Research Mapping
of input, process, and learning in
primary and lower secondary schools

Technical Report

by
Sven Erik Nordenbo
Anders Holm
Eyvind Elstad
Jaap Scheerens
Michael Søgaard Larsen
Michael Uljens
Per Fibæk Laursen
Trond Eiliv Hauge

Danish Clearinghouse
for Educational Research

Copenhagen 2009
Research Mapping of input, process, and learning in primary and lower secondary schools
The Danish Clearinghouse for Educational Research
is a unit at the School of Education, Aarhus University

Title: Research Mapping of input, process, and learning in primary and lower secondary schools

Copyright: © 2009 by Danish Clearinghouse for Educational Research

ISBN

Review group
Professor Eyvind Elstad, University of Oslo
Professor Trond Eliv Hauge, University of Oslo
Professor Anders Holm, University of Copenhagen
Professor Per Fibæk Laursen, Aarhus University
Professor Michael Uljens, Åbo Academy University
Professor Jaap Scheerens, University of Twente

Advisory group at Danish Clearinghouse
Professor, Director Sven Erik Nordenbo
Associate Professor, Deputy director Michael Søgaard Larsen
Communications consultant Mette Thornval

Scientific assistants
Anne Bang-Olsen
Majken Mosegaard Svendsen
Neriman Tiftikci
Rikke Eline Wendt
Trine Kløveager Nielsen

Section: Technical report

Authors
Anders Holm
Eyvind Elstad
Jaap Scheerens
Michael Søgaard Larsen
Sven Erik Nordenbo (editor)
Michael Uljens
Per Fibæk Laursen
Trond Eliv Hauge

Dansk Clearinghouse - reference number

Month and year of publication: April, 2009

This report shall be cited as

Contact address
Danish Clearinghouse for Educational Research
School of Education
Aarhus University
Tuborgvej 164
DK-2400 Copenhagen NV
Phone: +45 8888 9980
clearinghouse@dpu.dk
www.dpu.dk/clearinghouse
Foreword

This report presents a research mapping and research assessment of empirical research published internationally between 1990 and 2008 on the relationship between factors in primary and lower secondary schools (input and process) and pupils’ learning (output and outcome). The project was commissioned by the Danish Evaluation Institute (Danmarks Evalueringsinstitut) and was performed on behalf of the Nordic Indicator Workgroup (DNI). The project was carried out in the period 1.10.2008-28.04.2009.

Danish Clearinghouse wishes to express its warmest thanks to the Review Group, which not only accepted our invitation to participate in the project, but also – despite large workloads outside the project – devoted additional time and effort at critical moments in order to meet the fixed and rather tight deadlines.

Danish Clearinghouse also wishes to thank the National Library of Education, Denmark for exemplary assistance and for help in obtaining the many documents on which the report is based.

Finally, the Clearinghouse wishes to thank the five Nordic ministries of education (Skolestyrelsen in Denmark, Skolverket in Sweden, Utdanningsdirektoratet in Norway, Utbildningsstyrelsen in Finland and the Ministry for Education, Research, and Culture in Iceland), for setting the task, and especially the excellent working relationship with Special Advisor Signe Ploug Hansen, who acted as contact point to DNI.

This document was completed in April 2009.

Sven Erik Nordenbo

Copenhagen, April 28, 2009
Tables

Table 2.1: Searches performed........................................................................................................ 17
Table 2.2: Summary of the overall screening process ................................................................. 17
Table 3.1: Countries in which the studies took place ................................................................. 25
Table 3.2: Language in research reports....................................................................................... 26
Table 3.3: Educational setting of the studies .............................................................................. 27
Table 3.4: Phenomena/factor in school addressed in the studies .............................................. 27
Table 3.5: Curriculum area of the studies .................................................................................... 28
Table 3.6: Pupil result focus: Specific group of pupils ............................................................... 29
Table 3.7: Pupil result focus: Academic effects ......................................................................... 29
Table 3.8: Pupil result focus: non-academic effects .................................................................. 30
Table 3.9: Purpose of the study ................................................................................................. 30
Table 3.10: Design in studies ..................................................................................................... 31
Table 3.11: Methods applied in data collection in the studies .................................................... 32
Table 3.12: Quality of studies - reporting .................................................................................. 33
Table 3.13: Were users / relatives of users involved in the design or conduct of the study? .... 33
Table 3.14: Was the choice of research design appropriate for addressing the research ... 34
Table 3.15: Have sufficient attempts been made to establish the repeatability or reliability of ... 34
Table 3.16: Have sufficient attempts been made to establish the validity or trustworthiness of ... 34
Table 3.17: Have sufficient attempts been made to establish the repeatability or reliability of ... 35
Table 3.18: Have sufficient attempts been made to establish the validity or trustworthiness of ... 35
Table 3.19: To what extent are the research design and methods employed able to rule out any ... 35
Table 3.20: In light of the above, do the reviewers differ from the authors over the findings or ... 36
Table 3.21: Weight of evidence of the studies .......................................................................... 37
Table 4.1: Studies that have anchorage in theory or at least in a conceptual multi-level model. 41
Table 4.2: Weight of evidence ..................................Error! Bookmark not defined.

Figures

Figure 2.1: School effectiveness and indicators. Conceptually simplified relationship .. 16
Figure 2.2: Filtering of references from search results to mapping......................... 23
1 Introduction

1.1 Background and problem area

This report has been written on the basis of a contract between the Nordic Indicator Workgroup (DNI), represented by the Danish Evaluation Institute (EVA) on the one hand and Danish Clearinghouse for Educational Research.

DNI includes representatives from the school directorates in the five Nordic countries: Skolestyrelsen (Denmark), Skolverket (Sweden), Utdanningsdirektoratet (Norway), Utbildningsstyrelsen (Finland), and Ministeriet for Undervisning, Forskning og Kultur (Iceland).

DNI has been given the task of developing indicator systems for Scandinavian primary and lower secondary schools, related to inputs, ‘good organisations’ and good learning environments (processes) on the one hand, and to learning output on the other.

The research mapping presented in this report is the first phase in the process. The second phase, which lies outside the scope of this report, consists of building on the material collected in the research mapping and establishing evidence-based syntheses that address the relationship between the primary and lower secondary school’s efforts and its pupils’ learning. In continuation of the research mapping it is also our intention - after conclusion of the necessary contracts with the commissioners of the study - to answer the following question:

Which factors contribute most to the learning of primary and lower secondary school pupils?

The third phase, which DNI will be responsible for, will consist of establishing an indicator system. The results of the second phase will provide basic material for this task.

In an overall perspective, therefore, this research mapping is a step towards the realisation of DNI’s long-term goal of providing a set of relevant indicators that can be used for inspection and development etc. within the primary and lower secondary school sector in the Scandinavian countries.

1.2 Aims

On the basis of the previous section, the aim of this research assessment can be summarised in the question:

What empirical research has been carried out to examine the relationship between factors in primary lower secondary schools (inputs and processes) and the learning achieved by primary and lower secondary school pupils (outputs and outcomes)?

The question can be answered as follows:

By performing a systematic research mapping of the empirical research that has been carried out to examine the relationship between factors in primary and lower secondary schools (inputs and processes) and learning achieved by pupils (outputs and outcomes).
1.3 Review group

To carry out the task, Clearinghouse established a review group with the following members:

- Professor Eyvind Elstad, University of Oslo
- Professor Trond Eiliv Hauge, University of Oslo
- Professor Anders Holm, University of Copenhagen
- Professor Per Fibæk Laursen, Aarhus University
- Professor Michael Uljens, Aabo Akademi University
- Professor Jaap Scheerens, University of Twente

The review group participated with Danish Clearinghouse in the data extraction and coding of the research reports covered by this study. The final report was produced by the review group and Danish Clearinghouse in cooperation. There have neither been conflicts of interest for any member of the review group during the data extraction process nor during the preparation of the report. No review group member has participated in the coding of own reports.
2 Methods used in the research mapping

2.1 Design and method

This research mapping has been carried out following a standardised procedure described in Concept Note developed by Danish Clearinghouse for Educational Research (see http://www.dpu.dk/site.aspx?p=9864).

The procedure is described in a protocol established at the start of the project. The procedure is characteristic in utilising transparent and explicit methods in a series of steps. This is explained further in this report and also (briefly) in the Concept Note.

A special software tool was used, developed especially for this type of study: the EPPI-Reviewer. This is explained in more detail on the producer’s website: http://eppi.ioe.ac.uk

Data extraction from relevant and suitably qualified documents was carried out following the methodology and systematics of the EPPI-Reviewer. This procedure was developed by the EPPI-Centre at the Institute of Education, University of London. In this particular research mapping the procedure was adapted to the conceptual universe of the research in question - see Chapter 3.

The research mapping was carried out on the basis of codings and evaluations of the research reports by a review group working together with the staff of Danish Clearinghouse for Educational Research. The studies were characterized and their thematic relationships analysed.

2.2 Conceptual delimitation

The starting point of the research mapping was the review question:

What empirical research has been carried out to examine the relationship between factors in primary and lower secondary schools (inputs and processes) and the learning achieved by pupils (outputs and outcomes)?

The research mapping was intended to uncover factors relevant for pupils’ learning and emerging from a broad interpretation of the concept of ‘the good school’ - including physical layout, ways of teaching, teacher competences, administration etc., thus bringing in all the data about inputs (the factors determined by the school), processes (the school’s activities) and outputs (the pupils’ results), that might be relevant for the development of a reliable instrument for supervision and development etc. within the primary an lower secondary school sector.

The task has been to establish which factors or constellation of factors in the school are the most important for producing the desired results. Since the way in which the various factors interact is also important for the combined effect, we have searched for studies that describe synchronous effects.

This implies that studies on a single feature of the school, for example ‘teacher effectiveness’ or ‘the competence of school leaders’ were not included. Individual factors were included only where they were seen in relationship with other factors in the school, i.e. in a total perspective of the school. The approach adopted for this research mapping has been ‘school effectiveness’.
In this approach the school is seen as an institution, and concepts are employed that make it possible to state which factors in the school lead to effects on the short term (output) and/or on a longer term (outcomes). In this research mapping exercise, ‘the good school’ is therefore regarded as an empirical phenomenon. In other words, ‘the good school’ is a school that has proved that it lives up to certain desirable, explicit criteria, corresponding to those set up by the School Effectiveness Research Movement, a movement internationally anchored in the ‘International Congress for School Effectiveness and Improvement’ (ICSEI).

The following concepts, taken from the ‘ERIC Thesaurus’, will be used:

**School effectiveness**

Degrees to which schools are successful in accomplishing their educational objectives or fulfilling their administrative, instructional, or service functions

**Effective schools research**

Educational research focused on identifying unusually effective schools, studying the underlying attributes of their programs and personnel, and designing techniques to operationalise these attributes in less effective schools (Note: Use as a minor Descriptor for examples of this kind of research - use as a major Descriptor only as the subject of a document).

Research into effective schools is based on a theory that the results achieved by a school are based on (a) the individual abilities of the pupils, (b) the cultural, socio-economic and family background of the pupils and (c) what the pupil experiences at the school.

Effective schools research seeks information about factor (c), and must attempt to control and correct any influences arising from the other two factors. In effective schools research an analytical distinction is sometimes drawn between phenomena at the school level and at the classroom level (Creemers, 2008). The classroom level is admittedly a part of the school, but is only of interest for the current study if it is seen in the context of the school as a whole. For ‘good classrooms’ can also be found in ‘not very good schools’, and vice versa. In this study, the focus is ‘school effectiveness’, not ‘teaching effectiveness’.

The concept of ‘school effectiveness’ only gives meaning in relation to certain criteria that an effective school must meet. The question then is to define these criteria. In research into school effectiveness, these criteria are formulated as the desired effects expressed as ‘outputs’ or ‘outcomes’.

There is an indefinite number of possibilities. For the purposes of this study it has been decided that only effects on pupils have any interest. In the short term such effects might be e.g. the results achieved in specific school subjects, the acquisition of certain generally valued competences, or whether the pupils thrive in the school.

On a longer term, relevant effects might be the various functions or effects of the school seen from a societal viewpoint: economic effects, effects on the cohesiveness of local society, or effects on cultural life in the community. Such effects are not included in this analysis.

Initially it is unlikely to be the same basic factors in all schools that create such a diversity of effects. When the time comes to synthesise, it may be necessary to make addi-
tional conceptual distinctions in this area. In connection with this research mapping exercise, however, it is not necessary to introduce any other delimitation than stipulating that the effects must be relevant to the pupils.

Interest is also restricted to schools that in their nature are similar to the Scandinavian basic schools, i.e. schools internationally characterised as ‘primary and lower secondary schools’. The study only considers normal schools, not special schools or vocational schools.

Most other industrialised countries have school systems that differ organisationally from the Scandinavian system. Most industrialised countries divide their school system into ‘primary school’ and ‘secondary school’. Since this research mapping covers research on schools similar to the Scandinavian basic school, it includes research focusing on ‘primary school’, and research focusing on ‘lower secondary school’.

This study is only interested in schools in societies resembling Scandinavian societies. This means in practice that studies on 3rd world schools are not considered relevant to this study.

‘School’ is generally recognised to be a non-constant phenomenon. Thus, in principle, any school research from any period in time cannot be relevant. However, it can be difficult to stipulate one particular year since which research can be considered to be particularly relevant to the present day. In the first half of the 1990’s, however, the legal basis of many of the Scandinavian basic schools was changed considerably (Tjeldvoll, 1998). This might indicate that 1990 would be a good starting year for this research mapping exercise.

This cut-off year could also be defended from a research methodology viewpoint, since around 1990 school effectiveness research began to utilise a new research design that made research results more reliable. At this time the research tradition began to employ new statistical methods that permitted simultaneous analysis of hierarchical data. This is interesting, because what the pupils experience in the school takes place both a classroom level and at a leadership and organisational level (Willms, 1994; Creemers, 1992).

To this can be added that there are several thorough research reviews that cover research prior to 1990 in a competent manner (Scheerens, 1998; Teddlie, 2000; Townsend, 2007).

As an illustration of the relationship between the conceptual delimitations discussed here we can refer to Figure 2.1:
Figure 2.1: School effectiveness and indicators. Conceptually simplified relationship

The model attempts to indicate that there are at least three basic relationships contributing to what the pupil gets out of the school: (a) the individual abilities of the pupil, (b) the social background - in a broad sense - of the pupil and (c) the character of the school at which the pupil is taught. The present research mapping only looks at the outputs and outcomes that can be ascribed to the contribution of the school itself. This is achieved by correcting as much as possible for factors related to (a) and (b).

Similarly, as already mentioned, the establishment of an indicator system is a separate research task which can be undertaken after the conclusion of this research mapping and any subsequent research synthesis built on the studies identified in this research mapping exercise.

2.3 Searches

Searches were carried out by the Clearinghouse. The review group had the opportunity to discuss and correct both the sources to be searched and the search profiles. Both the search sources and the search profiles were explicitly described in the research mapping protocol set up in the initial phase of the project.

From the start the review group was also permitted to suggest additional references. During the project, six proposals were made by the review group and one by the commissioner of the project.
The core of the research mapping exercise has been ‘the Good School’, i.e. the characteristics of a school that creates the desired effects in its pupils. The special approach to school relationships adopted in school effectiveness research has also been used here.

<table>
<thead>
<tr>
<th>Source</th>
<th>Date of search</th>
<th>Number of hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEI (dialog)</td>
<td>21/11/2008</td>
<td>150</td>
</tr>
<tr>
<td>AEI (Dialog)</td>
<td>24/11/2008</td>
<td>500</td>
</tr>
<tr>
<td>Psychinfo(CSA)</td>
<td>24/11/2008</td>
<td>260</td>
</tr>
<tr>
<td>ERIC(CSA)</td>
<td>21/11/2008</td>
<td>1293</td>
</tr>
<tr>
<td>Evidensbasen</td>
<td>27/11/2008</td>
<td>21</td>
</tr>
<tr>
<td>Sociological abstracts(CSA)</td>
<td>25/11/2008</td>
<td>98</td>
</tr>
<tr>
<td>Fis Bildung</td>
<td>26/11/2008</td>
<td>801</td>
</tr>
<tr>
<td>CBCA Education (Proquest)</td>
<td>26/11/2008</td>
<td>107</td>
</tr>
<tr>
<td>Dansk Pædagogisk Base(DPB)</td>
<td>27/11/2008</td>
<td>29</td>
</tr>
<tr>
<td>forskningsdatabasen.dk</td>
<td>03/12/2008</td>
<td>10</td>
</tr>
<tr>
<td>Libris (Sverige)</td>
<td>27/11/2008</td>
<td>17</td>
</tr>
<tr>
<td>Skolporten.com</td>
<td>27/11/2008</td>
<td>2</td>
</tr>
<tr>
<td>Norbok (Norge)</td>
<td>01/12/2008</td>
<td>12</td>
</tr>
<tr>
<td>Bibsys Forskdom publikasjoner (Norge)</td>
<td>01/12/2008</td>
<td>52</td>
</tr>
<tr>
<td>Jykdok</td>
<td>01/12/2008</td>
<td>6</td>
</tr>
<tr>
<td>Swetswise</td>
<td>01/12/2008</td>
<td>122</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>03/12/2008</td>
<td>153</td>
</tr>
<tr>
<td>References from included studies</td>
<td>Continuous during review process</td>
<td>11</td>
</tr>
<tr>
<td>References from review group</td>
<td>Continuous during review process</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2.1: Searches performed

The professional universe of this review covers didactics and educational research, including more psychologically oriented and more sociologically oriented directions. It was therefore desirable to achieve the same breadth of scope in the sources that were searched and in the search profiles that were employed. The linguistic universe was initially defined as Danish, Swedish, Norwegian, German, French and English. The search process did not specify any restrictions with regard to research methodologies; this aspect was taken into account in the screening process - see section 2.4.

Sources and hits are shown in Error! Reference source not found.. All searches were uploaded in the software EPPI-Reviewer.
2.3.1 Search profiles

The searches covered material published during 1990-2008, as presented below. All search profiles were formed in accordance with the theme of the research mapping, paying particular attention to the subject data systems and professional content of the sources that were searched. All searches were done in November-December 2008.

Searches performed

**BEI (Dialog)**

(‘HIGH SCHOOLS’ OR ‘COMMUNITY SCHOOLS’ OR ‘ELEMENTARY SCHOOLS’ OR “INDEPENDENT SCHOOLS’ OR “MAINTAINED SCHOOLS’ OR “MIDDLE SCHOOLS’ OR “PRIMARY SECONDARY EDUCATION’ OR “SECONDARY EDUCATION’ OR ‘SECONDARY SCHOOLS’ OR ‘ELEMENTARY SCHOOLS’ OR ‘PRIMARY EDUCATION’ OR ‘JUNIOR SCHOOLS’ OR ‘PRIMARY SCHOOLS’) AND

(‘SCHOOL EFFECTIVENESS’) AND:

Year of Publication=('1990' OR.....’2008’)

**AEI (Dialog)**

AEI Subject Headings=(‘SECONDARY EDUCATION’ OR ‘ELEMENTARY SCHOOLS’ OR ‘JUNIOR PRIMARY SCHOOLS’ OR ‘PRIMARY EDUCATION’ OR ‘PRIMARY GRADES’ OR ‘PRIMARY SECONDARY EDUCATION’ OR “CENTRAL SCHOOLS’ OR “LOWER PRIMARY YEARS’ OR “MIDDLE PRIMARY YEARS’ OR “PRIMARY SCHOOLS’ OR “UPPER PRIMARY YEARS’ OR “YEAR 1’ OR “YEAR 2’ OR “YEAR 3’ OR “YEAR 4’ OR “YEAR 5’ OR “YEAR 6’ OR “YEAR 7’ OR “YEAR 8’ OR “YEAR 9’ OR “YEAR 10’ OR “HIGH SCHOOLS’ OR “SECONDARY SCHOOLS’ OR LOWER SECONDARY YEARS’ OR ELEMENTARY SCHOOLS’) AND


AEI subjects headings=(“SCHOOL EFFECTIVENESS’ OR “EFFECTIVE SCHOOLS PROJECTS’ OR “EFFECTIVE SCHOOL RESEARCH’)

**Psychinfo (CSA)**

(DE=(‘elementary schools’ or ‘high schools’ or ‘junior high schools’ or ‘middle schools’)) and (“effective* school*’ or “school* effective*’)

Limited to: Publication Year: 1990 -2008

**ERIC (CSA)**

((DE=‘effective schools research’) or (DE=‘school effectiveness’)) AND (PT=(142 reports: evaluative) or PT=(143 reports: research))

Limited to:

Publication year 1990-2008

And

Limited to:

Education level:

Elementary education or elementary secondary education or grade1 or grade 2 or grade 3 or grade 4 or grade 5 or grade 6 or grade 7 or grade 8 or grade 9 or grade 10 or high
schools or intermediate grades or junior high schools or middle schools or primary education or secondary education

**Evidensbasen**

Dk=37.3? and (ti=school? Eller ti=skol?)

**Sociological abstracts (CSA)**

Sociological abstracts searched 2008-11-25

(DE=(‘schools’ or ‘elementary schools’ or ‘private schools’ or ‘public schools’ or ‘secondary schools’) and((DE=‘effectiveness’) or(‘effective* school*’ or ‘school* effective*’))

**FIS-Bildung**

(Titelsuche: schul* ODER school) UND
(Slagwörter suche: Effizienz ODER effektivitaet) UND
(Jahr:>=1990)

**CBCA education (Proquest)**

Effective* W/2 school*

Limited to 1990-2008

Limited to scholarly journals

**Dansk pædagogisk base**

DK=37.3? and (skoleeffektivitet eller effektiv? eller ‘god skole’) and år=1990 til 2008

**Forskningsdatabasen.dk**


**Libris (Svensk bogfortegnelse)**


**Skolporten.com**

Under 'Forskning & utvickling'

Under 'Avhandlingar'

Browsing of all titles

**Norbok**

(DEWEY SØK: 3?0 OR 37? OR 370.193?) AND
(ORDSØK: bra OR god? OR effektiv?) AND
(ORDSØK: skol?) AND
Publication Year: 1990 - 2008
2.4 Screening

The searches were performed in such a way as to ensure that all relevant material would be found. However, not all that is found may be relevant to the study. All 3651 hits were therefore screened, and sorted according to their relevance.

The screening gave no weighting to research quality or the quality of the way in which the study was carried out and reported. Attention was given solely to whether the material belonged in the conceptual universe described above in section 2.2.

The screening process also looked at whether the reference reported primary research. Popular presentations, secondary research reporting and discussions of scientific methodology etc. were not included.

Prior to the screening process all duplicates were eliminated. As a natural consequence of the search process, duplicates must be expected to occur. 165 duplicates were removed. After this, the screening was carried out as a two-phased process:
2.4.1 Phase 1: Screening of references

After removal of duplicates, all the hits uploaded to EPPI-Reviewer were sorted into the 11 categories presented in Table 2.2. All references for which the information was deemed inadequate were regularly subjected to additional searches in order to supplement with an abstract or other additional information. This lack of information applied in particular to Scandinavian references.

This phase included everything that could not be excluded with confidence. Both ‘certain’ and ‘uncertain’ references were thus included at this stage. Only references with a high degree of certainty were excluded.

Exclusion was hierarchical, such that exclusion took place firstly on the grounds of ‘wrong scope’, then of ‘wrong paper’, then of ‘wrong research’ … etc. Since the exclusion criterion ‘wrong research design’ was deemed impossible to apply with certainty in the screening of references, this category was only introduced in the next phase of the screening process.

After the first screening phase there remained 353 references.

2.4.2 Phase 2: Full text screening

In Phase 2 the books, articles or reports that were the subject of all the remaining references were obtained and they were then screened on the basis of the full text.

The screening was carried out using the same criteria as in Phase 1 with the addition of the exclusion criterion ‘wrong research design’. This criterion was included so as to ensure that the included studies did in fact ascribe actual positive effects to the school on the basis of some form of control.

It is important to emphasise in connection with the screening process that reports from evaluations or innovative school experiments were not excluded solely on the grounds that they report evaluations or school experiments.

It is important to remember as a general point that research quality or reporting quality was not used as a basis for inclusion/exclusion.

2.5 Coding and data extraction

The EPPI-Centre at the Institute of Education, London University, was established in 1996. It has created a generalised coding and data extraction system for educational research. This is known as the EPPI-Centre data extraction and coding tool for education studies V2.0. This system has been used in a shortened and edited form for all coding and data extraction in this study. It is presented as Appendix 1, and in Chapter 3. The coding and data extraction system is an integrated part of the EPPI-reviewer.

The EPPI-reviewer was used to make a coding and data extraction of all the included in the study. A prerequisite for creating an overview or synthesis covering all the documents is that they are described using the same system. The principle of tertio comparisonis is employed here. That is to say, a comparison between two elements is made possible by introducing and comparing them with a third (common) element.

Coding and data extraction consists of answering questions about the studies in such a way that relevant data is drawn out for use in the comparison. The system is built up in
sections which are subdivided into questions which in turn are subdivided into multiple choice answers. At all points it is possible to insert notes and explanatory remarks linked to the selected multiple choice answer. In terms of content, the system covers the purpose of the study, its focus with respect to policy and practice, the factors investigated in the school, the focus on pupil performance, sampling considerations, results and conclusions, design and method, quality of research and reporting. The original EPPI questions have been modified considerably, as indicated in Appendix 1, in the light of the actual theme of this review.

Coding and data extraction was performed by the members of the review group in such a way that individual members were responsible for specific studies. The studies were also distributed to the scientific assistants at the Clearinghouse, who also were given responsibility for specific studies. The peer review principle was then applied systematically, and every study was examined by at least two people.

Special focus was given to ensuring the quality of the evaluation of the weight of evidence, which forms part of the coding and data extraction.

In this connection a procedure was employed to permit establishment of an ‘agreed version’: if there were differing opinions as to the evaluation of the four questions in the section concerning weight of evidence (cf. Appendix 1, Section N, Question 11-14), a dialogue took place between the member of the review group and the staff member of the Clearinghouse, in which explicit arguments for the differences were exchanged with a view to establishing agreement. If agreement could not be reached in this way, a third party was assigned the task of establishing an ‘agreed version’ on the basis of the presented arguments.

In this review differences were originally noted in connection with 105 out of 444 individual evaluations of weight of evidence (24 %). The disagreements applied to 57 of a total number of 116 studies (51 %). In connection with this review it was not necessary to employ the services of a third party in any single case.

An example of a complete coding and data extraction for one document is presented in chapter 5, Appendix 1.

The work of coding and data extraction provided the basis on which the research mapping could be carried out. The research mapping was performed using the analysis and reporting facilities available in the EPPI-Reviewer.

### 2.6 Summary of the review process

Figure 2.2 presents in graphic form the process from search to research mapping. The figure also indicates that a research synthesis can potentially be performed starting from the research mapping that has been carried out.
Figure 2.2: Filtering of references from search results to mapping
3 Research mapping and research assessment

This chapter gives a general description of all the 111 studies included in the survey. The studies are described cross sectionally and are evaluated in the light of the research assessment, so as to create a combined picture of current research, its character and quality.

First we examine the context of the studies: where and in what types of schools were they carried out? Next, we look at the content: what factors in the schools have been studied? Which subject areas were covered? Which pupils and what effects on pupils were looked at? Subsequently we look at the aims of the studies and their design and methodology. The chapter concludes with an analysis of the quality of the studies.

3.1 The context of the studies

School effectiveness research is an international research effort, and this manifests itself in the material in two different ways. Firstly, the 111 studies draw their data from a total of 38 different industrialised countries. Secondly, some of the investigations were in fact comparative educational studies that used data from a number of countries in one and the same research process.

<table>
<thead>
<tr>
<th>Countries</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>72</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>13</td>
</tr>
<tr>
<td>Netherlands</td>
<td>12</td>
</tr>
<tr>
<td>Australia</td>
<td>10</td>
</tr>
<tr>
<td>Belgium</td>
<td>8</td>
</tr>
<tr>
<td>France, Germany, Ireland, Canada</td>
<td>5 (from each country)</td>
</tr>
<tr>
<td>Cyprus, Norway, Hong Kong, Switzerland, Spain</td>
<td>4 (from each country)</td>
</tr>
<tr>
<td>Korea, Denmark, Greece, Hungary, New Zealand, Sweden, Portugal, Singapore</td>
<td>3 (from each country)</td>
</tr>
<tr>
<td>Austria, Czech Republic, Finland, Italy, Japan, Iceland, Thailand, Slovenia</td>
<td>2 (from each country)</td>
</tr>
<tr>
<td>Luxembourg, Taiwan, Slovak Republic, Russian Federation, Romania, Lithuania, Poland, Latvia</td>
<td>1 (from each country)</td>
</tr>
</tbody>
</table>

Table 3.1: Countries in which the studies took place
(N=111 studies; several categories permitted per study)

Table 3.1 shows the distribution of the studies amongst the various countries. It will be seen that 65 % of all the studies involve data from the USA. UK, Holland, Australia and Belgium account for 12 %, 11 %, 9 % and 7 % of the studies respectively.
It will also be seen from the table that some studies originate from Scandinavia: Norway (4 studies), Denmark and Sweden (3 studies each) and Finland and Iceland (2 studies each).

Because of the comparative studies there are more than 111 countries involved in the studies.

The reports of the studies are almost all in English (96 % of all studies), as seen in Table 3.2. Even though the searches covered a much larger linguistic universe, only a few non-English studies were found and included.

<table>
<thead>
<tr>
<th>Language</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>107</td>
</tr>
<tr>
<td>German</td>
<td>3</td>
</tr>
<tr>
<td>Danish</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.2: Language in research reports (N=111 studies)

As already described in Chapter 2, studies that investigated schools of types not corresponding to the Scandinavian ‘basic school’ were excluded. However, this still permitted several possible types of school in the studies. The distribution amongst school types is shown in Table 3.3.

This table shows that 43 % of the studies were made in schools covering the first six school years only (‘primary school’). 37 % of the studies concerned schools covering 7th to 12th grade (‘secondary school’). 28 % of the studies were made in schools corresponding exactly to the Scandinavian type (‘Primary and lower secondary’). 17 % of the studies took place in schools comprising grades 7 to 9 or 10 (‘lower secondary school’).

A number of studies examine both ‘primary school’ and ‘secondary school’. This means that the number of school types indicated in the table is greater than the number of studies.
<table>
<thead>
<tr>
<th>School</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower secondary school</td>
<td>19</td>
</tr>
<tr>
<td>Primary and lower secondary school</td>
<td>31</td>
</tr>
<tr>
<td>Primary school</td>
<td>48</td>
</tr>
<tr>
<td>Secondary school</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 3.3: Educational setting of the studies  
(N=111 studies; several categories permitted per study)

3.2 **School and pupil factors studied**

The inclusion criteria for this systematic review included the requirement that a given study had to examine at least two different school factors in order to be included. The range of actually studied school factors is shown in Table 3.4.

<table>
<thead>
<tr>
<th>School factor/phenomena</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class size</td>
<td>11</td>
</tr>
<tr>
<td>Curriculum/scheduling</td>
<td>41</td>
</tr>
<tr>
<td>Ethnic composition of the pupils in the schools</td>
<td>31</td>
</tr>
<tr>
<td>Leadership</td>
<td>55</td>
</tr>
<tr>
<td>Management</td>
<td>37</td>
</tr>
<tr>
<td>Other</td>
<td>43</td>
</tr>
<tr>
<td>Physical environment</td>
<td>8</td>
</tr>
<tr>
<td>School culture</td>
<td>68</td>
</tr>
<tr>
<td>School size</td>
<td>20</td>
</tr>
<tr>
<td>Socio-economic composition of the pupils in the schools</td>
<td>71</td>
</tr>
<tr>
<td>Staff development</td>
<td>17</td>
</tr>
<tr>
<td>Support systems</td>
<td>15</td>
</tr>
<tr>
<td>Teacher</td>
<td>64</td>
</tr>
<tr>
<td>Teacher teams</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 3.4: Phenomena/factor in school addressed in the studies  
(N=111 studies)

Here the full breadth of the studies becomes obvious: there are studies covering every one of the previously defined categories of phenomena and factors in the school. The most frequently investigated factors are the socio-economic composition of the pupils at the school, the school culture, the teacher and leadership. These are covered by 64 %, 61 %, 58 % and 50 % of the studies respectively. Factors such as physical school environ-
ment, class size, support systems, staff development, school size and teacher teams are not so frequently included in the studies, being covered by 7 %, 10 %, 14 %, 15 %, 18 % and 19 % of the studies respectively.

Even though the studies address school effectiveness as such, this may often be combined with other (school) subjects, either by viewing the school's effectiveness in relation to a subject success criterion such as performance in mathematics, or in the form of a study of special professional aspects of the school's activities, such as how reading is taught. This is shown in Table 3.5. It is seen here that 62 % of all studies have an inbuilt mathematical aspect, while 51 % look at literacy in the mother tongue. 24 % of the studies make no reference to specific factors in the curriculum.

<table>
<thead>
<tr>
<th>Curriculum area</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-curricular</td>
<td>2</td>
</tr>
<tr>
<td>Environment</td>
<td>3</td>
</tr>
<tr>
<td>General</td>
<td>2</td>
</tr>
<tr>
<td>Geography</td>
<td>2</td>
</tr>
<tr>
<td>Hidden</td>
<td>1</td>
</tr>
<tr>
<td>History</td>
<td>4</td>
</tr>
<tr>
<td>Literacy - first languages</td>
<td>57</td>
</tr>
<tr>
<td>Literacy - further languages</td>
<td>6</td>
</tr>
<tr>
<td>Literature</td>
<td>8</td>
</tr>
<tr>
<td>Maths</td>
<td>69</td>
</tr>
<tr>
<td>N/A (not on a specific curriculum area)</td>
<td>27</td>
</tr>
<tr>
<td>Other</td>
<td>9</td>
</tr>
<tr>
<td>Phys. Ed</td>
<td>1</td>
</tr>
<tr>
<td>Religious Ed.</td>
<td>1</td>
</tr>
<tr>
<td>Science</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 3.5: Curriculum area of the studies  
(N=111 studies; several categories permitted per study)

The studies can also be sorted by asking whether they examine the effect of the school on the pupils in general, or its effect on specific groups of pupils. This is shown in Table 3.6. 46 % of the studies examine the effects on pupils in general, 39 % and 23 % examine effects on pupils with low socio-economic status and pupils from ethnic groups respectively. On the other hand, there are very few studies looking at gender differences, differences in competence, and handicaps. Several studies include a number of different pupil groups. As a result, the number of studies listed under the various pupil groups is greater than the total number of studies.
### Table 3.6: Pupil result focus: Specific group of pupils  
(N=111 studies; several categories permitted per study)

<table>
<thead>
<tr>
<th>Group of pupils</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No specific group</td>
<td>51</td>
</tr>
<tr>
<td>Pupils with high competence</td>
<td>5</td>
</tr>
<tr>
<td>Pupils with low competence or handicaps</td>
<td>3</td>
</tr>
<tr>
<td>Yes girls</td>
<td>5</td>
</tr>
<tr>
<td>Yes, boys</td>
<td>6</td>
</tr>
<tr>
<td>Yes, other specific groups</td>
<td>12</td>
</tr>
<tr>
<td>Yes, pupils from ethnic groups</td>
<td>26</td>
</tr>
<tr>
<td>Yes, pupils with low SES</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 3.7 shows the distribution of studies sorted by the effect on the pupils, taken in a narrow academic context. Only 3 % of the studies made no reference at all to this aspect. 96 % include this focus and interpret it as pupil performance, usually measured by achievement or examination performance. 5 % look at the academic effect in other ways, for example as a successful transition to the next stage in the educational system. 4 of the latter studies also include performance measurements. As a result, the combined number of studies listed in the various categories is greater than the total number of studies.

### Table 3.7: Pupil result focus: Academic effects  
(N=111 studies; several categories permitted per study)

<table>
<thead>
<tr>
<th>Focus on academic effects</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without such focus</td>
<td>3</td>
</tr>
<tr>
<td>Yes, achievement or performance</td>
<td>107</td>
</tr>
<tr>
<td>Other academic effects</td>
<td>5</td>
</tr>
</tbody>
</table>

A number of the studies include an evaluation of effects on pupils apart from the academic effect. This includes topics such as the pupils’ wellbeing. Table 3.8 shows the distribution: 70 % of the studies did not include such effects, whilst 30 % of the studies did. A number of studies examined a number of non-academic effects on the pupils. For this reason the combined number of studies listed by effects examined is slightly greater than the total number of studies.
### Purpose, design and methodology of the studies

The following section gives a short description of the studies seen from a research viewpoint.

The aims of the various studies are listed in Table 3.9. Here it is seen that several studies have more than one purpose. For this reason, the combined number of studies listed by purpose is greater than the total number of studies. The table also shows that explorations of relationships and description are the most frequent purposes, covering 56% and 51% of the studies respectively. The aim 'what works' appears in only 10% of the studies. Methods development, here understood as research methodological development, is an aim in only 8% of the studies.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>57</td>
</tr>
<tr>
<td>Exploration of relationships</td>
<td>62</td>
</tr>
<tr>
<td>What works?</td>
<td>11</td>
</tr>
<tr>
<td>Methods development</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3.9: Purpose of the study  
(N=111 studies; several categories permitted per study)

When choosing the appropriate design for a study, this question is usually linked to the purpose of the study. The designs that were actually used in the studies are listed in Table 3.10. Here again, a number of studies can be assigned to more than one category, and therefore the combined number of studies listed by design is greater than the total number of studies.

The primary impression is one of considerable breadth in choice of design. The most frequently used designs are secondary data analysis (46%), cross-sectional studies (39%) and studies of views (34%). There is also a considerable number of studies using a case study design (28%) and cohort design (27%). Studies using an experimental or adapted experimental design are rare.
The data collection procedures in the studies were also diverse, as shown in Table 3.11. Many of the studies employed several data collection methodologies. Thus the combined number of studies listed by data collection method is greater than the total number of studies. The most frequently employed methods for data collection were: self-completion questionnaires (68%), use of secondary, already existing data (41%), and curriculum-based assessment or measurement (40%). One-to-one interviews (36%) and observation (31%) were also employed in a considerable number of studies.
<table>
<thead>
<tr>
<th>Data collection method</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum-based assessment</td>
<td>44</td>
</tr>
<tr>
<td>Examinations</td>
<td>9</td>
</tr>
<tr>
<td>Focus group interview</td>
<td>18</td>
</tr>
<tr>
<td>Not stated/ unclear</td>
<td>1</td>
</tr>
<tr>
<td>Observation</td>
<td>34</td>
</tr>
<tr>
<td>One-to-one interview (face to face or by phone)</td>
<td>40</td>
</tr>
<tr>
<td>Other documentation</td>
<td>18</td>
</tr>
<tr>
<td>Please specify any other important features of data collection</td>
<td>9</td>
</tr>
<tr>
<td>Practical test</td>
<td>1</td>
</tr>
<tr>
<td>Psychological test (e.g. I.Q test)</td>
<td>7</td>
</tr>
<tr>
<td>School/ college records (e.g. attendance records etc)</td>
<td>21</td>
</tr>
<tr>
<td>Secondary data such as publicly available statistics</td>
<td>45</td>
</tr>
<tr>
<td>Self-completion questionnaire</td>
<td>76</td>
</tr>
<tr>
<td>self-completion report or diary</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3.11: Methods applied in data collection in the studies
(N=111 studies)

3.4 Quality of studies

A quality assessment of research is a necessary step in the process of establishing an overview of what the research actually shows. Only studies carried out to a sufficiently high standard can be viewed with confidence. For this reason, all the studies included in this mapping have been assessed in relation to a broad range of questions concerning their quality, cf. Chapter 5, Appendix 1. For each individual study an assessment is made of the evidence that the study can provide. In section 2.5, a description was given as to how peer review was employed in the assessment process, with at least two different persons responsible for each assessment.

Table 3.12 displays how a number of relevant factors were evaluated concerning the adequacy of the description of the study that was available in the report. Here we see that the fewest problems were encountered with regard to the description of the context and aims (12% and 19% respectively). The greatest number of problems were encountered in the descriptions of avoidance of selective reporting bias (39%). In addition, 37% of the studies would not be replicable on the basis of the description in the report. Since the material includes a considerable number of qualitative studies this situation is not especially remarkable.
<table>
<thead>
<tr>
<th>Wording of question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the context of the study adequately described?</td>
<td>98</td>
<td>13</td>
</tr>
<tr>
<td>Are the aims of the study clearly reported?</td>
<td>90</td>
<td>21</td>
</tr>
<tr>
<td>Is there an adequate description of the sample used in the study and how the sample was identified and recruited?</td>
<td>81</td>
<td>30</td>
</tr>
<tr>
<td>Is there an adequate description of the methods used in the study to collect data?</td>
<td>87</td>
<td>24</td>
</tr>
<tr>
<td>Is there an adequate description of the methods of data analysis?</td>
<td>82</td>
<td>29</td>
</tr>
<tr>
<td>Is the study replicable from this report?</td>
<td>70</td>
<td>41</td>
</tr>
<tr>
<td>Do the authors avoid selective reporting bias? (e.g. do they report on all variables they aimed to study, as specified in their aims/research questions?)</td>
<td>68</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 3.12: Quality of studies – reporting  
(N=111 studies; several categories permitted per study)

Table 3.13 to Table 3.20 indicate the distributions of answers to a number of core assessments of the quality of the individual studies. These assessments, together with the assessments in Table 3.12, serve as the basis for the weight of evidence assigned to the individual studies.

Only a minority of the studies indicate problems of a research ethical nature concerning the involvement of participants or relatives of participants. Table 3.13 shows that this was only a problem in 8 (7 %) of the studies.

<table>
<thead>
<tr>
<th>Answer</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No, but involvement would be desirable</td>
<td>2</td>
</tr>
<tr>
<td>No, involvement is not relevant</td>
<td>71</td>
</tr>
<tr>
<td>Yes, however users/relatives are not appropriately involved</td>
<td>6</td>
</tr>
<tr>
<td>Yes, users/relatives are appropriately involved</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 3.13: Were users / relatives of users involved in the design or conduct of the study?  
(N=111 studies)

Moving to the choice of research design in the individual studies, it is our opinion, cf. Table 3.14, that this was only completely satisfactory in 54 % of the studies. In the remaining studies, there were considered to be major or minor problems in the design employed. Here it must also be recalled that studies that do not control for obvious alternative reasons than the school for the success of the pupil have not been included. This question is addressed in section 2.2.
An evaluation of the attempts made by the studies to establish reliability and repeatability of data collection is presented in Table 3.15. Here, 79% of the studies have made a good attempt or at least some form of attempt, while 21% have made no attempt to ensure reliability and repeatability.

<table>
<thead>
<tr>
<th>Answer</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No, none</td>
<td>23</td>
</tr>
<tr>
<td>Yes, good</td>
<td>47</td>
</tr>
<tr>
<td>Yes, some attempt</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 3.15: Have sufficient attempts been made to establish the repeatability or reliability of data collection methods or tools? (N=111 studies)

The attempts made in the studies to ensure the validity of the data collection procedures are analysed in Table 3.16. 77% have made a good attempt or some form of attempt, while 23% of the studies have made no attempt to ensure the validity of their data collection procedures.

<table>
<thead>
<tr>
<th>Answer</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No, none</td>
<td>26</td>
</tr>
<tr>
<td>Yes, good</td>
<td>45</td>
</tr>
<tr>
<td>Yes, some attempt</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 3.16: Have sufficient attempts been made to establish the validity or trustworthiness of data collection tools and methods? (N=111 studies)

The reliability and repeatability of the data analysis has been adequately established in 60% of the studies. Major or minor problems of data analysis were noted in 40% of the studies, cf. Table 3.17.
Table 3.17: Have sufficient attempts been made to establish the repeatability or reliability of data analysis? (N=111 studies)

<table>
<thead>
<tr>
<th>Answer</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>44</td>
</tr>
<tr>
<td>Yes</td>
<td>67</td>
</tr>
</tbody>
</table>

Table 3.18 shows that 73% of the studies have made good or adequate attempts to ensure the validity of their data analysis. 27% of the studies made no apparent attempt to do this.

Table 3.18: Have sufficient attempts been made to establish the validity or trustworthiness of data analysis? (N=111 studies)

<table>
<thead>
<tr>
<th>Answer</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No, none</td>
<td>30</td>
</tr>
<tr>
<td>Yes, good</td>
<td>42</td>
</tr>
<tr>
<td>Yes, some attempt</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 3.19 indicates whether the chosen design and methods have been capable of ruling out other explanations than the one arrived at in the study itself. 24% of the studies were found to be designed in such a way that they could rule out alternative explanations to a great extent. 48% of the studies were found to be designed so that they could rule out alternative explanations to a limited extent. 28% of the studies were not capable of ruling out alternative explanations at all.

Table 3.19: To what extent are the research design and methods employed able to rule out any other sources of error/bias which would lead to alternative explanations for the findings of the study? (N=111 studies)

<table>
<thead>
<tr>
<th>Answer</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A little</td>
<td>53</td>
</tr>
<tr>
<td>A lot</td>
<td>27</td>
</tr>
<tr>
<td>Not at all</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 3.20 examines whether the authors of this review arrived at different findings and conclusions from the authors of the studies in questions. This was found to be the case in 33% of the studies, while for 67% of the studies the reviewers were in agreement with the authors.
<table>
<thead>
<tr>
<th>Answer</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not applicable (no difference in conclusions)</td>
<td>74</td>
</tr>
<tr>
<td>Yes</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 3.20: In light of the above, do the reviewers differ from the authors over the findings or conclusions of the study? (N=111 studies)

The combined assessment of the contributions of the individual studies to the weight of evidence is shown in Table 3.21.

Weight of Evidence A indicates whether the individual study has been carried out in good agreement with its own declared aims, design, methods and results. It is a combined result based on how the study has been scored in all the assessments presented in Table 3.12 to Table 3.20. The distribution turns out to be fairly even: 33 % of the studies have a high weight of evidence, 38 % have a medium weight of evidence and 28 % have a low weight of evidence.

Weight of evidence B indicates whether the design employed by the individual study has been appropriate for providing an answer to the review question on which this mapping is based. Here the studies are distributed with 27 % in the high category, 43 % in the medium category and 30 % in the low category.

Every study has its own focus and its own way of viewing phenomena and their context. Weight of evidence C addresses the relevance of each study’s focus with respect to the review question of this research assessment. Here 41 % of the studies are found to have a high weight of evidence and 59 % a medium weight of evidence. Low weight of evidence is not included as an optional response category, since studies with such low weight of evidence were removed in the screening process, cf. section 2.4.

The position of the studies with respect to weight of evidence D, the combined weight of evidence, decides whether the studies should be included in a research synthesis covering the results emerging from the research within this field. Studies with a high weight of evidence D (15 %) and medium weight of evidence D (50 %) qualify for inclusion in the synthesis. Studies with low weight of evidence D (35 %) should not be included.
<table>
<thead>
<tr>
<th>Question</th>
<th>Antal studier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight of evidence A: Taking account of all quality assessment issues, can the study findings be trusted in answering the study question(s)?</td>
<td>37 43 31</td>
</tr>
<tr>
<td>Weight of evidence B: Appropriateness of research design and analysis for addressing the question, or sub-questions, of this specific systematic review.</td>
<td>30 48 33</td>
</tr>
<tr>
<td>Weight of evidence C: Relevance of particular focus of the study (including conceptual focus, context, sample and measures) for addressing the question, or sub-questions, of this specific systematic review</td>
<td>46 65 -</td>
</tr>
<tr>
<td>Weight of evidence D: Overall weight of evidence</td>
<td>17 56 38</td>
</tr>
</tbody>
</table>

Table 3.21: Weight of evidence of the studies (N=111 studies)
4 General characterisation of the studies analysed

This chapter presents a selection of the most important features emerging from the research review, based on those studies ranked as having high, medium and low research quality in sections 4.1, 4.3 and 4.4, and those studies ranked as having high and medium in section 4.2. This means that whereas chapter 3 presented a mapping and evaluation of all the 111 studies covered by the review, that section 4.2 is based on only 73 of the 111 studies, cf. Table 3.21.

4.1 Anchorage of the Studies in Theory

As a whole the body of studies that was reviewed shows a relatively limited coverage of theoretical issues. More than once the field of school effectiveness research has been accused of being empiricist, paying little attention to the theoretical foundations. In order to document this overall impression, all studies were categorised on criteria relevant for theoretical anchorage. A more in-depth analysis of the potential role of theory in research concerning “the good school” can be given in the second phase of the study. In this section we will just give a precursory introduction to the categories that we use, in order to document the degree to which studies are embedded in theories, models and conceptual analyses.

4.1.1 Categories indicative of the theoretical and conceptual scope of the studies

Review of the school effectiveness research literature
Most of the studies contain a more or less extensive review of earlier school effectiveness research and school effectiveness review studies. In itself this can hardly be taken as an indicator of theoretical anchorage, but it shows at least that a study is placed in a certain research tradition and could in this way be better positioned to yield knowledge accumulation.

In depth review of core factors
The factors that are studied in school effectiveness research are sometimes rather broad concepts like leadership and school climate. Some studies provide more in-depth conceptual analysis and definition of one or two factors that are assigned a central position. One could say that such conceptual analyses provide building blocks for further theoretical development.

Dealing with foundational issues in school effectiveness research
Foundational issues of school effectiveness research deal with the stability, scope and conceptual integrity of the overall concept of school effectiveness. Foundational questions are: whether a school that is effective in year 1, is still effective in year 1 +x (stability); whether a school that is effective in the final grade is also effective in the middle and early grades (scope); whether a school that is effective in one subject area, or outcome dimension, is also effective on other outcome dimensions (consistency); whether a school that is generally effective for low SES students is also effective for high SES students, and vice versa (differential effectiveness); whether malleable school factors interact with composition effects; the degree to which school effects can be ex-
plained by classroom effects etc. Foundational issues concern the demarcation of re-
search into “good schooling” as a coherent research program.

**Use of conceptual and path-analytical models**

Quite a few studies arrange variables according to a multi-level input, process, output
and context model. Sometimes this is a mere ordering of variables, in other cases such
models are actually tested by means of path-analytical and multi-level statistical mod-
els. Such models can be the basis of a more elaborate causal ordering of factors as pri-
mary causal, intermediary or intervening variables.

**Established theory as a basis for developing and interpreting research findings
about school effectiveness**

After the question “what works”, comes the question “why does it work”? For the latter
question one could try and connect to more established theory, either instruc-
tional/learning theory or management theory. The issue is to explain findings and con-
struct hypotheses on the basis of more established theoretical principles. As we shall
see, cf. Table 4.1, this was done relatively rarely in the reviewed studies; put in a dif-
ferent way, relatively few studies could be seen as “theory driven”.

**Conclusion**

Less than a quarter of the studies appear to have some anchorage in theory. Of those,
only a few can be seen as theory driven, in the sense that a conceptual model or theory
is formulated and then tested - in the better examples, on the basis of explicit hypothe-
ses. The best examples of these latter studies are: Campbell et al. (2000), Griffith
(2003), and Opdenakker et al., (2007).

The kinds of models and theories used in the sub-set of studies mentioned can be sum-
marized in three groups:

- **Organization theoretical frameworks.** Examples are models by Bolman and Deal,
  Quinn and Rohrbauch, Hoy and Miskel and Parsons. These models emphasize ma-
jor organizational dimensions such as: having a human relations emphasis in
school policy, being responsive to parents and other external constituencies, be-
ing goal oriented. The theory of open systems was also used in one study as well
as Parson’s social system theory. All these theories state what they understand as
the basic components of organizational functions, which when optimized are ex-
pected to raise organizational performance.

- **Multi-level models that define some more general dimensions indicative of effec-
tive school performance.** The most elaborate of these is the model developed by
Creemers, and further elaborated by Kyriakides and Creemers (2008). Similar
models are the ones by Teddlie and Slater (Teddlie and Stringfield, 1993) and
Scheerens (Young, Fraser et al., 1992). Core dimensions that are expected to op-
erate at different levels are: opportunity, time and quality of structures and proc-
esses. Consensus, continuity and collaboration are additional underlying dimen-
sions, sometimes used in these models.

- **Path-analytical models that, in their more advanced forms, have also a multi-
level structure.** Such models make distinctions between contextual and direct and
indirect malleable conditions of schooling. The set of studies by Van Damme et al.
(2002) (including Opdenakker) are a pronounced example of studies guided by such elaborate statistical models.

<table>
<thead>
<tr>
<th>N</th>
<th>Reference</th>
<th>Theory/model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Booker et al., 2007</td>
<td>Bronfenbrenner, ecological</td>
</tr>
<tr>
<td>2</td>
<td>Stringfield et al., 2008</td>
<td>High Reliability org.</td>
</tr>
<tr>
<td>3</td>
<td>Calaff, 2008</td>
<td>Multiple Worlds</td>
</tr>
<tr>
<td>4</td>
<td>Campbell et al., 2000</td>
<td>Creemers, Caroll model</td>
</tr>
<tr>
<td>5</td>
<td>Coates, 2003</td>
<td>Micro economic theory</td>
</tr>
<tr>
<td>6</td>
<td>Coco et al., 2004</td>
<td>Social constructivism, activity theory</td>
</tr>
<tr>
<td>7</td>
<td>Griffith, 2003</td>
<td>Quinn &amp; Rohrbauch model</td>
</tr>
<tr>
<td>8</td>
<td>Hoy et al., 1990</td>
<td>Coleman’s functional community theory</td>
</tr>
<tr>
<td>9</td>
<td>Kyriakides &amp; Creemers, 2008</td>
<td>Conceptual measurement model</td>
</tr>
<tr>
<td>10</td>
<td>Kyriakides &amp; Tsangaridou, 2008</td>
<td>Creemers, Caroll model</td>
</tr>
<tr>
<td>11</td>
<td>Opdenakker et al., 2007</td>
<td>Model driven study</td>
</tr>
<tr>
<td>12</td>
<td>Tarter, et al., 2004</td>
<td>Bolman &amp; Deal, Hoy and Miskell models</td>
</tr>
<tr>
<td>13</td>
<td>Teddlie &amp; Stringfield, 1993</td>
<td>T. &amp; Slater’s typology of school effectiveness &amp; leadership</td>
</tr>
<tr>
<td>14</td>
<td>Reezigt et al., 1999</td>
<td>Creemers, Caroll model</td>
</tr>
<tr>
<td>15</td>
<td>Ross &amp; Gray, 2006</td>
<td>Indirect effect model of leadership</td>
</tr>
<tr>
<td>16</td>
<td>Senkbeil, 2006</td>
<td>Path model</td>
</tr>
<tr>
<td>17</td>
<td>Silins and Mulford, 2004</td>
<td>Indirect effects model of leadership</td>
</tr>
<tr>
<td>18</td>
<td>Sweetland &amp; Hoy, 2000</td>
<td>Path Model</td>
</tr>
<tr>
<td>19</td>
<td>Van Damme et al., 2002</td>
<td>Implicit multi level model</td>
</tr>
<tr>
<td>20</td>
<td>Van der Werf, 1997</td>
<td>Creemers, Caroll, Bloom</td>
</tr>
<tr>
<td>21</td>
<td>Webster &amp; Fisher, 2003</td>
<td>Conceptual and path models</td>
</tr>
<tr>
<td>22</td>
<td>Young &amp; Fraser, 1992</td>
<td>Scheerens’ integrative model</td>
</tr>
</tbody>
</table>

Table 4.1: Studies that have anchorage in theory or at least in a conceptual multi-level model (N=111)

More detailed analyses concerning school effectiveness theory are considered very important in getting a more balanced idea of the potential of this field of study as an advancing research program. Such analyses could form part of the second phase of this project.
4.2  **Malleable school factors addressed in the study**

The following description of the results of the studies concentrates on those factors in the schools which — in contrast to social background, for example — can be influenced and altered (‘malleable factors’). The chapter concludes by considering from an overall perspective whether this research mapping and assessment provides a satisfactory basis for a research synthesis.

The following preliminary description of results is based on the 73 studies evaluated to be of either high or medium overall weight of evidence. Since one of the screening criteria of this review has been to exclude studies that are not dealing with at least two malleable school factors at the same time, the same study will occur in relation to several factors.

4.2.1  **School size**

Not many studies (N=14) analyse the relationship between school size and student learning, but when this factor is in fact studied, school size seemed to affect student outcomes positively (e.g. Martin et al., 2001; Opdenakker & van Damme, 2007; Rumberger & Palardy, 2005). This effect appears to be mediate by school practice (more teacher cooperation in bigger schools) which, in turn, affects school climate (better and more learning-oriented climate where teachers cooperate). According to other studies included in the review, however, school size had no significant effects (e.g. Lamb & Fullarton, 2002). Some of the studies furthermore point to the fact that there seems to be a trade-off between school size and drop-out rates, as larger schools tend to do better on achievement growth but not on preventing drop-outs (e.g. Rumberger & Palardy, 2005; Franklin et al., 1992).

- **2 studies with high overall weight of evidence examine this** (Opdenakker & Van Damme, 2007; Rumberger & Palardy, 2005))

- **12 studies with medium overall weight of evidence examine this** (Bondi, 1991; Silins, 2004; Fullarton; 2004; Lamb & Fullarton, 2002; Foley et al., 2007; Heck, 2007; Picucci et al., 2002; Florida State Department of Education, 1994; Mandeville et al., 1993; Franklin et al., 1992; Postletwhaite et al., 1992; et al., 2000).

4.2.2  **Class size**

Only nine studies examine the relationship between class size and student learning. According to several of the studies, smaller classes seem to have a clear positive effect on student outcome, especially among low-SES students (e.g. Bain et al., 1990; Coates, 2003). The study by Coates (2003) even concludes that increasing the average class size has detrimental effects on student learning, as larger class size directly affects achievements in a negative way (perhaps through classroom management effects), and because the effectiveness of any given amount of instructional time decreases when class size is increased (increasing instructional time will have larger benefits for students in small classes than for students in large classes). Yet again, other studies conclude that class size or type has no significant influence on student test scores (e.g. Franklin, 1992; Lamb & Fullarton, 2002). Only one study states that there is a positive relationship between (large) class size and student achievement (Woessmann, 2003).

- **1 study with high overall weight of evidence examine this** (Woessmann, 2003)
8 studies with medium overall weight of evidence examine this (Thomas, 1995; Lamb & Fullarton, 2002; Coates, 2003; Florida State Department of Education, 1994; Franklin, 1992; Bain et al., 1990; Martin et al., 2001; Ringsmose & Mehlbye, 2004)

4.2.3 Leadership

A large number of the studies (n=38) examine the relationship between leadership and student learning. The findings of these studies are however not unambiguous. Still, it seems to be the case that the great majority of studies that do in fact address this school factor identify leadership as an important predictor of student achievement (e.g. Hill et al., 1993; Hofman, 1996; Mosenthal et al., 2001; Pressley et al., 2004; Ringsmose & Mehlbye, 2004; Ross et al., 2006; Sammons et al., 1997; Teddlie et al., 1993; Traufler, 1992; Waxman et al., 2008). For example, the study by Ross et al. (2006) concludes that schools with higher levels of transformational leadership had higher collective teacher efficacy, greater teacher commitment to school mission, school community and school community-partnership as well as higher student achievement. Related to the findings of this study, a number of other studies conclude that school leadership exerts a positive influence on student achievement, but only in an indirect way (e.g. Hoy et al., 1990; Silins & Mulford, 2004; Tarter et al., 2004). Yet other studies found that school leadership did not have a significant effect on student achievement (e.g. Opdenakker & Van Damme, 2007; Rogers et al., 2006). Finally, some studies documented that educational leadership was negatively related to effectiveness, meaning that schools with low effectiveness manifested more educational leadership (e.g. Van der Werf, 1997).

6 studies with high overall weight of evidence examine this (Opdenakker & Van Damme, 2007; Rogers et al., 2006; Tarter et al., 2004; Van der Werf, 1997; Teddlie et al., 1993; Ross et al., 2006)

32 studies with medium overall weight of evidence examine this (Hofman, 1996; Webster, 2003; Silins & Mulford, 2004; Lamb & Fullarton, 2002; Waxman et al., 2008; Foley et al., 2007; Heck, 2007; Perez et al., 2007; Lindsay et al., 2006; Bottoms et al., 2006; Kitchen et al., 2006; Pressley et al., 2004; Hofman et al., 2002; Towns et al., 2001; Mosenthal et al., 2001; Sweetland et al. 2000; Texas Education Agency, 2000; Opdenakker & Van Damme, 2000; Zigarelli, 1996; Grissay, 1994; Hill et al., 1993; Kennedy et al., 1993; Postlethwaite et al., 1992; Hoy et al., 1990; Witte et al. 1990; Pressley et al., 2007; Lassen et al., 2006; Traufler, 1992; Sammons et al., 1997; Kyriakides et al., 2008; Stringfield et al., 2008; Ringsmose & Mehlbye, 2004)

4.2.4 Management

The effect of management on student learning is addressed in 23 studies. In most of these studies management is closely linked to school leadership (e.g. Hofman, 1996; Ringsmose & Mehlbye, 2004; Sammons et al., 1997; Griffith, 2003). This is the case, for instance, in the study by Hofman (1996), which concludes that a responsive management style by school leaders and school boards is central to school effectiveness. Some of the studies address the management factor in a more explicit way, however, relating management to the overall school conditions and/or the degree of school autonomy. As an example Woessmann (2003) finds that students in schools with autonomy in determining
teachers’ salaries performed significantly better than schools that did not enjoy this kind of autonomy (the same study however concludes that in some geographical areas such as Hong Kong, the results are the opposite).

- **4 studies with high overall weight of evidence examine this** (Opdenakker & Van Damme, 2007; Tarter et al., 2004; Teddlie & Stringfield, 1993; Woessmann, 2003)

- **19 studies with medium overall weight of evidence examine this** (Opdenakker & Van Damme, 2007; Tarter et al., 2004; Teedlie & Stringfield, 1993; Woessmann, 2003)

4.2.5 Staff development

Only seven of the studies examine the effects of staff development. A review of these, however, give an unclear picture. For example, some findings supported the main assumptions of Creemers' model of educational effectiveness, but were simultaneously difficult to interpret since, for instance, the quality of teaching did not have an effect at classroom level (e.g. Kyriakides & Tsangaridou, 2008). Yet other studies (e.g. Ross et al., 2006; Kitchen et al., 2006) showed that provision of extensive professional development opportunities had positive effects on pupil achievement.

- **2 studies with high overall weight of evidence examine this** (Ross et al., 2006; Ross & Gray, 2006)

- **5 studies with medium overall weight of evidence examine this** (Silins & Mulford, 2004; Kitchen et al., 2006; Hill et al., 1993; Kyriakides & Tsangaridou, 2008; Stringfield et al., 2008).

4.2.6 Curriculum/scheduling

A total of 26 studies examine different types of curriculum planning. The studies examine e.g. the effect of a broad curriculum (Lindsay & Muijs, 2006) or a challenging, problem-based curriculum (Kitchen et al., 2006), and the effect of the impact of teaching in the morning and systematic assessment of pupil progress (e.g. Rogers et al., 2006). However, a large proportion of the studies focus on the effect of an academically-oriented curriculum, and confirm that high-achieving schools have a focus on learning content and development of pupils’ cognitive skills (e.g. Opdenakker & Van Damme, 2007; van der Werf & Weide, 1996; Ross et al., 2006).

- **6 studies with high overall weight of evidence examined this** (Ross et al., 2006; Opdenakker & Van Damme, 2007; Rogers et al., 2006; Reezigt et al., 1999; Van der Werf, 1997; Van der Werf & Weide, 1996)

- **20 studies with medium overall weight of evidence examined this** (Lamb & Fullarton, 2002; Dumay & Dupries, 2007; Perez et al., 2007; Lindsay & Muijs, 2006; Kitchen et al., 2006; Pressley et al., 2004; Picucci et al., 2002; Mosenthal et al., 2001; Texas Education Agency, 2000; Opdenakker & Van Damme, 2000; Thomas & Collier, 1997; Zigarelli, 1996; Hill et al., 1993; Mandeville & Kennedy;
4.2.7 Teacher

A large proportion of the studies, 46 in all, examine the impact of the teacher. Of these, a large number measure pupils’ outcome in regard to teachers’ gender, experience, salaries and educational level (e.g. Hill et. al., 1993; Meelissen & Luyten, 2008; Rumberger & Palardy, 2005; Wößmann, 2003). Another cluster of the studies examines teachers’ instruction:

- **2 studies of teacher’s use of test and assessments and time on task** (Opdenakker & Van Damme, 2007; Pressley et al., 2004)
- **2 studies of teacher’s expectations** (Rumberger & Palardy, 2005; Kitchen et al., 2006)
- **2 studies of teacher’s providing high opportunity to learn and developing of home communication** (Taylor & Walpole, 2000; van der Werf & Weide, 1996)

In particular, teachers’ gender, educational level, high expectations and pupils’ opportunity to learn are considered by some studies to have a significantly positive effect on pupil outcome.

- **8 studies with high overall weight of evidence examine this** (Ross et al., 2006; Meelissen, M., & Luyten, H., 2008; Opdenakker & Van Damme, 2007; Rumberger & Palardy, 2005; Yu & White, 2002; Taylor et al., 2000; Teddlie & Stringfield, 1993; Woessmann, 2003)
- **38 studies with medium overall weight of evidence examine this** (Hofman, 1996; Thomas, 1995; Webster & Fisher, 2003; Fullarton, 2004; Lamb & Fullarton, 2002; Waxman et al., 2008; Dumay & Dupriez, 2007; Foley et al., 2007; Heck, 2007; Perez et al., 2007; Lindsay & Muijs, 2006; Bottoms et al., 2006; Choi & Kim, 2006; Kitchen et al. 2006; Pressley et al., 2004; Coates, 2003; Hofman et al., 2002; Mosenthal et al. 2001; Opdenakker & Van Damme, 2000; Zigarelli, 1996; Florida State Dept. (1994); Grisay, 1994; Hill, 1993; Mandeville, 1993; Kennedy et al., 1993; Franklin & Crone, 1992; Postlethwaite & Ross, 1992; Bain & Jacobs, 1990; Hoy et al., 1990; Pressley et al., 2007; Traufler, 1992; Senkbeil, 2006; Meijnen et al., 2003; Martin 2001; Kyriakides, 2008; Kyriakides, 2008; Papanastasiou (2008); Ringsmose & Mehlbye, 2004)

4.2.8 School culture

A large number of studies, 46 in all, examine school culture and its impact on pupils’ outcome. The studies examine e.g. discipline and atmosphere of the schools. A large proportion of the studies confirm that climate has a significant impact on pupils’ outcome; a safe, orderly and quiet atmosphere correlates positively with pupil outcome (e.g. Florida State Dept., 1994), while others are pointing to an academic climate, a learning climate or a work-oriented school climate (e.g. Rumberger & Palardy, 2005; Hoy et al., 1990).

- **8 studies with high overall weight of evidence examined this** (Ross et al., 2006; Opdenakker & Van Damme, 2007; Rogers et al., 2006; Rumberger &
38 studies with medium overall weight of evidence examined this (Hofman, 1996; Silins & Mulford, 2004; Fullarton, 2004; Young, 2001; Lamb & Fullarton, 2002; Waxman et al., 2008; Dronkers & Robert, 2008; Foley et al., 2007; Lindsay & Muijs, 2006; Bottoms et al., 2006; Kitchen et al., 2006; Craig et al., 2005; Pressley et al., 2004; Griffith, 2003; Hofman et al., 2002; Picucci et al., 2002; Towns et al., 2001; Sweetland & Hoy, 2000; Smyth, 2000; Zigarelli, 1996; Florida State Department of Education, 1994; Grisay, 1994; Hill et al., 1993; Mandeville & Kennedy, 1993; Kennedy et al., 1993; Hoy et al., 1990; Witte & Walsh, 1990; Pressley et al., 2007; Lassen et al., 2006; Griffith, 2002; Willis, 1996; Traufler, 1992; Senkbeil, 2006; Martin et al., 2001; Sammons, 1997; Kyriakides & Tsangaridou, 2008; Stringfield et al., 2008; Ringsmose & Mehlbye, 2004).

4.2.9 Teacher teams

14 studies examine the impact of teacher teams. The teacher teams are effective both socially, in creating a good atmosphere among the staff, and professionally, in developing teaching. Eight of the studies more or less strongly emphasize the importance of teacher teams in regard to positive pupil outcome (Opdenakker & Van Damme, 2007; Ringsmose & Mehlbye, 2004; Stringfield & Schaffer, 2008; Sammons et al., 1997; Opdenakker & Van Damme, 2000; Texas Education, 2000; Towns et al., 2001; Kitchen et al., 2006).

- 4 studies with high overall weight of evidence examined this (Ross et al., 2006; Opdenakker & Van Damme, 2007; Teddlie & Stringfield, 1993; Ross & Gray, 2006)

- 10 studies with medium overall weight of evidence examined this (Kitchen et al., 2006; Towns et al., 2001; Texas Education Agency, 2000; Opdenakker & van Damme, 2000; Florida State Department of Education, 1994; Witte & Walsh, 1990; Senkbeil, 2006; Sammons et al., 1997; Stringfield et al., 2008; Ringsmose & Mehlbye, 2004)

4.2.10 Support systems

A small part of the studies - with only medium weight of evidence - examine the effect of support systems at schools. Some studies are concerned with support for the pupils, others with support for the teachers. A strong partnership with parents also correlates positively with pupils’ outcome (Lindsay & Muijs, 2006). With disadvantaged pupils, attention towards individual students and provision of extra services and supports - beyond those traditionally offered by schools - correlate positively with pupil outcome (Picucci et al., 2002; Kitchen et al., 2006).

- 8 studies with medium overall weight of evidence examined this (Perez et al., 2007; Lindsay & Muijs, 2006; Bottoms et al., 2006; Kitchen et al., 2006; Picucci et al., 2002; Florida State Department of Education, 1994; Hill et al., 1993; Pressley et al., 2007)
4.2.11 Physical environment

Only a few studies examined different physical aspects of the school such as school resources i.e. library, available books in the school and class, IT equipment and the effect from the physical environment on student learning outcomes. One study underlines, that the shortage of materials is not always correlated to worse performance, especially in Japan where shortage in fact correlates positively with performance. (Woessmann, 2003). Another study by Pressley et al. 2007 shows that the students’ reading and writing competence are positively affected by the context of the school environment, including the students’ experience of many books in the classrooms.

- 1 study with high overall weight of evidence examined this (Woessmann, 2003)
- 3 studies with medium weight of evidence examined this (Bain et al., 1990; Pressley et al., 2007; Kyriakides & Tsangaridou, 2008)

4.2.12 Socio-economic composition and ethnic composition of the pupils in the schools

In school effectiveness research the socio-economic status (SES) and the ethnicity background of the student are utilized in two ways. On one hand, the socio-economic status (SES) and the ethnicity background of the student are utilized as control variables for the purpose of isolating the effects of student background characteristics in order to establish the relative importance of context, school and classroom factors. On the other hand, the socio-economic status (SES) and the ethnicity background of the student are seen as malleable variables, e.g. adjusting student composition through bussing or change of school districts.

In general, a range of studies examine the effects of student composition on student learning. A study shows that student composition has a direct effect on school practice, and thereby an indirect effect on student learning. Teachers tend to cooperate more in schools with high ability students (Opdenakker et al., 2007). In relation to these findings another study shows that class composition and process variables are correlated which again lead to positive school effects (Dumay, et al., 2007). Additionally, a study shows that as the percentage of students on free lunch increases, student participation decreases (Franklin et al., 1992).

According to a study by Mosenthal et al., 2001 two factors, SES and the nature of literacy instruction, did not play an explanatory role in literacy achievement test scores in the successful schools.

A study underlines that achievements are higher in schools with students with a large percentage of white students and staff members, though different factors are highlighted as being of importance as well, such as a climate focusing on academic achievement and concern for grades, leaders and teachers with high expectations of students (Kennedy et al., 1993). Furthermore, a study examines the percentage of low SES children in the school in relation to school location (suburban), and the stability of the student population, and its finding shows that these are significant predictors of change over time (Mandeville & Kennedy, 1993). In line with this, another study underlines that the achievement gaps were reduced in school settings that included students of different socioeconomic and ethnic background (Grisay, 1994).
Studies on socio-economic composition of the pupils in the schools

- **12 studies with high overall weight of evidence examined this** (Ross et al., 2006; Opdenakker & Van Damme, 2007; Rogers et al., 2007; Rumberger et al., 2005; Tarter & Hoy, 2004; Van Damme et al., 2002; Yu & White, 2002; Taylor et al., 2000; Reezigt et al., 1999; van der Werf, 1997; Young & Fraser, 1992; Ross & Gray, 2006)

- **37 studies with medium weight of evidence examined this** (Hofman, 1996; Thomas, 1995; Bondi, 1991; Silins & Mulford, 2004; Fullarton, 2004; Lamb & Fullarton, 2002; Dronkers & Robert, 2008; Dumay & Dupriez, 2007; Foley, 2007; Heck, 2007; Lindsay & Muijs, Bottoms et al., 2006; Kitchen et al., 2006; Hofman et al., 2002; Picucci et al., 2002; Towns et al., 2001; Mosenthal et al., 2001; Sweetland & Hoy, 2000; Smyth, 2000; Florida State Department of Education, 1994; Grisay, 1994; Hill et al., 1993; Mandeville & Kennedy, 1993; Kennedy et al., 1993; Franklin & Crone, 1992; Bain et al., 1990; Hoy et al., 1990; Lassen et al., 2006; Willis, 1996; Senkbeil, 2006; Meijnen et al., 2003; Martin et al., 2001; Sammons et al., 1997; Kyriakides & Creemers, 2008; Kyriakides & Tsangaridou, 2008; Papanastasiou, 2008; Ringsmose & Mehlbye, 2004)

Studies on ethnic composition of the pupils in the schools

- **4 studies with high overall weight of evidence examined this** (Rumberger & Palardy, 2005; Reezigt et al., 1999; Van der Werf, 1997; Young & Fraser, 1992)

- **16 studies with medium weight of evidence examined this** (Hofman, 1996; Thomas, 1995; Lamb & Fullarton, 1995; Heck, 2007; Lindsay, 2006; Hofman et al., 2002; Picucci et al., 2002; Towns et al., 2001; Grisay, 1994; Hill et al., 1993; Mandeville & Kennedy, 1993; Kennedy et al., 1993; Lassen et al., 2006; Willis, 1996; Meijnen et al., 2003)

4.2.13 Other factors and phenomenon

The school factor categorized as ‘other’ has been marked in 49 studies. Overall, the category contains a range of different aspects connected to the school effectiveness that are mentioned in several studies. Among these are e.g. the school type, school location, gender of the pupils and home/school relations.

Concerning the school type, a central point in the relationship between student achievement and the types of the school i.e. public or private schools, seem to be the willingness of the schools to let the school community members, including parents, influence decisions of the schools (Hofman R.H et al 1996/2002).

The relationship between student achievement and the location of the school in urban/suburban area has also been investigated. According to some studies, the school location is a less consistent predictor of achievement (e.g. by Choi et al., 2006 and Martin et al., 2001). According to another study, suburban schools are significant predictors of change over time (Mandeville, G. K et al., 1993). Furthermore, the most prominent indicator seems to be doing daily homework. Amount of homework is positively and significantly related to performance in different subjects in different countries (Martin, M. et al. 2001), (Woessmann, 2003).
11 studies with high overall weight of evidence examined this (Campbell et al., 2000; Meelissen & Luyten, 2008; Rogers et al., 2006; Rumberger & Palardy, 2005; Van Damme et al., 2002; Yu & White, 2002; Reezigt et al., 1999; Van der Werf, 1997; Van der Werf & Weide, 1996; Young & Fraser, 1992; Woessman, 2003)

21 studies with medium weight of evidence examined this (Hofman, 1996; Thomas, 1995; Bondi, 1991; Webster & Fisher, 2003; Young, 2001; Lamb & Fullarton, 2002; Choi & Kim, 2006; Coates, 2003; Hofman et al., 2002; Picucci et al., 2002; Opdenakker & Van Damme, 2000; Zigarelli, 1996; Grisay, 1994; Mandeville & Kennedy, 1993; Franklin & Crone, 1992; Postlethwaite & Ross, 1992; Witte & Walsh, 1990; Traufler, 1992; Senkbeil, 2006; Meijnen et al., 2003; Martin et al., 2001)

4.3 Research designs used in the studies

The review question has an underlying causal aspect: the idea that "something" contributes to the pupils' learning. In addition it is implied that many factors contribute to learning. In this review we focus on which factors contribute the most. However, the review question touches on some basic questions in educational research, about which there is considerable disagreement. In educational research it is not possible to identify causal relationships with the status of scientific laws governing empirical relationships between specific causal factors and a pupil's learning. A heated iron bar is not conscious, and the effects one would expect to observe (the laws of thermodynamics) can be examined and explained on the basis of a single factor (heating) while all other factors are held constant. The predictions in this situation will be fulfilled with a high degree of precision. The situation is rather different when studying phenomena in schools. Unlike iron bars, teachers and pupils are conscious individuals with values and motives. Merely knowing that a research project is in progress can influence the measured outcomes. The behaviour of individuals can not be predicted; it can at best be retrodicted. When the complexity of the research situation is so high, as it usually is, and when so many possible contextual factors can affect the dependent variable simultaneously, it is legitimate to ask whether causal attribution is possible at all in educational research. There is some justification for arguing that educational science is "the hardest science of all" (Berliner, 2002). On the other hand, the cases that are studied are not so unique as to rule out an ambition to generalize, i.e. to draw conclusions about validity extending over a wider range than the empirical material in the study. Such generalisations will often lead to questions about the limits of validity. A basic requirement for valid conclusions about factors in schools contributing to learning is: which strategic research decisions underpinned the study, i.e. what was the research design.

In order to study what affects a specific phenomenon (explanandum), it is advantageous to isolate all other factors than the one factor (explanans) that will be varied under controlled conditions. Experimental educational research of this type does exist, but there are no truly laboratory-like experiments amongst the studies included in this review, cf. Table 3.10. On the other hand, a quasi-experimental research design does occur. One

---

1 "Which factors/phenomena in school contribute most to primary and lower secondary pupils' learning?"
example is Sterbinsky et al. (2006). The premise for this study is that one compares a control group with an experimental group of schools that are as similar as possible to the controls, apart from the one explanatory variable that is of theoretical interest (in the study referred to, this variable is defined as the Comprehensive School Reform, CSR) and the dependent variable (pupils’ learning). The effects are studied over a length of time (3 years) during which the ‘treatment condition’ is expected to manifest itself in the dependent variable. The researchers manipulate the actual situations and phenomena in order to gain insight into the real, i.e. causal, relationships. Since in educational research one can not achieve a laboratory-like situation - except in very limited experiments - the researchers must compensate in other ways for factors that might influence the effect of “the treatment”. In the study referred to, this is done as follows:

“Each school was individually matched and compared to a demographically similar control school on measures of school climate, teacher satisfaction, observed classroom teaching methods, and student achievement on a battery of 4 individually administered reading tests” (Sterbinsky et al., 2006, 367).

A number of research-strategical techniques can be applied "to strengthen and verify matching fidelity". If the measures for achieving a reasonable degree of control over all other factors while studying the single factor of interest (CSR) are successful, this type of research design will provide a high weight of evidence when answering the causally-influenced review question. Some questions still remain, however, about the range of validity of the conclusions: what was the reference population of the study? Are the principal conclusions of the study (“effects of CSR in 10 schools”) only valid for schools from “a four-state region in the south-east USA”? Here the authors themselves state that the conclusions are not “necessarily representative of those occurring nationally for the individual models concerned” (Sterbinsky et al., 2006, 388), but many would consider the result of the study interesting in relation to Scandinavian schools. On the other hand, culture-specific factors can be significant. An example here is Opdenakker & Van Damme 2007 who, to their surprise, found that school leadership did not “affect the school practice much”, to which they add that “a possible explanation could be the lack of a strong participative professionally-oriented leadership in the majority of Flemish secondary schools” (Opdenakker & Van Damme, 2007, 195).

Whilst longitudinal quasi-experimental research designs appear to be acceptable as an adequate research strategy for evaluating causal relationships, cross-sectional comparisons of factors in low-achieving and high-achieving schools appear to be more controversial. In a number of studies, attempts are made to compensate for the socio-economic status of pupils (for example by comparing low-achieving and high-achieving schools within the same geographical area in which pupils have similar social backgrounds). Rudd et al. 2002 and Florida State Dept. of Education 1994 are examples of studies in which this research strategy was followed, and the question arises: What causes the difference between these groups of schools? A research strategy of this type has a number of built-in weaknesses that are difficult to compensate for (amongst other things, because distortions crop up in the correlations of the pupils’ social background and their

---

2 The authors’ expectation that school leadership affects practice has been verified in other research, especially in the USA.
test results). For this reason, this type of research design provides weaker evidence when weight of evidence is being assessed.

A related but rather different research strategy is to look at a relatively homogenous group of schools (for example "small, successful high schools" in Foley et al. 2007, or "six high-performing schools in Tennessee" in Craig et al. 2005). In the first step one identifies the schools on the basis of school statistics regarding pupil backgrounds and test results. The second step in the strategy consists of examining the successful schools in greater detail by studying school documents combined with observations and interviews and/or survey-type investigations. Often the aim of this type of study will be to explore one factor or another - for example, perceived success factors.

Ethnographic studies of schools usually look at individual schools. This may for example be done by examining how a successful school prepares a special group of pupils for starting College studies (e.g. Calaff, 2008) or by examining a school in a socially challenged environment (Jones, 2004). In certain cases, this type of approach can be regarded as a pre-study leading merely to unconfirmable conclusions (Yin, 1981:97). Others consider this type of case-study to be a truly scientific method in which detailed descriptions of individual situations are studied in depth (Marshall & Rossmann, 1995). For the purpose of this systematic review, studies that only lay weight on detailed descriptions of the workings of the school provide weak weight of evidence, even though the quality, judged in relation to the purpose of the study, may in itself be good. Amongst the studies captured in this survey one also finds studies combining qualitative and quantitative approaches: detailed descriptions of the activities in the school merge into the complexity of school effectiveness studies. Some researchers consider that observations and interviews can make school effectiveness studies more complete than studies that are based solely on tests and questionnaires.

In this systematic review the individual studies have been evaluated using the question "How appropriate is the design selected for this study?" Ranked into three groups with high, medium or low weight of evidence, the distribution of the studies is High: 27 %, Medium: 43 % and Low: 30 %.

4.4 Research quality

The quality of educational research is a much discussed issue in many countries, and the discussion is often based on systematic reviews of the research activities of institutions: some university institutes have been threatened with closure on the basis of expert evaluations of the quality of research at the institute. In Scandinavian countries, too, these types of expert evaluations of research quality have been carried out. Leading researchers have also issued statements about the perceived quality of educational research. For example, it has been claimed that "it is an open secret that [educational research] is a professionally weak discipline, not only in Norway, but internationally" (Elster 2006: 296). One comment to this claim is that the quality of educational research may vary, just like research quality in any other discipline. The question of quality must however be seen in connection with the quality assurance mechanisms that each contri-

---

3 This claim was put forward by one of Norway's most respected social scientists, Jon Elster, today a professor at Collège de France (http://www.prio.no/CSCW/News/NewsItem/?oid=83550).
bution has been subjected to on its way to publication. From this systematic review there appears to be a certain connection between the quality of the studies and whether or not they were published through channels that utilise anonymous "peer-review". There seems to be a tendency for studies assessed as having high weight of evidence to have been published in "good" academic journals. Some of the studies were carried out at the request of an educational authority and in a few cases performed by an entity which itself has carried out or participated in the deployment of an educational reform. In a few of the studies it seems that the clarity and precision of conclusions, which in the research community are considered to be virtues,\(^4\) seem to have been downplayed in order to be able to give clear recommendations. Commissioned research and commercially based research can be motivated by other incentives in this respect than researchers from the academic community. Some researchers might claim that a research review based on the premises underlying this systematic review will lay undue emphasis on measurement of achievement as a basis for quality measurement, and too little emphasis on other considerations that schools are also expected to take.

The question of the quality of the research contributions covered by this systematic review can be considered on the basis of the self-declared aims of the studies. Here the assessors considered that 33 % of the studies had a high weight of evidence, 38 % had a medium weight of evidence and 28 % had a low weight of evidence. When studies score low in this type of evaluation this indicates a lack of quality measured by the factors that educational researchers and the research community value highly. It is remarkable that study reports, doctoral dissertations, and articles in academic journals can all be found amongst the reports in the lowest quality category. This result can be interpreted as an indication that the quality assurance mechanisms have not operated in a completely satisfactory way.

The question of quality can also be considered in the light of the evidence that the studies contribute in relation to the review question framed in this systematic review. Here the distribution was: 15% of the studies had high, 50% had medium, and 35 % had low weight of evidence.

4.5 Overview: Prospects for systematic syntheses based on the research mapping completed

Outside Scandinavia, several syntheses have already been made of research into the influence of schools on pupils’ learning (e.g. Scheerens, 1997; Teddlie, 2000; Townsend, 2007; Creemers & Kyriakides, 2008). In countries such as the USA and England, these syntheses have had a significant influence on educational policy and teaching theory.

4.5.1 The need for a synthesis with a Scandinavian perspective

To some extent, of course, the already existing syntheses from abroad can be used with respect to Scandinavia. But there is a great and unfulfilled need for research to be ana-

\(^4\) Gulbrandsen & Langfeldt (1997).
lysed and summarised in a Scandinavian context. Even though this research review has concentrated exclusively on research about schools in developed Western countries, there are significant differences between the structure, leadership and educational methods in these school systems, and this means that part of the research concerns topics or has been carried out under circumstances that are not necessarily relevant in a Scandinavian situation. For example:

- A significant proportion of the foreign research was carried out on school systems in which lower secondary schools are subdivided into academic, "general" and practically-oriented types of schools. We know that an early allocation of pupils into general and practical streams affects their behaviour and academic achievement: see e.g. Bauer and Riphahn (2006) or Hanushek and Woessmann (2005). These results have been questioned by Manning and Pischke (2006), however. In any event, to be relevant in a Scandinavian perspective, a synthesis must focus on research carried out in school systems of the unified type in order to avoid the influence of effects resulting from the early sorting of pupils.

- The Anglo-Saxon countries, where a large part of the relevant research is carried out (cf. Table 3.1), have different rules and traditions for school leadership from the Scandinavian countries. The influence of the leadership factor on the effect of schools is one of the most important themes in this research area. Anglo-Saxon research must be considered in the light of the fact that it has been carried out within a tradition that differs from the Scandinavian with respect to school leadership. On the other hand, it is clear that there is also a great need for knowledge on this topic in Scandinavia, since there is only limited research on this subject within Scandinavia.

That school conditions in Scandinavian countries differ in several respects from those in other countries could at first sight be an argument for restricting a synthesis to research carried out in Scandinavian countries only. However, as this research review reveals, there are only a limited number of studies available from the Scandinavian countries (cf. section 3.1.), and a synthesis based solely on these studies would be inadequate. It is therefore necessary to include international research and then analyse the results in a Scandinavian perspective.

4.5.2 The conditions for a synthesis are fulfilled

Not all themes or research areas within educational research can be summarised into a workable research synthesis. Within some research areas, the research is so heterogeneous with regard to topics, concepts applied and theoretical perspective that it would hardly be meaningful to try to create a synthesis. But in the case of school effectiveness research, the principal conditions for creating a synthesis are fulfilled - which of course is also indicated by the fact that a number of foreign syntheses already exist.

The principal conditions are:

*Research is more or less cumulative.* As previously mentioned (cf. section 4.1) most studies take their point of departure in some form of review of existing research. This means that new research projects to a certain extent build on already existing knowledge, and thus the total research effort becomes more or less cumulative. That is to say, it moves towards a clarification about which there is at least partial consensus
amongst the researchers. This is a necessary condition for establishing a meaningful syn-
thesis.

A reasonable degree of consistency in the factors examined. The cumulative character of research into “the good school” is additionally underlined by the fact that there is a relatively high degree of overlap between the studies with respect to which factors are examined, and in particular which outcomes the activities of the schools are related to. The great majority of studies examine the effect of school factors on the academic skills of the pupils. As mentioned, the socio-economic background of the pupils, the school culture, the significance of the teacher and of school leadership are also included in more than half of the studies included in this research review (cf. section 3.2). The ef-
fect of these factors can thus be considered to have been thoroughly examined.

Differences in theoretical perspective play only a limited role. Some areas of educa-
tional research are affected by competing and incompatible theories, each with their various concepts and approaches, which makes it difficult to establish a synthesis, even of the empirical studies. Research into “the good school”, however, generally builds onto existing empirical research, and divergent theoretical approaches play only a minor role (cf. section 4.1).

The school level has independent significance. One contentious question has been about whether the "school level" has any real effect on the pupils' progress as compared with the "classroom level". If the "school level" were without importance, a synthesis of re-
search would only have limited interest. Firstly, however, newer and more advanced research has been able to relate both to the "school" and "classroom" levels, and sec-
ondly there is a clear tendency in the research results showing that the "school level" is in fact important. Therefore it is meaningful to make a synthesis focusing on "the school level", while also including "the classroom level".

4.5.3 Which factors can a synthesis be expected to clarify the importance of?

A research synthesis can naturally be expected first and foremost to shed light on the effects of those factors that are addressed by many studies. These are primarily the fol-
lowing (cf. section 3.2 og 4.2):

The socio-economic background of the pupils. The effect of the socio-economic back-
ground of the pupils is included in one half of the studies (cf. Error! Reference source not found.), and the subject is so well studied that there is a solid basis for a synthesis - although, on the other hand, strikingly new insights are unlikely to emerge. This could be the case, however, with regard to the following factors:

Leadership. Many studies - although not all - indicate that leadership or the school leader is one of the school factors having the greatest influence on the school "output" (cf. section 4.2.). Other studies, however, suggest that leadership only plays a modest role - or no role at all - in the results achieved by the school. An explanation of this re-
sult could be that school leaders operate so uniformly in the different schools that sta-
tistical analysis can not reveal any differences in effects when comparing the schools. A synthesis will hopefully be able to contribute to illuminating this question about what good school leadership consists of, and what effect it has on pupils' learning.

Staff development and the teachers. Research about "the good school" has displayed particular interest in the effects of the teachers' activities outside the classroom itself.
For example, several studies focused on the effects of supplementary training and staff development, and on the importance of the fact that the teachers collectively have an agreed approach to their educational practice and classroom leadership. The question of the effect of a greater or lesser degree of teacher cooperation and teamwork is also a theme in numerous studies.

School culture. Several aspects of the concept of school culture may be studied, and amongst those frequently included are the extent of consensus amongst the school leaders and teachers about a common "mission" or common values. Furthermore, many studies have addressed the question of the effect on pupils of high academic expectations to them, and this question too can very probably be elucidated as part of a synthesis. More diffuse factors such as the school's "atmosphere" or "spirit" have also been examined in a number of studies.

Other factors. A number of additional factors such as cooperation between home and school, school policy with regard to pupil behaviour, and a teaching environment emphasising calm working conditions have also been addressed by a number of studies, and a research synthesis may be expected to shed further light on these issues.

The above-mentioned factors, which a research synthesis can be expected to elucidate, relate to central problems in current educational policy and debate.

4.5.4 Conclusion

Syntheses of research into "the good school" in other countries have previously shown themselves to be illuminating and influential, and the necessary conditions are present to permit the creation of a new synthesis of research in this area. If a synthesis is to be relevant and penetrating with respect to the Scandinavian countries, it is necessary to carry out a synthesis with a Scandinavian perspective.
## 5 Appendix 1: EPPI-Centre tool for education studies V2.0 - editable version

### 5.1 EPPI-Centre tool for education studies V2.0 - editable version

**Item:** Van Damme, Jan; De Fraine, Bieke; Van Landeghem, Georges; Opdenakker, Marie-Christine; Onghena, Patrick (Dec 2002) A Study on Educational Effectiveness in Secondary Schools in Flanders: An Introduction.

#### Section A: Administrative details

<table>
<thead>
<tr>
<th>A.1 Name of the reviewer</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jaap Scheerens</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A.2 Date of the review</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The 23rd of January 2009</td>
</tr>
</tbody>
</table>

| A.3 Please enter the details of each paper which reports on this item/study and which is used to complete this data extraction. | Paper (1) Journal article  
Unique Identifier: 1650885 Van Damme  
Authors: Jan Van Damme; Bieke De Fraine; Georges Van Landeghem; Marie-Christine Opdenakker; Patrick Onghena.  
Title: A Study on Educational Effectiveness in Secondary Schools in Flanders: An Introduction.  
Paper (2) With respect to filling in this EPPI-reviewer, three articles are considered:  
Authors: Paper 1: Jan Van Damme; Bieke De Fraine; Georges Van Landeghem; Marie-Christine Opdenakker; Patrick Onghena. Paper 2: Jan Van Damme; Bieke De Fraine; Georges Van Landeghem; Marie-Christine Opdenakker; Paper 3: Jan Van Damme; Bieke De Fraine; Georges Van Landeghem; Marie-Christine Opdenakker; |
<table>
<thead>
<tr>
<th>A.4 Main paper. Please classify one of the above papers as the 'main' report of the study and enter its unique identifier here.</th>
<th>Unique Identifier: 1650885 Van Damme</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.5 Please enter the details of each paper which reports on this study but is NOT being used to complete this data extraction.</td>
<td>Specific focus of this data extraction (please specify) The study consists of public (state-run) schools and private (catholic schools, city school) schools. The data extraction is focused on the public schools.</td>
</tr>
<tr>
<td>A.6 If the study has a broad focus and this data extraction focuses on just one component of the study, please specify this here.</td>
<td></td>
</tr>
<tr>
<td>A.7 Language (please specify)</td>
<td>Details of Language of report English</td>
</tr>
</tbody>
</table>

### Section B: Study Aims and Rationale

<table>
<thead>
<tr>
<th>B.1 What are the broad aims of the study?</th>
<th>Explicitly stated (please specify) Our aim is to search for interesting correlates (student, class, or school characteristics) of the noncognitive outcomes at the end of the second grade that have some predictive power on top of the background characteristics and the group composition variables.</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.2 What is the purpose of the study?</td>
<td>A: Description Please edit</td>
</tr>
<tr>
<td>B.3 When was the study carried out?</td>
<td>Explicitly stated (please specify) 1991-1992</td>
</tr>
</tbody>
</table>
B.4 What are the study research questions and/or hypotheses?

| Research question with respect to the non-cognitive outcomes; | The first research question inquires about the ‘raw’ effect of the school and of the first-and second grade class on the noncognitive outcomes; The second research question asks about net effects of the secondary school and classes; Our third objective is achieved simultaneously, namely a description of the relationships between the student’s background characteristics and the noncognitive outcomes at the end of the second grade. A more strict brand of net effects is defined by additionally controlling for group composition in terms of the background characteristics (fourth research questions) |
| Research question with respect to the outcomes in mathematics: | (1) How important is the class/teacher and the school in explaining differences in mathematics achievement at the end of the second grade? | (2) To what extent are differences between classes within schools with respect to mathematics achievement attributable to differences in student intake? |
| Language: 1: | Does a student’s language achievement depend upon the school where he or she is taught? | 2: Are class and teacher more important than school with regard to the language achievement? |
|  | 3: To what extent are differences between schools and between classes a result of differences in student intake? | 4: What characteristics of schools and classes are linked to the language achievement of the students? |
|  | 5: Are some schools or classes more effective for particular groups of students (with respect to students’ ethnic background, gender and ability)? |  

Section C: Study Policy or Practice Focus

| C.1 What is the curriculum area, if any? | N/A (not on a specific curriculum area) |
| Please edit |
C.2 What is/are the educational setting(s) of the study?

- Lower secondary school
- Secondary school

C.3 In which country or countries was the study carried out?

Explicitly stated (please specify)
- Belgium

C.4 Please describe in more detail the specific phenomena, factors, services or interventions with which the study is concerned.

Details
The effect of schools and classes upon mathematics + literacy achievement in the second grade of second education is addressed. The reported data in this contribution stem from the LOSO-project.

Section D: Phenomena/Factors in School Addressed in the Study

D.1 Which phenomena/factors in school are addressed in the study?

- School culture
- Socio-economic composition of the pupils in the schools
- Other
  Please edit other variables addressed are gender, classroom and school climate, (to be seen as aspects of school culture), performance feedback and ability grouping

Section E: Pupil Result Focus
<table>
<thead>
<tr>
<th><strong>E.1 Are academic effects involved?</strong></th>
<th>Yes, achievement performance. Please specify</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mathematics achievement at secondary school level Dutch language achievement in Belgium secondary schools</td>
</tr>
<tr>
<td><strong>E.2 Are non-academic effects involved?</strong></td>
<td>Yes, psychical. Please specify</td>
</tr>
<tr>
<td></td>
<td>-the degree to which students feel at home in the school environment, - the extent to which the student does his/her best for the school work; - academic self concept - social integration in the class</td>
</tr>
<tr>
<td><strong>E.3 Does the study focus on effects on a specific group of pupils?</strong></td>
<td>Yes, pupils with low SES</td>
</tr>
<tr>
<td></td>
<td>Yes, boys</td>
</tr>
</tbody>
</table>

**Section F: Actual sample**

<table>
<thead>
<tr>
<th><strong>F.1 Who or what is/ are the sample in the study?</strong></th>
<th>Schools (please specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In total: 57 secondary schools With respect to the study of the mathematics achievement: A subsample of 2,552 students following the general track in the second grade of secondary education (and who did not have to retake the first grade), belonging to 150 mathematics classes and to 57 secondary schools, was used.</td>
</tr>
<tr>
<td></td>
<td>teachers (please specify)</td>
</tr>
<tr>
<td></td>
<td>In every school, a representative sample of 15 teachers in the first cycle completed a school characteristics questionnaire. Language: in the large sample there are 155 Dutch teachers and 275 second grade classes, so some teachers taught more than one class.</td>
</tr>
<tr>
<td></td>
<td>Pupils (please specify)</td>
</tr>
<tr>
<td></td>
<td>6411 students</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>F.2 What was the total number of participants in the study (the ACTUAL sample)?</strong></th>
<th>Explicitly stated (please specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With respect to the study of the mathematics achievement: A dataset of 2105 students belonging to 147 classes and 56 schools was available. Analyses on the effect of class and school variables are based on a dataset with 1,119 students, 74 classes and 33 schools. The analysis with the relevant student, class, and school variables included in the</td>
</tr>
</tbody>
</table>
model of mathematics achievement is based on a dataset with 1,936 students, 131 classes and 47 schools. With respect to the study of the noncognitive outcomes: The LOSO-cohort of 6,411 students who started secondary education in September 1990 contains a subsample of 4,759 students in 57 schools who were enrolled in the first grade in 1990-1991 (in 291 classes) and in the second grade of the general track in 1991-1992 (276 classes). Of those 4,759 students, 150 changed schools during their first 2 years in secondary education. Our analyses are based on the remaining 4,609 students. With respect to the study of language achievement: A subsample of 2569 students in 152 Dutch classes in 55 schools.

F.3 Please specify any other useful information about the study participants.

Details
With respect to the study of the mathematics + language achievement and noncognitive outcomes: Five student-level explanatory variables, measured at the start of the first grade, were used: initial cognitive ability (COGN), socioeconomic status of the family (SES), achievement motivation (AM), immunity to stress (STRESSIMM), and sex and language spoken at home (DUTCH-HOME). Prior mathematics achievement is based on a dataset with 1,936 students, 131 classes and 47 schools. Specifically, to the noncognitive outcome the age of the start of secondary education were also described. At class level, the following group composition variables were used: mean initial cognitive ability (CL-COGN), mean SES (CL-SES), mean achievement motivation (CL-AM), mean immunity to stress (CL-STRESSIMM), proportion of girls in the class (CL-SEX), and proportion of students who speak Dutch at home (CL-DUTCHHOME). At he school level, group composition variables comparable with the described group composition variables at the class level were used to describe the student population of a school.

Section G: Results and conclusions

G.1 What are the results of the study as reported by the authors?

Details
The study stands out for its attention for...
school and class composition effects, in terms of SES, initial cognitive ability, achievement motivation, language spoken at home and the age at the start of secondary education. These variables are used as individual student background control variables, but also as compositional effects at class and school level. Results are presented in three areas: mathematics, Dutch language and non-cognitive outcomes. Mathematics - a learning climate that is focused on learning and cohesive with a teacher that has positive expectations towards the achievement of students (the effects of other school and classroom process variables like opportunity to learn, disappeared after taking the compositional effects into consideration; climate was also overlapping with composition in terms of average SES and average initial ability in explaining mathematics achievement) - paying attention to differences between students had a negative effect - consultation between teachers had also a negative effect NO effects of structured teaching and performance feedback - ability grouping was found to be positive for low achieving students Language (also for language high effects of student background variables and compositional effects; specifically a pronounced effect of gender composition: a high proportion of girls is indicative of high achievement); - learning climate was the only school factor that had an effect over and above the individual background variables and compositional effects Non cognitive outcomes +Learning climate had a positive significant effect on three of the four non cognitive effect measures: environment, work, and peers - Feedback had a negative significant effect on self image

G.2 What do the author(s) conclude about the findings of the study?

Please edit The study provides a wealth of information on school composition effects, particularly the effects of mean initial ability level and mean SES (at school and classroom level). Very interestingly are the interaction or joint effects of school composition and school climate. AN orderly work oriented climate was the most relevant school factor, operational at classroom and at school level.
G.3 Which answer(s) does the study offer to the review question?

<table>
<thead>
<tr>
<th>Please specify</th>
</tr>
</thead>
<tbody>
<tr>
<td>The substantive school factor that had the expected positive effect on all outcome variables was a work oriented school climate. Relevant is also that strong candidates from other studies like: structured teaching, opportunity to learn and performance feedback had none or negative effects.</td>
</tr>
</tbody>
</table>

Section H: Study Method

<table>
<thead>
<tr>
<th>H.1 Study Timing</th>
<th>Prospective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please edit</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H.2 What is the method used in the study?</th>
<th>Cohort study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please edit</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H.3 Study design summary</th>
<th>Please specify</th>
</tr>
</thead>
<tbody>
<tr>
<td>The mathematics and the language (i.e., Dutch) achievement was measured by means of curriculum relevant multiple choice tests at the start of the secondary school and at the end of the first, the second, the fourth, and the sixth grade. As already mentioned, the LOSO-cohort was followed through secondary school, but also afterwards. This makes it possible to consider another type of effectiveness criterion by studying the effects of secondary schools upon dropout. Data about the students' primary school career were also collected because some school effectiveness studies indicate that the primary school can have long-term effects upon achievement in secondary school. They are though not considered in these articles.</td>
<td></td>
</tr>
</tbody>
</table>

Section I: Methods-groups

<table>
<thead>
<tr>
<th>I.1 If Comparisons are being made between two or more groups*, please specify the basis of any divisions made for making these comparisons</th>
<th>Not applicable (not more than one group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please edit</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I.2 How do the groups differ?</th>
<th>Not applicable (not in more than one</th>
</tr>
</thead>
<tbody>
<tr>
<td>Please edit</td>
<td>group)</td>
</tr>
<tr>
<td>I.3 Number of groups</td>
<td>Not applicable (not more than one group) Please edit</td>
</tr>
</tbody>
</table>

### Section J: Methods - Sampling strategy

| J.1 Are the authors trying to produce findings that are representative of a given population? | Explicitly stated (please specify) To all the Flemish secondary schools. |
| J.2 What is the sampling frame (if any) from which the participants are chosen? | Explicitly stated (please specify) Flemish secondary schools |
| J.3 Which method does the study use to select people, or groups of people (from the sampling frame)? | Implicit (please specify) The set of schools is to a certain extent representative of the Flemish secondary schools. |
| J.4 Planned sample size | Not stated/unclear (please specify) Please edit |
| J.5 How representative was the actual sample (as recruited at the start of the study) in relation to the aims of the sampling frame? | Medium (please specify) With respect to the mathematics and language achievement: The sample size can be reduced throughout the analyses due to missing values on some of the variables involved. With respect to the noncognitive outcomes: Of the 4,759 students, 150 changed schools during their first 2 years in secondary education. Our analyses are based on the remaining 4,609 students. |
| J.6 If the study involves studying samples prospectively over time, what proportion of the sample dropped out over the course of the study? | Implicit (please specify) Not clearly specified, though, it is mentioned that the noncognitive data have an inherent three-level structure and the noncognitive outcomes of all 4759 students are analysed. |
| J.7 For studies that involve following samples prospectively over time, do the authors provide any information on whether, and/or how, those who dropped out of | Yes (please specify) With respect to the noncognitive outcomes: Of the 4,759 students, 150 changed schools during their first 2 years in secondary education. Our analyses are based on the remaining 4,609 students. |
| **J.8 If the study involves following samples prospectively over time, do authors provide baseline values of key variables, such as those being used as outcomes, and relevant socio-demographic variables?** | Yes (please specify)  
*intelligence, initial achievement and SES* |

### Section K: Methods - Data Collection

| **K.1 Which methods were used to collect the data?** | Curriculum-based assessment  
*Please edit*  
Self-completion questionnaire  
*Please edit*  
Because of the large sample size, data on schools, classes, teachers and students had to be collected by means of questionnaires. All the noncognitive outcomes in this study were derived from one questionnaire (104 five-point items), the well-being questionnaire. The questionnaire were administered four times during the students' secondary school career: at the end of the first, second, fourth and sixth grade.  
Coding is based on: Author's description  
*Please edit* |

| **K.2 Do the authors' describe any ways they addressed the repeatability or reliability of their data collection tools/methods?** | Details  
The mathematics achievement test used at the end of the first grade (MATH1) covers set and relations theory, theory of numbers and geometry. The reliability is 0.76. The mathematics achievement test used at the end of the second grade (MATH2) covers theory of numbers and geometry. The reliability (Cronbach’s a) is 0.70 |

| **K.3 Do the authors describe any ways they have addressed the validity or trustworthiness of their data collection tools/methods?** | Details  
The level of content validity is high, because of the several items asked in the tests. The aggregation included all students in a class and not only those students belonging to the LOSO - cohort. |
### Section L: Methods - data analysis

<table>
<thead>
<tr>
<th>L.1 Which methods were used to analyse the data?</th>
<th>Explicitly stated (please specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The data were analysed by means of correlations and multilevel analysis. Three levels were identified: the student level, the school level, and an intermediate level that combines the mathematics teacher and the class group. The multi-level analysis is done by the MLwiN-software. Student-level measures were aggregated and used as descriptive indicators of the composition of the group of students in the classroom. A mean score is calculated for each class separately: mean initial cognitive ability (CL-COGN), mean SES (CL-SES), mean achievement motivation (CL-AM), mean immunity to stress (CL-STRESSIMM), proportion of girls in the class (CLSEX) and proportion of students who speak Dutch at home (CL-DUTCHHOME). The group means are calculated over all students for which the particular variable is available. The aggregation included all students in a class and not only those students belonging to the LOSO-cohort. Students with missing data are mostly omitted from the multilevel analyses, but in the aggregation process even students with values on other variables are included. On the other hand, we did not calculate group composition scores that are based on too small fraction of the group. If less than 50% of the scores on a student level variable in a class was available, the aggregated score for that class was not calculated. This procedure is expected to reduce the overall measurement error on the independent variables.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L.2 Which statistical methods, if any, were used in the analysis?</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>the tests were converted into IRT-scores</td>
<td></td>
</tr>
<tr>
<td>Noncognitive outcomes are analysed by factor analysis.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L.3 Do the authors describe strategies used in the analysis to control for bias from confounding variables?</th>
<th>Yes (please specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only students who did not have to retake the first grade were included in the study. That is only students from the A-stream that is students that stayed 2 consecutive years were considered.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L.4 Do the authors describe any ways they have addressed the re-</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>lanuage: Internal consistency was high</td>
<td></td>
</tr>
</tbody>
</table>
peatability or reliability of data analysis?

(0.90). Students with missing data were omitted from the multilevel models, but in the aggregation process even students with missing values on other variables were included. If less than 50% of the scores on a student-level variable in a class was available, the aggregated score for that class was not calculated. This procedure was expected to reduce the overall measurement error on these independent variables.

L.5 Do the authors describe any ways that they have addressed the validity or trustworthiness of data analysis?

Details

The content validity of the achievement tests for mathematics and language was assessed by teacher ratings of the extent to which students have had the opportunity to learn the content represented in the individual test items. A test item is scored "1" by the mathematics/Dutch teacher when the item is not covered in the curriculum. A score of "2" refers to items that students should be able to solve on the basis of the content covered, although the formulation of the item differs from the usual presentation in the class. A score of 3 indicates that the item is a typical question for the student in the class (would have appeared in a regular examination). The aggregation included all students in a class and not only those students belonging to the LOSO - cohort. Students with missing data are mostly omitted from the multilevel analyses, but in the aggregation process even students with missing values on other variables are included. On the other hand, we did not calculate group composition scores that are based on too small a fraction of the group. If less than 50% of the scores on a student-level variable in a class was available, the aggregated score for that class was not calculated. This procedure is expected to reduce the overall measurement error on the independent variables.

L.6 If the study uses qualitative methods, how well has diversity of perspective and content been explored?

Details

Not a study with qualitative methods.

L.7 If the study uses qualitative methods, how well has the detail, depth and complexity (i.e. the richness) of the data been con-

Details

Not a study with qualitative methods.
Section M: Quality of study - reporting

| M.1 | Is the context of the study adequately described? | Yes (please specify) | The study context is described in Van Damme et al., 2002 Sampling is explicitly described |
| M.2 | Are the aims of the study clearly reported? | Yes (please specify) | The study has attempted to be a fully fledged longitudinal school effectiveness study, paying particular attention to student background variables and compositional effect |
| M.3 | Is there an adequate description of the sample used in the study and how the sample was identified and recruited? | Yes (please specify) | Yes, the sample is described in Vab Damme et al 2002 p 386 |
| M.4 | Is there an adequate description of the methods used in the study to collect data? | Yes (please specify) | Questionnaires and scales administered to teachers, student assessment in reading, math and non cognitive outcomes |
| M.5 | Is there an adequate description of the methods of data analysis? | Yes (please specify) | Yes, multi level modelling was used, applying among others three level analyses, |
| M.6 | Is the study replicable from this report? | Yes (please specify) | Yes, procedures are described explicitly |
| M.7 | Do the authors avoid selective reporting bias? (e.g. do they report on all variables they aimed to study, as specified in their aims/research questions?) | Yes (please specify) | There is no sign of biased reporting, also none effects and counter intuitive effect were reported |

Section N: Quality of the study - Weight of evidence
<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
</table>
| N.1 Were users / relatives of users involved in the design or conduct of the study? | No, involvement is not relevant (please specify)  
No, involvement in never relevant |
| N.2 Was the choice of research design appropriate for addressing the research question(s) posed? | Yes, completely (please specify)  
The design was completely appropriate. In fact this series of studies is among the best in the field, from a methodological perspective |
| N.3 Have sufficient attempts been made to establish the repeatability or reliability of data collection methods or tools? | Yes, good (please specify)  
Yes, procedures are explicit, reliabilities are reported |
| N.4 Have sufficient attempts been made to establish the validity or trustworthiness of data collection tools and methods? | Yes, good (please specify)  
See previous point. Psychometric qualities of instruments are reported |
| N.5 Have sufficient attempts been made to establish the repeatability or reliability of data analysis? | Yes (please specify)  
Yes, see previous points |
| N.6 Have sufficient attempts been made to establish the validity or trustworthiness of data analysis? | Yes, good (please specify)  
Yes, see previous answers |
| N.7 To what extent are the research design and methods employed able to rule out any other sources of error/bias which would lead to alternative explanations for the findings of the study? | A lot (please specify)  
As stated before, in its longitudinal design, its elaborate controls and state of the art analysis this is an exemplary study in its field |
| N.8 How generalisable are the study results? | Details  
The sample might be representative for Flemish secondary schools |
| N.9 In light of the above, do the reviewers differ from the authors over the findings or conclusions of the study? | Not applicable (no difference in conclusions)  
No difference with the conclusions of the authors |
| N.10 Weight of evidence A: Taking account of all quality assessment issues, can the study findings be trusted in answering the study question(s)? | High trustworthiness  
Again this is a very good set of studies |
| **N.11 Weight of evidence B: Appropriateness of research design and analysis for addressing the question, or sub-questions, of this specific systematic review.** | **High**  
*The studies considered relevant school and class factors, only one or two factors had a positive effect. The study is of great relevance, since it is the best example of applying controls in term of background and compositional effects that I am aware of.* |
| --- | --- |
| **N.12 Weight of evidence C: Relevance of particular focus of the study (including conceptual focus, context, sample and measures) for addressing the question, or sub-questions, of this specific systematic review** | **High**  
*The study is very severe in controlling for non malleable variables, as a result the yield in terms of policy amenable variables is very limited. This could be seen as a "negative" outcome, but in fact it is not. School effectiveness is an academic field of research, and falsifying expecations is as important as finding support in the expected direction. The studies teach us that policy ameneable variables have much smaller impact than "given" background variables* |
| **N.13 Weight of evidence D: Overall weight of evidence** | **High**  
*Again: very high level research* |
6 Complete overview of references included in the research mapping

This is the total list of the 149 references included. Two references, indicated with *, were not available in time for them to be subject to a coding. The remaining 147 references refer to 111 studies analysed in this report, i.e. in many cases several references report different aspects of the same study.


Bearden, D., & et al. (1995). Effective Schools: Is There a Winning Combination of Administrators, Teachers, and Students? (No. (ED386801)).


Cepeley, P. E. (1999). Implementation of Title I Schoolwide Programs in Four Rural Virginia Schools (No. (ED438993)).


Coe, P., Keyes, M., Meehan, M., Orletskey, S., Lewis, S., Rigney, S., et al. (1999). Development and Validation of Successful Writing Program Indicators Based on Research in Continuously Improving and Continuously Declining Schools: Report of the Kentucky State Writing Project. Study of Writing Instruction in Kentucky Schools (No. (ED442602)).


Grisay, A. (1994). Effective and Less Effective Junior Schools in France: A Longitudinal Study on the School Environment Variables Influencing the Student's Academic Achievement, Study Skills, and Socio-Affective Development (No. (ED380864)).

Heck, R. H. (2007). Examining the Relationship between Teacher Quality as an Organizational Property of Schools and Students' Achievement and Growth Rates. Educational Administration Quarterly, 43(4), 399-432.


Kennedy, E., & et al. (1993). A Multilevel Analysis of Phase II of the Louisiana School Effectiveness Study (No. (ED361902)).


Mosenthal, J., Lipson, M., Mekkelsen, J., Russ, B., & Sortino, S. (2001). Elementary Schools Where Students Succeed in Reading (No. (ED459443)).


Picucci, A. C., Brownson, A., Kahlert, R., & Sobel, A. (2002). Driven To Succeed: High-Performing, High-Poverty, Turnaround Middle Schools. Volume I: Cross-Case Analysis of High-Performing, High-Poverty, Turnaround Middle Schools (No. (ED476107)).

Picucci, A. C., Brownson, A., Kahlert, R., & Sobel, A. (2002). Driven To Succeed: High-Performing, High-Poverty, Turnaround Middle Schools. Volume II: Case Studies of High-Performing, High-Poverty, Turnaround Middle Schools (No. (ED476108)).


Postlethwaite, T. N., & Ross, K. N. (1992). Effective Schools in Reading: Implications for Educational Planners. An Exploratory Study (No. (ED360614)).


Tanner, H., Jones, S., & Treadaway, M. (2000). The role of the middle managers in raising standards in Mathematics, Conference of Australian Association for research in education (pp. 17). Sydney: Australian Association for research in education.


Taylor, B., Pearson, D., Clark, K., & Walpole, S. (2000). Beating the Odds in Teaching All Students To Read: Lessons from Effective Schools and Accomplished Teachers (No. (ED450352)).


Texas Education Agency (2000). The Texas Successful Schools Study: Quality Education for Limited English Proficient Students. Texas Education Agency, Austin (No. (AD00-300-01))


Thomas, S. (1997). Differential secondary school effectiveness: examining the size, extent and consistency of school and departmental effects on GCSE outcomes for different groups of students over three years. School Field, 8(1-2), p57-95.


Young, D. J. (1999). Student Progress in Australian Schools: A Multilevel Multivariate Model (No. (ED431020)).

Young, D. J. (1999). The Usefulness of Value-Added Research in Identifying Effective Schools (No. (ED440810)).


Young, D. J., & Fraser, B. J. (1992). School Effectiveness and Science Achievement: Are There Any Sex Differences? (No. (ED356946)).

Yu, L., & White, D. B. (2002). Measuring Value Added School Effects on Ohio Six-Grade Proficiency Test Results Using Two-Level Hierarchical Linear Modeling (No. (ED481653)).

7 References

Listed here are only those references that are included in the textual commentary of the research mapping. A complete list of references included in the research mapping is given in Chapter 6.

**Bauer and Riphahn (2006)**


**Manning amd Pischke (2006)**


