Educational Administration Quarterly
Vol. 39, No. 3 (August 2003) 398-425

Educational Leadership and Student Achievement: The Elusive Search for an Association

Bob Witziers
Roel J. Bosker
Meta L. Krüger

This study revisits the existing scholarly debate on the possible impact of the principal's leadership on student achievement. Both "direct effect" and "indirect effect" models are discussed. A quantitative meta-analysis examines to what extent principals directly affect student outcomes. The small positive effects found in this meta-analysis confirm earlier research findings on the limitations of the direct effects approach to linking leadership with student achievement. Finally, lines of future research inquiry are discussed.

Keywords: *meta-analysis; educational leadership effects; school leadership and student achievement*

In the last 20 years, much attention has been paid to educational leadership and its impact on student outcomes. However, generally researchers concur that the effects are indirect if not difficult to measure (Hallinger & Heck, 1996, 1998; Leithwood & Jantzi, 2000). To be sure, one can find literature defending the position that principals matter. From certain early research into school effectiveness (Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979; Rutter, Maugham, Mortimore, Ouston, & Smith, 1979) and an early review of school leadership studies (Leithwood & Montgomery, 1982), the effective principal comes to the fore as an instructional or educational leader who affects school climate and student achievement. Other more recent examples of the statement that principals matter can be found in reviews on school effectiveness research conducted by Levine and Lezotte (1990) as well as Sammons, Hillman, and Mortimore (1995). Also one can point to other articles such as the one about the future direction of preparation programs for

DOI: 10.1177/0013161X03253411

school leaders in which Bredeson (1996, p. 255) notes that "there is ample evidence in the literature that effective leadership can and does positively affect school and student outcomes."

On the other hand, there are authors that doubt whether educational leadership effects exist and even if they do, whether these are important. Murphy (1988), for example, concludes in his review that the existing knowledge base fails to offer proof that educational leadership matters. He argues that (a) not much research is conducted in this area and that (b) most studies in this field are of poor quality. Since the publication of Murphy's article, the number of quantitative studies has increased drastically. Nevertheless, there are still doubts about the presumed positive effects of educational leadership. For instance, Hallinger and Heck (1996, p. 1) conclude that "despite the traditional rhetoric concerning principal effects, the actual results of empirical studies in the U.S. and U.K. are not altogether consistent in size or direction." Moreover, most of the Dutch studies into educational leadership failed to come up with a positive and significant relationship with student achievement (Bosker & Witziers, 1996; Scheerens & Bosker, 1997; Van de Grift, 1990; Van de Grift & Houtveen, 1999). Given the divergence in these results, the question of whether school principals matter remains unresolved. Another unresolved question is how the effects on student outcomes might be mediated.

This article aims to contribute to the existing scholarly debate on the question of whether school leadership matters. The study reported here used a quantitative meta-analysis to estimate the effect size of educational leadership on student achievement among multinational research reports. The study also attended to which factors (or moderators) might account for the variation in effect sizes. A well-known fact is that effect sizes of studies involved in a meta-analysis vary due to differences in procedures, instrumentation, study contexts, and treatments (Raudenbush, 1994). Therefore, paying attention to such factors and to the impact they have on effect sizes gives an even clearer insight into the potential impact of leadership because it might clarify the conditions under which leadership is effective.

This particular approach sets this meta-analysis apart from other syntheses of research into educational leadership (e.g., Hallinger & Heck, 1996, 1998; Pitner, 1988). However valuable these syntheses are in providing an answer to the question of whether educational leadership matters, they do not give insight into the specific issues addressed here.

We must note that the studies selected for our meta-analysis are all based on the direct effect models, because very few studies employing the indirect effect model have been conducted to date (Hallinger & Heck, 1998).

Although some researchers advocate abandoning direct effect approaches, studies based on the indirect effect model have not yielded unequivocal evidence supporting a relationship between leadership and student achievement (Hallinger & Heck, 1998). This leads us to examine once again, through a revised meta-analysis, the question To what extent does educational leadership *directly* affect student achievement?

Following the presentations of the meta-analysis results, a discussion compares this study with the results of the few studies using the indirect effect models. Finally, we conclude with suggestions with regard to the way educational leadership could be conceptualized and investigated in future research.

BACKGROUND

Educational researchers and practitioners hold different views regarding ways that school principals improve educational outcomes. Whereas some researchers found that school principals matter to student achievement, others found no effects of leadership on student outcomes. Which factors cause such contradictory results? Paradoxically, researchers using direct effect models produced different results from those using indirect models. Another reason might be that over the years, educational leadership has been conceptualized and operationalized in many different ways, thereby making the results hardly complementary and difficult to compare. Before discussing these factors, it must be noted that the relevance of our questions is not merely academic. Internationally, school principals increasingly are held accountable for educational quality in the belief that students' success or failure is determined by the way a school is run (Fullan & Watson, 2000; Leithwood & Menzies, 1998; Wildy & Louden, 2000). In the Netherlands, policy reforms aimed at deregulation and decentralization have gone hand-in-hand with efforts aimed at restructuring schools in such a way that principals are better able to manage the school's educational structure and by providing them with the necessary training programs. These efforts are guided by a belief among policy makers in school principals' capacity to improve student outcomes (Imants, 1996; Krüger, 1995). The burgeoning accountability policies for education represent an international interest in answering the question of the degree to which the expectation that school leaders influence student outcomes is a valid expectation. In contrast to this simple and hopeful view, the literature reveals much complexity.

Principal's Impact: Direct and Indirect Effect Models

Research into school effectiveness is considered the starting point for examining educational leadership and its impact on student outcomes (Brookover et al., 1979; Edmonds, 1979; Rutter et al., 1979). The results of this research suggest that educational leadership is an important characteristic of effective schools. According to these findings, principals should have high expectations of teachers and student achievement, supervise teachers, coordinate the curriculum, emphasize basic skills, and monitor student progress. In school effectiveness studies of the seventies and eighties, researchers were mostly looking for direct effects of instructional leadership on student outcomes.

Bossert, Dwyer, Rowan, and Lee (1982) severely criticized this approach. They developed an alternative model in which the characteristics of leadership were not the central focus. Instead they suggested studying how instructional leadership is strategically shaped. In their model, instructional leadership is broadly interpreted as the principal's orientation toward primary processes in school. The principal is highlighted as acting intentionally and from an overall perspective, taking the school context into account. The principal's routine behaviors create links between characteristics of school organization and instructional climate, which in their turn affect student achievement.

This perspective on indirect effects also occurs in more recent and more complex models for research into principal leadership. Leadership is no longer proposed as having a direct influence on learning outcomes but as having an indirect influence through the way it has an impact on school organization and school culture.

From such a perspective, Hallinger and Heck (1998) examined the empirical literature on principal effects that emerged during a period between 1980 and 1995. In the 40 studies they reviewed, they identified different models used to investigate the relationship between school leadership and student achievement. First, the direct effect model, which suggests that leaders' practices can have effects on school outcomes and that these can be measured apart from other related variables. Some researchers include antecedent factors, but these are not hypothesized as variables interacting with leadership or mediating its effects on the selected outcomes. Second, the mediated effect model, which hypothesizes that leaders achieve their effect on school outcomes through indirect paths: The leader's contribution is mediated by other people, events, and organizational and cultural factors. Finally, the reciprocal effect model, in which it is suggested that relationships between the principal

and features of the school and its environment are interactive. This model implies that school leaders adapt to the organization in which they work, changing their thinking and behavior over time. Researchers testing this model assert that causal relationships may be multidirectional, change over time, and even be nonlinear. In most of the reviewed studies, direct effects models were employed. More recently, researchers have applied mediated effect models. According to Hallinger and Heck (1998), studies in which indirect effect models are used show a greater impact of school leadership on student performance than do studies employing direct models. Although Hallinger and Heck (1998) state that their review "revealed several paths that begin to describe the means by which principal leadership influences learning outcomes" (p. 187), they also emphasize that

> even as a group the studies do not resolve the most important and practical issues entailed in understanding the principal's role in contributing to school effectiveness. These concern the means by which principals achieve an impact on school outcomes as well as the interplay with contextual forces that influence the exercise of school leadership. (p. 186)

Different Conceptualizations of Educational Leadership

Another reason for the contradicting results in the field of educational leadership is that educational leadership has been conceptualized and operationalized in many different ways. Pounder, Ogawa, and Adams (1995) stated that research in this field has led neither to an integrated concept of school leadership nor to a better understanding of the impact of leadership on schools' performance. The lack of conceptual congruence provides an empirical caution to the study of school leadership regardless of statistical models or other methodology.

What conclusions can be drawn from research when scholars do not agree on the meaning of the concept of educational leadership? For a long time, it has been a source of debate whether administrative management can be distinguished from educational leadership (Krüger, Witziers, Sleegers, & Imants, 1999). In the effectiveness studies of the eighties, educational leadership, mostly called instructional leadership, was seen as strictly separate from administrative management. It was usually defined as the performance of relatively simple tasks related to education. Some scholars still make a sharp distinction between educational and administrative leadership, whereas others criticize this distinction, claiming that a school policy that aims to achieve educational objectives demands an organizational structure that integrates administrative management and educational leadership

(Hughes, 1985; Krüger, 1995; Sleegers, 1991). In the view of these critics, a distinction can be made between management and leadership, but they cannot be separated. In this sense, educational leadership is seen as developing strategies so that a variety of management instruments can be used to achieve a school's most important primary task: the desired student results (Krüger, 1995). An educational leader then is someone whose actions (both in relation to administrative and educational tasks) are intentionally geared to influencing the school's primary processes and, therefore, ultimately students' achievement levels. Hallinger's (1989, 1994) work can primarily be considered an illustration of refining the concept of educational leadership. Based on Edmonds (1979) and Bossert et al. (1982), Hallinger and his colleagues constructed an instrument to measure principals' instructional management, the Principal Instructional Management Rating Scale (PIMRS) (Hallinger, 1989, 1994). They identified three dimensions of educational leadership: "defining the school mission," "managing the instructional program," and "promoting a positive school learning climate." From these dimensions they deduced 10 instructional leadership functions, such as "frames goals," "coordinates curriculum," "monitors progress," and "sets expectations." These functions are in turn translated into behaviors, which comprise the items in their questionnaire. Their concept of instructional management refers to the principal's mental and physical actions, which drive others to perform tasks in such a way that the school goals in terms of student performance are achieved.

Another example of refining the concept can be found in the work of Hoy and Miskel (1991). In their framework for studying leadership, they include leadership functions and traits, leader behavior, and situational characteristics, including leader role as well as organizational and personal effectiveness.

Finally, Leithwood's concept of transformational leadership can be seen as an elaboration of the concept of educational leadership (Leithwood, 1992; Leithwood & Jantzi, 1990). In this school principals represent change agents. This role arises from the notion that principals should not only perform tasks related to coordination and evaluation of the educational system but also in relation to further developing the educational system via transformation of the school culture. The underlying assumption is that school cultures, based on norms of autonomy and isolation, characterize many schools, which in turn block educational reforms. Instead, the importance of shared school cultures is stressed. Collegiality, "empowered" teachers, collaborative planning, and continuous improvement efforts characterize such cultures. One of the main tasks of school principals is to help create a working environment in which teachers collaborate and identify with the school's mission and goals.

Authors who emphasize this concept have widely differing views on how leaders should improve educational outcomes from that which is implied by the more traditional concept of instructional leadership. Nevertheless, they do not question the ability of principals to improve educational outcomes per se (Leithwood, Tomlinson, & Genge, 1996; Silins, 1994).

In summary, it appears that research of the last two decades has raised even more questions about the relation between educational leadership and student outcomes, rather than clarifying the issues involved. Research still does not give conclusive answers to the question Bossert et al. (1982) already posed in the early eighties: How do school principals steer the primary processes in their schools; and how can they become effective?

METHOD

Our quantitative meta-analysis focuses on studies into the direct effects of educational leadership on student achievement conducted between 1986 and 1996. The year 1986 was chosen due to concomitant developments in methodology and statistics, such as multilevel modeling for examining variables across level relationships (such as school leader behavior and student achievement). The current study provides an international perspective on the status of the direct effects model between 1986 and 1996.

Selection of Studies

A systematic search of documentary databases containing abstracts of empirical studies was conducted. Of particular importance were Educational Resources Information Center (ERIC) documents and database, *School Organization and Management Abstracts*, *Educational Administration Abstracts*, and the *Sociology of Education Abstracts*. Although these abstracts cover the most important scholarly journals, they do not cover all. Therefore, we paged through volumes of relevant educational peer-reviewed journals not covered by these (e.g., *Journal of School Effectiveness and School Improvement*, *School Leadership and Management*, *Journal of Educational Administration*, etc.). Moreover, reviews and handbooks were examined for references to empirical studies. Finally, all selected studies were examined for references to studies as yet not uncovered.

The next step consisted of selecting studies from those collected in the first stage. Two criteria for including studies were used. First, we only selected studies that had been expressly designed to examine educational

leadership, that is, research in which there was a clear conceptualization as well as reliable and valid measurement of educational leadership. Second, studies had to include explicit and valid measures of student achievement, thus excluding those that focused on other outcomes of student achievement. In total, 37 studies were selected (see Appendix A). In addition, data collected in the study of the International Association for the Evaluation of Educational Achievement (IEA) on reading literacy in 25 countries (Postlethwaite & Ross, 1993) could be used for meta-analytical purposes. These data were reanalyzed using multilevel statistical techniques, and the results on educational leadership found in this reanalysis were included in our study.

The Meta-Analysis: Different Studies, Different Analyses

Three different kinds of meta-analyses were conducted on the same research studies. First, the results of all studies were analyzed simultaneously. If a study, for example, provided empirical evidence for the relationship between three subdimensions of educational leadership and student achievement, the average effect size was used. A second meta-analysis was done on a subsample of all studies. This meta-analysis contains those studies that only used one measure for educational leadership. The implicit assumption underlying such studies is that educational leadership is a one-dimensional concept. Our last meta-analysis as a matter of fact consists of series of small meta-analyses, one for each subdimension of educational leadership. For this analysis, the PIMRS instrument developed by Hallinger (1989, 1994) was used as a framework to categorize the principal behaviors. Although this instrument does not give an exhaustive list of school leader behaviors, it was used in our study as a heuristic tool to categorize specific leadership behaviors. Thus, an important criterion for categorizing leadership behaviors was whether these were conceptualized and operationalized in a way that fitted Hallinger's framework and operationalization. If not, studies (or specific leadership behaviors) were discarded. Specifically, we estimated effect sizes regarding the following leadership behaviors:

1. defining and communicating mission
2. supervising and evaluating the curriculum
3. monitoring student progress
4. coordinating and managing curriculum
5. visibility
6. promoting school improvement and professional development
7. achievement orientation

The literature review showed that many researchers assessed school leader behaviors such as giving support or advice to teachers, discussing educational matters with them, and so forth. In our view, however, these behaviors could not be placed within Hallinger's (1989, 1994) framework. Thus, in addition to the seven aforementioned categories, we included one more category in our analysis, labeled "advice and support."

Within most categories (except visibility) there was no equivalency of instrumentation. Consistency in the way concepts are operationalized is not the strongest feature of leadership research, although sometimes researchers have good reasons for operationalizing concepts differently. For instance, on one hand, many Dutch researchers rely on instruments developed within the American context but, on the other, adapt these to improve the ecological validity of the instrument. In this respect, we believe that the categorized studies, although (sometimes slightly) different in instrumentation, do have something in common in the sense that they address the same phenomenon.

Statistical Procedure

The multilevel model suggested by Raudenbush and Bryk (1985) was applied for our statistical analyses. In this, the selected studies are considered to be a sample from the population of studies into the relationship between school leadership and student outcomes. Nested under each study are the secondary units: the schools. Each study can then be viewed as an independent replication. This concept could be used but does not solve the problem of multiple results from one study, such as when effects are reported for mathematics and language achievement in one study while using the same sample of schools and students. To deal with this problem, the two-level model for meta-analysis was generalized to a three-level one, in which the highest level of the studies is referred to as the across-replication level and the multiple results within a study as the within-replication level. The main advantage of the statistical meta-analysis employed here is that the information from each study is weighted by the reliability of the information, in this case the sample size. A more in-depth treatment of the statistical modeling technique is given in Appendix C of this article.

As we have already noted, one of the goals of our analysis was to decide which factors (or moderators) are responsible for the variation in effect sizes. Usually this means that differences in reported effect sizes are modeled as a function of study characteristics. One of the relevant characteristics deals with the question of whether studies have used either a language or a mathematics test score, or a composite score, to assess student achievement. Apart from assessing the impact of the type of test employed, we also assessed the

effects of study characteristics in relation to the country in which the study was conducted (the United States, the Netherlands, and other countries), the sector in which the study took place (primary or secondary education), the study design employed (whether or not the study adjusts for covariates at student level—this will be referred to as "value added"), the statistical modeling technique used to assess the relationship between leadership and student achievement (multilevel or not) and the data source employed by the study (whether teacher or principal data were used).

It is important to note that the use of the statistical model in which the effects of study characteristics are estimated is problematic when the sample size is small. This means that in some cases only the bivariate relationships between moderators and effect sizes are presented. In particular, this is the case when analyzing the relationship between moderators and specific leadership behaviors. The results of these analyses were only used to check the robustness of the findings focusing on the bivariate relationships. Finally, the analyses were conducted twice. The first one included all studies; in the second one, the so-called sensitivity analysis, the outliers were removed from the samples to check the robustness of the findings.

The Computation of Effect Sizes

To indicate the effect of educational leadership, Fisher's Z transformation of the correlation coefficient was used. Not all studies presented their results in terms of correlations, so all other effect size measures were transformed into correlations using formulae presented by Rosenthal (1994). For small values of the correlation coefficient, Z_r and r do not differ much, but it should be remembered that all tables that follow refer to Z_r. The most well-known effect size coefficient, Cohen's d, is approximately twice the size of the correlation coefficient (when the latter is small, say $r < .35$). The meta-analysis was conducted using MLn (Rasbash & Woodhouse, 1995), following and generalizing a procedure suggested by Lamberts and Abrams (1995).

RESULTS OF THE META-ANALYSIS

The Overall Impact of Leadership on Student Achievement

The next section presents the results of the analyses. We begin with the results of the first analysis based on an overall perspective on leadership. Table 1 shows the estimated effect size (using equations (1) through (3) of

TABLE 1
Estimated Effect Size for School Leadership and Standard Deviation Across and Within Replications (Total Sample $N_a = 61$, $N_w = 377$, Without IEA $N_a = 37$, $N_w = 281$)

	Total Sample		*Without IEA*	
	Total	p *Value*	*Estimate*	p *Value*
Mean effect size	.02	.07	.04	.02
Standard deviation across replications	.06	.00	.11	.00
Standard deviation within replications	.08	.00	.00	.50

NOTE: IEA = International Association for the Evaluation of Educational Achievement; N_a = sample size only across replications; N_w = sample size within and across replications.

Appendix C) of the variable leadership and the standard deviation across studies with and without the results from the IEA reading study. This latter distinction is made because the results of the IEA study form a substantial part of our data set because each country is considered a unique study into the association between educational leadership and achievement (the IEA Reading Literacy study thus delivers 25 cases in the analysis). This fact might genuinely affect our results, particularly when it is taken into consideration that compared to other studies, the operationalization of independent variables in general and the variable educational leadership in particular has certain flaws in this study due to the condition that each item should have meaning for each of the 25 countries. Hence, it seemed wise to perform two analyses, one with and one without IEA data, to gain a clearer view of the overall impact of educational leadership on student outcomes.

The results suggest that school leadership does have a positive and significant effect on student achievement. However, the effect sizes in terms of Z_r are very small ($Z_r = .02$ for the total sample, $Z_r = .04$ for the sample without the 25 IEA cases).

Another important question is whether effect sizes vary across and within studies. Table 1 gives information on whether the standard deviation in effect sizes across and within replications differs significantly from zero. Low p values indicate that the amount of variation within and between replications differs significantly from zero, whereas the reverse is true of course for high p values. Because the p values in Table 1 are low, the conclusion is that there is a large variation in effect sizes within and across studies, but if we then disregard the 25 IEA cases all variation appears to be situated at the across-studies level. If we add the two standard deviations, then square the result (this then is the variance) and then take the square root, the result is the standard deviation of the effect size estimate. For the sample where the IEA studies are excluded

this results in .11, indicating that the effect sizes expressed as Z_r vary from −.18 to .26. All the more reason to look into the possible causes of this variability. The next question is Which moderators can be held responsible for this variation? Table 2 shows the results of the analyses in which it is tried to predict differences between effect sizes with such study characteristics as subject matter, sector, country, design, sector, statistical technique, and respondent.

Our results indicate that discarding the IEA data, the effect of leadership on (uncorrected) student outcomes is .11 for studies in primary education using composite outcome variables, conducted in the United States, without using covariates, based on respondents other than the school leader himself or herself, and using multilevel modeling techniques. Taking the IEA data into account, the results show that the effect size of educational leadership is about .07.

Looking at country differences, it appears that there are large discrepancies between several educational contexts. That is, the aforementioned results are only true for studies conducted outside the Netherlands. The results, both for the total sample and the sample without IEA data, show that in the Netherlands the effect size is about zero. The only other significant effect pertains to sector differences. On average, school leadership effects are absent in secondary education, whereas they are related to student achievement in primary schools. Moreover, our results indicate that the predictors subject, value added, respondent, and statistical technique hardly explain any of the variation between studies. Finally, the results of a sensitivity analysis showed that our findings were robust. The same results occurred when the outliers were removed from the sample.

Results From Further Analyses on the Impact of Leadership on Student Achievement

The main results of the second and third analyses are shown in Table 3. The second analysis dealt with studies using a single instrument to measure educational leadership. The third analysis included studies using a multidimensional concept of leadership. This analysis tried to assess the impact of several leadership behaviors on student outcomes. As such, Table 3 contains the results concerning the estimated mean effect sizes.

The results of the first analysis show that studies using a single instrument, implying that educational leadership is a one-dimensional concept, fail to come up with positive and significant relationships between this concept and measures related to student outcomes: The effect size is close to zero (.01).

TABLE 2
Predicting Differences in Educational Leadership Effect Sizes

	Total Sample		*Without IEA*	
	Estimate	p *Value*	*Estimate*	p *Value*
Intercept	.07	.21	.11	.07
Secondary vs. primary	−.08	.03	−.07	.12
Math vs. composite	.00	.96	.02	.53
Language vs. composite	.00	.98	.03	.45
The Netherlands vs. United States	−.08	.05	−.14	.06
Other countries vs. United States	.00	.97	.00	.99
Value added vs. uncorrected	.01	.61	.00	.86
School leader is respondent vs. other	.01	.41	.04	.17
Technique is monolevel vs multilevel	−.01	.89	−.04	.22
Standard deviation across replications	.05	.01	.12	.00
Standard deviation within replications	.08	.00	.00	.50

NOTE: IEA = International Association for the Evaluation of Educational Achievement.

The results concerning specific leadership behaviors, that is, subdimensions of educational leadership, show that, in general, effect sizes are small. However, our results also suggest that some leadership behaviors have a significant ($p < .10$) and positive relationship with student outcomes. More specifically, this is the case for four out of the nine behaviors under review. These positive significant relationships range from .02 to .19 and relate to the following leadership behaviors: supervision and evaluation ($Z_r = .02$), monitoring ($Z_r = .07$), visibility ($Z_r = .07$), and defining and communicating mission ($Z_r = .19$).

Defining and communicating mission thus seems to be the most relevant leadership behavior in terms of improving student outcomes and confirms Hallinger and Heck's (1998) conclusion that this is one of the most important aspects of school leadership. It must be noted, however, that unlike the results concerning the other subdimensions, this specific leadership behavior result is not very robust. The results of our sensitivity analysis show that the effect size reduces from .19 to .08 when the outliers are removed from the sample. Although this implies that there is still a positive and significant relationship between this subdimension and student outcomes, the indicator loses much of its relevance.

One specific leadership behavior, namely, conducting activities aimed at improving and developing the school, appears to have a negative relationship with student achievement. This finding illustrates that one should take caution to interpret the results in terms of causes and effects because almost all studies under review are cross-sectional by design. A possible explanation

TABLE 3
Leadership Behaviors: Mean Effect Sizes and Standard Deviation Across and Within Replications

	Estimate	SE	p *Value*
Main dimension: Educational (single measure) leadership (total sample: $N_a = 45$, $N_w = 111$)			
Mean effect size	.01	.03	.38
Standard deviation across replications	.10	.07	.15
Standard deviation within replications	.13	.07	.05
Subdimension: Defining and communicating mission (total sample: $N_a = 6$, $N_w = 31$)			
Mean effect size	.19	.13	.07
Standard deviation across replications	.30	.24	.32
Standard deviation within replications	.08	.08	.22
Subdimension: Supervision and evaluation (total sample: $N_a = 32$, $N_w = 78$)			
Mean effect size	.02	.01	.02
Standard deviation across replications	.02	.03	.47
Standard deviation within replications	—	—	—
Subdimension: Monitoring (total sample: $N_a = 5$, $N_w = 22$)			
Mean effect size	.07	.03	.01
Standard deviation across replications	—	—	—
Standard deviation within replications	—	—	—
Subdimension: Coordinating and managing curriculum (total sample: $N_a = 10$, $N_w = 33$)			
Mean effect size	.02	.03	.31
Standard deviation across replications	.05	.06	.36
Standard deviation within replications	—	—	—
Subdimension: Advice and support (total sample: $N_a = 11$, $N_w = 31$)			
Mean effect size	.02	.03	.23
Standard deviation across replications	—	—	—
Standard deviation within replications	—	—	—
Subdimension: Visibility (total sample: $N_a = 3$, $N_w = 7$)			
Mean effect size	.08	.05	.08
Standard deviation across replications	—	—	—
Standard deviation within replications	—	—	—
Subdimension: School improvement (total sample: $N_a = 8$, $N_w = 18$)			
Mean effect size	−.05	.03	.05
Standard deviation across replications	—	—	—
Standard deviation within replications	.05	.07	.44
Subdimension: Achievement orientation (total sample: $N_a = 13$, $N_w = 46$)			
Mean effect size	.02	.02	.15
Standard deviation across replications	.02	.04	.62
Standard deviation within replications	—	—	—

NOTE: If a dash is used as an entry in the table, this indicates that the parameter could not be estimated due to small sample size and/or lack of variability across or within studies. N_a = sample size only across replications; N_w = sample size within and across replications.

for this negative relationship thus could be that principals in schools with low achievement levels feel rather compelled to take action to improve their schools. Teddlie and Stringfield (1993), for example, provide empirical evidence for such an explanation. They found that when principals reported to work with teachers to improve the school's education, this was associated with low expected achievement levels of students. Therefore, unequivocal conclusions cannot be drawn from these results.

Another problem is that some leadership behaviors are not often investigated. Visibility is a typical example. Only a very small number of all studies have included this variable in their research design, so that conclusions concerning this variable are premature. Nevertheless, we do believe that these findings show some evidence for the statement that educational leadership (really) matters for student achievement. However, its direct contribution to student outcomes is small. Also important in this context is the fact that positive and significant effect sizes are primarily found in studies that do not take other variables indicating school and teacher quality into account. Then the question arises whether these results hold up when these variables are included in the research designs.

The Influence of Moderators

The results concerning the question of whether effect sizes vary across and within studies are also given in Table 3. The high p values show little variation in effect sizes between and within studies.

Although the lack of variation within and between studies indicates that it is unlikely that moderators have a profound influence on the estimated effect sizes, we nevertheless tried to predict differences between effect sizes. The results of these analyses are shown in Table 4, which also deals with the bivariate relationships between moderators and effect sizes.

The results show that only in a few cases moderators have a significant relationship with the effect size. Moreover, when moderators do have a significant relationship with the mean effect size more than once, the results are not altogether consistent in direction. For instance, the effect sizes reported in studies conducted in countries other than the United States and the Netherlands are sometimes significantly higher, sometimes significantly lower. The notable exception concerns the sector in which the study has taken place. The results show that in three out of nine cases, studies conducted in secondary schools produced significantly lower effect sizes than those conducted in primary schools. This suggests that secondary school leaders may have less opportunity to directly affect student outcomes than primary school leaders. However, more important in our view are findings concerning the question of

TABLE 4
Leadership Behaviors: Bivariate Relationships Between Moderators and Effect Sizes

	Instructional Leadership	*Achievement*	*Orientation*	*School Improvement*	*Visibility*
Do studies in secondary schools report higher/lower effect sizes than studies in primary schools?	Lower	Lower			
Do studies using math tests report higher/lower effect sizes than studies using composite measures?					
Do studies using language tests report higher/lower effect sizes than studies using composite measures?					
Do studies conducted in the Netherlands report higher/lower effect sizes than studies conducted in the United States?	Higher				
Do studies conducted in other parts of the world than the Netherlands and the United States report higher/lower effect sizes than studies conducted in the United States?		Higher			
Do studies correcting for student intake report higher/lower effect sizes than studies not correcting for student intake?			Higher		
Do teacher studies relying on principal data report higher/lower effect sizes than studies relying on teacher data?	Lower[a]				
Do studies using multilevel procedures report higher/lower effect sizes than studies using "conventional" procedures?			Lower		

(continued)

TABLE 4 (continued)

	Advice and Support	Managing Coordinating Curriculum	Monitoring	Supervision	Mission
Do studies in secondary schools report higher/lower effect sizes than studies in primary schools?				Lower	
Do studies using math tests report higher/lower effect sizes than studies using composite measures?					
Do studies using language tests report higher/lower effect sizes than studies using composite measures?					
Do studies conducted in the Netherlands report higher/lower effect sizes than studies conducted in the United States?					
Do studies conducted in other parts of the world than the Netherlands and the United States report higher/lower effect sizes than studies conducted in the United States?				Higher	Lower[a]
Do studies correcting for student intake report higher/lower effect sizes than studies not correcting for student intake?			Higher	Lower	Lower[a]
Do teacher studies relying on principal data report higher/lower effect sizes than studies relying on teacher data?					
Do studies using multilevel procedures report higher/lower effect sizes than studies using "conventional" procedures?				Higher	

a. Result is robust: Analyses modeling all moderators show similar results.

whether studies correcting for student intake characteristics produce different effect sizes from those that do not. Our results indicate that this study characteristic significantly affects the effect size in four out of nine cases. For the other five cases, this implies that whereas studies correcting for student characteristics usually come up with lower effect sizes, the reported effect sizes are remarkably similar within and across studies. Moreover, for two leadership behaviors (supervision and evaluation as well as defining and communicating mission), adjustments for intake differences produce lower effect sizes, whereas for two other leadership behaviors (school improvement and monitoring) these adjustments produce higher effect sizes. However, these findings do not dramatically alter the findings presented in Table 3. For example, whereas the mean effect size of school improvement was about –.05, this value changes to –.01 when adjustment of student intake characteristics is taken into account. For the other three behaviors, these values change into .09 (monitoring), .02 (supervision and evaluation), and .15 (defining and communicating mission).

Finally, the results show that most findings are not robust. Findings for the analysis on bivariate relationships were hardly replicated in the analysis in which the impact of the moderators was assessed simultaneously. This suggests that these findings should be treated with caution.

The Results Summarized

Is educational leadership related to student achievement? In answering this question, first of all the results of a rigorous statistical meta-analysis of studies that sought evidence for direct effects of educational leadership on student achievement were presented. These suggest that in general, effect sizes are small. That is, correlations between leadership and student achievement are below .10, which implies a maximum effect size expressed as Cohen's d of .20. In Cohen's (1989, p. 25, 79) terminology, this is a small effect. Although it indicates that not more than 1% of the variation in student achievement is associated with differences in educational leadership, one should bear in mind that the measures used in the studies are far from perfectly reliable and thus may lead to an underestimation of the association. In organizational studies on the relation between leadership and student performance, one is studying how one individual affects many others. In this sense, a small effect may still be very relevant.

More refined analyses show that there is no evidence for a direct effect of educational leadership on student achievement in secondary schools. Moreover, studies conducted in the Netherlands on average show no effects of educational leadership. The effects found in the various studies appeared not to

be related to characteristics of the research design. When looking into specific leadership behaviors, "defining and communicating mission" has the largest effect (Cohen's *d*), ranging from .30 to .38. However, these latter findings are not very robust: The exclusion of outliers (studies with extreme low or high effects) led to significantly lower average values. Analyses in which differences between countries are modeled do not give clear indications that leadership matters more in the United States than in other countries (except, as already indicated, for the Netherlands, where it does not seem to matter much).

CONCLUSION

A variety of explanations account for why our further test of the direct effects model has been inconclusive. Better conceptualization of the phenomenon of educational leadership is needed. Context and intermediate factors should be taken into account in future research. In our view, the competing values framework developed by Quinn and Rohrbaugh (1983) offers an interesting approach for further thinking on these matters. One of the reasons is that this framework has proved to be fruitful in investigating school cultures. Different school cultures can indeed be distinguished with different consequences for student outcomes (Maslowski, 2001). What is needed is more insight into the role of school leaders in developing and sustaining these cultures. This notion is embedded in the framework inasmuch as it assumes an association between particular leadership values and behaviors on one hand and the existence of a specific culture on the other. One of the advantages of this framework is that it provides the opportunity to derive and test clear hypotheses, including the notion of reciprocity. The framework suggests that leaders not only shape cultures but also adapt to them. In other words, it provides a theoretical grounding for the call for reciprocal models in educational leadership studies (Hallinger & Heck, 1998).

A second advantage is that the model emphasizes the relationship between values and behaviors. In this respect, the framework is in line with present theoretical notions about leadership arguing that research should not only pay attention to behaviors but also to *why* principals act as they do (Leithwood, 1995). A last advantage lies in the fact that the employment of multiple outcomes is embedded in the framework. School effectiveness studies in general and educational leadership studies in particular have always been criticized for focusing on cognitive student outcomes only. The framework meets this objection because it implies that the organization's innovative capacity, teachers' working conditions, and smooth internal

organizational functioning are also worthwhile outcomes of leadership behavior. Moreover, thinking in terms of multiple outcomes allows researchers to draw a more balanced and thoughtful picture of the effective school by taking multiple indicators into account and by exploring their interrelationships.

Improved Methodology

Another question with regard to future research is which methodology should be used to investigate the relationship between educational leadership and student outcomes. Most studies consist of surveys that depend on naturally occurring variation. This leads by definition to small effects. Given the restrictions in variability in leadership behaviors and the fact that school effects are by definition small (achievement differences are best explained by student characteristics), samples of schools tend to be too small to detect significant effects.

Therefore, especially in working with mediated effect or reciprocal models, the need to use longitudinal data is often expressed to detect the real impact of leadership. However, these kinds of studies are time-consuming, expensive, and difficult to conduct. That is why we plead for intervention studies. Competency profiles for school principals are being developed in many countries nowadays (Davies & Ellison, 1997). The profiles not only play an important role in the selection and assessment of school principals, they also are quite directive in determining the contents of professional development programs for principals. This affords an excellent opportunity for researchers to set up true or quasi-experiments comparing various effects of competency trained versus untrained groups.

More Research on Indirect Effects

Despite repeated calls by researchers for a new approach to the study of leadership (Bossert et al., 1982; Hallinger & Heck, 1998; Pitner, 1988), research based on the indirect effect model is hard to find in peer-reviewed journals.[1] Nevertheless, we found five studies investigating the indirect effects of educational leadership on student achievement (see Appendix B for the studies involved). For instance, Hallinger and Heck (1996, 1998) reanalyzed data of two studies. They found a small but significant direct effect of principals' efforts on improved learning climate as well as a moderate indirect effect of principals' instructional efforts on student learning outcomes. Hallinger, Bickman, and Davis (1996) explored the nature and extent of the principal's effects on reading achievement in a sample of 87 U.S.

elementary schools. Their results indicated a direct effect of leadership on the existence of a clear school mission, which in turn influenced student opportunity to learn and teachers' expectations for student achievement. This constellation of instructional climate variables had a positive subsequent effect on student achievement gain in reading. Hill, Rowe, and Holmes-Smith (1995) used a multilevel path analysis to study the impact of leadership on student progress in English. Their results suggested that educational leadership is mediated by teacher practices and attitudes (in this study teacher-student interaction and professional culture). Bosker, De Vos, and Witziers (2000) combined multilevel and structural equation modeling to investigate the causal structure of school effectiveness in primary and secondary schools in the Netherlands. Their results show that a model assuming an indirect impact of school-level factors fitted the data better than a direct model. The specific implications of their study for educational leadership, one of the school-level conditions within their study, is renewed support for a model in which educational leadership affects student outcomes (reading and math) indirectly through, in this case, teachers' job satisfaction, teachers' achievement orientation, and evaluation and feedback practices.

It is obvious that these studies yield indications for the direction of future leadership studies. The empirical evidence reported in these five studies support the tenability of the indirect effect model, and comparisons of the direct with the indirect model all favor the idea of mediated effects. One study even shows that once indirect effects are modeled properly, direct effects of educational leadership are absent. These studies demonstrate that educational leadership is related to school organization and culture as well as to teacher behavior and classroom practices and these factors are related in turn to student achievement. Certainly, the evidence presented by the indirect effects model of educational leadership may not alter the conclusions that the tie between leadership and student achievement is weak. However, the studies suggest some new routes for future research including the indirect model.

APPENDIX A
Leadership Studies Into Direct Effects Under Review

Bamburg, J. D., & Andrews, R. L. (1991). School goals, principals and achievement. *School Effectiveness and School Improvement, 2*(3), 175-191.

Bedford, B. (1993). *School effectiveness. Characteristics and student achievement: A study of relationships in Georgia Middle Schools.* ERIC document No. EA 020 722.

Blank, R. K. (1987). The role of principal as leader: Analysis of variation in leadership of urban high schools. *Journal of Educational Administration, 81*(2), 69-80.

Brandsma, H. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs* [Characteristics of primary schools and school effectiveness]. Groningen, the Netherlands: RION.

Brewer, D. J. (1993). Principals and student outcomes. *Economics of Education Review, 12*(4), 281-292.

Cantu, M. (1994). A study of principal instructional leadership behaviors in successful and non-successful urban elementary schools. (Unpublished doctoral dissertation, University of Texas at Austin, 1995). *Dissertation Abstracts International, 55*(7).

Couch, J. C. (1991, November). *A study of student achievement and how it relates to the principal in the role of instructional leader.* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Lexington, NY.

Durland, M., & Teddlie, C. (1996, April). *A network analysis of the structural dimensions of principal leadership in differentially effective schools.* Paper presented at the annual meeting of the AERA, New York.

Ellet, C. D., Logan, C. S., & Naik, N. S. (1991). *Development, reliability and validity of a teacher instrument to assess organizational coupling structure of schools: Implications for the conception of coupling as "paradox."* Paper presented at the annual meeting of the AERA, Chicago.

Friedkin, N. E., & Slater, M. R. (1994). School leadership and performance: A social network approach. *Sociology of Education, 67*, 139-157.

Gouldring, E. B., & Pasternak, R. (1994). Principals' coordinating strategies and school effectiveness. *School Effectiveness and School Improvement, 5*(3), 239-253.

Grisay, A., (1995, January). *Effective and less effective junior schools.* Paper presented at the annual meeting of the ICSEI, Leeuwarden, the Netherlands.

Guillemard, L., Palmer, D. J., & Wilson, V. L. (1994, April). *Discrimination of average, effective and ineffective intermediate and secondary schools from ratings of school characteristics by parents, students, teachers and principals.* Paper presented at the annual meeting of the AREA, New Orleans.

Hallinger, P. (1989, April). *What makes a difference? School context, principal leadership and student achievement.* Paper presented at the annual meeting of the AERA, San Francisco.

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

Heck, R. H. (1993). School context, principal leadership and achievement: The case of secondary schools in Singapore. *Urban Review, 25*(2), 1993.

Hill, P. W., Rowe, K. J., & Holmes-Smith, P. (1995, January). *Factors affecting students' educational progress: Multilevel modeling of educational effectiveness.* Paper presented at the annual meeting of the ICSEI, Leeuwarden, the Netherlands.

Hofman, R. (1994). *Schoolbesturen en de kwaliteit van het onderwijs* [School boards and school quality]. Groningen, the Netherlands: RION.

Houtveen, T., Vermeulen, C., & Van der Grift, W. (1993). *Bouwstenen voor onderzoek naar de kwaliteit van scholen* [Towards a research model into school quality]. Utrecht, the Netherlands: ISOR.

Hoy, W. K., Tater, J. C., & Bliss, J. R. (1990). Organizational climate, school health, and effectiveness: A comparative analysis. *Educational Administration Quarterly, 26*(3), 260-279.

Hunter, C. (1994). Los Angeles Unified School District middle schools principals' instructional behaviors and academic achievement. (Unpublished doctoral dissertation, Peppardine University, 1995). *Dissertation Abstracts International, 55*(12).

Krug, S. E. (1992). *Instructional leadership, school instructional climate and student learning outcomes*. Department of Education, Washington, DC: Office of Educational Research and Improvement.

Leitner, D. (1994). Do principals affect student outcomes. *School Effectiveness and School Improvement, 5*, 219-238.

Mandeville, G. K., & Kennedy, E. (1991). The relationship of effective school indicators and changes in the social distribution of achievement. *School Effectiveness and School Improvement, 2*, 14-33.

Ming, T. W., & Cheong, C. Y. (1995, January). *School environment and student performance*. Paper presented at the annual meeting of the ICSEI, Leeuwarden, the Netherlands.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Somerset, UK: Open Books.

Postlewaithe, N., & Ross, K. (1993). *Effective schools in reading: Implications for educational planners*. Den Haag, the Netherlands: IEA.

Reezigt, G., Guldemond, H., & Creemers, B. (1995). *The empirical validity of the school effectiveness model*. Paper presented at the annual meeting of the ECER, Bath, UK.

Rowan, B. P., Chiang, F. S., & Miller, R. J. (1996, April). *Using research on employee performance to study teaching effectiveness: Results from an analysis of teachers effects on student achievement in mathematics using NELS: 88 data*. Paper presented at the annual meeting of the AERA, New York.

Rymemans, R., Geudens, V., Coucke, H., Van den Bergh, H., & Daems, F. (1996). *Effectiviteit van Vlaamse scholen* [Effectiveness of Flemish schools']. Antwerpen, Belgium: Universiteit van Antwerpen.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference. Lessons learned from a 10 year study of school effects*. New York: Teachers College Press.

Van de Grift, W. (1987). Zelfpercepties van onderwijskundig leiderschap en gemiddelde leerprestaties [Self perceptions of education leadership and average learning achievement]. In J. Scheerens, & W.G.R. Stoel (Eds.), *Effectiviteit van onderwijsorganisaties*. Lisse: Swets & Zeitlinger.

Van de Grift, W. (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement, 1*(3), 26-40.

Van der Werf, M. P. C., & Weide, M. G. (1993). Effectieve voorrangskenmerken in de school en klas [Characteristics of effective schools and classes for the educationally disadvantaged]. *Pedagogische Studiën, 70*, 108-221.

Vermeulen, C. J. (1987). De effectiviteit van 17 Rotterdamse stimuleringsscholen [The effectiveness of 17 Rotterdam schools]. *Pedagogische Studiën, 64*, 49-58.

Witziers, B. (1992). *Coördinatie van het onderwijs* [Coordination of education]. Enschede, the Netherlands: Universiteit Twente.

Yelton, B. T., Miller, K., & Ruscoe, G. C. (1994, April). *The stability of school effectiveness: comparative path models*. Paper presented at the annual meeting of the AERA, New Orleans.

APPENDIX B
Studies Into Indirect Effects of Leadership Under Review

Bosker, R. J., Vos, H. J., Witziers, B., & Scheerens, J. (2000, April). *Theories and models of educational effectiveness*. Paper presented at the AERA, New Orleans.

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, April). *The principal's role in school effectiveness: An assessment of methodological progress, 1980-1995*. Paper presented at the AERA, New York.

Hallinger, P. & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement, 9*(2), 157-191.

Hill, P. W., Rowe, K. J., & Holmes-Smith, P. (1995, January). *Factors affecting students' educational progress: Multilevel modeling of educational effectiveness*. Paper presented at the annual meeting of the ICSEI, Leeuwarden.

APPENDIX C
Statistical Modeling Issues

The multilevel model for the meta-analysis starting with the within-replications model is (cf. Raudenbush & Bryk, 1985; Raudenbush, 1994)

$$d_{rs} = \delta_{rs} + e_{rs} \tag{1}$$

The effect size d in replication r in study s (d_{rs}) is an estimate of the population parameter (δ_{rs}) and the associated sampling error is e_{rs} (because in each replication only a sample of schools is studied). The between-replications model is

$$\delta_{rs} = \delta_s + u_{rs} \tag{2}$$

In this model, the true replication effect size is a function of the effect size in study s and sampling error u_{rs}. Finally, the between-studies model is formulated as follows:

$$\delta_s = \delta_0 + v_s \tag{3}$$

Expressed in words: The true unknown effect size as estimated in study s (δ_s) is a function of the effect size across studies (δ_0) with random sampling error v_s (because the studies are sampled from a population of studies).

To assess the effects of study characteristic, we extended model (2) to

$$\begin{aligned}\delta_{rs} = \delta_0 &+ \gamma_1 \text{ subject-math}_{rs} + \gamma_2 \text{ subject-lang}_{rs} + \gamma_3 \text{ sector}_s \\ &+ \gamma_4 \text{ country-United States}_s + \gamma_5 \text{ country-NL} + \gamma_6 \text{ design}_s \\ &+ \gamma_7 \text{ respondent}_s + \gamma_8 \text{ statistical technique employed}_s + u_{rs} + v_s\end{aligned} \tag{4}$$

where

subject-math	0 = composite score for math and language, 1 = math only

subject-lang	0 = composite score for math and language, 1 = language only
sector	0 = primary education, 1 = secondary education
country-United States	0 = else, 1 = United States
country-the Netherlands	0 = else, 1 = The Netherlands
design	0 = gross, 1 = value added (correction for prior achievement and/or background variables)
respondent	0 = teacher, 1 = school leader
statistical technique employed	0 = multilevel, 1 = monolevel

Thus, in equation (4) δ_0 is the estimated effect size for studies where all predictors have value 0.

NOTE

1. This is particularly the case when we limit ourselves to those studies using student achievement as an endogenous variable in the model or those being conducted in or after 1995 and fulfilling a requirement inherent to the indirect model in the sense that modeling techniques are used that allow the testing of assumptions of causality among multiple variables.

REFERENCES

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

Bosker, R. J., & Witziers, B. (1996, April). *The true size of school effects*. Paper presented at the AERA, New York.

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

Bredeson, P. V. (1996). New directions in the preparation of educational leaders. In K. Leithwood, J. Chapman, D. Corson, P. Hallinger, & A. Hart (Eds.), *International handbook of educational leadership and administration* (pp. 251-277). Dordrecht, the Netherlands: Kluwer Academic.

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.

Cohen, J. (1989). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Davies, B., & Ellison, L. (1997). *School leadership for the 21st century: A competency and knowledge approach*. London: Routledge.

Edmonds, R. R. (1979). *A discussion of the literature and issues related to effective schooling*. Cambridge, MA: Harvard Graduate School of Education, Center for Urban Studies.

Fullan, M., & Watson, N. (2000). School-based management: Reconceptualization to improve learning outcomes. *School Effectiveness and School Improvement, 11*(4), 453-473.

Hallinger, P. (1989, April). *What makes a difference? School context, principal leadership and student achievement*. Paper presented at the annual meeting of the AERA, San Francisco.

Hallinger, P. (1994). *A resource manual for the principal instructional management rating scale*. Nashville, TN: Vanderbilt University, Center for Advanced Study of Educational Leadership.

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, April). *The principal's role in school effectiveness: An assessment of methodological progress, 1980-1995*. Paper presented at the AERA, New York.

Hallinger, P., & Heck, R. H. (1998). Exploring the principals' contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement, 9*(2), 157-191.

Hill, P. W., Rowe, K. J., & Holmes-Smith, P. (1995, January). *Factors affecting students' educational progress: Multilevel modeling of educational effectiveness*. Paper presented at the annual meeting of the ICSEI, Leeuwarden, the Netherlands.

Hoy, W. K., & Miskel, C. G. (1991). *Educational administration. Theory, research, and practice* (4th ed.). New York: McGraw-Hill.

Hughes, M. (1985). Leadership in professionally staffed organizations. In M. Hughes, P. Ribbins, & H. Thomas (Eds.), *Managing education* (pp. 262-290). London: Holt, Rinehart & Winston.

Imants, J. (1996). *Leiding geven aan onderwijs! Onderwijskundig leiderschap en taakdifferentiatie in basisscholen* [Leadership in education! Educational leadership and task differentiation in primary schools]. Leiden, the Netherlands: DSWO Press.

Krüger, M. L. (1995). Schoolleiderschap en schooleffectiviteit: Op weg naar een model voor onderzoek [School leadership and school effectiveness: Towards a research model]. *Pedagogische Studiën, 72*, 404-423.

Krüger, M. L., Witziers, B., Sleegers, P. J. C., & Imants, J. (1999). Onderwijskundig leiderschap in moderne onderwijsinstellingen [Educational leadership in modern educational organizations]. In B. P. M. Creemers, M. L. Krüger, & C. A. Van Vilsteren (Eds.), *Handboek schoolorganisatie en onderwijsmanagement* (pp. C4120.1-26). Alphen a/d Rijn, the Netherlands: Samson H. D. Tjeenk Willink.

Lamberts, P. C., & Abrams, K. R. (1995). Meta-analysis using multilevel models. *Multilevel Modelling Newsletter, 7*(2), 17-19.

Leithwood, K. (1992). The move to transformational leadership. *Educational Leadership, 49*(5), 8-12.

Leithwood, K. (1995). Cognitive perspectives on school leaders. *Journal of School Leadership, 5*, 115-135.

Leithwood, K., & Jantzi, D. (1990). Transformational leadership: How principals can help reform school cultures. *School Effectiveness and School Improvement, 1*(4), 249-280.

Leithwood, K., & Jantzi, D. (2000). Principal and teacher leadership effects: A replication. *School Leadership & Management, 20*(4), 415-434.

Leithwood, K., & Menzies, T. (1998). Forms and effects of school-based management: A review. *Educational Policy, 12*(3), 325-346.

Leithwood, K., & Montgomery, D. (1982). The role of the elementary principal in program improvement. *Review of Educational Research, 52*, 309-339.

Leithwood, K., Tomlinson, D., & Genge, M. (1996). Transformational leadership. In K. Leithwood, J. Chapman, D. Corson, P. Hallinger, & A. Hart (Eds.), *International handbook*

of educational leadership and administration (pp. 785-840). Dordrecht, the Netherlands: Kluwer Academic.

Levine, D. U., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, WI: National Center for Effective Schools Research and Development.

Maslowski, R. (2001). *School culture and school performance. An explorative study into the organizational culture of secondary schools and their effects*. Enschede, the Netherlands: Twente University Press.

Murphy, J. (1988). Methodological, measurement and conceptual problems in the study of instructional leadership. *Educational Evaluation and Policy Analysis, 4*(4), 290-310.

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.

Postlethwaite, N., & Ross, K. (1993). *Effective schools in reading: Implications for educational planners*. Den Haag, the Netherlands: IEA.

Pounder, D. G., Ogawa, R. T., & Adams, E. A. (1995). Leadership as an organization-wide phenomenon: Its impact on school performance. *Educational Administration Quarterly, 31*(4), 564-588.

Quinn, R. E., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management Science, 29*(3), 363-377.

Rasbash, J., & Woodhouse, G. (1995). *MLn command reference*. London: University of London, Institute of Education, Multilevel Models Project.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-323). New York: Russell Sage.

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10*, 75-98.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-245). New York: Russell Sage.

Rutter, M., Maugham, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours. Secondary Schools and their effects on children*. Somerset, UK: Open Books.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

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

Silins, H. (1994). The relationship between transformational and transactional leadership and school improvement outcomes. *School Effectiveness and School Improvement, 5*(3), 272-298.

Sleegers, P. J. C. (1991). *School en beleidsvoering* [Schools and policy making]. Nijmegen, the Netherlands: Katholieke Universiteit Nijmegen.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference. Lessons learned from a 10 year study of school effects*. New York: Teachers College Press.

Van de Grift, W. (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement, 1*(3), 26-40.

Van de Grift, W., & Houtveen, A. A. M. (1999). Educational leadership and pupil achievement in primary education. *School Effectiveness and School Improvement, 10*(4), 373-389.

Wildy, H., & Louden, W. (2000). School restructuring and the dilemmas of principals' work. *Educational Management & Administration, 28*(2), 173-184.

Bob Witziers *is an assistant professor at the University of Twente. Research interests include school leadership, school effectiveness, school improvement, and educational policy. His most recent publication, in* School Leadership and Management *(with J. Imants and P. Sleegers), is a 2001 article titled "The Tension Between Organizational Sub-Structures in Secondary Schools and Educational Reform."*

Roel J. Bosker *is a full professor at the University of Groningen. Research interests include educational effectiveness, educational evaluation, multilevel models, and comparative education. His most recent publication, in* Educational Research and Evaluation *(with H. P. J. M. Dekkers and G. Driessen), is a 2000 article titled "Complex Inequities of Educational Opportunities. A Large-Scale Longitudinal Study on the Relation Between Gender, SES, Ethnicity, and School Success."*

Meta L. Krüger *is an assistant professor at the University of Amsterdam. Her research interests include educational management and leadership, leadership and gender, school organization, school cultural and school improvement, and educational policy. Her most recent publication is "Images of Leadership in Research, Policy, and Practice" in* Educational Leadership for the New Century *(Kosovo: KEDP Press).*