

# The Evidence Base for School Inspection Frameworks\*

Jaap Scheerens<sup>1</sup> - Melanie Ehren<sup>2</sup>

<sup>1</sup> University of Twente - Enschede (the Netherlands)

<sup>2</sup> UCL Institute of Education - London - Department of Learning and Leadership (UK)

doi: 10.7358/ecps-2015-012-sche

j.scheerens@utwente.nl

m.ehren@ioe.ac.uk

---

## I SISTEMI DI ISPEZIONE SCOLASTICA BASATI SULL'EVIDENZA EMPIRICA

### ABSTRACT

*This article describes how Inspectorates of Education operationalize different inspection goals (control, improvement, liaison) in their inspection indicator frameworks. The paper provides an overview and examples of the indicators used across a number of countries and how these are incorporated in inspection frameworks to evaluate and assess schools with the purpose of control, improvement and liaison. We shall describe the inspection and assessment of the processes and results of schooling (which includes making expert judgements), and compare and contrast them with inspection frameworks that focus on controlling input requirements and checking compliance to legislation. We will discuss the value and adequateness of different frameworks in the light of recent school effectiveness research. The results collected suggested that teaching/instruction level conditions, such as high expectations, a challenging teaching approach, an orderly learning environment and clear and structured teaching are more important than school level conditions in improving student achievement. Most inspectorates however do not explicitly evaluate teaching or teachers on a classroom/subject or grade level, preferring instead to evaluate school level conditions and general instruction characteristics or teaching patterns such as learning time, school leadership and school climate. A final «council» to further and future developments of national school inspection frameworks is therefore the relative emphasis on school organizational as*

---

\* A preliminary version of the text was published in M. Ehren (Ed.), *Accountability and educational improvement*. Berlin: Springer, 2014.

*compared to teaching and learning, or didactic standards. A prudent warning could be for designers and adaptors of inspection frameworks not to lose sight of the primary process of teaching and learning.*

*Keywords:* European countries, Evidence base research, Inspection indicator, School effectiveness, School inspection.

---

Purposes of school inspection include the control, support and liaison with teachers, schools and education systems. These three purposes reflect, according to De Grauwe (2007), different types of systems in the quality concept that is being monitored and the focus that they have. Control of schools often includes an evaluation of input indicators and schools' compliance to legislation, while Inspectorates of Education who aim to support and motivate school improvement will often evaluate educational processes and output of schools. Liaison particularly involves the brokerage role of Inspectorates of Education in transferring knowledge and information to relevant stakeholders in the education system. Thematic evaluations, in which Inspectorates of Education collect information about a specific, policy-relevant topic (e.g. how new policy is implemented or the performance of the education system in a specific area), are examples of how Inspectorates of Education can have such a liaison role.

## 1. CONTROL OF INPUT, RULES AND REGULATIONS

De Grauwe (2007) and Eddy Spicer *et al.* (2014) describe how inspection and monitoring systems can emphasize school inputs, such as the number of text books per pupil, teacher qualifications, number of pupils per class, etc. Such systems are particularly about controlling compliance as its first goal is to make sure that schools comply with predetermined norms fixed by law and administrative rules and regulations, such as the availability and use of procedures, policies and protocols concerning for example, admission policies or safety regulations and increasingly the satisfactory completion of school self-evaluation documents. Examples are the Swedish Inspectorate of Education checking the extent to which schools provide equal access to education for all students, and the Dutch and Irish Inspectorate of Education checking whether schools schedule and offer a minimum number of lesson

hours (Ehren *et al.*, 2013). According to De Grauwe (2007), this type of control is the oldest bureaucratic type of monitoring: checking that rules and regulations are respected. The classic inspectorate system combined with several forms of administrative self-reporting by schools (filling out forms) is the main device on which this type of monitoring relies.

## 2. EVALUATION AND SUPPORT OF EDUCATIONAL PROCESSES

Educational processes include the quality of the teaching in the school, the classroom-level interactions amongst teacher-students-curriculum and the «administrative» organizing processes of the school. These processes have become an increasingly more important part of inspection frameworks as there is a general consensus that process variables are more important than input variables in explaining differences in school quality, and information about school quality is needed to improve the quality of schools. This is particularly the case in high income countries where there is little variation in school inputs (see Hanushek, 1986).

An overview of Van Bruggen (2010) for example shows how eighteen European Inspectorates of Education have indicators and criteria on «the organisation and management in the school», and «the teaching and learning» in their frameworks to ensure a national perspective on quality education and to evaluate schools against a common set of criteria representing a national perspective on quality education. Many of these frameworks are inspired by school effectiveness research according to Ehren *et al.* (2013). Their comparative study of inspection frameworks in six European countries indicates a strong focus on educational processes such as opportunity to learn and learning time, achievement orientation, clear and structured teaching, challenging teaching approaches and orderly learning environment. These indicators are to some extent part of the inspection frameworks of all the six countries in their study (England, the Netherlands, the province of Styria in Austria, the Czech Republic, Ireland, and Sweden).

## 3. EVALUATION OF SCHOOL OUTPUT

The increasing availability and use of student achievement data has resulted in an increased focus of school inspections on output of schools. Some of the Inspectorates of Education (e.g. the Netherlands) also focus on output

of schools to respect the schools' autonomy in shaping their educational processes. An increased focus on the output of schools often goes hand in hand with a broader governmental drive towards performance measurement and league tables to create competition between schools (Tolofari, 2005).

The goal of this type of monitoring is to ensure that pupils learn what they are supposed to learn and to evaluate the output of the school that relate to student learning, such as student achievement in a variety of subjects and graduation and/or dropout rates. Evaluation of academic achievement may also prevent goal displacement when schools focus on producing protocols and procedures to inform Inspectorates of Education about their educational processes. Student achievement results on national standardized tests are aggregated to evaluate the performance of schools and in some cases to publish league tables of schools.

More recently a number of Inspectorates of Education (e.g. in Norway, the Netherlands, Scotland) have also started to evaluate social outcomes of schools. Social outcomes are defined by Ehren and Dijkstra (2014) as «the individual and collective benefits of education for interpersonal interaction in the noneconomic spheres of life». At the level of the school, the social outcomes of education consist, according to these authors, of the competences to live with others, the social competences that people need to realize their goals and to relate to others in all kinds of situations, both at work and elsewhere. It also concerns the civic competences required to make a contribution to society, democracy and the social networks in which people live. Social outcomes are included in inspection frameworks to provide a broader picture of school output and to prevent a narrow focus of schools on teaching only mathematics, reading and writing. Test results on cognitive subjects are considered to provide an incomplete picture of young people's competences and many countries feel that a wider range of competences and skills need to be part of what students learn and what they need to be as active members of society and the labour market (Dijkstra & De la Motte, 2014).

The most common monitoring devices used to evaluate output of schools are the regular measurement of learner achievement by standardized tests and examinations, combined with the publication of league tables and systematic (external) auditing of schools. These tests particularly include cognitive subjects, such as mathematics and literacy while social competences are more often measured through observations in real-life situations (Dijkstra & De La Motte, 2014). Only recently have some countries, such as the Netherlands, started to develop national standardized tests to measure social competences. As only a limited number of schools have administered those tests, the benchmarks and targets to compare and evaluate schools on these measures are also still limited.

A number of countries have recently also started to use school output indicators to more deliberately target their inspection visits of schools in «risk-based inspections». The assumption underlying these risk-based inspection models is that school output (as measured via student achievement results) is an adequate predictor of the teaching quality in the school and can be used in early warning analyses to identify potentially failing schools. Students' results (corrected for their socio-economic backgrounds) on national standardized tests and examinations in the Netherlands are for example used to classify schools into one of three categories; schools in the «green» category are considered to have no risks of failing, «orange» schools have potential risks of failing, whereas «red» schools have high risks of failing. Schools in the «green» inspection category receive a «basic» inspection treatment, which means there is no further inspection activity in the school that year. Additional desk research of self-evaluation reports and other school documents is scheduled for schools in the orange category, while schools in the red category are immediately scheduled for inspection visits.

The increased focus on school output has warranted a more refined and sophisticated analyses and use of student achievement data to improve the accuracy of inspection assessments as well as the predictability of early warning analyses. A number of Inspectorates of Education have recently therefore started to develop value-added measures to ascertain the impact that individual schools and teachers have on the quality of education provided. Value-added measures employ mathematical algorithms in an attempt to isolate the school's contribution to student learning from all the other factors that can influence academic achievement and progress – e.g., individual student ability, family income levels, the educational attainment of parents, or the influence of peer groups. The goal of these models, which are also referred to as Value-Added Assessment (VAA) Models, is to estimate effects of individual teachers or schools on student achievement while accounting for differences in student background (ASA, 2014, p. 1).

Hamilton and Koretz (2002, p. 23) distinguish two types of reporting of test scores to understand (when aggregated to the school level) performance of schools. *Norm-referenced reporting* involves the description of the performance of an individual school in terms of its position in a distribution of scores of other schools. Such reporting can be based on:

1. Percentile rank: indicating the percentage of a reference group (often, the national population of students in schools) who obtained lower scores than a given school. Thus, a school with an average National Percentile Rank (NPR) of 75 scored higher than 75 percent of a national sample of schools.
2. Standard score: expressing a school's performance in terms of how far the school's test score is from the mean. The scores are transformed to have

a specific mean and standard deviation (or SD – a measure of the spread of scores). Examples are z-scores (mean = 0, SD = 1) T-scores (mean = 50, SD = 10), and normal curve equivalents (or NCEs – mean = 50, SD = 21.06). Thus, a school with a T-score of 60 is one standard deviation above the mean, which is roughly a percentile rank of 84.

3. Grade Equivalent (GE): expressing a (group of) student's performance in terms of the grade level at which that performance would be typical. GEs are generally expressed in decimal form, such as 5.7, in which the first number is the grade and the second is the month (for ten academic months, with zero representing the performance of students first entering that grade level). A student who scores a 5.7 on a fourth-grade test has the same level of performance as would a median student in the seventh month of fifth grade if that student took the same test. GEs are a developmental scale designed to examine growth. In any subject and at any level, the median increase in performance over a year of growth is 1.0 GE.

The alternative to norm-referenced reporting of (aggregated) test scores is criterion-referenced or standards-based reporting. This type of reporting does not include a comparison to other groups of schools or schools but compares the performance of a school to one or more fixed levels of performance. Such fixed levels of performance typically include targets on minimum test scores and material students are expected to master in specific content areas.

Hamilton and Koretz (2002) distinguish between two broad approaches to setting targets. The first is referred to as «status», while the second one is about «change» measures. The status measure compares a unit's performance at one point in time with a single standard, which may be a performance criterion set by the Inspectorate of Education, the average performance of similar schools (e.g. with similar student populations), or a historical average (e.g. the average of a group of schools over a period five years).

Change measures on the other hand compare a unit's performance at one time with some measure of prior performance. Change can be measured using a cross-sectional approach in which this year's fourth graders are compared to last year's fourth-graders, a quasi-longitudinal approach in which this year's fourth graders are compared with last year's third-graders, and a longitudinal approach in which individual student scores are used to compare students with themselves over time. Targets would quantify the amount of change expected of schools.

Test scores can, according to Hamilton and Koretz (2002), be reported on the level of schools, classrooms, subjects or specific student groups. Decisions about whether to report school-level, classroom/subject-level, or student-level scores, and whether to disaggregate for specific groups, should be, according to these authors, informed by the purposes for which scores

will be used and the desire on the part of stakeholders for specific types of information. In each of the models, test scores can be adjusted for a number of school and student characteristics, taking into account the strong relationships between student achievement and socioeconomic status and other aspects of student background.

Examples of Inspectorates of Education using value added measures can be found in the Netherlands where student achievement data are classified into separate performance bands on the basis of level of disadvantage (mainly using parental educational level). This classification is used to evaluate and grade school output as well as in the early warning analyses to identify potentially failing schools for inspection visits. The use of value added measures can also be seen in inspection frameworks in England, Canada and the Ireland. For example, in Ireland the Department of Education and Skills drafted a strategy to improve literacy and numeracy standards and suggested using a benchmarking data analysis tool referred to as «Schools Like Ours» which is prescribed as allowing a school to «have access to its own data as well as the data from the ‘matched’ schools». In the case of Canada, the Literacy and Numeracy Secretariat of the Ontario Ministry of Education developed a benchmarking module, also called «Schools Like Ours». Its purpose is to «find similar schools to any selected school», using any combination of the available indicators, such as similar demographics but higher achievements.

The promotion of value added indicators within school inspection frameworks is the need to make accurate use of student achievement results in judging school practice. Scheerens, Glas and Thomas (2003) for example assert that having more information about individual students, sub-groups of students, and all students in a school as well as comparative data across a whole population (or representative sample) of schools allows for a more reliable and informative analysis of student achievement results (Scheerens, Glas, & Thomas, 2003, ch. 13.3, § 1).

However, as Donaldson and Johnson (2010) state, there is still a great degree of uncertainty about the value that schools actually add to student learning and such models are still under development and therefore prone to error. Nonetheless, many Inspectorates of Education see the benefits of developing value-added measures to improve the reliability and validity of their judgments, particularly when compared to their current more crude methods of comparing school performance to the average raw score of a population or using free school meal bands or other data on socio-economic backgrounds of students to classify and compare schools into separate, similar performance bands.

#### 4. FIT FOR PURPOSE: SCHOOL EFFECTIVENESS MODELLING AND THREE SCHOOL INSPECTION FUNCTIONS

Many Inspection systems have shifted their purpose in recent years to improving teaching and learning. Such a purpose of school improvement has become more important over the last years as a result of an increased policy of making schools more autonomous and self-governing. High levels of school autonomy are counterbalanced in some countries by systematic evaluations of schools to assure the quality and effectiveness of school level decisions. Declining student achievement results, as measured in international surveys such as PISA and TIMSS, have also often spurred an increase in evaluation and control of schools even in supposedly decentralized education systems.

If inspection is to be fit for purpose, the nature of inspection, and particularly the standards in inspection frameworks should be matched to its intended objectives of improved teaching and learning, and ultimately student achievement. The educational and school effectiveness literature is an important source to define what a good school is, and to critically reflect on the extent to which inspection standards in different countries are supported by research evidence. In the most general sense «educational effectiveness» refers to the level of goal attainment of an educational system. An educational system could be a national education system, a school, a group of students or even an individual student. Given the current topic of school inspection, we shall concentrate on schools and school effectiveness research as the focal level.

School effectiveness research attempts to deal with the causal aspects inherent in the effectiveness concept by means of scientific methods. Not only is an assessment of school effects, in terms of outcomes, considered, but particularly the attribution of differences in school effects to malleable conditions, both inputs and processes. Usually, school effects are assessed in a comparative way, e.g. by comparing average achievement scores between schools. Achievement scores in core subjects, established at the end of a fixed program are the most probable «school effects», although alternative criteria like the responsiveness of the school to the community and the satisfaction of the teachers may also be considered. In order to determine the «net» effect of malleable conditions, like the use of different teaching methods or a particular form of school management, achievement measures have to be adjusted for intake differences between schools. For this purpose student background characteristics like socioeconomic status, general scholastic aptitude or initial achievement in a subject are used as control variables. This type of statistical adjustment in research studies has an applied parallel in striving for «fair comparisons» between schools, known under the label of «value-added» (see Scheerens, 2013, p. 4) and the previous section.



The connection of educational effectiveness research and the knowledge base that this kind of research has yielded to the business of school inspection, is, first of all an interest in educational outcomes as the ultimate quality standard. Yet, the most important relevance of educational effectiveness research to school inspection is to provide a scientifically grounded rationale for the choice of input and process indicators, by providing empirically supported information on which malleable school conditions matter most in influencing educational outcomes.

When we take the three basic functions of school inspection, which were put forward in earlier paragraphs of this chapter as a point of departure, *improvement* would seem to have the closest connection to the above stated rationale of connecting educational effectiveness to school inspection. When school inspections report not only on school outcomes, but also on input and process indicators, feedback on these indicators could be expected to provide direct handles for school improvement actions. For example, if a school process indicator, like the connection of the school curriculum to the assessments or examinations (often indicated as «opportunity to learn»), would have a low value, improving the match between teaching content and assessment content would be a likely course of action to improve school performance.

When considering *monitoring* as a function of school inspection, the connection with educational effectiveness is a bit more complex. Firstly, compliance monitoring has no connection to educational effectiveness, as far as living up to standard procedures and regulations is concerned. Compliance monitoring is however more in line with the effectiveness logic if basic school inputs are evaluated that have straightforward implications for educational outcomes. Such inputs are readily available, and examples are teacher qualifications, pupil teacher ratios and formally prescribed teaching time. The degree to which such input measures make a difference depends, among others, on the context. Such inputs usually have higher effects in developing than in industrialized countries as generally all schools in industrialized countries have these basic inputs in place and there is little variation between schools in the qualification of teachers or scheduled teaching time (Hanushek, 1997). Hanushek (1986, p. 1161) for example shows that only the variable «teacher experience» shows some consistency, in that 30% of estimated coefficients appeared to be statistically significant. Hanushek's overall conclusion is that as yet educational expenditure is not consistently related to achievement and it would take greater variation in inputs to expect important effects.

As far as the *liaison* function of school inspection is concerned, there is only a more theoretical connection in the sense that multi-level models of

educational effectiveness are pre-occupied with alignment between facets and elements that operate at different levels (e.g. Scheerens, 2007; Creemers & Kyriakides, 2008). When educational systems are seen as hierarchies, school effectiveness can be distinguished from instructional effectiveness and from «system effectiveness». The latter term is less common, and refers to a more recent strand of research that is strongly stimulated by the upsurge of international assessment studies. In such studies policy amenable conditions at the national system level can be associated with student outcomes; examples are policies of enhancing school autonomy, accountability and choice. Instructional effectiveness focuses on the classroom level and on effective teaching on the classroom level. The distinguishing characteristic of this stream of educational research is the fact that process characteristics of education are studied at the teacher or classroom level. So, when we are considering variables at this level that have been found to be associated with achievement, we are really delving into the primary processes of schooling. School effectiveness on the other hand focuses on the conditions of an effective school and which conditions «add value» to achievement of students; the aim is generally to discover school characteristics that are positively associated with school output, usually measured as students' achievement.

Educational effectiveness refers to the union of the effectiveness research on these three levels. Conceptual contributions to this line of work depict schools as a set of «nested layers» (Purkey & Smith, 1983), where the central assumption is that higher organizational levels facilitate effectiveness enhancing conditions at lower levels (Scheerens & Creemers, 1989). Multilevel analysis has contributed significantly to the development of such integrated school effectiveness models (Scheerens, 2013, p. 4).

These notions of educational effectiveness and «nested layers» of classroom levels within the school level, within a national education system relates to the liaison function of school inspections in thinking about the specific connections that inspection can make between the different levels in their evaluation of teaching, schools and the education system and in the information they provide to actors on these different levels. Inspectorates of Education could enhance alignment and coupling of the nested layers, for example by motivating coherence of national curriculum frameworks with evaluation, inspection and assessment frameworks. However, theoretical educational effectiveness models also recognize that many education systems have loose coupling between the layers of educational system; also the degree to which national inspectorates are expected to function as instruments of the central administration and enhance strong coupling differs between countries. In some cases inspectorates are expected to function independently or semi-independently and set their own agenda for school evaluation. To the extent

that this kind of systemic alignment has been studied (e.g. Mourshed, 2010; Scheerens *et al.*, 2015), the functioning of inspectorates of education has not been addressed. Although an analysis of the *liaison* function of inspectorates of education is very interesting, it is beyond the scope of this chapter which focuses only on the school level.

## 5. IDENTIFICATION OF EFFECTIVENESS ENHANCING SCHOOL CONDITIONS; CONSENSUS AMONG REVIEWS

The core of educational effectiveness research is the identification of effectiveness and improvement oriented conditions. In this section recent and earlier research reviews will be cited, and considered for consensus on the main effectiveness enhancing conditions. Such a summary allows us to compare and contrast the school effectiveness research base with inspection frameworks in a subsequent section.

Scheerens (2014) summarizes results of review studies that were carried out in the 1990s, and more recent review studies by Reynolds *et al.* (2014), Muijs *et al.* (2014) and Hopkins *et al.* (2014). The older review studies are those by Purkey and Smith (1983), Scheerens (1992), Levine and Lezotte (1990), Sammons *et al.* (1995), Cotton (1995). These earlier review studies mention the following conditions as contributing to high student achievement:

- *Achievement orientation and high expectations*: a productive school climate, a school mission focused on achievement, shared vision and goals, high expectations that all students can achieve.
- *Cooperative atmosphere and an orderly climate*: cooperative planning, a learning oriented atmosphere consensus, orderly climate.
- *Clear goals on basic skills*: focus on student learning, concentration on teaching.
- *Frequent evaluation*: appropriate monitoring, evaluative potential of the school, assessment.
- *Professional development*: staff development, in-service training, a learning organization.
- *Parental involvement*: parent support, home school partnership.
- *Strong leadership*: educational leadership, school management and organization, improvement oriented leadership.
- *Effective instructional arrangements*: classroom management, time on task, structured teaching, opportunity to learn, coordination in curriculum and instruction.

Consensus among the authors of the earlier review studies is largest with respect to the factors on achievement orientation (which is closely related to «high expectations»), co-operation, educational leadership, frequent evaluation, time, opportunity to learn and «structure» as the main instructional conditions.

More recent reviews by Reynolds *et al.* (2014), Muijs *et al.* (2014) and Hopkins *et al.* (2014) also provide an overview of the most relevant conditions in educational effectiveness research (SER) and teaching effectiveness research (TE). The review from (Hopkins *et al.*, 2014) adds an improvement component (SSI) to this research and aims to enhance our understanding of effective interventions or improvement programmes and of the conditions in schools that contribute to effective school improvement. The summary of these studies, as provided in Table 2 (cited from Scheerens, 2014) indicates that there is clearly consensus about the main conditions of schooling and teaching over time. The five factors on which closest consensus was seen among the earlier reviews, are still present in these more recent ones. The most important development is the addition of teaching strategies inspired by «constructivism» in the review on teaching effectiveness; these are shown in italics in the second column of Table 1.

Table 1. – Summary of recent reviews on effectiveness and improvement oriented conditions.

EER	TE	SSI
Effective Leadership	Opportunity to learn	Dimensions of organizational health
Academic focus	Time	School based review
A positive orderly climate	Classroom management	School development planning
High expectations	Structuring and scaffolding, including feedback	Comprehensive school reform
Monitoring progress	Productive classroom climate	Facets of educational leadership (transformational, instructional, distributed)
Parental involvement	Clarity of presentation	Effective systemic reform; among others, student achievement and teaching quality emphasis
Effective teaching (time)	<i>Enhancing self regulated learning</i>	
Staff professional development	<i>Teaching meta-cognitive strategies</i>	
Pupil involvement	<i>Teaching modeling</i>	
	<i>More sophisticated diagnosis</i>	
	<i>Importance of prior knowledge</i>	

Effectiveness enhancing conditions referred to in the review studies by Reynolds *et al.* (2014), Muijs *et al.* (2014) and Hopkins *et al.* (2014). Source: Scheerens, 2014.

## 6. LESS CONSISTENCY OF EFFECT SIZES IN QUANTITATIVE RESEARCH SYNTHESSES

The previous section showed an extensive overlap in effectiveness enhancing conditions found in both qualitative and quantitative reviews. The quantitative reviews and meta-analyses however also indicate important differences across meta-analyses in the specific effect sizes of each of the key variables. These differences are illustrated in the Table 2, below, cited from Scheerens, 2013, p. 14. The differences in effect sizes reported by Hattie (2009) on the one hand and the other meta-analyses, which were more Europe based, on the other is quite striking. It should be noted that Hattie expressed effect sizes by means of the d-coefficient and the other authors report correlation, which roughly can be converted to one another by considering that the correlations are half of the d-coefficients.

Table 2. – Results from recent meta-analyses.

### *School level variables*

	Scheerens <i>et al.</i> , 2007	Hattie, 2009	Creemers & Kyriakides, 2008
Consensus & Cohesion	.02	–	.16
Orderly climate	.13	<b>.34</b>	.12
Monitoring & evaluation	.06	<b>.64</b>	.18
Curriculum/OTL	.15	–	.15
Homework	.07	<b>.30</b>	–
Effective Learning Time	.15	<b>.34</b>	–
Parental involvement	.09	<b>.50</b>	–
Achievement orientation	.14	-	-
Educational leadership	.05	<b>.36</b>	.07
Differentiation	.02	<b>.18</b>	–

### *Teaching level variables*

	Scheerens <i>et al.</i> , 2007	Hattie, 2009	Seidel & Shavelson, 2007
Time and OTL	.08	<b>.34</b>	.03
Classroom management	.10	<b>.52</b>	.00
Structured teaching	.09	<b>.60</b>	.02
Teaching learning strategies	.22	<b>.70</b>	.22
Feedback & monitoring	.07	<b>.66</b>	.01

Results from recent meta-analyses (coefficients are based on the Fisher Z transformation of correlations); as Hattie presents effect sizes in terms of d, these are indicated in bold.

As our aim is a comparison between inspection frameworks and school effectiveness research, we include the following table with an average effect size of key effectiveness conditions as described in meta-analyses of Marzano (2003), Scheerens *et al.* (2007) and Hattie (2009). Although averaging the effect sizes is a bit of a rough procedure, it nevertheless provides an impression of the relative importance of these core effectiveness enhancing conditions. The average effect sizes indicate that «exposure» to educational content (opportunity to learn and instruction time) is the most important condition in schools' contribution to high student achievement. Organizational factors like school leadership and cooperation on the other hand have relatively small effect sizes (Table 3).

Table 3. – Rank ordering of school effectiveness variables according to the average effect sizes.

	Marzano, 2003	Scheerens <i>et al.</i> , 2007	Hattie, 2009	Average effect size
Opportunity to learn	.88	.30	.39*	.523
Instruction time	.39	.30	.38	.357
Monitoring	.30	.12	.64	.353
Achievement pressure	.27	.28	.43**	.327
Parental involvement	.26	.18	.50	.313
School climate	.22	.26	.34	.273
School leadership	.10	.10	.36	.187
Cooperation	.06	.04	.18***	.093

Rank ordering of school effectiveness variables according to the average effect sizes (d-coefficient) reported in three reviews/meta-analyses; \* = operationalized as «enrichment programmes for gifted children»; \*\* = operationalized as «teacher expectations»; \*\*\* = operationalized as «team teaching». Source: Scheerens (2013, p. 24).

Another important topic to consider is the consistency of these effects across individual schools. Consistency in the estimation of school effects across grades, teachers and subjects, and stability of school effects across years is an important underlying assumption of school inspections. Inspection assessments of school quality are generally made once in every 3 to 5 years and are expected to remain relatively unchanged until the next inspection visit.

Several school effectiveness studies address the potential (in)consistency of school effectiveness by means of an analysis of a correlation matrix of subject – and cohort (or grade) level effects, and computing the magnitude of a general school factor. Typically the rank ordering of the (value-added) mean achievement of schools is correlated across years. Bosker *et al.* (1989) found correlations that declined according to the time interval from one to

four years from .74 (one year), .62 (two years), .49 (three years) and .49 (four years) in a study of Dutch secondary schools. Gray *et al.* (1995) looked at time intervals of one, two and three years in English secondary schools and found correlations of .94, .96 and .81. Thomas *et al.* (2010) analyzed school data over a period of 11 years in the Lancashire district. They concluded that there was a fair stability in school effects. Still, when schools were categorized as average, over – or underachieving there were many changes in categories; over a period of 11 years, 50% of the schools had changed category. Moreover continuous progress was rare:

For the majority of schools three years of upward movement seems to have been the typical limit. In short, our evidence from the non-linear modelling suggests that, whilst there were undoubtedly changes, these were not very «continuous» and in many cases could have occurred by chance. This finding contrasts starkly to government ideals of continuous school improvement. (Thomas *et al.*, 2010, p. 280)

Less stability was again also found in a recent Dutch study, where of the highest scoring secondary schools only 15% were still in the top category after three years (Vermeer & Van der Steeg, 2011). As a caution against instability it would make sense to assess the position of schools in accountability and reward schemes over a certain period of time (e.g. three years) and compare schools on their average achievement across a number of years (adapted from Scheerens 2013, pp. 9-10). The findings from different countries also indicate that the number of years of averaging results may differ per country as the stability of school effects seems to vary across countries and potentially reflects the homogeneity of education systems.

## 7. A CLOSER LOOK AT THE MEANING OF THE KEY FACTORS IN EDUCATIONAL EFFECTIVENESS

In order to look a little bit deeper, «behind» the labels of the main factors that have appeared in the tables of the previous sections, main characteristics and sub-components of these factors will be described in this section.

*Achievement orientation* – This factor expresses outcome oriented ambition and a positive, optimistic outlook on the competences of all students to achieve. Data sources are planning documents, like school development plans, or mission statements, questionnaire responses from school heads and teachers, and administrative evidence on record keeping of student

achievement. Achievement orientation is often detailed into subcomponents on «clear focus on the mastering of basic subjects», «high expectations expressed at school and teacher level», and «keeping records on pupils' achievement».

*Educational leadership* – In many operational definitions and instruments that represent educational leadership, there is a strong focus on leadership roles directed at the primary process of teaching and learning and organizational conditions that are seen in support of this primary process, including coaching of teachers, and providing guidance on curricular choices. Often a connection is also made with student assessment and progress monitoring. In addition to educational leadership, focused at the primary process of teaching and learning, «transformational leadership», is more directed at school organizational improvement activities. «Distributive leadership», and even «teacher leadership», emphasize that parts of school leadership may be delegated to teachers. Relevant sub-components of educational leadership are: general leadership styles, leadership roles concerning coordination, orchestrating participative decision making and providing information, meta-control of classroom processes, facilitation of staff professionalization.

*Staff cooperation, cohesion and consensus* – In early applications there was a certain emphasis on measurable facets of cooperation (like frequency of meetings) and personal satisfaction. More recently, enforced by conceptions of schools as professional learning communities, and «peer learning», cooperation is more closely focused on school level improvement initiatives on the one hand and discussing teaching and learning on the other hand. «Team teaching» is also sometimes used as an indicator of teacher cooperation.

Staff cooperation, cohesion and consensus is generally measured through types and frequency of meetings and consultations, satisfaction about cooperation, task related facets of cooperation, consistency on teaching goals and methods.

*Curriculum quality and opportunity to learn* – Curriculum quality is mostly measured in the sense of systematic planning processes and experienced satisfaction with the curriculum. The concept of opportunity to learn addresses the alignment between educational objectives, teaching and student assessment. The basic question is the correspondence between the content that is taught and the content that is tested. In more recent studies «test preparation» is a new way to look at opportunity to learn. At classroom level «instructional alignment» is another more recent interpretation of opportunity to



learn. Over the years, curriculum quality and opportunity to learn have been operationalized in sub-components of systematic setting and monitoring of curriculum priorities, choice and application of methods and textbooks and opportunity to learn in the sense of «content covered».

*School climate* – An orderly and safe school climate is the one school organizational conditions that have obtained relatively positive support in international assessment studies, like PISA. It has also been «on the scene» in school effectiveness research since the very beginning. Clearly the achievement oriented facet of the school climate is closely associated with «achievement orientation», and «achievement oriented school policy», as treated in the above. Internal relations that are part of the «relational school climate» are relationships between teachers and students, teachers and head teachers, and teachers and students, among themselves. Relevant sub-components of school climate are discipline, achievement orientation and good internal relationships between school staff and students and staff.

*Evaluation and monitoring* – Evaluation as an effectiveness enhancing condition is about the presence or absence of evaluation orientations at school, classroom and student level. Also the frequency of application is being measured, as well as the staff's satisfaction with evaluations and the use that is made of the evaluation results to improve the school. At a basic level, evaluation and monitoring is measured by checking whether a school uses a systematic school self evaluation procedure, a pupil monitoring system, and/or other types of testing and student assessment. More intensive measures of evaluation and monitoring in schools encompass an analysis of task related collaboration between teachers and whether (both formal standardized and informal forms of) teacher and teaching evaluation have a place in «peer learning». Sub-components of evaluation and monitoring are school evaluation, classroom evaluation and student assessment.

*Parental involvement* – Main components of parental involvement are the voice of parents in determining school policies, active involvement and support by parents in school matters, either for assistance with practical matters or concerning teaching and learning, while the most ambitious way is for the school to try and influence the pedagogical climate of the home. Parental involvement is often measured by asking schools about the emphases in school policy on parental involvement, the frequency of contacts with parents, and the satisfaction of relevant actors (teachers, parents and school heads) with parental involvement. The concept of parental involvement may be enlarged to «community involvement» with the school.

*Classroom climate* – Classroom climate mirrors to some extent the previous description of school climate, particularly in the description of sub-components. Both definitions include a disciplinary part and emphasize good relations, where classroom climate focuses on the relationship between teachers and students and school climate also include relations between school staff. Classroom climate additionally also includes a notion of cognitive and emotional support of students, in the sense of clear explanations and help with assignments, as well as stimulating engagement and a sense of self-efficacy of students. Sometimes a «fun factor» of classroom climate is also included, asking students about their sympathy with the teacher, whether the teacher chats about non school activities, and whether there were any jokes or laughter.

*Effective learning time* – Learning time can be measured holistically or in more detail by distinguishing allocated learning time (official lesson hours), net teaching time (the part of a lesson that teachers are actually involved with teaching, subtracting time for organizing the lesson and distractions), and time on task (the amount of time a students are actively engaged). Classroom management is often defined in terms of maximizing net teaching time and time on task. Another important distinction is between teaching time at school and time spent on doing homework. Studies on teaching time often include information on student absenteeism and suspended lessons.

*Structured teaching* – Structured teaching is associated with the cognitive support facet of classroom climate and includes the extent to which teachers give clear explanations and support students with assignments. The general idea of structured teaching is the application of frequent interventions to support the learning process. Examples of these include: stating educational objectives clearly, dividing the total subject matter that must be learned into relatively small units, providing a well-planned sequence of these units, providing many opportunities for pupils to do exercises, giving cues and hints, frequent questioning and testing to monitor progress, and giving feedback. Relevant subcomponents are the setting of clear objectives, preparing structured sequences of teaching and learning activities, providing clear explanations, the use of questioning and feedback in instruction, as well as the monitoring of student progress.

*Constructivist teaching and independent learning* – Constructivist teaching and independent learning appear to be opposites of the more behaviouristic and guided practice approaches of structured teaching. Constructivist teaching and independent learning emphasize the learning processes of stu-

dents and the teaching of general and/or subject-specific learning strategies. «Cognitive activation» is an important element of constructivist teaching and refers to providing sufficient depth in content presentation, aiming for understanding at a higher level where students are able to understand and use authentic applications and use concepts in different contexts. The notion of «scaffolding», where the amount of students' self-regulation of their learning is gradually increased as they master subject content combines the notions of structured and constructivist teaching and places these two approaches on a continuum.

*Differentiation* – Differentiation recognizes individual differences between students, and tries to provide room for variation in teaching that is adapted to these differences. Schools and teachers can differentiate the teaching by means of streaming students into classrooms that work at different ability levels, grouping students in different ability groups within one classroom, pacing instruction (allowing students to cover subject matter in different time schedules) and individualizing instruction within relatively heterogeneous classrooms. Differentiation also includes special programmes for and/or additional teaching and support of pupils at risk and providing extra challenges to high achieving students.

## 8. THE MULTI-FACETTED NATURE OF SCHOOL EFFECTIVENESS ENHANCING VARIABLES

It is important to note that these concepts are not mutually exclusive and that there are several cases of conceptual overlap between them. Achievement orientation is described in terms of direct school policies, but also as a relevant orientation in the school's climate. Next, it is important to see the essential place of assessment and evaluation as a means to shape the achievement orientation in the school. Achievement orientation has an orientation on student achievement results in common with educational leadership.

Apart from the conceptual overlap between these main indicators, schools are also likely to combine a number of these indicators in an overall strategy for (improving the) teaching and learning in the school. Opportunity to learn and time on task for example are essentially part of an overarching strategy to increase students' exposure to content, while «focused teaching» (a term coined by Seashore Louise *et al.*, 2010) combines direct instruction and constructivist teaching strategies. A third example is the way evaluation and monitoring are inherently related to ensuring opportunity to learn, expressed

in Popham's credo «test better, teach better» (Popham, 2003). Finally, evaluation and assessment can also be seen as strong levers in more structured teaching approaches, where students' progress records are needed to improve their «time on task», to set learning objectives and provide structured support and guidance to students (this approach is currently piloted in the Netherlands, under the heading of «result oriented work»: Visscher & Ehren, 2011).

Many of the factors, discussed in the above, such as climate, achievement orientation, evaluation and monitoring and opportunity to learn can also be measured on both the classroom and school level and may have different interpretations on those two levels. For example individual teachers may make limited use of student monitoring in informing their teaching while the head teacher has strong systems in place to monitor student progress across the school to inform school-level improvement policies

The multi-faceted nature of school effectiveness, the conceptual overlap between the conditions and the multiple ways to describe and measure each sub-component indicates the complexity in designing effective inspection frameworks and calls for a thoughtful reflection on which indicators to include in inspection frameworks, as well as how to measure them. The next section includes such a reflection of six inspection frameworks. These frameworks are from the Inspectorates of Education in the Netherlands, England, Ireland, Sweden, the province of Styria in Austria and the Czech Republic, which were studied by Ehren *et al.* (2013) in an EU funded project. De Volder's dissertation (2012) and country profiles on the website of SICI, the European Association of Inspectorates of Education<sup>1</sup> were used to complete the table. The description of the inspection frameworks refers to 2010.

## 9. ANALYZING INSPECTION FRAMEWORKS

Table 4 below provides an overview of the inspection frameworks used in six European countries. The countries represent a broad variety of types of school inspections. They vary in using a low stakes capacity-building inspection approach (e.g. Ireland), to test-based early warning inspections to control schools (e.g. the Netherlands), and range from very centralized national Inspectorates of Education (e.g. England) to inspection agencies that operate at the level of the provinces (Austria). In Table 4 we summarized the inspection standards and thresholds these Inspectorates of Education use to evaluate schools. We also provided a brief summary on the hierarchical structure

---

<sup>1</sup> <http://www.sici-inspectorates.eu/Members/Inspection-Profiles>.

of these frameworks and the types of data collection to inform inspection assessments. It is important to note that any visual overlap in standards in inspection frameworks will likely hide quite a diversified gamma of operational definitions and specific measurement instruments. The summary below and subsequent comparison with school effectiveness research can therefore only provide very general comments and suggestions on how to potentially improve inspection frameworks.

The summary in Table 4 indicates that the six Inspectorates of Education all focus on malleable conditions and processes, while some also evaluate the output of the school, and only one Inspectorate of Education specifically assesses elements of the school's input (the Czech Republic). Interestingly, two countries (Ireland and England) collect information on input (e.g. quality of school building) but only use this information if school inspectors feel that those inputs specifically affect the quality of school processes. These Inspectorates do not grade the school's input separately. In choosing such an approach they seem to recognize the complex and interrelated nature of the input and process conditions in explaining school quality and high student achievement.

Such a perspective however seems to be lacking in how the six Inspectorates of Education generally assess school processes and output. Table 4 indicates that overall assessments of school quality include a set of standards and substandards with underlying detailed criteria, where strict rules are applied on grading schools on a scale in a similar manner for all the schools. The assessment protocols and guidelines require an assessment of conditions as «present» or «absent», or an assessment on a 3 or 4 point scale, ignoring the potential interrelatedness of the conditions.

The way in which information on conditions at the classroom level are aggregated to evaluate school level effectiveness also discounts the fact that these conditions have different meanings at different levels of the school hierarchy. Observations of «achievement orientation» in lesson observations (at the classroom level) are for example often simply averaged to come to an assessment of the school's quality in this area. Achievement orientation on the school level however may also include an assessment of school policy in this area and monitoring systems the school has in place to support teachers' orientation on high student achievement. Only some of the Inspectorates of Education (e.g. Sweden and Austria) seem to steer away from this approach of treating school and classroom conditions similarly by emphasizing a more holistic approach to the evaluation of schools and providing schools with an overview of strengths and weaknesses instead of using rudimentary thresholds to single out failing schools. The downside of this approach, as will be discussed in the next chapter, is however the potential lack of accuracy and transparency in inspection assessments.

Table 4. – Standards in inspection frameworks used in regular full inspections in six European countries (as in 2010).

STANDARDS					
	<i>Input</i>	<i>Process</i>	<i>Output</i>	<i>Threshold for failing schools</i>	<i>Data collection and scoring guidelines</i>
The Netherlands	NA	<ul style="list-style-type: none"> <li>The school has a system for assuring the quality of its education.</li> <li>The subject matter offered to pupils prepares them for continuing education.</li> <li>The students get enough lesson time to learn the subject matter.</li> <li>The school systematically assesses the progress of students.</li> <li>The school climate is safe and stimulating.</li> <li>The pedagogical behaviour of teachers meets the basic requirements.</li> <li>De didactical behaviour of teachers meets the basic requirements.</li> <li>Children with specific educational needs receive the care they need.</li> </ul>	<ul style="list-style-type: none"> <li>The students' results reach a level that may be expected (taking the characteristics of the student population into account).</li> </ul>	<ul style="list-style-type: none"> <li>Overall assessment of very weak, weak, sufficient, or good.</li> <li>A school is weak when the student achievement results at the end of primary education have been insufficient for more than three years.</li> <li>A school is assessed as very weak when the results at the end of primary education have been insufficient for more than three years and the teaching and learning or the pupil care is insufficient.</li> </ul>	<ul style="list-style-type: none"> <li>Framework includes the following hierarchical levels: 5 domains, 10 quality aspects, 46 indicators 10 of which are included in the threshold).</li> <li>There are decision rules for the assessment of each indicator as sufficient or insufficient and decision rules for aggregating the assessment of indicators to an assessment of quality aspects, to the assessment of domains and to the overall assessment of the effectiveness of the school.</li> <li>Data collection during visits include lesson observation (using detailed scoring protocol), interviews with school staff, parents and students (using interview guideline), and document analyses (including an analysis of the school's self-evaluation).</li> <li>Output of the school is measured via analyses of national standardized tests and exit examinations, as well as other indicators such as drop-out, throughput.</li> </ul>
England	Reporting on condition of buildings if these affect quality of education provided or health and safety.	<ul style="list-style-type: none"> <li>Quality of provision, including teaching and learning, curriculum and care, guidance and support for pupils.</li> <li>Leadership and management.</li> </ul>	<ul style="list-style-type: none"> <li>Achievement and standards (progress of pupils and standards achieved in national tests, taking context into account).</li> <li>Pupils' personal development and well-being.</li> </ul>	<ul style="list-style-type: none"> <li>Schools are evaluated as «outstanding», «good», «satisfactory» or «inadequate», with «inadequate» schools being subdivided into «notice to improve» and «special measures,</li> </ul>	<ul style="list-style-type: none"> <li>Framework includes 6 main aspects, which are detailed into 7 criteria for pupils' outcomes, 3 criteria about the effectiveness of the provision, and 8 criteria on the effectiveness of the leadership and management of the schools.</li> </ul>

STANDARDS					
	<i>Input</i>	<i>Process</i>	<i>Output</i>	<i>Threshold for failing schools</i>	<i>Data collection and scoring guidelines</i>
		<ul style="list-style-type: none"> <li>The overall effectiveness of the school.</li> <li>The school's capacity to improve.</li> </ul>		and around 40% of «satisfactory» schools being categorised as requiring a monitoring inspection.	<ul style="list-style-type: none"> <li>Data collection during visits includes: lesson observation (using structured protocol), interviews with school staff, governors and students (using interview guideline), and document analyses (including an analysis of the schools' self-evaluation and questionnaires filled in by parents before the visit).</li> <li>Output of the school is measured via an analysis of performance data, looking at progress of pupils over time and standards achieved in national tests and examinations, taking into account contextual factors (proportions of pupils from deprived backgrounds, mother tongue other than English) and an analysis of students' written work and performance data provided by the school.</li> </ul>
Sweden	NA	<ul style="list-style-type: none"> <li>The main focus is legal conformity and the purpose is to ensure the right of each individual in relation to the Education.</li> <li>Results, standards of achievement, learning and teaching: how the schools are working to increase the standards of achievement, how teachers adapt their teaching to the individual</li> </ul>	<ul style="list-style-type: none"> <li>Results, standards of achievement, learning and teaching: what students learn in view of the objectives in the National Curriculum.</li> </ul>	<ul style="list-style-type: none"> <li>Assessment guideline is being developed, currently holistic approach of reliance on inspector's expertise to evaluate whether schools are complying with legislation or not.</li> </ul>	<ul style="list-style-type: none"> <li>Framework includes: 3 main areas and 16 indicators.</li> <li>Data collection includes: lesson observations (notes are made on set points in an assessment document, no detailed scoring protocol), surveys administered to governing body, analysis of documents (statistics, quality reports, any previous supervisory decisions, and reports from quality inspections, self-evaluation reports), interviews with staff, including school leaders, teachers, and school nurses, as well as pupils, parents, and the politicians in charge.</li> </ul>

STANDARDS					
	<i>Input</i>	<i>Process</i>	<i>Output</i>	<i>Threshold for failing schools</i>	<i>Data collection and scoring guidelines</i>
Ireland	Information is collected on the allocation of resources for pupils with special educational needs is also obtained, as is information relating to building programmes and plans for the improvement or upgrading of facilities (not separately assessed).	<p>At primary level:</p> <ul style="list-style-type: none"> <li>the quality of the school management;</li> <li>the quality of school planning;</li> <li>the quality of teaching and learning;</li> <li>the quality of support for pupils.</li> </ul>	<ul style="list-style-type: none"> <li>No output standard, although test results are analysed during the inspection visit, but not reported on. The results of analyses of test scores with school principals and/or teachers and help to inform the inspectors' judgements regarding the quality of teaching and learning, but statistical data are not included in the inspectors' reports.</li> </ul>	<ul style="list-style-type: none"> <li>No threshold. A single overall rating is not applied to the evaluation at present, a series of recommendations for future action by the school management or by subject teachers is given.</li> </ul>	<ul style="list-style-type: none"> <li>Framework includes: 5 areas of evaluation, subdivided into 143 «themes for self-evaluation». Schools are required in theory to gather evidence and then make judgments about their own performance on a four-part rating scale in respect to each theme. This process of self-evaluation then informs the work of a visiting team of inspectors that carries out «whole school evaluations» (WSE).</li> <li>Data collection includes an analysis of documents (school information form filled in by principal, documentation from the school on enrolments, curriculum provision, planning, policies, etc.) questionnaires to principals and teachers who coordinate subject departments, interviews with school staff, lesson observations.</li> </ul>



STANDARDS					
	<i>Input</i>	<i>Process</i>	<i>Output</i>	<i>Threshold for failing schools</i>	<i>Data collection and scoring guidelines</i>
Austria (province of Styria)	NA	<ul style="list-style-type: none"> <li>• Teaching and learning.</li> <li>• Classroom and schools as a space for living.</li> <li>• Partnerships in schools and external relationships.</li> <li>• School management.</li> <li>• Professionalism and personnel development.</li> </ul>	<ul style="list-style-type: none"> <li>• NA. There are standard-based performance tests but Inspectorate is not allowed to view or use these to measure output.</li> </ul>	<ul style="list-style-type: none"> <li>• No threshold, an overview of strengths and weaknesses is provided, and consequences and necessary actions are reported to the school and to relevant authorities.</li> </ul>	<ul style="list-style-type: none"> <li>• Output data is collected, inspectors (at primary level) compare the achievements of pupils in standardised tests in English and Mathematics with national norms. Inspectors (at primary level) analyze ongoing testing used by teachers in all subjects for both formative and summative purposes, and standardized tests for English and Mathematics, developed by an external research agency and purchased and administered by teachers in the school.</li> <li>• Framework includes 5 school quality areas, detailed in 16 criteria, but focus is on a critical analysis of the schools' self-evaluation procedures and results as laid down in the school programme. Only, if this desk research («methodical-substantive analysis of the self-evaluation») reveals deficiencies (with respect to the procedures of self-evaluation) or show that «essential requirements and standards» (this relates to the results of selfevaluation) are not achieved, a «specific examination» has to take place. Measures of counselling and support will precede such an examination.</li> <li>• Data collection includes: document analysis of «inspection folder» (put together by the school) including</li> </ul>

STANDARDS					
	<i>Input</i>	<i>Process</i>	<i>Output</i>	<i>Threshold for failing schools</i>	<i>Data collection and scoring guidelines</i>
Czech Republic	<ul style="list-style-type: none"> <li>• During a state check auditors examine whether legal regulations relating to the provision of education and school services are met (additional to institutional inspection).</li> <li>• Personnel conditions.</li> <li>• Material prerequisites.</li> <li>• Financial prerequisites.</li> </ul>	<ul style="list-style-type: none"> <li>• Equal opportunities for education.</li> <li>• School education programmes.</li> <li>• School management.</li> <li>• Effective organization of education.</li> <li>• Effective support of personality development of children.</li> <li>• Partnership.</li> <li>• Effective support of development of key competencies of children, pupils and students.</li> </ul>	<ul style="list-style-type: none"> <li>• Systematic evaluation of individual and group education results of children.</li> <li>• Evaluation of overall results in education at school.</li> </ul>	<ul style="list-style-type: none"> <li>• Inspectorate evaluates schools on a four-level scale: A, B, C, D. Evaluation outcome A may lead to removal of a school from the Register of Schools. Evaluation D is considered good practices.</li> </ul>	<p>all previous school development plans of the past 5 years, the school's programme for individual promotion of all students, a list of all teachers (including their age), timetables, list of students, lesson observation (using structured protocol which is also provided to teachers), interview with teachers, students parents, principal and mayor (using structured guidelines).</p> <ul style="list-style-type: none"> <li>• Framework includes 7 main areas, which are detailed in 22 indicators which are detailed into 60 subindicators.</li> <li>• Data collection includes lesson observation (using observation protocol), analysis of outcomes of self-evaluation of schools, inspection analyses of school documents, on-the-spot inspections, observations of students and teachers in classes and standardized interviews with pedagogical staff and contact with school founders.</li> <li>• Output is measured through upper secondary leaving examinations which are the main responsibility of individual schools with no use of national tests, no tests in primary education.</li> <li>• Standardized tests were introduced after 2011.</li> </ul>

Remaining on the topic of the global or more specifically operationalized nature of inspection standards, it should be noted that the examples from quantitative indicator systems presented in this chapter, are mostly more specific than the inspection standards, listed in Table 1. But, of course, in this respect one should not lose sight of the fact that systematic quantitative indicator systems differ from school inspection, in the sense that the latter have the great asset of expert professional quality *judgment*, while the former depend on *measurement* techniques. Both evaluation approaches (quantitative indicators and inspections) present a different kind of evaluation procedure. Indicator sets are applied by means of standardized data collection procedures and research methods, while in school inspections, the inspection standards and check-lists are more to be seen as tools and «extensions» of the professional expert judgements. In this sense inspection frameworks can purposefully be more global than quantitative indicator systems. Nevertheless, the more extensive indicator sets could be used as a resource in the development of inspection frameworks, and possibly partially be copied as a basis for structured classroom observation schedules, and to «scaffold» professional judgements of school inspectors. There are many options and choices of instruments to inform and design inspection frameworks. As an illustration of a relevant set of inspection standards, the set of indicators developed by Scheerens *et al.* (2011) is cited in the Annex.

However, some Inspectorates of Education (such as Ofsted in England) have also abandoned such detailed check-lists as they turned into standard and scripted recipes for school improvement. Schools and other stakeholders (e.g. school improvement partners and developers of school self-evaluations) developed so called «Ofsted-approved» school organization and teaching models and these were increasingly copied and pasted by schools, without any reflection or consideration of whether these practices are fit for purpose for the specific classroom and school context in which they are implemented.

Table 4 also summarizes how a number of Inspectorates of Education include school output in their evaluation of school quality. The description of how test data is analysed to assess school output indicates that these evaluations are still rather rudimentary and make limited use of the more sophisticated value added models of analysing and reporting on school output as described earlier in this chapter. The Inspectorates of Education that have access to student achievement data (the Netherlands and England) take into account potential instability in the data by calculating averages (generally across three years) when assessing output of schools. Most Inspectorates of Education however do not seem to have detailed and high quality performance data available to make such analyses as there is no national standardized testing in place, or as they are not allowed to access such data. A number of Inspectorates of Education instead analyse teacher assessments or

students' work to get a sense of the output of the school. Such analyses are, interestingly, in some cases (e.g. Ireland) also used to evaluate the quality of processes (e.g. quality of teaching) and, as a result, act as a proxy for the quality of educational processes in the school.

Looking at Table 4, it is also remarkable to see that, despite the recent focus and recognition of teaching quality as the main condition for school quality, no Inspectorate of Education included teacher characteristics in their framework of inspection standards. Even though the Dutch inspection framework incorporates classroom-level criteria, such as about pedagogical and didactic behaviour of teachers, they are only assessed on the school-level by averaging the scores of a selection of classroom observations. The previous section however suggested that teaching/instruction level conditions, such as high expectations, a challenging teaching approach, an orderly learning environment and clear and structured teaching are more important than school level conditions in improving student achievement. Most inspectorates however do not explicitly evaluate teaching or teachers on a classroom/subject or grade level, preferring instead to evaluate school level conditions and general instruction characteristics or teaching patterns such as learning time, school leadership and school climate.

A final «council» to further and future developments of national school inspection frameworks is therefore the relative emphasis on school organizational as compared to teaching and learning, or didactic standards. A prudent warning could be for designers and adaptors of inspection frameworks not to lose sight of the primary process of teaching and learning.

## ANNEX

A more elaborated set of output, input and school process indicators (Scheerens *et al.*, 2011).

Table 1. – Overview of educational outcome indicators.

MAIN CATEGORIES OF OUTCOME INDICATORS	SUB-CATEGORIES	TECHNICAL ISSUES
<i>Output indicators</i>	<i>Achievement measures</i> <ul style="list-style-type: none"> <li>• Subject matter based.</li> <li>• Literacy (reading, mathematical, scientific).</li> <li>• Competencies (e.g. learning to learn).</li> </ul>	<ul style="list-style-type: none"> <li>• Value-added effect measures; growth curves.</li> <li>• Assessment methodology (ranging from multiple choice tests to authentic assessment).</li> <li>• Criterion versus norm-referenced testing.</li> </ul>

MAIN CATEGORIES OF OUTCOME INDICATORS	SUB-CATEGORIES	TECHNICAL ISSUES
<i>Outcome/attainment indicators</i>	<i>Attainment measures</i> <ul style="list-style-type: none"> <li>• Graduation rates.</li> <li>• Proportion of students graduated without delay.</li> <li>• Drop-out rates.</li> <li>• Class repetition rates.</li> </ul>	<ul style="list-style-type: none"> <li>• Controlling for selection oriented school policies.</li> </ul>

Table 2. – School level financial and material resources indicators.

SCHOOL LEVEL FINANCIAL AND MATERIAL RESOURCES
<ul style="list-style-type: none"> <li>• Proportion of the school's budget that is acquired through other than public funding.</li> <li>• School building facilities.</li> <li>• Classroom equipment (furniture, computers, etc.).</li> <li>• School supplies like pencil and paper, chalk board, flipchart.</li> <li>• Availability of textbooks in the major school subjects.</li> <li>• Basic services like separate toilets for girls and boys, water, electricity, heating, telephone, provision of ancillary services, regarding nutrition, health and transportation.</li> </ul>

Table 3. – Overview of examples of process indicators of school functioning.

PROCESS INDICATORS DEFINED AT SCHOOL LEVEL
<p><i>Community involvement</i></p> <ul style="list-style-type: none"> <li>• The degree of actual involvement of parents in various school activities (the teaching and learning process, extra-curricular activities and supporting activities).</li> <li>• The percentage of the total annual school budget that is obtained from the local community.</li> <li>• The amount of discretion local school boards have in the conditions of labour of teachers (possible operationalizations in EDUCO project – El Salvador).</li> </ul>
<p><i>School financial and human resources</i></p> <ul style="list-style-type: none"> <li>• Average years of teachers' experience per school.</li> <li>• School level pupil teacher ratio.</li> <li>• Average class size per school.</li> <li>• Proportion of formally qualified teachers per school.</li> <li>• School managerial «overhead» (principal and deputy-principal fte per 1000 students).</li> </ul>
<p><i>Achievement oriented policy</i></p> <ul style="list-style-type: none"> <li>• Whether or not schools set achievement standards.</li> <li>• The degree to which schools follow (education) careers of pupils after they have left the school.</li> <li>• Whether or not schools report achievement/attainment outcomes to local constituencies.</li> </ul>
<p><i>Educational leadership</i></p> <ul style="list-style-type: none"> <li>• The amount of time principals spend on educational matters, as compared to administrative and other tasks.</li> <li>• Whether or not principal's appraise the performance of teachers.</li> <li>• The amount of time dedicated to instructional issues during staff meetings.</li> </ul>

---

PROCESS INDICATORS DEFINED AT SCHOOL LEVEL

---

*Continuity and consensus among teachers*

- The amount of changes in staff over a certain period.
- The presence or absence of school subject-related working groups or departments (secondary schools).
- Frequency and duration of formal and informal staff meetings.

*Orderly and safe climate*

- Statistics on absenteeism and delinquency.
- Ratings of school discipline by principals, teachers and pupils.

*Efficient use of time*

- Total instruction time and time per subject matter area.
- Average loss of time per teaching hour (due to organization, moving to different rooms, locations, disturbances).
- Percentage of lessons «not given», on an annual basis.

*Opportunity to learn*

- Teacher or student ratings of whether each item of an achievement test was taught or not.

*Evaluation of pupils' progress*

- The frequency of use of curriculum specific tests at each grade level.
- The frequency of use of standardized achievement tests.
- The actual use teachers make of test results.

*Ratings of teaching quality*

- Quality of instruction as rated by peers (other teacher).
  - Quality of instruction as rated by students.
- 

Table 4. – Overview of effective teaching and learning variables.

---

EFFECTIVE TEACHING VARIABLES

---

*Main teaching factors*

- Opportunity to learn.
- Structuring and scaffolding (cognitive structuring).
- Stimulating engagement (motivational structuring).
- Climate aspects: task orientation;  
mutual respect;  
orderliness, safety.
- Monitoring and questioning.
- Feedback and reinforcement.
- Modeling learning and self-regulation strategies.
- «Authentic» applications.
- Adaptive teaching.

*Learning strategies of students*

- Overt: engaged learning time;  
student use of resources;  
cooperative learning.
  - Covert: self-regulatory capacity;  
auto-control;  
meta-cognitive «actions»;  
learning styles.
-

## REFERENCES

- American Statistical Association (2014). *ASA statement on using value-added models for educational assessment*. Alexandria, VA: Author. Retrieved (January 2015) from: [https://www.amstat.org/policy/pdfs/ASA\\_VAM\\_Statement.pdf](https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf)
- Bosker, R. J., Guldemon, H. G., Hofman, R. H., & Hofman, W. H. A. (1989). De stabiliteit van schoolkwaliteit. In J. Scheerens & J. C. Verhoeven (Hg.), *Schoolorganisatie, beleid en onderwijskwaliteit*. Lisse: Swets and Zeitlinger.
- Cotton, K. (1995). *Effective schooling practices: A research synthesis* (1995 update). Northwest Regional Educational Laboratory, School Improvement Research Series.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness*. London - New York: Routledge.
- De Grauwe, A. (2007). *Module 7. Alternative models in reforming school supervision*. [http://www.iiep.unesco.org/fileadmin/user\\_upload/Cap\\_Dev\\_Training/Training\\_Materials/Supervision/SUP\\_Mod7.pdf](http://www.iiep.unesco.org/fileadmin/user_upload/Cap_Dev_Training/Training_Materials/Supervision/SUP_Mod7.pdf)
- De Volder, I. (2012). *Externe schoolevaluaties in Europa. Een vergelijkend onderzoek*. Antwerpen: Grant uitgevers.
- Dijkstra, A. B., De la Motte, P. I., Ehren, M. C. M., & Eilard, A. (2014). Discussion. School inspections and school improvement in the social domain. The assessment of social outcomes of education. In A. B. Dijkstra & P. I. De la Motte (Eds.), *Social outcomes of education: The assessment of social outcomes and school improvement through school inspections* (pp. 189-215). Amsterdam: Amsterdam University Press.
- Donaldson, M. L., & Johnson, S. M. (2010). The price of misassignment: The role of teaching assignments in Teach for America teachers' exit from low-income schools and the teaching profession. *Educational Evaluation and Policy Analysis*, 32(2), 299-323.
- Eddy Spicer, D., Ehren, M., Bangpan, M., & Khatwa, M. (2014). *Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised? A realist synthesis of school accountability in low- and middle-income countries. Protocol*. London: EPPI-Centre, Social Science Research Centre, Institute of Education, University of London. <http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=x-3hzguBXhY%3D&tabid=3174>
- Ehren, M. C. M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on school improvement: Describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, 25(1), 3-43. <http://dx.doi.org/10.1007/s11092-012-9156-4>
- Gray, J., Jesson, D., Goldstein, H., Hedges, K., & Rasbash, J. (1995). A multi-level analysis of school improvement: Changes in school's performance over time. *School Effectiveness and School Improvement*, 6, 97-114.

- Hamilton, L. S., & Koretz, D. M. (2002). Tests and their use in test-based accountability systems. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education*. Santa Monica, CA: Rand cooperation. [http://www.rand.org/pubs/monograph\\_reports/MR1554/](http://www.rand.org/pubs/monograph_reports/MR1554/)
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economics Literature*, 24(3), 1141-1177.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19, 141-164.
- Hattie, J. (2009). *Visible learning*. Abingdon: Routledge.
- Hopkins, D., Stringfield, S., Harries, A., Stoll, L., & Mackay, T. (2014). School and system improvement. *School Effectiveness and School Improvement*, 25, 257-281.
- Levine, D. K., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, WI: National Centre for Effective Schools Research and Development.
- Marzano, R. J. (2003a). *What works in schools*. Alexandria, VA: ASCD.
- Mourshed, M., Chijioko, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. London: McKinsey. <http://mckinseysociety.com/how-the-worlds-most-improved-school-systems-keep-getting-better/>
- Muijs, D., Creemers, B., Kyriakides, L., Van der Werf, G., Timperley, H., & Earl, L. (2014). Teaching effectiveness. A state of the art review. *School Effectiveness and School Improvement*, 24, 231-256.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Purkey, S. C., & M. S. Smith (1983). Effective schools: A review. *The Elementary School Journal*, 83, 427-452.
- Reynolds *et al.* (2014). Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., Van Damme, J., Teddlie, C., & Stringfield, S. (2012). Educational Effectiveness Research (EER): A state of the art review. *School Effectiveness and School Improvement*, 25, 197-230.
- Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.
- Scheerens, J. (1992). *Effective schooling, research, theory and practice*. London: Cassell.
- Scheerens, J. (2013). *What is effective schooling? A review of current thought and practice*. Paper for the International Baccalaureate Organization, Washington, DC.
- Scheerens, J. (2014). School, teaching and system effectiveness: Some comments on three state of the art reviews. *School Effectiveness and School Improvement*, 25, 282-290.
- Scheerens, J., & Creemers, B. P. M. (1989). Conceptualizing school effectiveness. *International Journal of Educational Research*, Special Issue: *Development in School Effectiveness Research*, 13(7).



- Scheerens, J., Glas, C. A., Thomas, S. M., & Thomas, S. (2003). *Educational evaluation, assessment, and monitoring: A systemic approach*, Vol. 13. London: Taylor & Francis.
- Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organisation and Management.
- Scheerens, J., Luyten, H., Van den Bergh, S. M., & Glas, C. A. W. (2015). Exploration of direct and indirect associations of system level policy amenable variables and reading literacy performance. Accepted for publication in *Educational Research and Evaluation*.
- Scheerens, J., Luyten, H., & van Ravens, J. (2011). *Perspectives on educational quality. Illustrative outcomes on primary and secondary education in the Netherlands*. Research Briefs. Dordrecht - Heidelberg - London - New York: Springer.
- Seashore Louis, K., Leithwood, K., Wahlstrom, K., & Anderson, S. (2010). *Learning from leadership: Investigating the links to improved student learning*. University of Minnesota, Center for Applied Research and Educational Improvement - University of Toronto, Ontario Institute for Studies in Education.
- Thomas, S. M., Peng, W. J., & Gray, J. (2010). Modeling patterns of improvement over time: Value added trends in English secondary schools. *Oxford Review of Education*, 33, 261-295.
- Tolofari, S. (2005). New public management and education. *Policy Futures in Education*, 3(1), 75-89. <http://firgoa.usc.es/drupal/files/tolofari.pdf>
- Van Bruggen, J. C. (2010). *Inspectorates of Education in Europe: Some comparative remarks about their tasks and work*. Standing International Conference of Inspectorates of Education in Europe (SICE).
- Vermeer, N., & Van der Steeg, M. (2011). *Onderwijsprestaties Nederland in Internationaal Perspectief. CPB Achtergronddocument bij CPB Policy Brief 05, 2011*. Den Haag: CPB.
- Visscher, A., & Ehren, M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken (analyse in opdracht van de Kenniskamer van het Ministerie van Onderwijs, Cultuur en Wetenschap)*. <http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/rapporten/2011/07/13/de-eenvoud-en-complexiteit-van-opbrengstgerichtwerken/visscher-ehren-eenvoud-en-complexiteit-van-opbrengstgericht-werken-def-1-7-11.pdf>

## RIASSUNTO

*Questo articolo descrive il modo in cui gli Ispettorati della Pubblica Istruzione rendono operativi gli obiettivi dell'ispezione (controllo, miglioramento, collegamento). Lo studio fornisce una panoramica, ed alcuni esempi, degli indicatori utilizzati in diversi Paesi per*

*monitorare e valutare le scuole. Descriveremo l'ispezione e la valutazione dei processi e dei risultati delle attività scolastiche (inclusi i giudizi espressi da esperti), confrontandoli con i sistemi di ispezione che si concentrano sul controllo dei requisiti di ingresso e sulla verifica del rispetto della normativa. Discuteremo il valore e l'adeguatezza dei differenti sistemi alla luce delle recenti ricerche sull'efficacia della scuola. I risultati raccolti suggeriscono che le condizioni di insegnamento/apprendimento come avere grandi aspettative, avere un approccio didattico stimolante, un ambiente di apprendimento ordinato e un insegnamento chiaro e strutturato, sono più importanti delle condizioni generali della scuola per il miglioramento dei risultati degli studenti. La maggior parte degli ispettorati tuttavia non valuta in modo esplicito le condizioni di insegnamento a livello del rapporto in aula, preferendo invece valutare le condizioni generali della scuola e dell'istruzione o i modelli di insegnamento e i tempi di apprendimento, la leadership scolastica e il clima scolastico. Un «consiglio» finale per eventuali e futuri sviluppi dei sistemi ispettivi della scuola è quindi quello di relativizzare l'enfasi sulla dimensione organizzativa della scuola in rapporto alle dimensioni dell'insegnamento e dell'apprendimento, o agli standard didattici. Un avvertimento prudente potrebbe essere, per chi progetta e adatta i sistemi ispettivi, di non perdere di vista il processo primario dell'insegnamento e dell'apprendimento.*

*Parole chiave:* Efficacia della scuola, Indicatori dell'ispezione, Ispezione scolastica, Paesi Europei, Ricerca basata sull'evidenza.

*How to cite this Paper:* Scheerens, J., & Ehren, M. (2015). The evidence base for school inspection frameworks [I sistemi di ispezione scolastica basati sull'evidenza empirica]. *Journal of Educational, Cultural and Psychological Studies*, 12, 43-76. doi: 10.7358/ecps-2015-012-sche