DOCUMENT RESUME

ED 392 147 EA 027 376

AUTHOR Bosker, Roel J.; Witziers, Bob
TITLE A Meta Analytical Approach Regarding School Effectiveness: The True Size of School Effects and the Effect Size of Educational Leadership.
PUB DATE 95
NOTE 19p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Effective Schools Research; *Effect Size; Elementary Secondary Education; Error of Measurement; *Leadership; Outcomes of Education; *School Effectiveness; Statistical Analysis; Statistical Data

ABSTRACT

School-effectiveness research has not yet been able to identify the factors of effective and noneffective schools, the real contribution of the significant factors, the true sizes of school effects, and the generalizability of school-effectiveness results. This paper presents findings of a meta analysis, the Dutch PSO programme, that was used to answer the above questions. The paper relates results to the size of reported school effects and to the effect size of the variable, educational leadership. Some persistent problems in the measurement of school effects are identified: specifically, measurement error, specification in relevant levels, and the choice of covariates. Studies conducted in the United States show a significant positive relationship between educational leadership and achievement, which raises the question why leadership is an important variable in the United States and not in other countries. Eight tables are included. Appendices contain statistical tables. (Contains 31 references.) (LMI)

A META ANALYTICAL APPROACH REGARDING SCHOOL EFFECTIVENESS*:

the true size of school effects and the effect size of educational leadership

Roel J. Bosker**
Bob Witziers

Department of Education
Division of Education Administration
University of Twente

*The research reported in this paper was financially supported by SVO (SVO-project "Modelling school effectiveness")

**Correspondence should be directed to: Dr. R.J. Bosker, Department of Education, University of Twente, PO Box 217, 7500 AE Enschede
E-mail address: R.J.Bosker@edte.utwente.nl

A meta-analytical approach regarding school effectiveness: the size of school effects and the effect size of educational leadership

ABSTRACT

In the field of school effectiveness research there is a growing uncertainty about questions like which factors are responsible for differentiating between effective and non-effective schools, what is the real contribution of the significant factors, what are the true sizes of school effects and the extent of generalizability of school effectiveness results. One of the possibilities to address these questions is performing a meta-analysis. This paper deals with background of a meta-analytical approach as conducted by the University of Twente. Furthermore some preliminary results are presented. These results are related to the size of reported school effects and to the effect size of the variable educational leadership. With regard to first results will be presented for different sectors and subjects with an estimation of the boundaries where between the "true" effects may be. Prior to the presentation of these results some persistent problems in the measurement of school effects will be dealt with. More specifically attention will be paid to problems like measurement error, specification in relevant levels and the choice of covariates. With regard to educational leadership, attention will be paid to the question whether educational leadership does have a significant relationship with student achievement and what the effect size of this variable might be.

INTRODUCTION

The history of school effectiveness research has by now a long tradition. A popular view is to look at school effectiveness research as a reaction to the quite pessimistic views on teachers, schools and education in general brought forward by disappointing results of research, in particular those results of the work of influentiai researchers like Coleman et. al. (1966) and Jencks (1972). In this respect the work of Edmonds (1979) and Brookover, Beady, Flood, and Schweitzer (1979) in the United States and of Rutter, Maughan, Mortimore and Ouston (1979) in the United Kingdom are often seen as important starting points for school effectiveness research.

In particular in the United States a great deal of work has been done by researchers building on the work of Edmonds and Brookover and associates. Around the mid-1980 these studies in turn led to reviews of school effectiveness research, in which frequently (five or more) factors were cited as being responsible for differences between effective and non-effective schools (Purkey & Smith, 1983; Wilson & Corcoran, 1983). In a sense, these reviews and the factors mentioned in these reviews formed one of the basis tenets of the school effectiveness community of the 1980's.

However, things have drastically change in the last decennium, mainly due to the increased internalization of the school effectiveness community since the 1980's. Where eight or ten years ago generally isolated communities of researchers in different cultures (especially in the United Kingdom, the United States, Australia, the Netherlands, Canada, Scandinavia) were working on the subject of school effectiveness, nowadays there is an international network making use of each other's concepts and building on results stemming from different cultures than their own.

The effect of the internationalization of the field has not merely been the affirmation of the validity of the existing knowledge base of the discipline. In the United States the results of research are the most consistent with the 'original' knowledge base. Most clearly this can be deducted from the research review conducted by Levine and Lezotte (1990;1992). Drawing on a large body of studies in the field of school effectiveness and school improvement they are able to note the consistent tendency for certain school effectiveness 'correlates' or factors to appear in virtually all studies reviewed as being linked with school

effectiveness. These factors are shown in table 1.

Table 1 Characteristics of unusually effective schools (Levine & Lezotte, 1990)

Productive school climate and culture
- Orderly environment
- Staff commitment
- Problem solving orientation
- Staff cohesion, collaboration, consensus, communications
- Staff input into decision making
- Schoolwide emphasis on recognizing positive performance

Focus on student acquisition of central learning skills
- Maximum availability and use of time for learning
- Emphasis on mastery of central learning skills

Appropriate monitoring of student progress

Practice orientated staff development

Outstanding leadership
- Vigorous selection and replacement of teachers
- Maverick orientation and buffering
- frequent monitoring of school activities
- High expenditure of time and energy for school improvement actions
- Support for teachers
- Acquisition of resources
- Superior instructional leadership
- Availability and effective utilization of instructional support personnel

Salient parental involvement

Effective instructional arrangements
- Successful grouping and related organizational arrangements
- Appropriate pacing and alignment
- Active/enriched learning
- Effective teaching practices
- Emphasis on higher order learning in assessing instructional outcomes
- Coordination in curriculum and instruction
- Easy availability of abundant, appropriate instructional materials
- Classroom adaption
- Stealing time for reading, language and maths

High operationalized expectations and requirements for students

Other possible correlates
- Student sense of efficacy
- Multi-cultural instruction and sensitivity
- Personal development of students
- Rigorous and equitable student promotions policies and practices

However, a different picture emerges from research on school effectiveness conducted in other countries. In the Netherlands since the 1980's there is a growing body of studies dealing with factors associated with school effectiveness. The results of these studies are summarized in table 2. These results show clearly that there is a consistent inability of Dutch researcher to establish in their schools the importance of the school effectiveness correlates mentioned by Levine and Lezotte.

Table 2 School effectiveness factors supported by Dutch studies (n=42)

	no. of studies addressing factor	*no. of studies showing positive significant relations*
factor		
structured lessons/feedback	17	6
instructional leadership	17	3
orderly climate	15	6
student evaluation	18	6
whole class/differentiation	10	3
achievement orientation	24	9
team stability/teacher cooperation	15	1
time/homework	17	8

More or less the same holds true for school effectiveness research in the United Kingdom. Discussing the British research tradition in the field of school effectiveness Reynolds (1992) states that many of the early certainties in the British research paradigm have eroded as the field has developed. More specifically he points to the fact that in the United Kingdom the field of school effectiveness research has been affected by uncertainties related to the size of school effects and their consistency over time, the interrelationship of outcome variables and the precise factors responsible for differentially effective school processes.

The Dutch-PSO programme

The above delineated development of increasing uncertainty about the validity of the knowledge base of school effectiveness research has led to a growing awareness of the relevancy of questions like what factors are really responsible for differences in effectiveness between schools, what is the size of school effects and the generalizabilty of results of school effectiveness research. One can point for instance to the ISERP-project (Reynolds, Creemers, Bird & Farrel, 1994). This project aims to build on existing models of 'good practice' in terms of research design and to avoid the variation in national studies' research designs that limits transferability within and between countries. It does so by utilisation of standard measures of inputs, processes and outcomes, common methods of data analysis and common methods of data collection. As such, this project tries to answer the question which factors are generalizable across countries and what is the influence of the context on these factors. Another approach that tries to address questions about which factors are relevant with respect to school effectiveness, the size of school effects and the generalizability of school factors is the Dutch PSO-programme. In this paper the background of this programme and some results will be presented.

The Dutch PSO-programme, undertaken by the University of Twente and the University of Groningen, builds on the notion that there is a growing uncertainty with respect to the

school effectiveness knowledge base. The underlying assumption is that some "hard" questions should be asked with respect to the existing school effectiveness models. These models are not only seen as general and vague as to the internal relationships of factors responsible for difference in effectiveness between schools, but also uncertain as far as the significance of the factors that are supposed to cause achievement are concerned. One of the contentions of the programme is trying to put the next step forward in school effectiveness research by a number of activities. One important activity is the appliance of a quantative meta-analysis on existing school effectiveness research and thus, apart from making the available knowledge base in our field more accessible, sharpening our knowledge on which factors are and which factors are not essential in explaining educational achievement. Furthermore, a second aim of conducting this analysis is to bear upon the relevant question on the reality of generalizable school effectiveness models versus their differential or context-specific nature. Furthermore, this programme aims at a theoretical reconstruction of school effectiveness models by means of analytic work using relevant theories focuses on evaluation practices within schools and at an exploration of alternative causal specifications of conceptual school effectiveness models using available empirical data basis.

In next sections of this paper some preliminary results of the meta-analytic approach will be presented. First will be dealt with the question about the true size of school effects. This section will start with an overview of the problems related to isolating the true effects of schools. Next to this the results of a quantative meta-analysis on the size of school effects will be presented. After the question of the true size of school effects, attention will be paid to the question whether educational leadership does have a positive and significant relationship with student achievement and what the effect size of this variable might be.

THE 'TRUE' SIZE OF SCHOOL EFFECTS

introduction

As a first stage in the project mentioned, we will consider the fundamental topic of the size of school effects. there are some problems worth mentioning in considering the size of school effects. In school effectiveness research one may wish to differentiate between four types of school effects:

1. the effect of a school on its pupils is their gross mean achievements score, expressed as the deviation form the grand mean (the mean school effect). so the predicted score for school j is the grand mean. Discussions about standards in education involve such a notion of school effects.
2. the effect of a school on its pupils is the mean progress these pupils make in a given time period. The predicted score in this case is based on the pupils initial achievement. This kind of operationalization is often referred to as 'learning gain'.
3. the effect of a school on its pupils is the mean overachievemnt of its pupils. The predicted score in this case is based on pupil background characteristics such as socio-economic status, mental abilities and the like, that are known to have a substantial effect on their achievement. This kind of operationalization is most widely used in school effectiveness research.
4. the effect of a school on its pupils is the mean net progress these pupils make in as far as this progress cannot be accounted for by relevant pupil background characteristics like socio-economic status and so on. the predicted score in this case is based on the pupils' background characteristics and their initial achievement. This combination is in effect used in the 'Junior schools'- project (Mortimore et al., 1988).

Qualitative reviews ignore the fact that studies in the field of school effectiveness vary with respect to the operational definition of "school effect". The only reconciliation is, that Bosker (1990) empirically demonstrated that the four operationalizations correlate high (.78

and more).

As an example of a quantitative approach to synthesizing the results of school effectiveness research will present a meta-analysis on studies that assessed gross school effects and/or school effects based on the idea of overachievement. We confine ourselves to the UK and the Netherlands, to primary and secondary schools, to mathematics and language as subject domains, and to those studies that used multi-level models or random effects ANOVA. In the appendix the studies are mentioned that are selected for the meta-analysis. The questions that we seek to answer is:

1. what is the size of the gross and overachievement based school effects?
2. does the size of the school effect vary across subject domains (mathematics and language), sectors (primary and secondary), and/or country (UK and the Netherlands)?

For the meat-analysis we apply the multilevel model as suggested by Raudenbush (1994). We consider the selected studies as a sample from the population of studies on school effects. Nested under each study are the secondary units: the schools. What we will consider as the size of the school effect is the estimated between school variance proportional to the total variance in achievement (within and between schools).
The multi-level model then, starting with the within-study model, is:

(1) $\text{size}_j = \textit{size}_j + e_j$

The effect size estimate in study j (size) is an estimate of the population effect size ($\textit{size}_j$) and the associated sampling error is e_j (since in each study only a sample of schools is studied).
The between-studies model is:

(2) $\textit{size}_j = \text{intercept} + \upsilon_j$

In words: the true unknown-effect size as estimated in study j ($\textit{size}_j$) is a function of the effect size across studies (intercept) with random sampling error υ_j (since the studies are sampled from a population of studies).
In assessing effects of subject domain, country, and sector model 2 is extended to:

(3) $\textit{size}_j = \text{intercept} + \gamma_1\text{subject}_j + \gamma_2\text{sector}_j + \gamma_3\text{country}_j + \upsilon_j$

Only a few of the studies reviewed mentioned standard errors for the estimated variance components (the size of which depends a.o. on the sample size used in the study), and when they did, it was not in all cases quite clear whether these standard errors had to do with the variance or the square root of it. For this reason we roughly calculated the standards errors from (cf Longford, 1994, 58):

(4) $\text{var}(\hat{\tau}^2) = 2\sigma^4/N * [1/(n-1) + 2\omega + n\omega^2]$

where τ^2 is the between school variance, σ^2 is the within school variance, N is the total sample size, n is the (average) number of pupils per school in the sample, and ω is the the variance ratio τ^2/σ^2

This approach to calculating the standard errors of the variance components is rather crude, since we have to assume balanced designs, and no predictors (which of course is

not so in the meta-analysis of the net between school variance).

results

The results of the meta-analysis on the gross school effects are presented in table 3.

Table 3: Results for the meta-analysis on gross school effects

	effect	s.e.	p-value
mean gross school effect	.1386	.0100	.000
variance across studies	.0039		.000
intercept gross school effect	.1594	.0134	.000
sector-effect (secondary)	-.0831	.0183	.000
subject-effect (mathematics)	.0261	.0164	.113
country-effect (uk)	.0323	.0224	.140
variance across studies (res)	.0020		.000

In the first two lines of the table the estimated mean gross school effect is presented: across all studies the size of the school effect is .1386, and the variance across studies indicates that the 95% confidence interval runs from .0162 to .2610. The remainder of the table presents the effects of subject, sector, and country on the size of school effects: for language achievement in primary education in the Netherlands the size of school effects in secondary education is .1594 - .0831=.0763.

The results with respect to the net size of school effects are presented in table 4:

Table 4: Results from the meta-analysis on net school effects

	effect	s.e.	p-value
mean net school effect	.0711	.0062	.000
variance across studies (res)	.0010		.000
intercept net school effect	.0727	.0111	.000
sector-effect (secondary)	-.0194	.0153	.177
subject-effect (mathematics)	.0268	.0123	.040
country-effect (uk)	.0006	.0148	.396
variance across studies (res)	.0008		.000

When the school effect is estimated after taking intake differences between schools into account the size of the school effect (that can now be interpreted as the net school effect) diminishes to .0711, meaning that only 7 percent of the variance in achievement can be accounted for by the schools that pupils attend. Furthermore the results indicate that the size of the school effect for mathematics is .0268 percent higher than for language. No sector effects and no country effects are found.

In appendix 2 the results of meta-analysis per sector are presented. The results from these analyses can be summarized as: differences in the size of net school effects can only be demonstrated in primary education: the estimated between school variance is higher for mathematics and for the UK.

EDUCATIONAL LEADERSHIP

introduction

In the foregoing section we have tried to answer the question what the 'true' size of school effects may be. The second step in this project consists of determining through means of a meta-analytical approach whether variables mentioned in the school effectiveness literature do have a positive relation with relevant output measures and what the estimated effect size of these variables might be. One of the important variables in this respect is (educational) leadership and in this section of this paper we will explore the question whether this variable has a significant relationship with measures of student achievement and what the effect size of this variable might be. Answering these questions is not an easy task. One of the reasons for this is the diversity in the way this concept is conceptualized and investigated. In this respect one can discern at least three approaches. From a meta-analytical point of view the first two approaches are the least problematic. These approaches use either a single concept of leadership or an overall (or 'latent') concept of leadership. In a meta-analytical approach then we can use the data mentioned in these studies about the nature and size of the relationship between the single or latent concept of leadership and the output measure used in these studies. Far more problematic in this respect however, are those studies that use different indicators of leadership, but do not discern an overall or 'latent' concept of leadership. In many cases these studies only report the size of relationship between the indicators and the relevant output measures, while they do not give any indications about the overall effect of these indicators on the output measures used. Things get even more complicated, when these studies only supply data about those indicators that do have a significant relationship with the output measures used and refuse to give any information about the nature and size of effect of indicators that do not have a significant relation with the output measure used. Examples in this respect are for instance the studies of Mortimore et al (1987) and the IEA-reading study (Postlewaithe & Ross, 1993).

To deal with these problems we applied two meta-analytical procedures. The first procedure consisted of applying the vote-counting procedure. This procedure makes use of the number of positive results in relation to the number of not-positive results (Bushman, 1994). We applied this procedure thereby in two ways. Firstly, we investigated all studies involved in this analysis (see appendix 3) from an 'overall' -perspective on leadership. For studies using a single or latent concept of leadership this implied we calculated the number of times these concepts had a positive and significant relationship with the output measures used in relation to the number of times these concepts were used. For studies using multiple indicators only,

we used the following decision rule. Leadership was thought to have a positive relation with the outcome measure used when at least half of all indicators had a positive relationship with the output measure used.

This procedures was repeated but then from an 'indicator'- perspective. For instance, we calculated the number of times the leadership indicator 'teacher evaluation' had a positive, significant result in relation to the number of times this indicator was studied.

The deficiencies of the vote-counting procedure, at least from the specific procedure we used in this study, are that it does not take into account sample size and does not provide an effect size estimate. To overcome these deficience a multi-level analysis was performed. Earlier in this paper we already dwelled upon this kind of analysis, so we will restrict ourselves here to the following formula:

effect size (j) = intercept + gamma (1) country + gamma (2) method +
gamma (3) math + gamma (4) lang + gamma (5) sector

where:

country	0=else, 1=USA
study design	0= gross, 1=value addes (correction for prior achievement and/or background variables)
math	0=composite score for math and language, 1=math score only
lang	0=composite score for math and language 1=language score only
sector	0= primary education, 1= secondary education

Important to note in this context are three things. Firstly, in this analysis only studies were used which supplied all relevant data. This remark is in particular relevant for studies using multiple indicators only; studies with missing data were not used. The second important remark is the fact that this analysis deals with an 'overall'-perspective on leadership. This implies that for studies using multiple indicators only we used the mean effect of the different indicators on the output measure in the study as indicator for the effect size in the particular study. However, when data about the relationships between the indicators themselves and the output measures were available, we repeated the analyses used in the study at hand, most of the time simple regression analyses, and used the amount of variance explained by these indicators to determine the 'overall' effect size. Finally, the effect size is expressed in r_{xy} (correlation between variable x and y), implying in most cases a transformation of the original data into this metric. For an overview of the formulas involved we refer to Rosenthal (1994).

results

The results regarding the vote-counting procedure are shown in table 5 and 6. From Table 5 it can be derived that is it is not very likely that leadership does have a positive relationship with output measures indicating student achievement. Most studies fail to come up with positive significant results.

Table 5: results vote counting procedure from an overall perspective on leadership :

	math	language	composite score	total
no. of positive results	3	2	2	7
no. of possible positive results	15	38	9	62
p-value	.98	1.00	.97	1.00
Signtest (H_0:π=.05; H_a:π>.5)	H_a in all case rejected			

This conclusion can also be easily reached, when we are looking from the indicator perspective. In most cases indicators relating to the concept of educational leadership do not have a positive relationship with measures giving insight in pupils' achievement levels,

at least when we look at the indicators most commonly used in leadership studies.

Table 6: results vote counting procedure from an indicator perspective on leadership

indicator	no of positive significant relations	no of possible positive significant relations (p-value)
teacher evaluation/supervising	3	36 (.00)
observation/class visits	2	8 (.03)
defining mission	2	12 (.03)
stressing academic standards	3	24 (.00)
involvement with instruction	7	39 (.00)
discussing objectives with staff	6	28 (.00)
managing curriculum	5	17 (.01)
monitoring students	4	13 (.04)
development teachers/school	7	41 (.00)
staff participation	1	5 (.03)
providing resources	1	1 (1.00)*
visibility	2	2 (1.00)*
keeping teacher morale high	2	2 (1.00)*
rewarding/punishing pupils	0	15 (.00)
pastoral care	4	27 (.00)
community/parents	3	28 (.00)
safe and orderly climate	2	2 (1.00)*

Signtest (H_0:π=.05; H_a:π>.5)
*=signicant at .05

The results shown in table 5 and 6 can be summarized by saying that it is not very likely that leadership is related to pupils' achievement. This conclusion is more or else confirmed by the results from the multi-level analysis. Table 7 shows the estimated effect size of the variabele leadership and the variance across studies.

Table 7: Estimated effect size and variance across studies

	effect	s.e.	p-value
mean effect-size	.0414	0.225	.075
variance across studies	.0106		.000

The estimated mean effect size across all studies, which is, as mentioned before expressed in r, is .0414, which is significant at the 10%-level (one-tailed).
The estimated variance across all studies is .0106, which indicates that the 95%-prediction

interval around the mean effect-size runs from r=-.1604 to r= .3443.
The results regarding analyses trying to predict differences between effect sizes with study characteristics (or moderators) as subject matter, sector, country, and method of analysis show that two predictors have a significant relationship with the effect size (see table 8).

Table 8: predicting differences

	effect	s.e.	p-value
intercept	-.1159	.0543	.043
country (USA)	.1388	.0452	.005
study-design (value added)	.1080	.0488	.037
variance across studies (res)	.0065		.006

This model shows that on average the effect of leadership on (uncorrected) student outcomes is -.1159. In studies where the effect of leadership is assessed after taking previous student achievement and/or background characteristics of students into account the effect is -.0079 (-.1159 + .1080). In words, after correcting for student background and/or previous achievement there is no effect at all. Study design thus influences the estimated effect size.
Looking at country differences, however, it turns out that US-studies show significantly higher effect sizes. If we take the value added studies, then the estimated effect size for the US is .1309 (-.0079 + .1388). In words, in the United States leadership seems to matter when one wants to differentiate between 'good' schools and 'bad' schools.
Finally, one can deduce from table 8 that after having taken country and study-design differences into account, there are no longer residual differences between the studies in the estimated effect-sizes for leadership effects on student achievement. This implies that the moderators subject and sector hardly explain any of the variation between studies.

DISCUSSION

We have argued that the results of school effectiveness research can not be generalized other then with great caution. A qualitative approach led to the conclusion that the empirical basis for the effectiveness enhancing school factors is poor. The reasons why new analyses of existing datasets and /or quantitative meta-analyses should be undertaken are clear. To illustrate the meta-analytical approach used in our project both the size of school effects was estimated and the effect size of school leadership was estimated in a quantative way.
Regarding the school effect size we found that gross school effects had an estimated magnitude of 14 and net school effects an estimated magnitude of 7 percent. It turned out that sector, country and subject domain affected the estimated size of school effects.
What then do we know about the importance of schools? First of all we have to be aware of three pertaining problems:
1) the effect size is underestimated, since measurement error in the achievement tests shows up as within school variance; if we would take count of the, say, 20 percent "noice", the ration of the between school variance to the total "true" variance would improve from 14 to 18 percent.

2) the effect size is overestimated, since the important intermediate level of the classroom is ignored. Including the intermediate level would lead to an estimated decline in the "true" size of the gross school effect to approximately half of it. Misspecification in this respect leads to a statistical artificial increase in the size of school effects (e.g. Rowe and Hill, 1994).
3) we have assumed stable school effects, but other research (e.g. Bosker, 1992; Luyten, 1994) has shown that there is considerable variation between subjects, between cohorts, between grades, between classes, and between groups of pupils with different backgrounds. A further reduction of the true effect size thus seems plausible.
All in all the true gross effect may be something like 10 percent, and the true net effect something like 6 percent. Is this "much a do about nothing?". A technical answer can be given following the conceptual idea that one school affects all of its pupils. The importance of the school effect then can be assessed by looking at the school total of deviations, to which the within school variance and the between school variance relatively contribute 1: $n\tau^2/\sigma^2$ (Longford, 1994, 27-28). The tric lies in the premultiplication with n. This may be the number of pupils per school in the sample, or a value deemed important a priori (e.g. total number of pupils in a cohort, or even better: total number of pupils leaving the school over a number of years). The net between school variance is then as important as the within school variance if we consider a small class of 20 pupils per school. But if we consider consistent stable performing secondary schools, that serve 1,000 pupils over a period of 5 years each, the relative importance of the school is 50 times as high the within school variation. If thus seems a matter of taste to judge something as important of not. Our contention would be: much a do about something, and quite rightly so!
More problematic in this respect are of course the results regarding leadership. Since the overall contribution of educational leadership to student achievement is about zero, we might argue that all the fuzz about educational leadership must be based on ideological rather than on empirical grounds. However, since our results also show that there are large differences between educational contexts (US-studies showing a positive relationship of r= .1309 between educational leadership and achievement, while this result can not be found in other countries), our conclusion might be that we should not forget about this variable when thinking about effective schools. In this respect the most valid conclusion then is that educational leadership does matter in certain educational contexts, but that this effect is not generalizable to other educational contexts. This in turn leads of course to the more fundamental question why leadership is an important variable in the United States and not in other countries of the world.
The last remarks made in this paper will deal with the status of our conclusions. Important to note in this respect is that our conclusions are based on preliminary findings. For instance, our results about the size of school effects deal only with studies conducted in two countries. Furthermore, there are still some technical problems to be solved regarding our meta-analytical approach. Here one can think for instance of questions like whether or not it is possible to adjust our results for unreliability of the (in)dependent variables used in the studies at hand and whether or not we should apply procedures that adjust our meta-analytical results for bias due to publication bias.

References

AITKIN, M. & LONGFORD, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. **The Journal of the Royal Statistical Society. Series A (General) 149, Part 1,** 1-43.
ANDERSON, L.W. & MANDEVILLE, G.K. (1986). **Towards a solution of problems inherent in the identification of effective schools.** AERA-paper. San Fransisco.
BOSKER, R.J. (1992). **De stabiliteit en consistentie van schooleffecten in het basisonderwijs.** Enschede: UT.
BOSKER, R.J., GULDEMOND, H., HOFMAN, R.H. & HOFMAN, W.H.A. (1988). **Kwaliteit in het voortgezet**

onderwijs. Groningen: RION.
BOSKER, R.J. & SCHEERENS, J. (1989). Issues in the interpretation of the effects of school effectiveness research. **International Journal of Educational Research, 13, 7**, 741-751.
BROOKOVER, W.B., BEADY, C & P.FLOOD (1979). **School social systems and student achievement - schools can make a difference**. New York: Praeger Publishers.
BRYK, A.S. & RAUDENBUSH, S.W. (1992). **Hierarchical linear models.** New York: Sage.
COLEMAN, J.S., CAMPBELL, E.,HOBSON, C., MCPARTLAND, J., MOOD, A.,WEINFELD, F & YORK, R. (1966). **Equality of Educational Opportunity**. Washington D.C.:US Government Printing Office.
BUSHMAN, B.J. (1994). Votecounting prodcedures in Meta-analysis. In H. Cooper & L.V. Hedges (Eds.). **The Handbook of Research Synthesis.** New York: Russell Sage Foundation.
EDMONDS, R.R. (1979). **A discusssion of the literature and issues related to effective schooling**. Cambridge: Center for Urban Studies. Harvard Graduate School of Education.
FITZ-GIBBON, C.T. (1992). School effects at A Level: Genesis of an Information System?. In D.Reynolds & P.Cuttance (Eds.), **School effectiveness: research, policy and practice**. London [etc.]: Cassell.
GOOD, T.L. & BROPHY, J.E. (1986). School effects. In: M.C. Wittrock (Ed.): **Handbook of research on teaching.** New York: McMillan Inc.
JENCKS, C. (1972). **Inequality.** New York: Basic Books.
LEVINE D.U. (1992). An interpretive review of US research and practice dealing with unusually effective schools. In D.Reynolds & P.Cuttance (Eds.), **School effectiveness: research, policy and practice**. London [etc.]: Cassell.
LEVINE, D.U. & LEZOTTE, L.W. (1990). **Unusually Effective Schools: a review and analysis of research and practice**. Madison Wisconsin: National Center for Effective Schools Research and Development.
LOCKHEED, M.E. (1990). A multilevel model of school effectiveness in a developing country. In: S.W. Raudenbush & J.D. Willms (Eds.): **Schools, classrooms, and pupils: international studies of schooling from a multilevel perspective.** Edinburgh: Academic Press.
LONGFORD, N.L. (1994). **Random coefficient models.** Oxford: Clarendon Press.
LUYTEN, J.W. (1994). **School effects: stability and malleability.** Enschede: UT.
MORTIMORE, P., SAMMONS, P., STOLL, L., LEWIS, D. & ECOB, R. (1988). **The junior school project.** Somerset: Open Books.
PURKEY, S.C. & SMITH M.S. (1983). Effective schools: a review. **Elementary School Journal, 4**, 427-452.
RALPH, J.H. & FENNESEY, J. (1983). Science or reform: some questions about the effective schools model. **Phi Delta Kappan, 64**, 689-694.
RAUDENBUSCH, S.W.(1994). The random effects models. In H.Cooper & L.V.Hedges (Eds.), **The Handbook of Research Synthesis**. New York: Russell Sage Foundation.
RAUDENBUSCH, S.W. & BRYK, A.W. (1985). Empirical Bayes meta-analysis. **Journal of Educational Statistics, 10**, 75-98. REYNOLDS, D. (1992). School effectiveness and school improvement: an updated review of the Britisch literature. In D.Reynolds & P.Cuttance (Eds.), **School effectiveness: research, policy and practice**. London [etc.]:Cassell.
REYNOLDS, D., CREEMERS, B.P.M., BIRD, J.& FARREL, S. (1994). School effectiveness - the need for an international perspective. In D. Reynolds, B.P.M. Creemers, P.S. Nesselrodt, E.C. Schaffer, S.Stringfield, & C.Teddlie (Eds.), **Advances in school effectiveness research and practice**. Oxford: Pergamon Press.
ROSENTHAL, R. (1994).Parametric measures of effect size. In H. Cooper & L.V. Hedges (Eds.). **The Handbook of Research Synthesis.** New York: Russell Sage Foundation.
ROWAN, B. BOSSART, S.T. & DWYER, D.C. (1983). Research on effective schools. A cautionary note. **Educational Researcher, april**, 24-31.
ROWE, K.J. & HILL, P.W. (1994). **Multilevel modelling in school effectiveness research: how many levels?**. Paper presented at the International Congress for School Effectiveness and Improvement, Melbourne, 1994
RUTTER, M., MAUGHAN, B., MORTIMORE, P. & OUSTON, J. (1979). **Fifteen Thousand Hours. Secondary schools and their effects on children.** Somerset: Open Books Publishing Ltd.
SCHEERENS, J. (1989). **Wat maakt scholen effectief?** Den Haag: SVO.
SCHEERENS, J., VERMEULEN, C.J.A.J. & PELGRUM, W.J. (1990). Generalizability of instructional and school effectiveness indicators across nations. **International Journal of Educational Research, 13, 7**, 789-800.
WILSON, B.L. & CORCORAN, T. (1988). **Successful secondary schools: visions of excellence in American public education**. London [etc.]: Falmer Press.

Appendix 1: list of studies in the meta-analysis

nr	author(s)	schools	pupils	gross	var($\hat{\tau}^2$)	net	var($\hat{\tau}^2$)	sec	sub	cou
1	weide	45	12.24	.11	.00150	.03	.00188	p	a	nl
2	weide	45	12.24	.15	.00216	.07	.00198	p	1	nl
3	vandergrift-akker	73	3.42	.13	.00477	.06	.00459	p	1	nl
4	vandergrift-akker	73	3.42	.16	.00518	.13	.00457	p	a	nl
5	bosker	150	3.23	.14	.00262	.09	.00237	p	a	nl
6	bosker	150	3.23	.13	.00256	.03	.00245	p	1	nl
7	dejong	107	19.63	.11	.00045	.04	.00050	p	1	nl
8	dejong	107	19.63	.18	.00092	.11	.00066	p	a	nl
9	vandervelden	74	10.22	.09	.00089	.03	.00137	p	1	nl
10	vandervelden	74	10.22	.13	.00127	.07	.00142	p	a	nl
11	vanderwerf-weide	124	15.82	.17	.00080	.09	.00061	p	1	nl
12	blok-hoeksma	134	19.56	.20	.00087			p	1	nl
13	blok-hoeksma	134	19.56	.14	.00051			p	1	nl
14	blok-hoeksma	134	19.56	.20	.00087			p	1	nl
15	blok-hoeksma	134	19.56	.18	.00074			p	1	nl
16	blok-hoeksma	134	19.56	.14	.00051			p	m	nl
17	blok-hoeksma	134	19.56	.23	.00108			p	m	nl
18	blok-hoeksma	134	19.56	.20	.00087			p	m	nl
19	blok-hoeksma	134	19.56	.20	.00087			p	m	nl
20	blok-hoeksma	134	19.56	.18	.00074			p	m	nl
21	blok-hoeksma	134	19.56	.16	.00062			p	1	nl
22	kuhlemeier	43	7.98	.30	.00704			s	1	nl
23	kuhlemeier	43	7.98	.12	.00255			s	1	nl
24	vanderwerf	183	16.14	.18	.00058	.12	.00046	p	m	nl
25	reezigt	252	19.84	.19	.00042	.07	.00023	p	1	nl
26	reezigt	252	19.84	.24	.00062	.09	.00025	p	m	nl
27	hofman	75	40.88			.09	.00052	s	1	nl
28	hofman	75	40.88			.16	.00096	s	m	nl
29	luyten	1055	76.69	.04	.00001			s	1	nl
30	luyten	462	97.40	.02	.00			s	m	nl
31	luyten	470	74.47	.05	.00002			s	1	nl
32	luyten	1055	42.18	.15	.00005			s	m	nl
33	luyten	462	53.57	.10	.00006			s	m	nl
34	luyten	470	52.13	.06	.00003			s	m	nl
35	leseman	30	86.83	.10	.00081	.03	.00043	s	1	nl
36	leseman	30	86.83	.06	.00033	.01	.00039	s	1	nl

37 nl	leseman	30	86.83	.08	.00055	.05	.00053	s	l
38 nl	vanderwerf-weide	696	30.00	.21	.00016	.06	.00006	p	l
39 nl	vanderwerf-weide	696	30.00	.23	.00019	.11	.00008	p	m
40 nl	witziers	39	25.00	.20	.00276	.20	.00289	s	m
41 nl	bosker-hofman	72	4.94			.05	.00306	p	l
42 nl	bosker-hofman	72	4.94			.28	.00438	p	m

nr cou	author	schools	pupils	gross	var($\hat{\tau}^2$)	net	var($\hat{\tau}^2$)	sec	sub
43 uk	jesson-gray	120	40.83	.10	.00025	.10	.00035	s	l
44 uk	jesson-gray	120	37.50	.13	.00039	.11	.00040	s	l
45 uk	jesson-gray	120	49.17	.14	.00041	.13	.00043	s	l
46 uk	jesson-gray	40	25.00	.11	.00106	.07	.00119	s	l
47 uk	jesson-gray	20	50.00	.07	.00079	.03	.00107	s	l
48 uk	jesson-gray	30	150.00	.04	.00014	.05	.00038	s	l
49 uk	jesson-gray	30	126.67	.04	.00015	.04	.00036	s	l
50 uk	daly	30	47.23			.11	.00144	s	m
51 uk	daly	29	50.79			.07	.00098	s	l
52 uk	thomas-sammons	94	81.09			.04	.00016	s	l
53 uk	thomas-sammons	94	81.09			.05	.00018	s	m
54 uk	mortimore	50	40.00			.09	.00078	p	l
55 uk	mortimore	50	40.00			.13	.00112	p	l
56 uk	mortimore	50	40.00			.11	.00093	p	m
57 uk	fitz-gibbons	425	2.72	.38	.00188	.01	.00110	s	l
58 uk	fitz-gibbons	533	2.17	.10	.00154	.06	.00115	s	m
71 uk	fitz-gibbons	30	19.70	.13	.00203	.09	.00212	s	m
72 uk	willms-raudenbush	20	325.00			.01	.00016	s	l
73 uk	willms-raudenbush	20	325.00			.03	.00024	s	m
74 uk	goldstein	66	87.09			.05	.00024	s	l
75 uk	goldstein	66	87.09			.04	.00022	s	m

legenda: p/s: primary or secondary education; m/l: mathematics or language; nl/uk: the Netherlands or the UK.

Appendix 2: Meta-analyses per sector

Table 5: Results from the meta-analysis on gross primary school effects

	effect	s.e.	p-value
mean gross school effect	.1739	.0085	.000
variance across studies (res)	.0010		.000
intercept gross school effect	.1634	.0116	.000
subject-effect (mathematics)	.0219	.0167	.165
country-effect (uk)	not applicable (no uk-studies)		
variance across studies (res)	.0009		.000

Table 6: Results from the meta-analysis on net primary school effects

	effect	s.e.	p-value
mean net school effect	.0851	.0076	.000
variance across studies (res)	.0004		.002
intercept net school effect	.0618	.0061	.000
subject-effect (mathematics)	.0426	.0089	.000
country-effect (uk)	.0317	.0182	.089
variance across studies (res)	.0000[1]		.416

[1] the estimated value is: .0000029

Table 7: Results from the meta-analysis on gross secondary school effects

	effect	s.e.	p-value
mean gross school effec t	.1032	.0151	.000
variance across studies (res)	.0043		.000
intercept gross school effect	.0822	.0240	.003
subject-effect (mathematics)	.0274	.0352	.287
country-effect (uk)	.0305	.0313	.241
variance across studies (res)	.0043		.000

Table 8: Results from the meta-analysis on net secondary school effects

	effect	s.e.	p-value
mean net school effect	.0616	.0084	.000
variance across studies (res)	.0011		.000
intercept gross school effect	.0687	.0188	.002
subject-effect (mathematics)	.0230	.0194	.193
country-effect (uk)	-.0176	.0210	.274
variance across studies (res)	.0013		.000

Appendix 3: Leadership- studies under review

Andrews, R.L.. & Soder, R.L., Principal Leadership and student achievement Educational Leadership, 44 (6), 1987

Bamburg, J.D. & R.L. Andrews (1991, Vol 2, No 3, pp 175-191)

School Goals, Principals and Achievement, SESI, Vol. 2. no. 3, pp. 175-191

Bedford, Bill, School effectiveness Characteristics and student achievement: a study of realtionships in Georgia Middle Schools. ERIC document EA 020 722.

Brandsma, H (1993), Basisschoolkenmerken en de kwaliteit van het onderwijs, Groningen: RION

Chubb, J.E. & Moe , T.M. (1990) Politics, markets and America's Schools. Washington: The Brooking Institution.

Eberts, R.W. & Stone, J.A. (1988).Student achievement in public schools: do principals make a difference. Economics of Education Review, Vol. 7, No. 3. pp. 291-299.

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 AERA, Chicago, Ilinois, April 1991.

Goldring, E.B., Pasternak, R. (1994). Principals' Coordinating strategies and school effectiveness. SESI, Vol. 5, No. 3, pp 239-253.

Grift, W. van der & Akkermans, W. Onderwijskundig Leiderschap en leerlingprestaties in het basisondewijs. TOR, 16, no. 4., 244-257.

Grift, W. van der. (1990). Educational Leadership and Academic Achievement in Elementary Education. SESI, Vol.1. No. 3, pp 26-40.

Grisay, A., (1995). Effective and less effective junior schools. Paper held at the ICSEIR, Leeuwarden.

Heck, R., Marcoulides, Lang, P. (1991). Principal Instructional Leadership and School Achievement: The application of Discrimant Techniques. SESI

Hofman, R. (1994). Schoolbesturen en de kwaliteit van het onderwijs. Gronigen: RION.

Hoy, W.K. , Tater, J.C., Bliss, J.R. (1990). Organizational climate, school health, and effectiveness: a comparative analysis. Educational Administration Quarterly, vol. 26, no. 3, 260-279

Krug, S. (1992). Instructional Leadership, School instructional Climate and student learning outcomes. Department of Education, Washington, DC: Office of Educational Research and improvement.

Kuhlemeier, H., Bergh, H. van den (1995). Het belang van de vaksectie aardrijkskunde voor de aardrijkskundeprestatie. Arnhem: Cito

Leitner, D. (1994),Do principals affect student outcomes. SESI, vol. 5, pp. 219 - 238.

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

Postlewaithe, N., Ross, K. (1992). Effective schools in reading: implications for educational planners.IEA: The Hague.

Tam Wai Ming & Cheng Yin Cheong (1995). School environment and student performance. Paper presented at the ICSEI '95, Leeuwarden.

Vermeulen, C.J. (1987) De effectiviteit van 17 Rotterdamse stimuleringsscholen. Pedagogische Studiën, (64), 49-58.

Van der Werf, G., Nitert, E.H.M., Reezigt, G. (1995). Effectieve en minder effectieve scholen voor voor allochtone leerlingen. Groningen: RION.

Witziers, B. (1992). Coördinatie van het onderwijs. Enschede; Universiteit Twente.

Yelton, B.T., Miller, K., Ruscoe, G. C. (1994). The stability of school effectiveness: comparative path models. AERA-paper, New Orleans, 1994

Appendix 4: descriptive statistics of studies used in multi-level analyses to determine the effect size of leadership

STUDY DESIGN
Value Label

	Value	Frequency	Percent	Valid Percent	Cum Percent
gross	g	18	36.0	36.0	36.0
value added	v	7	14.0	14.0	50.0
net	n	25	50.0	50.0	100.0
	Total	50	100.0	100.0	

SUBJECT
Value Label

	Value	Frequency	Percent	Valid Percent	Cum Percent
composite	c	15	30.0	30.0	30.0
math	m	19	38.0	38.0	68.0
language	l	16	32.0	32.0	100.0
	Total	50	100.0	100.0	

SECTOR
Value Label

	Value	Frequency	Percent	Valid Percent	Cum Percent
primary	p	42	84.0	84.0	84.0
secundary	s	8	16.0	16.0	100.0
	Total	50	100.0	100.0	

COUNTRY
Value Label

	Value	Frequency	Percent	Valid Percent	Cum Percent
france	f	1	2.0	2.0	2.0
hong k	h	2	4.0	4.0	6.0
israel	i	1	2.0	2.0	8.0
netherlands	n	13	26.0	26.0	34.0
usa	u	33	66.0	66.0	100.0
	Total	50	100.0	100.0	