by researchers. This reinforces the important of bringing conceptual and methodological power to the study of leadership effects. (Hallinger & Heck, 1998, p. 167)

The Intermediate Variables

The question arises of through which intermediate variables school leadership can influence pupil achievement. In their review, Hallinger and Heck (1998) organise the results according to four domains, based on the theoretical models of Leithwood (1994) and Ogawa and Bossert (1995): (1) vision and goal setting, (2) organisational structure and social networks, (3) human capital, and (4) organisational culture.

The most consistent findings are found for the first domain: vision development and coherent goals (Hallinger, 2003; Witziers et al., 2003). These studies show the importance of the school leaders in making tangible the goals, mission, and vision of the school. Moreover, they have an influence on the consensus about these goals. These characteristics prove to be effectiveness enhancing.[1]

For the domain of organisational structure and social networks, there is empirical evidence of a positive effect of transformational leadership within different cultural contexts. Engagement, co-operation, and participation in decision-making seem to be key characteristics of effective schools (Heck et al., 1991; Leithwood, 1994).

A third domain that can play an intermediate role between leadership and achievement is the human capital of the school. According to Leithwood (1994), people form the cornerstone of the transformational leadership model, although effects are also found for instructional leadership. The essence is that effective schools show a stronger educational and social cohesion.

Recent theories emphasise the importance of the fourth domain: the organisational culture, or the importance of shared values and standards. There is only a small amount of empirical evidence for this domain, because that focus is still quite new within research on educational leadership.

Effects of Context Factors on Educational Leadership

For both the direct and the indirect effect model, there is an alternative model that includes antecedents. These alternative models assume that context factors have an

influence on educational leadership. In their descriptive model, Bossert et al. (1982) emphasised the influence of these context factors (macrolevel educational policy, school environment, the school leader's personal characteristics) on educational leadership.

Although there are many conceptual and methodological problems concerning research on the influence of these context factors, Hallinger and Heck (1996, 1998) and Hallinger (2003) find a constant in different effectiveness studies. Frequently, evidence is found for the hypothesis that socioeconomic factors in the school and the school environment affect the school leaders' behaviour and the effectiveness of the school.

Re-analyses With an Alternative Conceptual Model

Witziers et al. (2003) performed a multilevel meta-analysis on approximately 40 international effectiveness studies for the period 1986 to 1996. They focused on studies based on the direct effect model, because research based on the indirect effect model is scarce (Hallinger & Heck, 1996, 1998). They attempted to estimate the effect size of educational leadership on achievement, and to examine which factors interact with this effect size. These meta-analyses produced the following results:

- On average, educational leadership has a significant positive effect on achievement, although a very small one: This variable cannot explain much more than 1% of the total variance in pupil achievement. More refined analyses showed that there are differences according to the context in which the studies were carried out. For instance, in Dutch research no effect of school leadership has been found (Scheerens & Bosker, 1997; Van de Grift, 1990). Another important factor is the educational level: Different effects are recorded in secondary education versus those from primary education.
- Research studies that assume that educational leadership is a one-dimensional concept do not find a significant effect on pupil achievement.
- For four school leaders' practices from the PIMRS instrument a positive relationship with pupil achievement was found: supervision and evaluation, monitoring, visibility, and—the most relevant characteristic—defining and communicating a school's mission.

Based on these results, Witziers et al. (2003) conclude that a better conceptualisation of educational leadership is needed and that future research should focus on intermediate variables and context factors. Therefore they advise that effectiveness researchers should apply the indirect effect model (with antecedents). Re-analyses of research findings based on the direct effect model with an alternative conceptual model have highlighted more complex relations between context, leadership, intermediate variables, and pupil achievement (Bosker, De Vos, & Witziers, 2000; Hallinger, Bickman, & Davis, 1996; Hallinger & Heck, 1996, 1998). One study even shows that the direct effect of school leadership disappears when indirect effects are modelled (Witziers et al., 2003).

In this respect, the re-analysis Hallinger and Heck (1996) performed on the data of Van de Grift (1990), is confronting. They found that, after fine-tuning the concept, educational leadership had an indirect effect on pupil achievement, whereas Van de Grift (1990) had not been able to find a (direct) effect within the same data.

A Case Study

From this literature review, we conclude that several authors have outlined different models to test the effect of educational leadership on pupil achievement and that the lack of consistent findings can be due to this divergent conceptualisation. In this article, we offer empirical evidence for the importance of choosing an accurate conceptual model. The main purpose of this contribution is to demonstrate the importance of the chosen conceptual model on the research findings. The central question we want to address is: Does the conclusion whether educational leadership has an influence on pupil achievement or not depend upon the model used to test this effect?

To answer this question, we present a re-analysis of the data from a school effectiveness study performed in Belgian (Flemish) vocational and technical secondary education (De Maeyer & Rymenans, 2004). Based on this dataset, we examine four alternative conceptual models for the effect of ''integrated leadership''. For an extensive description of this study we refer to De Maeyer, Rymenans, Daems, Van Petegem, and Van den Bergh (2003).

Sample

In this study, pupils from the fourth grade (15 or 16 years of age) and the sixth grade (17 or 18 years of age) within the same school took tests on different output measures. The sample was representative for all Flemish schools providing technical education. We gathered data in 47 schools. Within each school, 20 pupils for both grades were randomly sampled over all classes. In the final sample we have data for 850 fourth graders and 847 sixth graders resulting in a response rate of approximately 90% for both populations. To measure certain characteristics of schools we surveyed 20 teachers in every school.

Variables

Output measures. In their study, De Maeyer et al. (2003) used four different output measures: mathematics achievement, mother tongue achievement, problem-solving behaviour, and well-being. In this re-analysis, we focus on the cognitive achievement measures (mathematics and reading). The influence of educational leadership on both achievement measures was tested within one model, because they strongly correlate both at pupil and school level. In previous research, we argued that neglecting this possible correlation can distort the research findings (De Maeyer & Rymenans, 2004).

For mathematics, functional proficiency was measured. The test contained 29 items that related to mathematical problems faced by a learner in daily practice (e.g., the measurement of a surface when preparing to lay a wooden floor). Reading proficiency was measured by means of a functional reading test that contained 26 items related to different types of texts (e.g., a newspaper article and a manual). For the analyses, we arbitrarily transformed both these variables to a variable with a mean score of 100 and a standard deviation of 15 points.

Educational leadership. In this analysis, we focused on the effect of ''integrated leadership'' as defined by Marks and Printy (2003). Based on a multilevel confirmative factor analysis, De Maeyer and Rymenans (2004) built a measurement model for integrated leadership based on the answers of 20 teachers in every school for 10 items with a 5-point measuring scale. For an elaborate description of the construction of this variable we refer to De Maeyer and Rymenans. The standardised factor scores derived from a multilevel confirmative factor analysis were used. These scores were calculated by the software package M*Plus* (Muthén & Muthén, 1998–2001).

Intermediate variables. Given the main purpose of this paper (to demonstrate the importance of the chosen conceptual model on the research findings, rather than testing a specific theoretical model on school effectiveness), this analysis examined the indirect effect of integrated leadership through only one of the four domains as described by Hallinger and Heck (1998), namely organisational culture. More specifically, we focused on the intermediate effect of an achievement oriented school climate, labelled as ''academic climate''. In schools scoring high on this characteristic, the school policy lays a strong emphasis on the performance of their pupils. This emphasis is shared by teachers and the principal. It results in shared values within the school's staff. This school characteristic was measured through a multilevel confirmative factor analysis on the answers of 20 teachers in every school for seven items. A more detailed description of this concept and the way in which it was measured in this study can be found in De Maeyer and Rymenans (2004). The scores for this variable ''academic climate'' are the standardised factor scores as derived by the M*Plus* package.

Background variables. Within the different conceptual models, we controlled for pupil background variables that could have possibly affected pupil achievement: gender, IQ, socioeconomic status (SES), linguistic ethnic background (LEB), and grade.

Gender and grade are two dummy variables with a score of one respectively for girls and for pupils in the sixth grade.

IQ is measured by a subset of the nonverbal PSB test (Horn, 1969). This test contains 40 items consisting of a series of signs in which one does not fit. Pupils have 5 min to complete the test. A pupil's score is the number of items correct. In the analysis, this variable is transformed to a z-score.

To create a variable for SES, five different indicators were combined: both father's and mother's educational attainments, both father's and mother's position on the

labour market, and father's unemployment history. To combine these five indicators in one underlying variable measuring SES, we did a categorical principal component analysis (PRINCALS) (Gifi, 1990), described in De Maeyer et al. (2003). For the analyses, the resulting variable was normalised: A high score indicates that a pupil has a high-SES background.

Language spoken at home and both parents' ethnic background cannot fully describe a pupil's ethnic background in the specific situation for Flanders.[2] Therefore, we created a new variable, ''linguistic ethnic background'' (LEB), which is a combination of three indicators: the language spoken at home, father's nationality, and mother's nationality. These indicators are combined in a PRINCALS analysis, on which more details can be found in De Maeyer et al. (2003). The PRINCALS scores were normalised before they were used in further analyses. Pupils scoring high on this variable were pupils whose parents both have Belgian nationality and who only speak Dutch at home. At the other end of the continuum were pupils with, say, Turkish parents who never speak Dutch at home.

Context variables. At school level, we controlled for four context variables: percentage of girls, mean IQ, mean SES, and mean LEB. These context variables will also be modelled as antecedent variables. In other words, in the antecedent model we assume that these context variables influence the school leaders' behaviour.

Conceptual Models

Four conceptual models were used to test the effect of integrated leadership on achievement:

1. The direct effect model (Model 1)
2. The indirect effect model (Model 2)
3. The direct and indirect effect model (Model 3)
4. The antecedents model (Model 4).

These four models will be described and visualised.

Model 1: The direct effect model. In the first model, we assume only a direct effect from integrated leadership on pupil achievement (see Figure 1). This model is very similar to the classic effectiveness studies that use the multilevel regression approach. One point of difference is the fact that in multilevel structural equation modelling (SEM) it can be taken into account that explanatory variables may correlate, which is impossible to model in the multilevel regression approach. In Model 1, we assume a correlation between explanatory variables at both the pupil and the school levels.

At pupil level, an effect of gender, IQ, SES, linguistic ethnic background (LEB), and grade was assumed. Moreover, we expected that SES and LEB would be positively correlated. Finally, a correlation between grade and IQ was modelled because the test used to measure IQ is not age independent (De Maeyer & Rymenans, 2004).

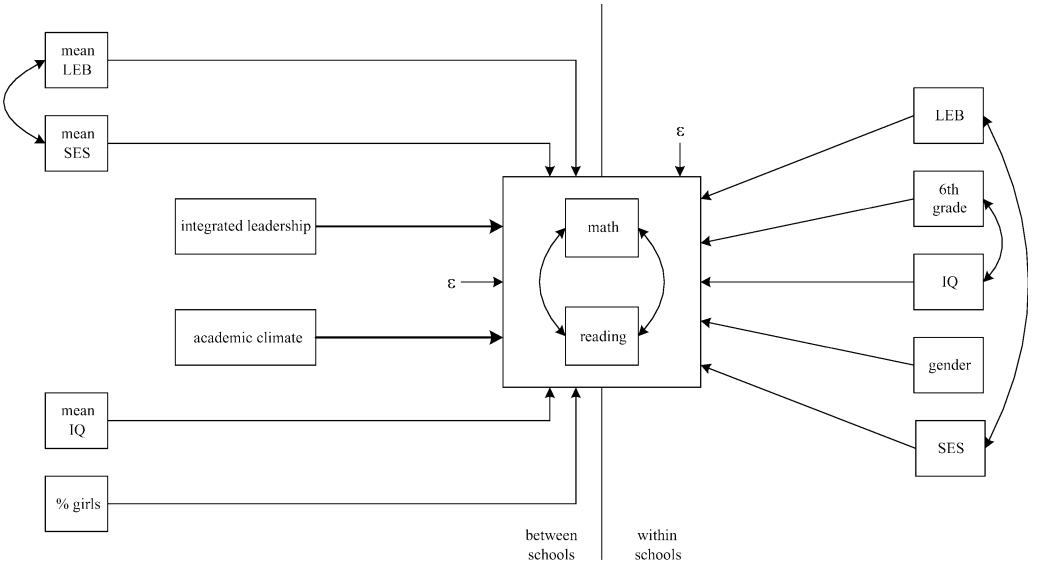

Figure 1. The direct effect model

The between-schools part models the effects we assumed from school characteristics. Figure 1 shows that an autonomous direct effect from integrated leadership is hypothesised, independent from academic climate. For this school characteristic, a direct effect on achievement was assumed too. In this model we controlled for some context factors: mean LEB score, mean SES score, mean IQ, and percentage of girls. For these context factors an autonomous direct effect on the pupils' achievement was expected. On the other hand, no relationship between these context factors and integrated leadership nor academic climate was modelled. Finally, we assumed a relationship between the mean SES and the mean LEB of the school.

With the double-pointed arrows between mathematics achievement and reading proficiency at both levels (between and within schools), a correlation between both dependent variables was assumed at both levels. Finally, two error terms were modelled (εs), indicating that we did not expect that the explanatory variable would explain all the variance in the dependent variables at both levels.

Model 2: The indirect effect model. In the indirect effect model, we did not expect a direct effect of integrated leadership on achievement, but only an indirect one. Put differently, we expected that integrated leadership would affect the academic climate which in its turn would affect pupil achievement. Alternatively, in this model integrated leadership only has an effect on achievement through the stimulation of an academic climate. Figure 2 gives a partial visualisation of this model; the relationships that are not shown in this figure are similar to those assumed in Figure 1.

Model 3: The direct and indirect effect model. Model 3 is based on the hypothesis that integrated leadership has both a direct and an indirect effect on pupil achievement

(see Figure 3). Introducing a direct effect, besides an indirect one, assumes that integrated leadership can also have a direct effect on pupil achievement. Note that this effect can be an indirect effect through other variables that are not measured in our study. All the other relationships are similar to those in Model 1.

Model 4: The antecedents model. This conceptual model is an extension of Model 3 (see Figure 4). It also assumes a direct and an indirect effect of integrated leadership. What makes this model different is the direct effect it assumes from the context

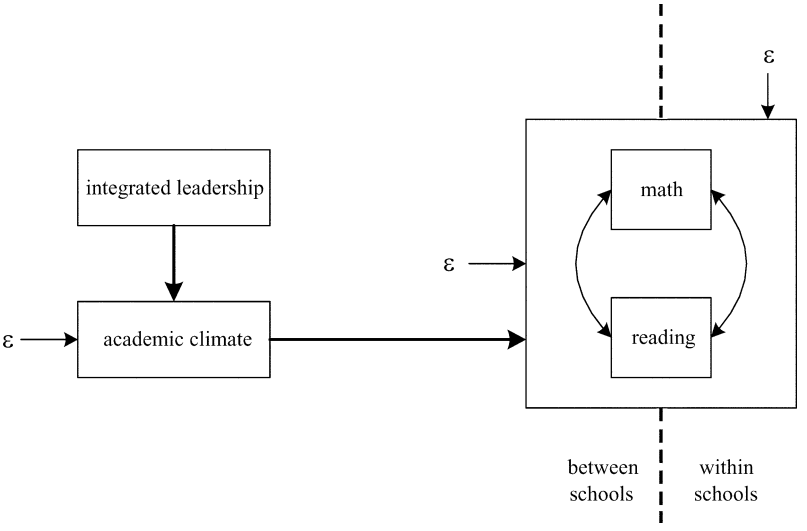

Figure 2. The indirect effect model (partial)

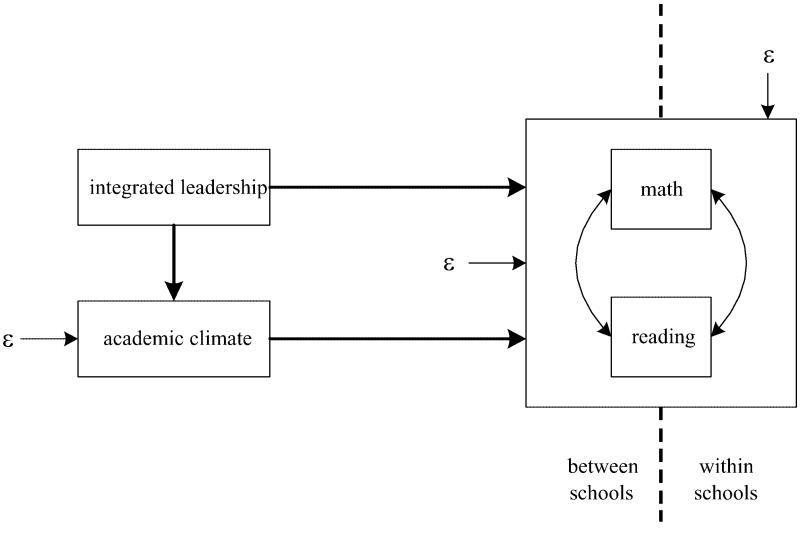

Figure 3. The direct and indirect effect model (partial)

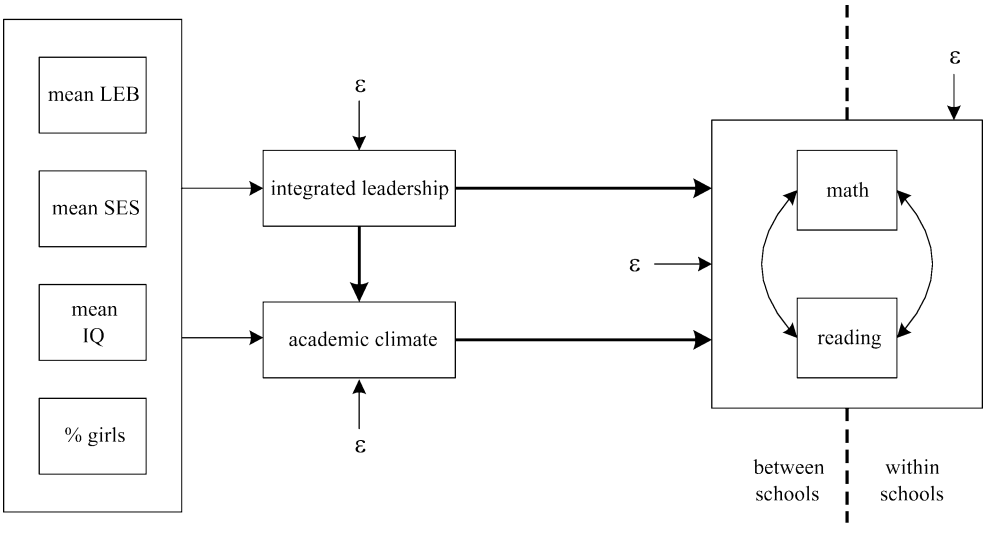

Figure 4. The antecedents model (partial)

factors on both integrated leadership and academic climate. In other words, in this model we assumed that integrated leadership and academic climate would also be themselves a function of the school context.

Method of Analysis

Classic multilevel regression analysis does not satisfy our needs to test these four conceptual models. Path analysis is a technique that cannot be estimated within this framework of classic multilevel analysis, as used in most school effectiveness techniques. SEM is another statistical approach that makes path analysis possible. A number of researchers have developed methods to apply SEM on multilevel data (e.g., Goldstein & McDonald, 1988; Muthén, 1989; Muthén & Satorra, 1989). Hox (2002) gives an overview of the different approaches to multilevel SEM. For this case study, we use the approach of Muthén (1994, 1997) through the use of the software package M*Plus* (Muthén & Muthén, 1998–2001).

The basic idea for this technique is the decomposition of the individual scores (e.g., pupils' scores) in an individual component (the deviation of the individual score from the groups' mean score) and a group component (the disaggregated group mean score) (Heck, 2001). This decomposition is used to calculate two independent covariance matrices: a between and a within matrix. To test a multilevel SEM model, both matrices were used. Different authors describe this technique more extensively (see Heck, 2001; Heck & Thomas, 2000; Hox, 1995, 2002; Muthén, 1994, 1997). In this case study, we used M*plus* to test the four conceptual models. The parameter estimates are the result of the Muthén's limited information maximum likelihood (MUML) estimation.

Results

The four models were estimated separately. First, we compare the model fit for the four models, then we describe in detail the model that best fits the data. As such, we are not interested in a generalising conclusion on the effect of integrated leadership on pupil achievement. Moreover, we are mainly interested in the comparison of the conclusions that can be drawn from these four estimated models. Therefore, this description of the best fitting model can be seen as giving a point of departure to compare the effects of integrated leadership in the other estimated models.

Model Fit for the Four Conceptual Models

Table 1 summarises the fit indices for the four conceptual models. The chi-square test shows that Model 4—the antecedents model[3]—fits the data best: the p value for the chi-square test is higher than .05, both the Tucker-Lewis Index (TLI) and the Comparative Fit Index (CFI) are higher than .95, the Root Mean Square Error of Approximation (RMSEA) is lower than .06 (see Hu & Bentler, 1999). The Standardized Root Mean Square Residual (SRMR) indicates that the model is well fitting at the lowest level, but that it does not fit that well at the school level (SRMR should be lower than .08, see Hu & Bentler, 1999). In comparison with the other models, this model fits best with the data, although the alternative models also fit relatively well. The chi-square test for the difference between both nested Models 3 and 4 shows that Model 4 fits the data significantly better (chi-square difference = 13.42 with 1 degree of freedom, $p < .005$).

The Best-Fitting Model

Figure 5 gives a visualisation of the best-fitting model and the estimated standardised parameters. Table 2 shows some of the relevant estimated parameters for this model

Table 1. Fit indices for the four conceptual models

	χ^2 test	RMSEA	CFI & TLI	SRMR
Model 1	value = 71.60 df = 30 p = .000	value = .030 $p < .05$ = 1.00	CFI = .94 TLI = .95	within = .17 between = .01
Model 2	value = 55.98 df = 31 p = .004	value = .023 $p < .05$ = 1.00	CFI = .97 TLI = .97	within = .13 between = .01
Model 3	value = 54.40 df = 29 p = .003	value = .024 $p < .05$ = 1.00	CFI = .97 TLI = .96	within = .14 between = .01
Model 4	value = 40.98 df = 28 p = .054	value = .018 $p < .05$ = 1.00	CFI = .98 TLI = .98	within = .12 between = .01

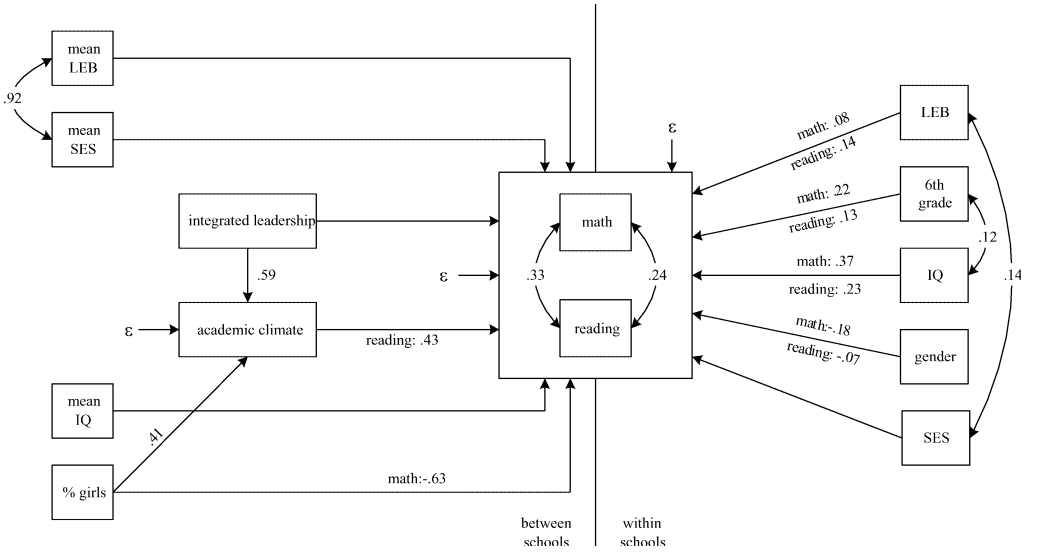

Figure 5. Significant standardised parameter estimates for Model 4

Table 2. Estimated parameters for Model 4, the antecedents model

Pupil level	Est.	S.E.	StdYX	Est.	S.E.	StdYX
		mathematics			reading	
6th grade	5.20	.54	.22	3.03	.57	.13
IQ	5.20	.33	.37	3.11	.34	.23
Gender	−5.98	.75	−.18	−2.16	.78	−.07
LEB	1.57	.47	.08	2.87	.50	.14
SES	−.26	.30	−.20	−.43	.32	−.03
R^2		.24			.10	
School level	Est.	S.E.	StdYX	Est.	S.E.	StdYX
		mathematics			reading	
Mean IQ	−.19	1.74	−.01	−1.63	1.51	−.17
% girls	−10.27	2.37	−.63	−3.12	2.06	−.26
Mean LEB	−3.19	5.23	−.26	2.27	4.48	.25
Mean SES	4.26	4.91	.04	1.37	4.20	.17
Academic climate	−.10	.79	−.02	1.39	.69	.43
Integrated leadership	−.66	.72	−.15	−.88	.62	−.27
R^2		.46			.29	
		academic climate				
% girls	1.50	.38	.41			
Integrated leadership	.59	.10	.59			
R^2		.52				

and the indices for model fit and explained variances. At the pupil level, the background variables explain 24% of the variance in mathematics achievement and 10% of the variance in reading proficiency. At the school level, the R^2s show

that this model has more predictive power for the differences between schools. For mathematics achievement, this model explains 46% of the variance between schools. Note that this percentage of explained variance is only the result of the effect of the percentage of girls on mathematics achievement. The between-schools variance for reading proficiency is lower: 29%.

The model shows that both output measures correlate at pupil level and at school level. In addition, there is a strong correlation between the mean SES score and the mean LEB score of a school. At the pupil level, there seems to be a positive correlation between the SES and the LEB on the one hand, and between the sixth grade and IQ on the other hand. Except for SES, all background variables have an influence on the pupils' achievement.

This antecedents model shows a positive effect from the percentage of girls on academic climate. In other words, in a school with relatively more girls the climate focuses more on pupils' achievement. The other context factors have no influence on the school climate or on integrated leadership. In addition, this model reveals a negative effect of the percentage of girls in a school on mathematics achievement. Put differently, in a school with relatively more girls, pupils score on average lower on mathematics than pupils in a school with relatively less girls. This is the only context factor that has a significant effect on pupil achievement.

According to this model, integrated leadership does not have a direct effect on pupil achievement. Given their standard errors, the estimates for the effect of integrated leadership on mathematics and reading do not differ significantly from zero. Indirectly, integrated leadership does have an effect on pupils' reading proficiency, as does the percentage of girls. Both variables have a positive effect on the academic climate, which in its turn has a positive influence on pupils' reading proficiency. For mathematics achievement no indirect effect of integrated leadership is found.

A Comparison of the Findings in the Four Models Concerning the Effect of Integrated Leadership

The central question in this article was whether the findings are dependent on the chosen conceptual model. Table 3 gives an overview of the estimated effects in the four conceptual models.

This table shows the influence of the chosen model on the possible conclusions. If we had only tested direct effects, we would have concluded that integrated leadership has no effect on pupil achievement but that academic climate has. From Model 2, we would have concluded that integrated leadership has only a positive effect on academic climate but that both school characteristics do not influence pupil achievement. If we had chosen to test only Model 3, we would have concluded that integrated leadership only has an indirect effect independent of the school context and that it has no direct effect on pupil achievement. The antecedents model adds that the percentage of girls has an autonomous influence on the academic climate. This clearly shows the importance of the choice for a valid conceptual model to test the effect of integrated leadership (and other school characteristics) on pupil achievement.

Table 3. Overview of the effects from integrated leadership and academic climate as estimated in Model 1, Model 2, Model 3, and Model 4

	Academic climate	Mathematics	Reading
Model 1			
Integrated leadership	−,−	0	0
Academic climate	−,−	0	+
Model 2			
Integrated leadership	+	0	0
Academic climate	−,−	−,−	−,−
Model 3			
Integrated leadership	+	0	0
Academic climate	−,−	0	+
Model 4			
Integrated leadership	+	0	0
Academic climate	−,−	0	+

−,− = not estimated.
0 = no effect.
+ = positive significant effect.

Conclusion and Discussion

In this article, we examined whether integrated leadership has an effect on pupil achievement. Research on the effects of educational leadership makes use of different conceptual models for the relationship between educational leadership and pupil achievement. Most school effectiveness research implicitly assumes a direct effect model that hypothesises that educational leadership directly influences pupils' achievement. However, recent research studies and re-analyses of existing data convincingly demonstrated the importance of an indirect effect model. Moreover, in research on the effect of educational leadership not only indirect effects but also antecedent effects should be included if the data and the technique of analysis allow it.

Using data gathered in Flemish school effectiveness research in technical secondary education, we examined whether and how integrated leadership influences two measures of pupil achievement: functional mathematics achievement and reading proficiency. This relationship was tested by means of four conceptual models: (1) the direct effect model, (2) the indirect effect model, (3) the direct and indirect effect model, and (4) the antecedents model. One variable that is often used as an intermediate between educational leadership and pupil achievement is school climate. In this case study, we used ''academic climate'', which means that schools scoring high on this characteristic put a strong emphasis on their pupils' achievement. As possible antecedent context factors affecting educational leadership and academic climate, we used the school composition concerning gender, IQ, the linguistic ethnic, and the socioeconomic background.

The analyses show that the antecedent model fits best with the data. This model is based on the assumption that integrated leadership has a direct, as well as an indirect, effect on pupil achievement. Furthermore, it was hypothesised that integrated leadership and academic climate are both influenced by the context characteristics of a school. From the analysis we learn that integrated leadership has only an indirect effect, through academic climate, on pupils' reading proficiency. The only context characteristic that influences the other school characteristics is the percentage of girls in the pupil body.

Most importantly, this case study has shown the impact of the choice of a conceptual model to test effects of educational leadership on pupil achievement. Depending on the model, we come to different conclusions concerning the effect of educational leadership. If we use the direct effect model, we would have concluded that integrated leadership has no effect on pupil achievement but that academic climate has. From the indirect effect model, we would have concluded that neither integrated leadership nor academic climate affect pupil achievement. The direct and indirect effect model would make us conclude that integrated leadership only has an indirect effect independent of the school context and that it has no direct effect on pupil achievement. The antecedents model adds that the percentage of girls has an autonomous influence on the academic climate.

The finding that we would come to different conclusions, depending on the chosen conceptual models, demonstrates the importance of this choice. This conclusion, which is the primary focus of this case study, supports the weight that other researchers recently have given to this choice (Hallinger & Heck, 1998; Witziers et al., 2003). It is understood that such a complex antecedents model, which assumes direct and indirect effects and an influence of the school context, is closer to the complex reality of education. This conclusion can be extended to school effectiveness research in general, with a focus on several school and class characteristics as predictors of pupil achievement. We want to argue that future school effectiveness research should give more attention to the conceptual model for testing the expected effects. Moreover, researchers in this domain should apply these more complex antecedents models, given the fact that the results of these models have a higher internal and external validity than the direct effect models.

With this recommendation for future school effectiveness research, we also want to emphasise the key role of a well-founded theory. These kinds of models assume a different approach. Most classic school effectiveness studies have an exploratory character. For a number of school characteristics that are often found to have an effect on pupil achievement, researchers explore in a data-driven way whether these characteristics also have an effect on their study. In these analyses, researchers no longer pay attention to how these school characteristics can influence pupil achievement. On the other hand, modelling more complex models like the antecedents direct and indirect effect model should go hand in hand with a theory on how these different factors relate and how they can improve pupil achievement. Based upon a theory, a (complex) model is built for which the researcher checks whether the theory can be or has to be rejected, given the data. This confirmative

approach demands a mental shift from researchers. Although we did not have the ambition to present a confirmative and theory-driven article as well, we want to argue, based upon our findings, that future school effectiveness research should be more confirmative rather than explorative. There are already a considerable number of explorative studies in different contexts and with several output measures. Theory building on how particular school characteristics have an influence on pupil achievement and confirmative research on new aspects of these resulting theories can be more informative for this field of research. Moreover, this type of research should focus on testing hypotheses deduced from, for example, theories in sociology, organisational theory, pedagogical sciences, learning psychology, and instructional strategy. Different authors recognised this lack of theory within school effectiveness research and made a plea for theory building in this field (Creemers, Scheerens, & Reynolds, 2000; Scheerens, 1997; Scheerens & Bosker, 1997). In their article, Luyten, Visscher, and Witziers (2005) give an extensive review of critics concerning the assumed theoretical limitations of school effectiveness research. Two relatively recent examples of research that test organisational and pedagogical theories are Reezigt, Guldemond, and Creemers (1999) and Imants (2002).

In our case study, as in examples of similar studies (Hallinger & Heck, 1996, 1998; Witziers et al., 2003), we used school climate as a possible intermediate variable. Until now, we have not gained insight into how educational leadership can influence pupil achievement through the primary instructional process that takes place in the classroom. The question to be answered is: How does educational leadership influence class practices which in their turn influence learning processes? This link between processes at school level, class level, and pupil level should receive close attention in future school effectiveness research.

Finally, we want to emphasise that this case study is meant as an example and that the findings cannot be generalised. In preparatory analyses, we found that the antecedent model did not hold for secondary Flemish vocational education, for example. It was not our prior intention to test a universal theory of the influence of educational leadership but to demonstrate the importance of the choice of a valid conceptual model. Many other researches emphasised the importance of a research context (e.g., country, educational level) on the results (Teddlie, Stringfield, & Reynolds, 2000; Witziers et al., 2003). Therefore, it is important to take this context into account in theory building and in choosing a valid conceptual model.

Notes

1. For a more extensive description of conceptual models about the relationship between goals, leadership, and achievement, we refer to Hallinger and Heck (2002).
2. In Belgium, there are three official languages (Dutch, French, and German). Therefore, the variable "language spoken at home" is not as such a good proxy-variable for the fact that a pupil comes from an ethnic minority group.

3. This is the model fit for the most parsimonious antecedents model. Testing the model that included all possible relationships between context factors and integrated leadership and academic climate, revealed that for only one of these relationships the parameter estimate differed significantly from zero. Given the fact that we are in fact not really interested in the best fitting model but in the comparison between estimates within different models, we describe this more parsimonious model as the best-fitting model. This is done because most researchers would take the same step and use this data-driven parsimonious model as the best-fitting model.

References

Bosker, R. J., De Vos, H., & Witziers, B. (2000). *Theories and models of educational effectiveness.* Enschede, The Netherlands: Twente University Press.

Bossert, S. T., Dwyer, D. C., Rowan, R., & Lee, G. V. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, *18*(3), 34–64.

Brookover, W. B., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School social systems and student achievement. Schools can make a difference.* New York: Praeger.

Creemers, B., Scheerens, J., & Reynolds, D. (2000). Theory development in school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 283–298). London/New York: Falmer Press.

De Maeyer, S., & Rymenans, R. (2004). *Onderzoek naar kenmerken van effectieve scholen. Kritische factoren in een onderzoek naar schooleffectiviteit in het technisch en beroepssecundair onderwijs in Vlaanderen* [Research on school effectiveness-enhancing factors. Critical factors in a research on school-effectiveness for technical and vocational secondary education in Flanders]. Gent, Belgium: Academia Press.

De Maeyer, S., Rymenans, R., Daems, Fr., Van Petegem, P., & Van den Bergh, H. (2003). *Effectiviteit van tso- en bso-scholen in Vlaanderen. Een onderzoek naar de effecten van schoolkenmerken op de prestaties en het welbevinden op school van tso- en bso-leerlingen* [Effectiveness of technical and vocational secondary education in Flanders. A research on the effects of school characteristics on the performance and well-being of pupils in technical and vocational secondary education]. Leuven, Belgium/Leusden, The Netherlands: Acco.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership*, *37*(1), 15–24.

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester, UK: Wiley.

Goldstein, H., & McDonald, R. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*, 455–467.

Hallinger, P. (1983). *Principal instructional management rating scale.* Palo Alto, CA: Stanford University.

Hallinger, P. (2003). Leading educational change. Reflections on the practice of instructional and educational leadership. *Cambridge Journal of Education*, *33*(3), 329–351.

Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership and student reading achievement. *Elementary School Journal*, *96*(5), 527–549.

Hallinger, P., & Heck, R. H. (1996). The principal's role in school effectiveness. An assessment of methodological progress, 1980–1995. In K. Leithwood, J. Chapman, D. Corson, P. Hallinger, & H. Hart (Eds.), *International handbook of educational leadership and administration* (pp. 723–783). Dordrecht, The Netherlands/Boston/London: Kluwer Academic Publishers.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, *9*, 157–191.

Hallinger, P., & Heck, R. (2002). What do you call people with visions? The role of vision, mission, and goals in school leadership and improvement. In K. Leithwood & P. Hallinger (Eds.), *Second international handbook of educational leadership and administration* (pp. 9–40). Dordrecht, The Netherlands/Boston/London: Kluwer Academic Publishers.

Heck, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 89 – 127). Mahwah, NJ/London: Lawrence Erlbaum Associates Publishers.

Heck, R. H., Marcoulides, G. A., & Lang, P. (1991). Principal instructional leadership and school achievement. The application of discriminant techniques. *School Effectiveness and School Improvement*, *2*, 115 – 135.

Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques.* Mahwah, NJ/ London: Lawrence Erlbaum Associates Publishers.

Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung (Handanweisung)* [Testing systems for school and developmental growth]. Göttingen, Germany: Verlag für Psychologie Hogrefe.

Hox, J. (1995). *Applied multilevel analysis* (2nd ed.). Amsterdam: TT-Publikaties.

Hox, J. (2002). *Multilevel analysis. Techniques and applications.* Mahwah, NJ/London: Lawrence Earlbaum Associates Publishers.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1 – 55.

Imants, J. (2002). The counterproductive effects of a national reform initiative. Reflections from organizational theory. *School Effectiveness and School Improvement*, *13*, 31 – 61.

Krüger, M. L., Witziers, B., Sleegers, P. J. C., & Imants, J. (1998). Onderwijskundig leiderschap in moderne onderwijsinstellingen [Educational leadership in modern educational institutes]. In B. P. M. Creemers, M. L. Krüger, P. J. C. Sleegers, & C. A. Van Vilsteren (Eds.), *Handboek schoolorganisatie en onderwijsmanagement. Leiding geven in bestel, school en klas* [Handbook of school organisation and educational management. Managing the establishment, school and class] (C4120, pp. 1 – 26). Alphen aan den Rijn, The Netherlands: Kluwer.

Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly*, *30*(4), 498 – 518.

Luyten, H., Visscher, A., & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, *16*, 249 – 279.

Marks, H. M., & Printy, S. (2003). Principal leadership and school performance. An integration of transformational and instructional leadership. *Educational Administration Quarterly*, *39*(3), 370 – 397.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557 – 585.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, *22*(3), 376 – 398.

Muthén, B. O. (1997). Latent growth modeling with longitudinal and multilevel data. In A. E. Raftery (Ed.), *Sociological methodology* (pp. 453 – 480). Boston: Blackwell.

Muthén, L. K., & Muthén, B. O. (1998 – 2001). *MPlus user's guide* (2nd ed.). Los Angeles: Muthén & Muthén.

Muthén, B. O., & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R. D. Bock (Ed.), *Multilevel analysis of educational data.* San Diego, CA: Academic Press.

Ogawa, R., & Bossert, S. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, *31*, 224 – 243.

Pitner, N. (1988). The study of administrator effects and effectiveness. In N. Boyan (Ed.), *Handbook of research in educational administration* (pp. 99 – 122). New York: Longman.

Reezigt, G. J., Guldemond, H., & Creemers, B. P. M. (1999). Empirical validity for a comprehensive model on educational effectiveness. *School Effectiveness and School Improvement*, *10*, 193 – 216.

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours. Secondary schools and their effects on children.* Cambridge, MA: Harvard University Press.

Scheerens, J. (1997). Conceptual models and theory-embedded principles on effective schooling. *School Effectiveness and School Improvement*, *8*, 269 – 310.

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

Teddlie, C., Stringfield, S., & Reynolds, D. (2000). Context issues with school effectiveness research. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 160–185). London/New York: Falmer Press.

Van de Grift, W. (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement*, *1*, 26–40.

Van de Grift, W. (1992).

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement. The elusive search for an association. *Educational Administration Quarterly*, *39*(3), 398–425.