THE MEANING OF SCHOOL EFFECTIVENESS

Prof. Jaap Scheerens

School effectiveness is defined in the way it is used in school effectiveness research. Basically this means that between school differences in students' performance are attributed to malleable school variables, after adjustments have been made for student background conditions. Research is guided by conceptual models in which school conditions are often defined at the level of the school context, the school as an organization, and the classroom. A brief summary will be given of the most important results from empirical school effectiveness research. Next the paper will look at conceptual and theory related interpretations that can be given to these results. The ambition of school effectiveness research is to establish malleable factors that are robust and generalizable across different contexts. Examples are: time, opportunity to learn and a structured approach to teaching that includes frequent monitoring of students' progress as classroom level conditions, and supportive ''educational'' leadership at school level. The theoretical basis for these factors will be analyzed by looking at more general theories as well as a few "middle range" theories and models. The balance on the appropriateness and usefulness of the school effectiveness research approach will be made up by addressing a few critical issues, such as: the alleged narrowness of the most frequently used effectiveness criteria and the usefulness for educational practice and policy.

Conceptualization

The elementary design of school effectiveness research is the association of hypothetical effectiveness enhancing conditions of schooling and output measures, mostly student achievement. The basic model from systems theory, shown in Figure 1 is helpful to clarify this basic design. The major task of school effectiveness research is to reveal the impact of relevant input characteristics on output and to "break open" the black box in order to show which process or throughput factors "work", next to the impact of contextual conditions. Within the school it is helpful to distinguish a school and a classroom level and, accordingly, school organizational and instructional processes.
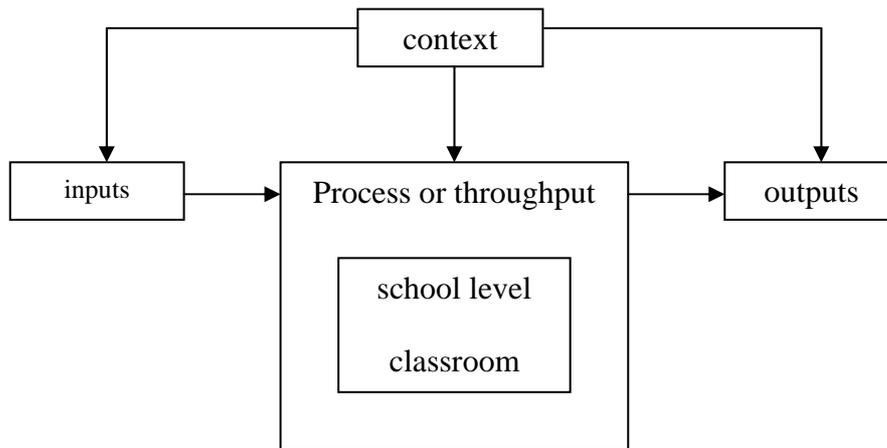
Figure 1: A basic systems model on the functioning of education

Research tradition in educational effectiveness varies according to the emphasis that is put on the various antecedent conditions of educational outputs. These traditions also have a disciplinary basis. The common denominator of the five areas of effectiveness research that will be distinguished is that in each case the elementary design of associating outputs or outcomes of schooling with antecedent conditions (inputs, processes or contextual) applies. The following research areas or research traditions will be considered in summarizing the research results obtained in developed countries:

1) *Research on equality of opportunities in education and the significance of the school in this.*
2) *Economic studies on education production functions.*
3) *The evaluation of compensatory programs.*
4) *Studies of unusually effective schools.*
5) *Studies on the effectiveness of teachers, classes and instructional procedures.*

In developing countries there is a strong predominance of studies of the education production function type. Relatively few of these have been expanded by including school organizational and instructional variables.

Below is a somewhat different presentation of the systems model that was introduced in the previous section. Here the central box is defined at the level of an organization, in our case, a school. The functioning of the organization is again seen inputs flowing into the central box into and by outputs being "somehow" produced (see Figure 2).

input → organisation as a black box → output

Figure 2: *The organisation as a black box.*

In Figure 2 it is assumed that within the black box *processes* take place that transform inputs into outputs. When it is attempted to further describe these processes in terms of which process characteristics are most effective in obtaining desired levels of outputs, the model of Figure 1 becomes more elaborate. This model is often used as a conceptual framework to summarise the results of school effectiveness research.

In Figure 3 an example of such an ordered summary is shown (cf. Scheerens, 1989).
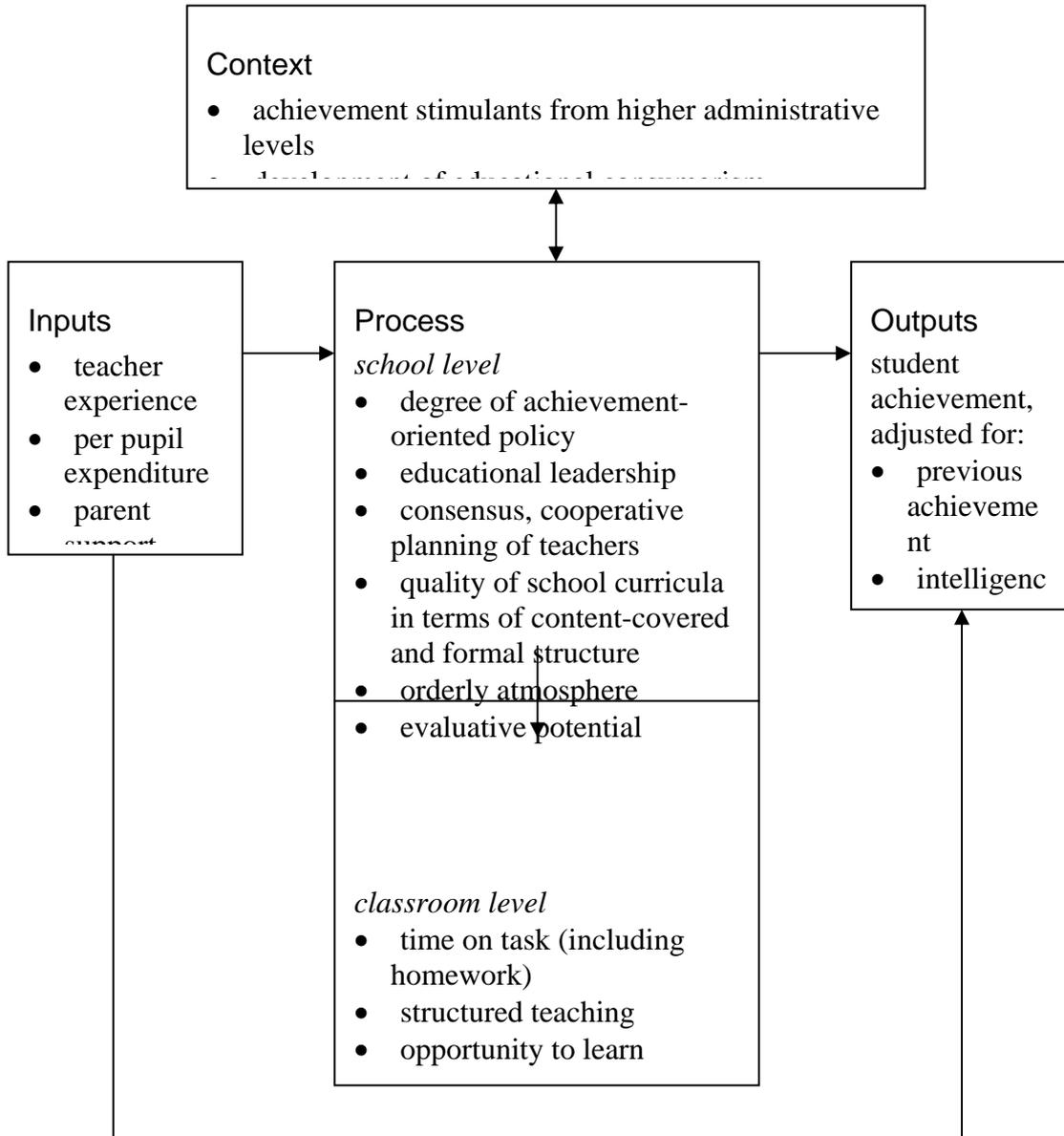
Figure 3:   *A summary of the findings from school effectiveness research, from Scheerens, 1989.*

The notion of quality inherent in integrated school effectiveness models like the one in Figure 3 is that:

a)   outputs are the basic criteria to judge educational quality;

b)   in order to be able to properly evaluate output, achievement or attainment measures should be adjusted for prior achievement and other pupil intake characteristics; in this way the value added by schooling can be assessed;

c)   in selecting variables and indicators to assess processes and context one should look for those factors that have been shown to be correlated with relatively high output, adjusted in terms of "added-value" as described above;

d) the model is a multi-level model, uniting effectiveness enhancing conditions at system, school, classroom and individual student level.

It should be noted that educational effectiveness models do not prescribe the types of outputs that should be used to assess quality. In principle all types of outputs, cognitive or non-cognitive could be inserted in the right-hand box of Figure 3. In the actual practice of school effectiveness research, however, cognitive outcomes, mostly in terms of achievement in core-subjects like reading, arithmetic, and language, have predominated. The process factors shown in the middle section of Figure 3, might well be somewhat different if non-cognitive outcomes or less subject-matter tied cognitive outcomes would have been used in the actual research studies.

It should also be noted that there is still quite a lot of uncertainty about the selection of process factors such as indicated in the figure. The available knowledge-base is far removed from a situation where it would be possible to make precise predictions on the likely added value of schooling, given the state of certain processes and inputs (in a subsequent section of this paper this will be documented further).

To the degree that educational effectiveness models provide an acceptable operational definition of quality, they can also be used as a guideline in the design of instruments for school evaluation and school self-evaluation. The previously mentioned points: *a* (focus on outcomes), *b* (proper adjustment of outcomes) and *c* (measure process characteristics associated with high added value) mentioned in the above can be read as guidelines to make choices with respect to instrumentation.

However, a broader perspective on quality can be considered. Such a broader perspective can be obtained from multiple orientations towards organisational effectiveness (cf. Scheerens, Glas and Thomas, 2003). In this context I will not go into these broader models of organisational effectiveness.

Qualitative research reviews on school effectiveness

Research known under labels like "identifying unusually effective schools" or the "effective schools movement" can be regarded as the type of research that most touches the core of school effectiveness research. In Coleman's and Jencks' surveys the inequality of educational opportunity was the central problem. In economic-related input-output studies the school was even conceived as a "black box". In the still to be discussed research on the effectiveness of classes, teachers and instruction methods, education characteristics on a lower aggregation level than the school are the primary research object.

Effective school research is generally regarded as a response to the results of studies like Coleman's and Jencks' from which it was concluded that schools did not matter very much when it came down to differences in levels of achievement. From titles such as "Schools can make a difference" (Brookover et al., 1979) and "School matters" (Mortimore et al., 1988) it appears that refuting this message was an important source of inspiration for this type of research. The most distinguishing feature of effective schools research was the fact that it

attempted to break open the "black box" of the school by studying characteristics related to organization, form and content of schools. The results of the early effective schools research converged more or less around five factors:

strong educational leadership;

emphasis on the acquiring of basic skills;

an orderly and secure environment;

high expectations of pupil attainment;

frequent assessment of pupil progress.

In the literature this summarizing is sometimes identified as the "five-factor model of school effectiveness". It should be mentioned that effective schools research has been largely carried out in primary schools, while at the same time studies have been largely conducted in inner cities and in predominantly working-class neighborhoods.

In more recent contributions effective schools research became more integrated with education production function and instructional effectiveness research, in the sense that a mixture of antecedent conditions was included, studies evolved from comparative case-studies to surveys and conceptual and analytical multi-level modeling took place to analyze and interpret the results. Numerous reviews on school effectiveness have been published since the late seventies. Early reviews are those by Anderson (1982), Cohen (1982), Dougherty (1981), Edmonds (1979), Murnane (1981), Neufeld et al. (1983), Purkey and Smith (1983), Rutter (1983), Good and Brophy (1986), Ralph and Fenessey (1983), Kyle (1985), and Sweeney (1982). More recent reviews are those by Levine and Lezotte (1990), Scheerens (1992), Creemers (1994), Reynolds et al. (1993), Sammons et al. (1995), and Cotton (1995).

The focal point of interest in the reviews is the "what works" question; typically the review presents lists of effectiveness enhancing conditions. There is a fairly large consensus on the main categories of variables that are distinguished as effectiveness enhancing conditions in the reviews, also when earlier and more recent reviews are compared. Table 3 summarizes the characteristics listed in the reviews by Purkey and Smith (1983), Scheerens (1992), Levine and Lezotte (1990), Sammons et al. (1995), Cotton (1995).

Table 1:   *Effectiveness enhancing conditions of schooling in five review studies (italics in the column of the Cotton study refers to sub-categories).*

Purkey & Smith, 1983	*Levine & Lezotte, 1990*	*Scheerens, 1992*	*Cotton, 1995*	*Sammons, Hillman & Mortimore, 1995*

Achievement-oriented policy; cooperative atmosphere, orderly climate	Productive climate and culture	Pressure to achieve, consensus, cooperative planning, orderly atmosphere	Planning and learning goals, curriculum planning and development	Shared vision and goals, a learning environment, positive reinforcement
Clear goals on basic skills	Focus on central learning skills		Planning and learning goals *school wide emphasis on learning*	Concentration on teaching and learning
Frequent evaluation	Appropriate monitoring	Evaluative potential of the school, monitoring of pupils' progress	Assessment (district, school, classroom level)	Monitoring progress
In-service training/ staff development	Practice-oriented staff development		*Professional development* collegial learning	A learning organization
Strong leadership	Outstanding leadership	Educational leadership	School management and organization, leadership and school improvement, leadership and planning	Professional leadership
	Salient parent involvement	Parent support	Parent community involvement	Home school partnership
Time on task, reinforcement, streaming	Effective instructional arrangements	Structured, teaching, effective learning time, opportunity to learn	Classroom management and organization, instruction	Purposeful teaching
High expectations	High expectations		Teacher student interactions	High expectations
				Pupil rights and responsibilities
			Distinct-school interactions	
			Equity	
			Special programs	
		External stimuli to make schools effective		

Physical and
material school
characteristics

Teacher
experience

School context
characteristics

Consensus is largest with respect to the factors: achievement orientation (which is closely related to "high expectations"); co-operation; educational leadership; frequent monitoring; time, opportunity to learn and "structure" as the main instructional conditions.

A quantitative review

In Table 6 (cited from Scheerens and Bosker, 1997) the results of three meta-analysis and a re-analysis of an international data set have been summarized. The results concerning resource input variables are based on the re-analysis of Hanushek's (1989) summary of results of production function studies that was carried out by Hedges, Laine & Greenwald, 1994. As stated before this re-analysis was criticized, particularly the unexpectedly large effect of per pupil expenditure. The results on "aspects of structured teaching" are taken form meta-analyses conducted by Fraser, Walberg, Welch and Hattie, 1987. The international analysis was based on the IEA Reading Literacy Study and carried out by R.J. Bosker (Scheerens & Bosker, 1997, ch. 7). The meta-analysis on school organizational factors, as well as the instructional conditions "opportunity to learn", time on task", "homework" and "monitoring at classroom level", were carried out by Witziers and Bosker and published in Scheerens & Bosker, 1997, Ch. 6. The number of studies that were used for these meta-analyses varied per variable, ranging form 14 to 38 studies.

The results in this summary of reviews and meta-analyses indicate that resource-input factors on average have a negligible effect, school factors have a small effect, while instructional have an average to large effect. The conclusion concerning resource -input factors should probably be modified and "nuanced" somewhat, given the results of more recent studies referred to in the above, e.g. the results of the STAR-experiment concerning class-size reduction. There is an interesting difference between the relatively small effect size for the school level variables reported in the meta-analysis and the degree of certainty and consensus on the relevance of these factors in the more qualitative research reviews. It should be noted that the three blocks of variables depend on types of studies using different research methods. Education production function studies depend on statistics and administrative data from schools or higher administrative units, such as districts or states. School effectiveness studies focussing at school level factors are generally carried out as field studies and surveys, whereas studies on instructional effectiveness are generally used on experimental designs. The negligible to very small effects that were found in the re-analysis of the IEA data-set could be partly attributed tot the somewhat "proxy" and superficial way in which the variables in question were operationalized as questionnaire items. An additional finding from international

comparative studies (not shown in the table) is the relative inconsistency of the significance of the school effectiveness correlates across countries, also see Scheerens, Vermeulen and Pelgrum, 1989 and Postlethwaite and Ross, 1992.

Table 2: *Review of the evidence from qualitative reviews, international studies and research syntheses*

	Qualitative reviews	International analyses	Research syntheses
Resource input variables:			
Pupil-teacher ratio		-0.03	0.02
Teacher training		0.00	-0.03
Teacher experience			0.04
Teachers' salaries			-0.07
Expenditure per pupil			0.20
School organizational factors:			
Productive climate culture	+		
Achievement pressure for basic subjects	+	0.02	0.14
Educational leadership	+	0.04	0.05
Monitoring/evaluation	+	0.00	0.15
Cooperation/consensus	+	-0.02	0.03
Parental involvement	+	0.08	0.13
Staff development	+		
High expectations	+	0.20	
Orderly climate	+	0.04	0.11
Instructional conditions:			
Opportunity to learn	+	0.15	0.09
Time on task/homework	+	0.00/-0.01 (n.s.)	0.19/0.06
Monitoring at classroom level	+	-0.01 (n.s.)	0.11 (n.s.)
Aspects of structured teaching:			
-cooperative learning			0.27
-feedback			0.48
-reinforcement			0.58
Differentiation/adaptive instruction			0.22

In Table 2 (cited from Scheerens and Bosker, 1997) the results of three meta-analysis and a re-analysis of an international data set have been summarized. The results concerning resource input variables are based on the re-analysis of Hanushek's (1989) summary of results of production function studies that was carried out by Hedges, Laine & Greenwald, 1994. As stated before this re-analysis was criticized, particularly the unexpectedly large effect of per pupil expenditure. The results on "aspects of structured teaching" are taken form meta-

analyses conducted by Fraser, Walberg, Welch and Hattie, 1987. The international analysis was based on the IEA Reading Literacy Study and carried out by R.J. Bosker (Scheerens & Bosker, 1997, ch. 7). The meta-analysis on school organizational factors, as well as the instructional conditions "opportunity to learn", time on task", "homework" and "monitoring at classroom level", were carried out by Witziers and Bosker and published in Scheerens & Bosker, 1997, Ch. 6. The number of studies that were used for these meta-analyses varied per variable, ranging form 14 to 38 studies.

The results in this summary of reviews and meta-analyses indicate that resource-input factors on average have a negligible effect, school factors have a small effect, while instructional have an average to large effect. The conclusion concerning resource -input factors should probably be modified and "nuanced" somewhat, given the results of more recent studies referred to in the above, e.g. the results of the STAR-experiment concerning class-size reduction. There is an interesting difference between the relatively small effect size for the school level variables reported in the meta-analysis and the degree of certainty and consensus on the relevance of these factors in the more qualitative research reviews. It should be noted that the three blocks of variables depend on types of studies using different research methods. Education production function studies depend on statistics and administrative data from schools or higher administrative units, such as districts or states. School effectiveness studies focussing at school level factors are generally carried out as field studies and surveys, whereas studies on instructional effectiveness are generally used on experimental designs. The negligible to very small effects that were found in the re-analysis of the IEA data-set could be partly attributed tot the somewhat "proxy" and superficial way in which the variables in question were operationalized as questionnaire items. An additional finding from international comparative studies (not shown in the table) is the relative inconsistency of the significance of the school effectiveness correlates across countries, also see Scheerens, Vermeulen and Pelgrum, 1989 and Postlethwaite and Ross, 1992.

More recent contributions to the study of instructional effectiveness

In this section the results of some more recent contributions and reviews briefly summarized, on the basis of work by Anderson, 1991, 2004, Brophy, 2001, Baumert et al., 2000, NCS, 2002, Muijs and Reynolds, 2001, OECD, 2003. In these reviews, a strong corroboration of the main characteristics of effective instruction as laid out in earlier reviews (Scheerens, 2003) can be discerned. In addition to this consolidation in the knowledge base there are a few additional newer trends. These are the following:

- a reconsideration of personal characteristics of effective teachers
- more attention to the teaching of higher order skills, self-regulated learning and "constructivist" approaches
- a strong re-statement of the fact that teaching is about facilitating learning, by considering learning activities and student engagement

In the United States the issue of effective teacher characteristics is receiving much attention in the debate about standards for teaching competency (Darling Hammond, 2000). Empirical studies indicate that subject matter mastery and verbal skills are important assets of teacher effectiveness. In the United Kingdom, Hay McBerr (2000, cited by Anderson, 2004) identified twelve characteristics, in the sense of relatively stable traits, associated with effective teachers. Among others he mentions characteristics like flexibility, trustworthiness, and commitment. An interesting feature in Hay McBerr's list is the "drive for improvement". This trait is similar to the "relentlessness" that is emphasized in Slavin's "Success for All" program, 1999) and what Anderson and Pellicer (1998) have called "zero tolerance to failure". The dimension of confidence is associated with the "high expectations" factor in the school and classroom climate, as one of the frequently identified factors of effective schooling. It seems very likely that these motivationally oriented personal characteristics of teachers have an impact on climate and culture of the school and the classroom.

There is one other dimension in which more recent contributions return to an aspect that was also present in the very early publications on teaching effectiveness, as Gagne's conception about "the conditions of learning" (Gagne, 1972) and the Caroll model, (Caroll, 1963). This is attention for student engagement and learning strategies as the ultimate "mediator" between teaching activities and student outcomes. In the OECD study on "student approaches to learning" a range of variables related to engagement is discerned, variables like "self-efficacy", "instrumental motivation" and subject matter interest (OECD, 2003). As learning strategies a distinction is made between memorization strategies, elaboration strategies and control strategies. Control strategies have a significant association with reading performance. Control strategies refer to students' ensuring that their learning goals are reached. "These strategies involve checking what one has learned and working out what one still has to learn, allowing learners to adapt their learning to the task at hand", (OECD, 2003, p.13). In a way these control strategies are the pendant of the main features of "structured teaching" and direct instruction, where it is the teacher who actively orders and controls the teaching and learning situation. When putting these two orientations, structured teaching on the one hand, and students effectively employing control strategies next to one another the following types of associations can be discerned:

- structured teaching happens as a *substitute* for student control strategies
- structured teaching happens as an *additional support* for student control strategies
- structured teaching happens as a *model and example* to enhance student control strategies
- structured teaching happens as a *suppressor* of student control, because students are not given sufficient leeway to develop and manifest this behavior themselves.

Weaker students in primary and secondary education are more likely to benefit from the first two alternatives, whereas the last two alternative combinations are more probable when dealing with better students in secondary education (where obviously the third alternative is a positive and the fourth a negative example).

The results of these more recent reviews, in the sense of the most important instructional conditions that were referred to, are summarised in the table 6 below. The table includes the main observation categories of a classroom observation schedule that is currently being used in an internationally comparative study of SICI, an international organisation of educational Inspectorates in Europe.

Table 3:   *Summary of recent reviews and the observation categories of the Dutch*
            Inspectorate

Teaching (Anderson)

enacted curriculum
classroom physical environment
classroom climate
classroom organisation & management
actual teaching
          pre-conditions (lesson planning)
          communication with students
          stimulating involvement

Brophy

opportunity to learn
curricular alignment
supportive classroom climate
achievement expectations
cooperative learning
goal-oriented assessment
coherent content; clear explanations
thoughtful discourse
establishing learning orientations
sufficient opportunities for practice and application
scaffolding student's task engagement

Dutch inspectorate

- learning time
- support in climate
- challenge in climate
- structure in teaching
- activating students
- teaching learning strategies
- attainment/teacher focus on attention

Baumert et al.

quantity and quality of instruction
teacher student relations
student student relations

Theoretical interpretation of the findings from educational effectiveness research

Preliminary issues

After the presentation of these lists of factors that result from reviews of the research literature on educational effectiveness one may feel the urge to make up the balance in a more concise way. This could take the form of trying to identify the more robust core factors that emerge from the research literature reviews. Or, even more ambitiously, to try and discover more basic principles of effective schooling that could explain "why" what works in education. For this latter task one would like to connect the most prominent findings to existing theories so that they could be explained by subsuming them under more general theory embedded principles.

Before going in this direction a few preliminary issues should be clarified, however.

First of all the "action orientation" inherent in the concept of school effectiveness should be addressed. The goal concept is central in the definition of effectiveness. As stated before, all sort of substantive choices could be made in defining goals. But in a research context one would need a certain clarity and constancy in the kinds of educational goals that serve to identify what would be called "effect criteria" in a research context. As a matter of fact this constancy is present in educational effectiveness research in the sense that the bulk of studies has concentrated on student achievement results in language or mathematics and reading, at primary and lower secondary level. From an epistemological point of view this represents a restriction in the generalizability of the outcomes and from a political and normative point of view it limits the relevance of the research outcomes for reforms that address a broader range of desired competencies. Apart from this issue of the constancy of effect criteria the second implication of the action orientation of educational effectiveness is the concept of instrumentality. The focus in educational effectiveness research is on malleable conditions, that is factors than can be changed, controlled or manipulated by policy planners, school managers, teachers or students. In the neo-positivist research tradition questions of instrumentality are considered to be formally identical to questions of causality. But this view is contested from other epistemological positions concerning "intentionality" (Von Weiss, 19..) and complexity (Morin, 19..). It is beyond the scope of this presentation to address these issues further and the position that instrumental questions can be seen as causal questions will be accepted at face value.

Secondly, a few things more should be said about the multi-level aspects of the relationships laid bare by empirical school effectiveness research. Apart from statistically controlling for dependencies resulting from units on a lower level being "part of" units at higher levels, more substantive interpretations are at stake. One could choose to decompose the context in which educational effectiveness is studied into different mono-level layers. But typically the more interesting aspects of educational effectiveness seem to be the interactions between units and actors at various levels because these seem to be more in line with action and control interpretations, where conditions at higher levels are assumed to facilitate processes at lower levels. For example, Bosker and Scheerens, (1995) discuss different interpretations of the relationship between a school manager and his staff (steering, facilitating, providing an

example that could be mirrored, and buffering). Similarly alternative multi-level causal models could be explored statistically, by comparing direct and indirect causal links, additive and interactive effects, and recursive relationships. In search of robust principles of effective schooling one might look for principles that apply at various levels and consequently analyze consistency in the application of corresponding malleable characteristics at various levels. For example, the principle of "retro-active planning" or evaluation driven control, has manifestations concerning evaluation and assessment at the level of national educational systems, at school level and at classroom level, which in principle could be mutually reinforcing. It seems to be more interesting to study educational effectiveness phenomena from a multi-level rather than from a uni-level perspective because the control perspective is better represented in this way.

A third issue relates to the generality versus specificity of the phenomena that are studied in educational effectiveness research. The preference for a multi-level perspective implies that research questions and theories will address macro-level issues of educational policy making and systemic approaches to education, meso-level issues of organizational effectiveness of schools, and micro level issues of effective instruction. But also at each level relevant questions could be asked about general versus more specific orientations. At the level of classroom teaching there is the issue of instructional strategies that are relevant across subject matter areas, and subject specific didactics. At school level the scope and stability of school effects are a case in point. A narrow interpretation of an effective school would be a school that manifests relatively high achievement in one subject matter area, at one grade level at one point in time. A more critical investigation would try to establish these observations at more than one grade level, several subjects and with a certain degree of stability over time. At system level one might critically want to assess the effect of a certain type of policy, for example an elaborated accountability policy, across school levels (e.g. primary, secondary and vocational). In my opinion educational effectiveness research has a primary interest in robust, generalizable factors. Findings of the kind that Walberg once indicated with the phrase: "what is good for the geese is good for the gander". However, to be able to arrive at such statements a range of more specifically oriented studies should ideally serve as an important part of the evidence base.

In the subsequent sections a set of theoretically embedded principles will be presented that might be seen as potentially explaining the findings of empirical educational effectiveness research. These principles vary from rather general principles from policy and management studies to more specific middle range theories from disciplines like didactics, psychology and micro-economics. Particularly with respect to these latter, more "micro" theories the selection is merely illustrative without the ambition of being exhaustive.

General control theoretical perspectives

The theoretical approaches that will be discussed in this section occur in planning theory and in management theory as far as their "procedural" characteristics are concerned. When

organizational structural implications are included, they are also related to organizational theories and images, such as the bureaucracy and the learning organization.
.

Overview of theories and core mechanisms	
Theory	*Core mechanism*
Rational control	Proactive structuring
Public choice	Market-mechanisms
Retroactive planning	Cybernetic principle
Contingency theory	Fit
Theory of autopoietic systems	Self-organization

In the summary table above some important managerial meta-theories and their associated core mechanisms are presented. Here a relatively brief synthesis is given, for a more elaborate treatment see Scheerens, 1997, Scheerens, Glas and Thomas, 2003.

The rationality model that seeks to establish an instrumental and scientific approach to planning and control is well-known. Its imperative being: think before you act. Planning in terms of explicit goals and methods that are most effective and efficient in reaching these goals is still a predominant orientation, also in the field of educational reforms (compare the so-called logical framework approach). At the same time this model has been criticized enormously by scholars in the field of public administration, particularly for its stringent information demands and implicit assumptions of social harmony and consensus about social goals and means. The prototype organizational model that uses this principle of pre-structuring and "uncertainty avoidance" is the bureaucracy. Modern applications are neo-institutional economics, with a similar emphasis on formally established institutions and enforcement mechanisms, and quality management systems like the ISO and EFQM models. In these latter models there is a central pre-occupation with formally established procedures and structural arrangements.

Public choice models take into account the fact that actors may not only pursue organizational goals, but also personal goals. The "royal road" to diminish the in-efficiency enhancing tendencies is to install market conditions and competition in public sector organizations, including education.

Retro-active planning starts from the premise that action often precedes thinking, and that rational behaviour might take the form of "rationalization" or rational reconstruction rather

than pro-active structuring and planning. In more practical terms this approach suggests a change in the starting point of the planning cycle: start with an assessment of the current situation, identify strong and weak points, and plan and implement corrective, improvement oriented action. Learning from (performance) feedback is central in this approach. Different forms of accountability, evaluation and assessment are important supportive instruments for retro-active planning.

Contingency-theory emphasizes the position that the effectiveness of organizational arrangements depends on situational conditions. Fitting approaches in teaching and instruction to characteristics of the target population, as in adaptive teaching, confirm to the contingency paradigm. The same could be said with respect to allowing schools a sufficient degree of autonomy to adapt to the local environment.

The theory of autopoietic systems (Maturana & Varela, 1980; Luhmann, 1995) emphasizes the importance of self-organization and self-reference in organizations. This theoretical approach might have important implications for a completely different view on issues of school effectiveness and school improvement. By emphasizing endogenous factors, and internal dynamics of organizations, it puts the whole issue of adaptation to the environment and external control on a different footing. In more practical terms the implications from this theoretical approach seem to be relevant to rethink issues of self-evaluation and feedback, the implementation of externally stimulated innovations (do innovations penetrate to the level of an organizations pattern of self-references?) and the whole issue of school autonomy.

More specific models and theories

In the summary table below more specific "middle" range theories or models are mentioned, in connection to the more general principles that were discussed in the previous sub-section.

Examples of more specific theories	
General theories and core mechanisms	*Specific theories*
Rational control and synoptic planning	Carroll model of effective instruction
Public choice and market mechanisms	Expectations and realizations of school choice
Retroactive planning and cybernetics	Theory of performance feedback
Contingency theory and fit	Adaptive teaching
Theory of autopoietic systems and self	Managing composition effects

organization	

Due to lack of space, not all of these more specific models and theories will be discussed in detail in this paper in much detail; some can only get a cursory treatment. The most detailed treatment will be given to the Carroll model, because it deals with the core process of education.

The Carroll model

The Carroll model (Carroll, 1963) is usually considered as the starting point of modelling instructional effectiveness. It consists of five classes of variables that are expected to explain variations in educational achievement. All classes of variables are related to the time required to achieve a particular learning task. The first three factors are directly expressed in terms of amounts of time, while the two remaining factors are expected to have direct consequences for the amount of time that a student actually needs to achieve a certain learning task. The five classes of variables are:
- *aptitude*; variables that determine the amount of time a student needs in order to learn a given task under optimal conditions of instruction and student motivation;
- *opportunity to learn*; the amount of time allowed for learning;
- *perseverance*; the amount of time a student is willing to spend on learning the task or unit of instruction.
- *quality of instruction;* when the quality of instruction is sub-optimal, the time needed for learning is increased;
- *ability to understand instruction*, e.g. language comprehension, the learners' ability to figure out independently what the learning task is and how to go about learning it (Carroll, 1963, 1989).

The model can be seen as a general, encompassing causal model of educational achievement. In an attempt to formulate an encompassing model of educational productivity (Walberg, 1984) the basic factors of the Carroll model remained intact, while an additional category of environmental variables was included.

Numerous research studies and meta-analyses have confirmed the validity of the Carroll model (see chapter 5). The Carroll model has also been the basis for Bloom's concept of mastery learning (Bloom, 1968) and is also related to "direct instruction", as described by Rosenshine (1983).
Characteristics of mastery learning are:
1) Clearly defined educational objectives.
2) Small discrete units of study.
3) Demonstrated competence before progress to later hierarchically related units.

4) Remedial activities keyed to student deficiencies.

5) Criterion-referenced rather than norm-referenced tests (Block & Burns, 1970).

Direct instruction also emphasizes structuring the learning task, frequent monitoring and feedback and high levels of mastery (success rates of 90 to 100% for initial tasks) in order to boost the self-confidence of the students.

The one factor in the original Carroll model that needed further elaboration was "quality of instruction". As Carroll pointed out himself in a 25-year retrospective of his model, the original formulation was not very specific about the characteristic of high-quality instruction "but it mentions that learners must be clearly told what they are to learn, that they must be put into adequate contact with learning materials, and that steps in learning must be carefully planned and ordered" (Carroll, 1989, p. 26).

The cited characteristics of mastery learning and direct instruction are to be seen as a further operationalization of this particular factor, which is of course one of the key factors (next to providing optimal learning time) for a prescriptive use of the model. It should be noted that Carroll's reference to students who must be put into adequate contact with learning materials, developed into a concept of "opportunity to learn" different from his own. In Carroll's original formulation, opportunity to learn is identical to allocated learning time, while now opportunity to learn is mostly defined in terms of the correspondence between learning tasks and the desired outcomes. Synonyms for this more common interpretation of opportunity to learn are: "content covered" or "curriculum alignment" (Berliner, 1985, p. 128). In more formal mathematical elaborations the variable "prior learning" has an important place (Aldridge, 1983; Johnston and Aldridge, 1985).

The factor allocated learning time has been further specified in later conceptual and empirical work. Karweit and Slavin (1982), for instance, divide *allocated learning time* (the clock time scheduled for a particular class) into *procedural time* (time spent on keeping order, for instance) and *instructional time* (subject matter related instruction) and *time on task* (the proportion of instructional time during which behaviour appropriate to the task at hand took place).

Ability to understand instruction can be seen as the basis for further elaboration in the direction of learning to learn, meta-cognition, etc. The comprehensiveness of the Carroll model is shown by this potential to unite two schools of instructional psychology, the behaviouristically inclined structured teaching approaches and the cognitivist school (cf. Bruner, 1966; De Corte & Lowyck, 1983). In this context the crucial factor seems to be the degree of pre-structuring and "scaffolding" of students' learning in comparison to providing space to independent learning. This variable could also be seen as the critical variable on which more behaviourist and "constructivist" orientations to teaching and learning could be compared for there effectiveness. This is clearly a matter in need of more documented treatment, but the current research evidence seems to support the structured approach. This issue may be used to illustrate the call for "robust" factors in educational effectiveness research. From a fundamental research perspective the question of structure and independence in learning arrangements would clearly call for studying interactions and contextual contingencies. From a more practice oriented perspective a more global assessment might be required. Brophy and Good's (1986) conclusion that highly

structured teaching works equally well for the acquisition of complicated cognitive processes in secondary education, and not just for rote learning in primary schools, is a case in point.

In summary one could say that the empirical research generated by the Carroll model and its off-springs yielded three core factors of effective instruction:

- effective teaching time
- opportunity to learn (also in the sense of content covered)
- a structured approach to teaching in which teaching closely monitors and "scaffolds" learning processes

On the side of the learner, the model underlines the importance of aptitudes, student motivation and engagement.

The Carroll model applies to the micro level of the teaching and learning situation. If one likes one could see similarities, however, between the issue of structure and independence at micro-level and the issue of centralization and decentralization at system and school level. International comparative studies have provided some evidence for a mix of structure and independence that is characterized by centralization in the curriculum domain and autonomy for schools in other domains, like financial and personnel management (Woessmann, 2000, 2003)

Choice

Public choice theory challenges the "synoptic" and integrative assumptions of rational planning, and focuses on the rationality of individual actors. This implies great doubts in the possibility to effectively manage organizations in the public sector on the basis of administrative control and monitoring. The standard remedy of economic theory, that is providing opportunity for market mechanisms, applied to the functioning of educational organizations, has lead to an interest in school choice as an effectiveness enhancing mechanism. The basic assumption being that free choice of schools by the parents of students, will force schools to become effective and efficient. In its turn this assumption presupposes that parents as rational actors will choose the qualitatively best schools. Empirical research studies that show generally better learning results of private than for public schools, could be seen as providing evidence in support of the choice model. But the claims of the choice protagonist need to be examined critically, for several reasons. The empirical evidence which favours private schools is contested, since it often appears that the advantage of private over public schools disappears once adjustments are made for the socio-economic background of the student intake (e.g. Scheerens and Visscher, 2004). More scepticism follows when results of empirical studies are considered on the way parents actually decide on the selection of a school for their child. First of all, in reality the actual choice options are few, due to geographical conditions, but also because of the fact that in industrialized countries the large majority of schools do not differ significantly in output quality. Empirical research further points out that parents tend to value other criteria more than the performance of schools in terms of student achievement. Examples are: the reputation of the schools with respect to safety; special profiles of the school and the geographical closeness of the school. (Bosker and Scheerens, 1999) From an equity perspective choice is likely to produce more segregation and inequality, because it is likely that particularly the better educated parents will make use of the

opportunities that choice offers. The "side effect" of the choice philosophy, namely the creation of better outcome based information on schools, as in school performance reporting, may be the only real effectiveness enhancing result. Not so much because this information is used well by parents (for making a rational school choice), nor as a tool for school improvement, but rather as a general stimulus for schools to become result oriented (Bosker and Scheerens, 1999).

Effective performance feedback

Frequent monitoring of student progress is one of the classic factors that school effectiveness research has put forward. More institutionalized forms of assessment, like examinations and educational testing in the service of administrative accountability, are also seen as important vehicles to maintaining and improving quality in education. At school level, school self-evaluation is seen as one of the levers of school improvement. There is some empirical evidence that developed evaluation and monitoring at system, school and classroom level is positively associated with performance (Scheerens, 2002- Silver Bullet presentation Toronto).

The assumption on the way evaluation works is that it provides feedback on actual performance, which can then be used to guide improvement oriented activity. Crucial assumption are that "the message gets across" and that evaluation results are actually used. Various strands of literature indicate that this cannot be taken for granted. Studies on the use of the results from program evaluation document many examples of non-use or politically biased use of evaluation results (cf Weiss, geef meest recente referentie.). Studies on "high versus low stakes testing" point at the importance of the motivation of the students who are doing the test (Beaton, 19..). Studies on feedback and performance feedback indicate that the effectiveness of feedback depends on certain conditions (Visscher and Coe, 200., Kluger and DeNissi, 1996). The most important positive and negative conditions concering evaluation use an effectiveness of feedback are summarized in the two tables below.

Evaluation use

Positive conditions	*Negative conditions*
• Quality of evaluation methods	• Doubts about quality and credibility
• Perceived quality of evaluation methods (accuracy, relevance)	
• Credibility of information providers	
• "moderate" stakes of the evaluees	• Stakes high and stakes very low
• Quality of communication aspects	• "Bad" communication e.g. incomprehensible reporting
• Partial confidentiality of evaluation results	• Unnecessary disclosure of detailed technical information
	• Strategic behavior (Hawthorne-like effects)

Table 4: Positive and negative conditions concerning the use of evaluation results

A few of the entrances in table XX need some further explanation.

- When stakes are very high evaluees are likely to show resistance and avoidance behavior as well as strategic behavior aimed at obtaining more favorable results. When stakes are extremely low the evaluees might not bother to try and do well on tests. Moderate stakes would be when evaluees see the evaluation results as relevant, experience procedures and standards as fair and believe that the results will not results in dramatic administrative decisions about their own position or reputation.
- Communication has many aspects; a lot of attention should be given to a proper introduction about the scope and nature of the evaluation; ongoing communication between evaluators and evaluees should be possible; written communication, reports etc. should be accessible and clear.
- In finding an optimum on the dimension of "openness" vs. confidentiality of evaluation results the following criterion might be considered: information that has detail about technical and operational factors is only to be disclosed to the actors that have direct responsibility for using this information. In this sense the existing patterns of centralization and decentralization are helpful in answering questions as to what information should be provided to what group of actors. Information on standard attainment could also be provided to higher organizational/administrative levels.

feedback

Positive conditions	*Negative conditions*
• feedback not just about standard attainment but also instrumental information, to the extend that feedback recipients can actually use this information	• feedback exclusively about standard attainment (stimulates a judgmental rather than an instrumental application of feedback)
• the above condition can reinforce the task related- (or *perceived* task related) nature of the feedback	• feedback is taken "personal" and purely judgmental
• standards experienced as realistic	• standards perceived as unrealistic
• feedback appeals to intrinsic motivation	• feedback appeals to extrinsic motivation (implication: doubts about

	incentive schemes in education)
• negative feedback is superior in stimulating incremental task related learning progress • positive feedback stimulates overall motivation if received from credible and relevant sources	

Table 5: Positive and negative conditions affecting effective use of feedback (sources, among others: Kluger and DeNisi, 1996)

The following points implied in table 2 are worth noting:
- the research findings on the effective application of feedback underline the importance of task related information, which does not only show which level of attainment has been reached, but also hints at background conditions that could explain why attainment was at a certain level;
- these research findings (as the results on evaluation use) again underline the sensitivity of people who are the "object" of evaluations or inspections of judgements and the importance of conditions where there is trust in the fairness and reliability of evaluations and inspections;
- these research findings stress the importance of setting standards at a level that it appropriate and attainable for the evaluees;
- the research findings speak in favor of stimulating intrinsic as compared to extrinsic motivation (in this sense this line of psychological research contrasts strongly with the economic paradigm on utilities, based on monetary incentives).

These research findings are not so much one coherent theory on the way the "cybernetic principal" works in social settings, as a set of results of pieces of more fundamental research. As such they challenge to look further than "just" simply applying good evaluation, assessment and monitoring tools. Instrumentally relevant information, "moderate stakes" and confidence in the agents or institutions that are responsible for monitoring and evaluation, seem to be the core conditions of effective evaluation and feedback.

Adaptive teaching

Contingency theory has a basic message that goes squarely against the idea of robust factors that work in all kinds of conditions. This idea, formulated at the level of organizations as "there is no one best way to organize", could be reconciled with the idea of robust factors somewhat if at least we could have discovered stable interactions in educational research.

Judging from the results of aptitude treatment interaction at the micro level of teaching and learning, (Snow, 19..) this is hardly the case. Nor have empirical studies on contingency theory applied at the way school organizations are structured shown stable and convincing outcomes (Scheerens and Bosker, 1997, pp 284- 290). This does not mean of course that adaptive teaching is not important, nor that the idea of organizations adapting to their environment would not be a good one. Interactive research of this nature, including what is known as "differential school effectiveness research", just does not seem to have stable outcomes. I should emphasize that this conclusion is based on reviews that are limited and somewhat dated, and in need of further investigation.

Adaptive teaching is more or less implied in the Carroll model, as it allows for "deficits" in one area being compensated by extra's in other areas; for example by allowing students with fewer capacities more time for fixed learning tasks. The research finding that weaker students generally require more structured teaching also tends to a basic and relatively robust factor of adaptive teaching. This is also in line with a result of a recent secondary analysis on the PISA 2000 data set, indicating that less advantaged students generally profited more from favourable climate and resource conditions than more advantaged ones (Scheerens and Visscher, 2004).

The contingency paradigm, applied to adaptive teaching, seems to indicate the border between what empirical research can clarify and what should be left to the "art" of professional work by teachers. The next theoretical paradigm, the one on autopoiesis and self-reference probes this border even more.

Managing composition effects

It is beyond the scope of this paper to try and explain theories on autopiesis and self-reference applied to the functioning of organizations. Basic assumptions are endogeneous development, that is development that comes from "within" the organization, and processes that are in some way self-enforcing (Scheerens, 2004, geaccepteerd article in Studies in Educational Evaluation). Examples of "positive" and "negative" circular processes are given in Argyris and Schoen's treatment of organizational learning and factors that inhibit organizational learning (Argyris and Schoen, 19..). Such self-referential dynamics could be seen as getting started by "something", an idea, or a specific grouping of elements and relationships (compare Luhmann's concept of "Sinn" -sense or meaning-, Luhmann, 19.. (Social systems).Elsewhere I have argued that the composition of an organization in the sense as the aggregate of some relevant characteristics of the members or the organization could provide a "kick-off" condition for self-referential processes.. Compositional effects, in the sense of aggregates of characteristics of individual students are usually treated as "given conditions", for which school effectiveness research studies seek to control and adjust. It should be noted, however, that school composition is principally also a malleable condition, that could be changed and managed by certain selection policies or way of grouping teachers and students. In some empirical investigations student composition effects even outweigh the impact of school organizational and instructional variables (Scheerens and Visscher, 2004). Care should be taken, however, in assuring that such strong composition effects are not to some extent

methodological artefacts, due to lack of reliability of the individual level student background characteristics, and/ or weak operationalisations of the school organizational and instructional variables (cf. Harker and Tymms, 2004 (SESI, vol 15 no. 2, 177 -200). Substantive interpretations of student composition effects could be in terms of peer effects, like in the case when a majority of students from homes with a favourable attitude to education positively influence a minority with less favourable background conditions, or in terms of interactive teaching effects, as in the hypothetical case where a teacher teaches at a pace that is adapted to a majority of slow learners at the expense of a minority of more gifted students.

Endogeneous processes and aggregation effects show the limits of control and management in schools, but also offer new ways to improve conditions by means of selection and grouping policies.

Conclusion, a fragile model of robust factors

Educational effectiveness, seen in a multi-level perspective of malleable conditions at system, school and classroom level has laid bare the most important variables that need to be monitored in maintaining and enhancing achievement in basic school subjects in primary and lower secondary education. The importance or relative importance of this conclusion depends, first of all, on the perceived broadness or "narrowness" of the orientation on basic school subjects. According to the OECD (OECD, 2003. Education at a Glance) these subjects take about 40% of the curriculum, generalized over countries. In my view this could well be seen as the core of the curriculum, particularly because it includes language and reading in the mother tongue, which is absolutely basic for many other subject matter areas, also outside this 40%.

At classroom level important factors are: effective learning time, opportunity to learn and a degree of structure in scaffolding learning tasks. At school level conditions are important that could be seen as supporting the primary process of teaching. In recent international studies, variables related to a supportive, achievement oriented and orderly climate, show larger impact than other categories, related to resources and school processes, like evaluation practices. Research reviews on educational leaderships show very small effects (Witizers, Krueger and Bosker, 2003). Of the major programmes of systemic reform during the last decades, decentralization on the one hand and accountability on the other, no dramatic impacts can be reported from empirical effectiveness studies. Research seems to indicate that decentralization is best approached in terms of functional decentralization, which implies that a certain mix of centralization in some areas and decentralization in other works best. A mix for which both empirical and also theoretical support was found comes down to centralization in the curriculum domain and decentralization in the domains of personnel management and financial management of schools (Woessmann, 2000). The expectations of accountability based on various forms of educational testing need to be modified with respect to motivational aspects and the way feedback is realized.

When it comes to effect sizes three decades of research have seen a gradual diminishing of the reported effects of malleable conditions, when the results have been adjusted for student background conditions and contextual effects. For a while consensus seemed to be put the effect of malleable school variables at about 10% of the total variation in student performance. In their more recent review, Scheerens and Bosker (1997, p 79), estimate this "net" effect as being between 9 and 4%, while in a recent review of the PISA data set the effect of malleable school variables, expressed as a percentage of the total variance, was no more than about 2%, generalized over countries (Scheerens and Visscher, 2004). When malleable conditions at macro- (system), meso- (school) and micro-level (classroom) are compared the micro level factors close to the actual teaching and learning process have relatively the largest impact, while the more distant factor at system level have the smallest impact.

The most interesting relatively recent contribution to the field of school effectiveness research is the consideration of compositional effects, and the interaction between student background variables and malleable conditions. Variables related to student background used to be treated as "control" variables, but are now gradually also being seen as interesting in their own right. So far school composition has mostly been studied in the sense of student composition, but it might also be interesting to study staff composition (cf Scheerens, 2004, article Studies in Educational Evaluation). The reader might have wondered why there is so little about teacher quality in this review on educational effectiveness. For one thing, this is because of the fact that in industrialized countries the more basic factors like teacher qualification and experience do not make much of a difference in comparative studies. (And after all, school effectiveness research only studies comparative effects). Industrialized countries all have fairly high levels of teacher training, so that the range of this factor as a variable is rather restricted. American studies in which teacher competencies could be assessed in more depth have indicated, however, that general intelligence, verbal skills and subject matter competence do make a difference (Versteegen and King, 19..). Studying aspects of a schools' culture on the basis of aggregates of relevant personality characteristics of teachers, in other word a kind of staff composition effect, might be another interesting way to put teachers back on the educational effectiveness research agenda.

What we can learn from reviews of educational effectiveness research seems to make a lot of sense. It is perhaps quite subjective, almost a matter of taste, whether one would decide that the factors that have been found in research are robust or not. Given the smallness of effects and the relatively poor consistency in "what works" as evident from international comparative studies, one would be inclined to say that all we have is just a fragile model. Many more studies aimed at checking interactions would be needed to have more certainty on claims of generalizability of the factors that seem to work.

References