School (Self) Evaluation and Student Achievement

HOFMAN, Roelande H.
GION, the Institute for Educational Research, University of Groningen
DIJIKSTRA, Nynke J.
GION, the Institute for Educational Research, University of Groningen
HOFMAN, W. H. Adriaan
Education at UOCG, University centre for learning and Teaching
University Groningen

Abstract: The last two decades have shown a transition in the governance philosophy of national governments combining devolution of authority with a strong emphasis on the quality of education. Most accountability policies combine central control with growing relative autonomy reserved for school governing bodies and individual schools. Educational systems have developed towards accountability policies in which schools maintain autonomy for their pedagogical, instructional and organizational practices (internal control). At the same time they are held accountable for the quality level of their schools' education to public authorities (external control). It is not clear whether these contrasting types of accountability policies contribute to the improvement of the schools' educational quality. In this study the general research question is whether a relationship exists between school (self) evaluation and student achievement. Using a database of 81 primary schools and 2099 students analysis of variance and multilevel analysis show what factors characterizing the (type) of school (self) evaluation contribute to students' cognitive achievement. The quality assessments of the national inspectorate shows positive relationships with progress in school self evaluation and school quality. Some school self evaluation perspectives seem positively related to student achievement.
Keywords: school effectiveness, quality assurance, accountability, achievement, multi-level

Introduction and research problem

New models of school regulations based upon accountability measures, evaluation practices, have lately been given considerably more attention. Such models include the development of national educational policies that include standards for schools' as prescribed by the National Inspectorates of Education, external student assessment, internal and external evaluations (audits) of schools and the development of examples of best practice (Hofman, Dijkstra, de Boom & Hofman, 2005; Broadfoot, 1996). The objective of these educational policies is to assure and enhance educational quality and improve the schools within an education system from a two-sided approach. More precisely these educational policies point at two different types of approaches to accountability: a so-called govern-based versus a market-based accountability approach.
In general, in the last two decades a transition has been noticeable in the governance philosophy of national governments combining devolution of authority with a strong emphasis on the quality of education (Hofman, Dijkstra, Hofman & De Boom, 2004a). According to Eurydice, a resourceful European database, the improvement of quality in education became the central concern of educational policy in many European countries (Wastiau-Schlüter, 2004). Several Western-European countries accomplished – or are working on – legislation and monitoring in the field of school (self) evaluation in which the schools' own responsibility for quality is

stressed (MacBeath, Meuret, Schratz & Jakobssen, 1999; Wilcox & Gray, 1996; Leithwood, Edge & Jantzi, 1999; Reezigt, 2001; National Inspectorate of Education, 2003a).

A system of school (self) evaluation can be understood from several positions depending on the school's goals, ranging from a restricted view that focuses purely on the school's outcomes (output), to a broad perspective in which the school's input, internal processes at school and classroom level and performance are assessed (e.g. the range may include context, input, processes and output) (Hofman, Dijkstra & Hofman, 2005). According to the findings of the Dutch Inspectorate of Education, not all schools develop an integrated and systematic way of school (self) evaluation. Some schools opt for a restricted form taking into consideration several bottlenecks and conditions in their schools' context (e.g. many ethnic minority pupils, very poor language performance of pupils). Furthermore, it seems that schools with unsatisfactory school self evaluation also might have poor school performance and a lacking quality of the teaching-learning process (Inspectorate of Education, 2003c). This leads to the following general research question:
  'Is there a relationship between school (self) evaluation and student achievement''?

Theoretical background
Two perspectives are fundamental to the research in this paper. The *first perspective* observes school (self) evaluation from the viewpoint of actors involved. It distinguishes two main actors in school evaluation: an internal actor (schools) and an external actor (the National Inspectorate of Education. The *second perspective* focuses on the actual practices, processes and viewpoints on school (self) evaluation. In this perspective research outcomes of theories of effective schools and effective management practices are investigated and those suggest that different approaches towards school (self) evaluation are expected depending upon the focus of a school (e.g. the school as a team of teachers, the optimizing of the pupils' school career, external pressure for school evaluation).

External versus internal school evaluation
Most accountability policies show an interesting combination of central control and steering by the countries' central government with relative autonomy reserved for school governing bodies and individual schools. The literature on the subject of school (self) evaluation makes a distinction between two major functions of school evaluation: an internal and an external one (Newmann, King & Rigdon, 1997; Wilcox & Gray, 1996; Wastiau-Schlüter, 2004). The *external function* focuses on the safeguarding of standards of quality of schools and in most European countries a National Inspectorate of Education is responsible for this task. In that respect the government (through the efforts of the Inspectorates) hold strategic control over the goals of the education system, based upon standards, objectives and criteria of success regarding the outcomes of a school. While at the same time the daily management practices are left to the local schools' responsibility. The *internal function* is the responsibility of the schools themselves; schools are supposed to determine, guarantee, and guard their quality and improve the teaching-learning process and their school performance (Hofman et al., 2004a). In general, several European countries acknowledge that the evaluation of their schools is at the very heart of the quality of schooling and this includes evaluation by an external Inspectorate, as well as internal procedures by the school community itself (Wastiau-Schlüter, 2004; Eurydice, 2004).

Newmann, King and Rigdon (1997) studied the connection between internal and external types of school accountability. They concluded that (a) external accountability seems to strengthen the internal monitoring and use of evaluation systems within schools, and (b) seems to encourage the search for success or failure in the schools' educational practices. Based on outcomes of their European Pilot project "Quality evaluation in school education', MacBeath et al. (1999) present the clear message that internal and external evaluation are complementary processes in which the relation between the two should be clearly articulated. Outcomes of this European pilot project also seem to indicate that more is to be expected from a system of internal school (self) evaluation than from an external focus: in general 37% of the schools in the project expect improvement in teaching from external and 55% from internal school (self) evaluation, 36% of the schools expect improvement of management from an external and 63% from an internal system (MacBeath et al., 1999).

Theories of school management and school (self) evaluation

School (self) evaluation includes the determination and judgment of the quality of a school and next to that, if necessary, the improvement of the school. These two sides of a medal are also fundamental to this research. Hofman & Hofman (2003) developed a framework for school (self) evaluation using relevant standards from an accountability perspective combined them with a school improvement perspective as well. This leads to the use of the so-called *CIPPO model* (context, input, processes at school level, processes at classroom level, output) which is an adapted version of the "Context-Input-Process-Output model" that has been widely used in research into school and classroom effectiveness (e.g. Bosker, 2001; Creemers, 1994; Hofman, 1993; Scheerens, 1989).

For the school improvement perspective, the framework joins theoretical organizational perspectives that focus on school development and improvement using a system of integral school (self) evaluation as a starting-point (e.g. Dalin 1993; Hofman & Hofman, 2003; Reezigt, 2001; Reynolds & Teddlie, 2000; Stoll & Wikeley, 1998; Deming, 1989). In this process, four implementation stages for improvement reflect the *Plan-Do-Check-Act cycle*; the stage of orientation and preparation (plan phase), implementation (do phase), evaluation (check phase) and finally the institutionalization or integration (act/adapt phase).

Studies into effective school improvement offer knowledge on the matters schools should commence in relation to self-evaluation. These lead to three general perspectives regarding how school (self) evaluation is developed or takes place in a certain school setting (a) school self evaluation within schools as high-reliability organizations (Stringfield & Slavin, 2001; Hofman, Hofman & Guldemond, 2000), (b) school self evaluation within schools as learning organizations (Leithwood, Aitken & Jantzi, 2001; Arts, Kok, Verbiest, Sleegers & De Wit, 2003), and (c) school self evaluation developed under pressure of external organizations (Hofman, et al, 2005). These theories focus on different groups and perceptions in the schools as input for school (self) evaluation.

The *first* theoretical approach views the school as a high-reliability organization focuses on the *pupils*. In this type of school all members strive for perfection and presume the principle *trial without error* with an optimal school career for all pupils as their goal (Stringfield & Slavin, 1992; LaPorte & Consolini, 1991). The idea is that an organization cannot permit itself to make

mistakes; the consequences of a mistake would be disastrous. Twelve factors are being considered to characterize a high-reliability school, according to Stringfield & Slavin (1992). However, the improvement of the quality of schools is related in particular to the following three factors: (a) monitoring and use of extensive rich data, (b) extensive recruiting of staff, training and retraining, and (c) mutual staff monitoring without loss of autonomy and confidence (Stringfield, Reynolds & Schaffer, 2001).

The second *theoretical* approach views the school as a learning organization that focuses on the *teacher of staff.* Leithwood & Aitken (1995) define this: as "a group of people pursuing common purposes (and individual purposes as well) with a collective commitment to regularly weighing the value of those purposes, modifying them when that makes sense, and continuously developing more effective and efficient ways of accomplishing those purposes". This definition approaches the learning organization as a dynamic process; the goal is reaching not a static finish, but a continuous accentuation of purposes and means. In a learning organization, schools need to adapt to their context and population while giving shape to five aspects that promote *collective learning* found by Leithwood, Jantzi & Steinbach (1998): (1) school vision and mission, (2) school culture, (3) school structure, (4) school strategies, and (5) school policy and means.

A *last* theoretical approach originates from contingency theory (Mintzberg, 1998) and includes the perspective that school (self) evaluation is stimulated by the external community that surrounds a school. For example schools can be stimulated or forced by the Local Educational Authority or by parents to evaluate and improve their quality. Reezigt (2001) uses the term external pressure to describe one of the most important factors to stimulate school (self) evaluation more specifically school improvement. Choice and competition in education have found growing support among policy makers recently. Yet evidence on the actual benefits of market-oriented reforms is at best mixed (Gibbons, et. al., 2006).

Research model, design and method

In the theoretical background of our research several theories, approaches and actors have been presented that are essential to a large research concerning the quality management of schools from an internal viewpoint and the relationship with the external assessments of the schools' quality by the National Inspectorate of Education. To make the possible associations between those essentials more clear we developed a conceptual research model.

            ---- insert here research model ------------

Sample

The research into school (self) evaluation is a study into the state of the art in Dutch elementary schools. The sample includes 939 primary schools of which the principal of the school has filled out a questionnaire about school (self) evaluation. Next to this, two other dataset have been linked to our study. The first one is a database of the National Inspectorate that includes school level data. The second one is a large-scale national research that includes school and students data as well. The overlap of both datasets have been used in the research described in this paper. The overlap includes 81 primary schools and 2099 students.

Both our samples (n=939) and (n=81) are representative of the schools and student population in primary education in the Netherlands with regard to factors such as achievement levels of the pupils, pupil population, number of pupils in the schools, degree of urbanisation.

Variables and scales: school (self) evaluation
The school (self) evaluation study concerns a baseline measurement among 939 school leaders using questionnaires that include four dimensions: (1) the perspective or focus of the school on school (self) evaluation, (2) the characteristics of the used school (self) evaluation-system, (3) the degree in which schools are actively implementing measures of accountability and school improvement, and (4) the role of interest groups, external support and the use of specific instruments for school (self) evaluation.

(ad 1) Three reliable scales were constructed to measure the focus of the school towards school (self) evaluation: the "schools as learning organizations" (LO) perspective focuses on the school staff i.e the expertise of the teachers; the "high-reliability schools" (HRO) school perspective focuses on the school career of the pupils, and the third "SSE influenced by external pressure" focuses on the involvement of actors surrounding the school (inspectorate, educstional authority, governing body, parents) that have influenced the development of the SSE-system .

(ad 2) Two reliable scales have been developed to describe the evaluation system (process) used in the school: the first one the degree to which the SSE-system includes an integrated system (including the information or input of several actors within the school evaluation process); the second one is a scale based upon the well-known PDCA cycle introduced by Deming (196??) that focuses on a cyclic approach of the school- improvement process using from a Plan-Do-Check-Act cycle in the SSE-process.

(ad 3) The survey includes subscales concerning 'Context / Input', Processes at school level', Processes at classroom level, and 'Output' reflecting the activities of the school in determining its current position (accountability). Next to that, school improvement subscale reflects the stage of improvement of the school also according to the items of the so-called CIPPO model, ranging from orientation, implementation, evaluation, to integration in the school. The two overall scales for accountability and school improvement both contain 27 items. These include four subscales each: 8 indicators for context/input, 7 indicators for processes at school level, 7 for processes at classroom level, and 5 items for output) to measure to what degree schools have implemented accountability and school improvement measures,

(ad 4) The last dimension includes scales that operationalise the possible influence of expertise of teachers, of interest groups, external support and the use of specific school self-evaluation instruments in their school (self) evaluation system.

Table 2. *Psychometric characteristics of dimensions and indicators of SSE*

	M	SD	Range	n items	Alpha
1. Perspective or focus on SSE Focus of SSE on schools as learning organization (LO)	3.33	.51	1-4	5	.70

Focus of SSE on pupils within a High Reliability Organisation (HRO)	3.11	.43	1-4	9	.70
2. Characteristics the SSE system					
Use of a Plan-Do-Ccheck-Act/Adapt cycle	2.93	.40	1-4	14	.86
Use an integrated and systematic SSE - system	2.57	.29	1-3	12	.72
3a. Degree of accountability (AC) overall	1.82	.16	1-2	27	.84
• Accountability context / input (C/I)	1.72	.24	1-2	7	.70
• Accountability processes school (PS)	1.87	.18	1-2	8	.88
• Accountability processes teachers (PL)	1.88	.19	1-2	7	.64
• Accountability outcomes / output (O)	1.82	.24	1-2	5	.63
3b. Degree of school improvement (SI) overall	2.66	.62	1-4	27	.93
• School improvement C/I	2.67	.77	1-4	7	.84
• School improvement PS	2.56	.76	1-4	8	.84
• School improvement PL	2.75	.74	1-4	7	.82
• School improvement O	2.72	.79	1-4	5	.78
4. Groups, support and instruments for SSE					
Expertise of teachers in SSE	2.64	.48	1-3	4	.88
Support by external community and network of schools	1.88	.53	1-4	7	.80
Use of instrument for SSE in general	2.62	.48	1-4	12	.88
Use of SSE-instrument for school improvement	2.34	.60	1-4	8	.88

AC = accountability-scale, SI = school improvement-scale: Pearson's correlation (n = no of schools)
* Correlation is significant op 0.05 level (2-tailed)   ** Correlation is significant op 0.01 level (2-tailed)

Variables and scales: the National Inspectorate of Education
   The National Inspectorate of Education has been quite forthcoming and helpful by making three data files available to our research project. These data files of the National Inspectorate include the results (judgments) of the supervision of the schools according to their supervision framework (see Table 1). The Inspectorate includes (standardized) assessments into three domains: (a) the management of the school self evaluation, (b) the quality of the teaching-learning process, and (c) the quality of school outcomes." .

   Inspectorates' scale Quality control (SSE). Based upon the indicators in the supervision framework and information of the data files of the Inspectorate reliable scales have been constructed (reliability > .80). These scales have been standardized and transformed into one samengestelde criterion variable, that indicates the degree to which the school has put into practice the indicators of quality control (see Table 1) in their school self evaluation system.

   Inspectorates' scale quality of the teaching-learning process. This scale has been developed based upon the indicators of the Inspectorates' supervision framework regarding the teaching and learning process. Using the data files of the Inspectorate three scales have been constructed

(reliability: .64, .86, .79). These scales have also been standardized and transformed into one samengestelde criterion variable, that indicates the degree to which the school has implemented the indicators of the quality of the teaching and learning process: curriculum, learning time, the pedagogical and didactic performances of teachers, the school climate, harmonization with the educational needs of pupils, an active and independent role of pupils and finally support and guidance for pupils (see Table 1)

Inspectorates' scale quality of the school outcomes. The constructed variable assesses to what degree the achievement levels of the schools' pupils over three school years at, under or above the level that may be expected of the schools pupil population. The score ranges from 1 (= under the expected level, through 2 to 3 (= at the expected level) and through 4 to 5 (= above the expected level (for more information on the construction process see Hofman, de Boom, Hofman & van den Berg, 2005). However, it must be said that this scale shows relatively very limited variance between the schools.

Variables and scales: covariates and cognitive outcomes
A large scale National data base has been used for the student data which include three cognitive output measures and student background data.
The cognitive performances at the pupil level are indicated by the criterium variable math. This standardized test has been developed by the National Institute for Test Development (CITO). The exact content of the questions can be found in the tests included in the documentation for PRIMA 5 (School and class characteristics primary education; Basic report PRIMA cohort study Fifth measurement 2002–2003, I. van der Veen, A. van der Meijden, G. Ledoux, SCO-Kohnstamm Institute, 2004).

Math achievement. This test has been developed by CITO to measure the general numeric skills in pupils aged 6-12. The tests consist of three parts, each with a duration of about 45 minutes, which the pupils can complete independently. The tests contain a lot of open questions. There are in principle two subscales: Numbers and Calculations in addition to Measurement, Time and Money. The arithmetic/mathematics test for pupils aged 11-12 contains 120 questions.

Student and school covariates
In order to determine a fair estimate of the achievement levels of students and the effectiveness of schools, we had to take into account the individual characteristics of students (covariates at pupil level) and the school's student body as a whole (school covariates). In this study at <u>pupils level</u> we use four variables as covariates:
- socio-economic background [SES]
- pupils' intelligence [IQ]
- pupils´sexe [sexe]

In this study at <u>school level</u> we use four variables as covariates:
- number of pupils in school [scale]
- place of residence of school [urbanisation]
- number of schools per governing board, one school versus more [onetype board]

Methods of data analysis
To answer our research questions two types of analyses have been concducted: (1) oneway analysis of variance and (2) multilevel analysis.

Oneway analysis of variance. This type of analysis has been used to investigate whether a relationship can be found between our independent variables (scales) that describe the school (self) evaluation system and processes within the school on the one hand with the assessments of the National Inspectorate of Education concerning the quality of the schools school self evaluation system and the quality of the teaching –learning process and the ultimate quality of student performance as well. As we make use of types of school (self) evaluation (see the following section) ANOVA Oneway analysis of variance has been used to investigate whether different types of SSE differ significantly from each other. Significance is based on F-testing with Pos-Hoc analysis of cluster deviation: --, -, o, +, ++ significant deviation between mutual clusters with p < .05.

Multilevel analyses.   In this study we carried out multilevel analyses, using a hierarchically structured database. The multilevel model reflects more realistically than is the case in other models (such as MLWIN) the nested or hierarchical nature of data found in school-effect studies. This model is particularly suitable for the identification of those school level attributes that are correlated with student outcomes, since it allows for the partition of total variance of estimates of within-unit parameter into parameter and sampling error components. The model used in this study has two analysis levels: student and school. In the analysis phase a number of models were constructed in a step-by-step manner. Typically, first an estimate is made of an unconditional ('empty') model and, using this, the proportion of total variation, that is parameter variation, can be assessed. The first model (0) that is always formulated with respect to each criterion variable is the *basic model*. No explanatory variables are included in this model; it only uses the estimates of the total variant components at every model level. The hierarchical linear model (Raudenbush & Bryk, 1986) takes the following simple form (stochastic variables underlined):

$$\underline{Y_{ij}} = \underline{\beta_{0j}} + \underline{\varepsilon_{ij}} \, (0)$$

student i in school j has a position Y_{ij}, which can be split into a school-specific part (the intercept β_{0j}) and a student-specific residue (ε_{ij}). This basic model can be used to estimate the variances at the school and student levels.

Our next step is to formulate our conditional (theoretical) models (the SSE types and scales) and determine the degree to which they account for true parameter variability (Kennedy & Mandeville, 2000). The analyses are carried out with the statistical software program MLWIN, which is able to handle two- and more-level databases (in this case school and student levels) adequately (Longford, 1993).

RESULTS
Can we determine configurations of school (self) evaluation?
A central assumption of this research is that "the total is more than the sum of parts" (Lammers, 1991: Mintzberg, 1979, 1983). As a starting point, we assume that apart from the main effects that point to effective management, the variation between the differences in the schools' effectiveness will more strongly be explained by the interactions of these main effects. We will attempt to uncover the joint effects of composed indicator variables (or configurations of school self evaluation indicators) on the students' performance. It is especially the line of

reasoning in terms of configurations or school types that can be considered as the basic design of this study.

We search for a limited number of basic types in school (self) evaluation. A method that respects the functioning of several indicators in combination is cluster analysis. *A* hierarchical type of cluster analysis is employed to create such configurations using our indicators of school self evaluation. This type of cluster analysis is known as Ward's method (Wishart, 1987). It starts with as many clusters as there are stimuli and in following cycles the most similar clusters are combined. We select the number of most relevant clusters based on the following three criteria (cf. Everitt, 1980):
a. the squared fusion coefficients must increase at substantial intervals;
b. the number of units per cluster has to be substantial;
c. the interpretation of the clusters has to be clear and consistent with the formulated hypotheses.

Table 3. *The best fitting cluster solution*

	Overall n = 939 (100 %)	Cluster 1 n = 73 (7.8 %)	Cluster 2 n = 309 (32.9 %)	Cluster 3 n = 286 (30.5 %)	Cluster 4 n = 271 (28.9 %)
Accountability C/I	1.72 (.24)	--	-	+	+
Accountability PS	1.87 (.18)	-	0	+	+
Accountability PC	1.88 (.19)	-	0	+	+
Accountability O	1.82 (.23)	--	0	+	+
School improvement C/I	2.67 (.77)	-	-	++	-
School improvement PS	2.56 (.75)	-	-	++	-
School improvement PL	2.75 (.74)	-	-	++	-
School improvement O	2.72 (.79)	0	-	++	-

Significance based on F-testing with Post-Hoc analysis
Deviation: --, -, o, +, ++ significant deviation between mutual clusters with p < .05.
Overall-column: mean M for subscales (correlation)

The cluster analysis (see Table 3) draws on the CIPPO model as a directive framework, with levels divided in: Context/Input, Processes at school level, Processes at classroom level, and Output. This model gives a summary of the most important aspects (for each level) that are significant for the determination of the quality of the school for both for the accountability scale and for the school improvement scale. The model covers the aspects that influence the quality of the school and on which research- and education experts agree in a fair degree. The survey data of the elementary schools combine four subscales of the CIPPO-model measuring accountability with four subscales for school improvement (z-scores): Context/Input (C/I), Processes at school level (PS), Processes at classroom level (PC), and Output (O). Based on these scales a typology of school (self) evaluation has been constructed.

Four empirically based types of management, using descriptive scores of the +, 0 and – type, based on the significant deviancy of the clusters per indicator (significance level of p<.05) are presented.

Cluster 1 is the smallest containing 73 of the 939 original schools, that is 8% of the elementary schools in our research. Compared with the other clusters, the implementation of accountability measures is very low; the schools have not determined their actual position concerning the four levels of school (self) evaluation. The same trend is visible for school improvement, with an exception for the scoring on the subscale Output. Schools in cluster 1 paid some attention to the improvement of their performance in terms of outcomes. In short, this cluster can be characterized as *hardly any school (self) evaluation,* with very low accountability (AC) measures and hardly any school improvement (SI) in the stage of implementation (acronym "AC-SI-").

Cluster 2 is the largest cluster and contains 33% of all elementary schools. The schools score average on the accountability subscales (see Table 2). This type of schools show accountability to some extent, with an exception for the below-average score on the subscale concerning the Context/Input measures. Their scoring on the school improvement subscales is comparable to the low scoring of cluster 1. Schools in cluster two can be typified as *average school (self) evaluation,* with some attention to accountability and school improvement that reaches the stage of implementation (acronym "AC0SI-").

Cluster 3 covers 30% of the Dutch elementary schools in the research. This cluster is the counterpart of the first cluster; the schools score the highest on all subscales of accountability and school improvement. Especially the high scores on the school improvement subscales catch the eye. This cluster characterizes as *advanced school (self) evaluation*; including higly implemented accountability measures and school improvement in the evaluation stage (acronym "AC+SI++").

Cluster 4 includes a group of almost 29% schools. This cluster distinguishes itself from the other clusters with an above average scoring on the accountability-scales and a extremely low score on the school improvement subscales. The scores on accountability are comparable with those of cluster 3; however remarkably, the attention for school improvement is considerably lower than the scoring of cluster 1. Therefore, schools in cluster 4 characterize as *mixed school (self) evaluation* (acronym "AC+SI-").

Is school (self) evaluation and the Inspectorate's assessment of school quality related?
We distinguish three domains that the national Inspectorate uses in its supervision framework to judge the quality of the school's educational processes. Based on these domains three scales are constructed (see the methods section). The first scale is the "Inspectorate Quality control scale", which reflects the judgment of the Inspectorate towards the involvement of the schools concerning quality control using school (self) evaluation at school level. The second one is called the "Inspecorate Quality of teaching-learning process scale". It reflects the quality of teaching-learning process of the schools according to the assessment of the Inspectorate. The third scale that has been constructed concerns the outcomes of the schools "Inspectorate Student achievement scale".

The intriguing matter is now whether a relationship exists between these three quality scales (based on judgements by the Inspectorate) and our clusters of school (self) evaluation (based on school directors judgements). Table 4 shows the results of this comparison.

Tabel 4: *Comparison of types of SSE and quality assessments of the Inspectorate*

Inspectieschalen	Kwaliteitszorgtypen				
	Cluster 1 'hardys' (7.8 %)	Cluster 2 'moderate' (32.9 %)	Cluster 3 'advanced' (30.5 %)	Cluster 4 'mixed' (28.9 %)	*Sig.*
Kwaliteit en kwaliteitszorgtypen					
• Inspectorate Quality control scale	-.08 0	-.07 0	.11 0	-.01 0	.149
• Inspecorate Quality of teaching/learningproces s scale	-.04 0	-.03 0	.19 +	-.16 -	.000
• Inspectorate Student achievement scale	.02 0	.06 0	.02 0	-.10 0	.294

Anova (F-toets met Post-hoc tests tussen de vier clusters)
Afwijking --, -, o, +, ++ significante afwijking tussen clusters onderling met significantiecriterium .05).

Table 4 shows that a significant difference is visible for one of the constructed quality scales especially between cluster 3 and cluster 4. It concerns the scale that indicates the quality process of the "teaching and learning process" in the schools. We observe a positive effect of cluster 3 (*advanced in school (self) evaluation*), versus the score in cluster 4 (mixed *in school (self) evaluation*) and cluster 1 and 2 as well. Schools characterized by well implemented accountability measures and who are already in the stage of evaluation of their school improvement measures have a significant better quality of the teaching-learning process than the other groups of schools.

This indicates that schools with an advanced school (self) evaluation system show a higher quality (according to the Inspectorate) regarding (see Table 1) the curriculum, the use the available learning time, the pedagogical and didactic performances of teachers, the school climate, harmonization with the educational needs of pupils, an active and independent role of pupils and finally a higher quality of support and guidance for pupils, in comparison to the rest of the schools in our study.

Although, significant differences have not been found for the other two scales of the Inspectorate the scale Quality control shows the same trend as the before mentioned outcome: the advanced SSE shows the highest score in this respect. Surprising is the fact that no significant differences have been found in regard to the student achievement of pupils in the schools of our sample. However, as was stated earlier this scale shows relatively very limited variance between the schools.

Does the type of School (Self) Evaluation have an impact on pupil cognitive achievement?
    The research question that is central to this section of our paper includes the testing of the central or general hypothesis: there is a relationship between types of school self evaluation and the cognitive achievement of pupils. This hypothesis has been tested based upon a sub sample of 81 schools of or original dataset hat include 2099 pupils in grade 8 of elementary schools (11-12 year olds). Our general hypothesis has been specified and divided in the following sub hypotheses.

    Sub hypothesis 1: Pupils in the enhanced type of school self evaluation show a higher achievement level for mathematics than pupils in the other three types of SSE.

    Sub hypothesis 2: Pupils in type 1 including hardly any school self evaluation score lower achievement levels for mathematics than pupils in the other three types and especially as compared to the type 3 schools with enhanced SSE..

    Table 5 shows to what degree significant differences have been found for the four types of school self evaluation on the measure for cognitive achievement.

Table 5 *Significant differences in cognitive achievement between SSE types (subset)*

	Total	Type 1 (hardly)	Type 2 (moderate)	Type 3 (advanced)	Type 4 (mixed)	Sig.
Math	71.38 (20.75)	65.44 (23.08)	71.81 (20.13)	70.76 (20.41)	72.44 (21.14)	.020

Anova (F-toets met Post-hoc tests tussen de vier clusters)
Afwijking --, -, o, +, ++ significante afwijking tussen clusters onderling met significantiecriterium .05).

    Significant differences are shown for the math achievement measure. However, our first *sub hypothesis must be rejected*: the third SSE type does not score highest on math achievement. Furthermore, Table 5 shows that the second *sub hypothesis is confirmed*: the schools that implemented hardly any school self evaluation measures (type 1) scores significantly lowest on mathematics achievement.

    However, to compare the four types of SSE in a fair way we have to take into account the pupil population and other characteristics of the schools. Multilevel analysis is a fitting model for such an analysis.

Multilevel analyses
    In this study we carried out multilevel analyses, using a hierarchically structured database. The multilevel model reflects more realistically than is the case in other models the nested or hierarchical nature of data found in school-effect studies. This model is particularly suitable for the identification of those school level attributes that are correlated with student achievement, since it allows for the partition of total variance of estimates of within-unit parameter into parameter and sampling error components. Typically, first an estimate is made of an

unconditional ('empty') model and, using this, the proportion of total variation, that is parameter variation, can be assessed. Our next step is to formulate our conditional (theoretical) models and determine the degree to which they account for true parameter variability (Kennedy & Mandeville, 2000). The analyses are carried out with the statistical software program VARCL, which is able to handle two- and more-level databases (in this case school and student levels) adequately (Longford, 1993).

First of all, we estimate which part of the total variance is situated at the school level and which part at the student level. By inclusion of student input characteristics in the second model and school input characteristics in the third multilevel model (the so-called covariate models), we provide fair effectiveness scores (value added) of schools. The 'value added' is the difference between the actual achievement and the predicted achievement. The prediction of achievement is generally drawn from covariates such as intelligence and social class. After identification of these models, we will represent other educational models that are theoretically based. These theoretical models aim to estimate the degree to which types and characteristics of school self evaluation account for differences in math achievement.

Results of the multilevel analyses
The analyses have been conducted with the statistical programme MLWIN. We use one measure of cognitive achievement: math outcomes of pupils in grade 8 in elementary schools. However, if no substantial between/school variance is found then a further search for predictors of achievement differences at school level makes no sense.

Table 5 shows the results of the multilevel analyses. Only significant effects have been included in the Table which are standardized beta's.

Table 5 *Multilevel analyses for math (8)*

Model 0: Variance components	Math
School level (n=81)	13.1 %
Student level (n=2099)	86.9 %
Model 1: Covariates students	ß (s.e)
SES	**0.199** (0.022)
IQ	**0.355** (0.019)
SEXE	**0.202** (0.035)
Model 2: Covariates school	
More schools 1 board	**0.327** (0.158)
Model 3:	

SSE Types	geen-	
Type 2		
Type 3		
Type 4		
Model 4: Indicators quality assurance		
Learning organisation	**0.113** (0.053)	
External pressure		

outcomes of model 4 with all variables/scales; standardised beta's
Bold: only significant effects

The results presented in Table 5 model 0 shows the existence of 13% between/school variance in math. This outcome indicates that it makes sense to search for predictors at the school level that could account for differences in achievement levels between schools.

Next, we include the student level covariates in the model. The students' intelligence, sexe and socio-economic status of the family show substantial positive relationships for math performance in grade 8 of our elementary schools. These results are in line with outcomes of studies into effective schools and classrooms that try to explain the role of context factors in school effectiveness research (Teddlie, Stringfield & Reynolds, 2000).

The following step concerns the inclusion of school level covariates. This step is crucial in preventing a situation in which the student population of a school determines the effectiveness of these same schools. Willms (1992: 41) states the following: 'The composition of a school's intake can have a substantial effect on pupils' outcomes over and beyond the effects associated with pupils' individual ability and social class'. Table 5 shows that the only school level covariate that exerts an influence on math achievement, even after inclusion of the student level characteristics, is the number of schools that a school board governs. Using boards with one school as the base line we observe substantial higher achievement levels for pupils in schools of school boards that govern more than one school.

We come to the heart of our research. The two theoretical models concerning characteristics of school self evaluation have been included in the multilevel analysis. First the four types of school self evaluation (with type 1 hardly any SSE as the baseline) have been included in the analysis and subsequently the indicators of the school self evaluation process: perspectives, focus, actors, external support, instruments, etc. see Table 2'.

The results presented in Tabel 5 kan be summarized as follows:
▪ Type of School (Self) Evaluation does not seem to matter. Students in schools with or without advanced SSE show no differences in math achievement.

- Schools who function as a single school under one board perform les well than schools who fall under boards that govern also other schools. Grade 8 students in the latter schools perform better in math.
- Schools who we can typify as 'learning organizations' perform significantly better in math.

Conclusion and discussion

Pressure for improvement may be stimulated by two approaches. The quasi-market model (Harris & Herrington, 2006) observes the pressure to improve from a rational consumer viewpoint in which competition among schools is perceived as the central channel to quality improvement (Chubb & Moe, 1997). Within this approach strong parental school choice empower them to (power by feet) to influence their childrens' school quality especially when resourced with published performance data on the academic achievements of their chosen schools" (Reynolds, Muys, & Treharne, 2003, p. 84).

The contrasting model, government-based accountability also includes a sort of regulation by results (Harris & Herrington, 2006) but this one is employed in education systems (several states in USA, France, Portugal) in which educational authorities do not officially promote, free school choice. In these settings quality improvement is stimulated through (more) external evaluation mechanisms. School reports give details on the basis of external pupils' assessment or on the basis of an internal audit of the organisational or educational processes within the school. This information are returned to the school staff in order to incite them to improve their performance. Pressure and regulation of schools are thus organised by (national or local) government and in particular through formal engagements with the schools. In England and in some American states (Texas for example) this type of public pressure can even lead to strong interventions from the government, such as a takeover of the school or a reconstitution of the staff.

These approaches to accountability must be regarded upon as signs of the significance and chosen priorities of the governments policy. However, in both cases it is crucial to investigate to what extent the chosen accountability policies are based only upon academic achievement within schools or, whether information is also included about the school's student population and classroom observations of the teaching and learning processes, as observed in the Netherlands and some American States (Hofman et al, 2005; Ellet & Teddlie, 2003),

However, although many countries employ certain accountability policies policy-makers and researchers lack evidence on the real impact of this line of reforms. In our research we observed specific conditions or contexts that influence the development of school self evaluation practices. These practices are related to the described accountability systems and we detected some trends on what seems to contribute to high quality schools.

Types of school self evaluation

When we began our research, we had two major concerns. First, we wanted to explore whether it was possible to distinguish different types of school (self) evaluation in Dutch primary schools. The key dimensions, which make up what we refer to as school (self) evaluation, are accountability (the determination and position of the quality of a school), and school improvement. This links with the definition of the Dutch Inspectorate of Educatio's' approach to accountability. Second, if such types were to be found the question was how these relate to the quality assessments of the National Inspectorate of Education.

Four clusters or types of school (self) evaluation were developed, ranging from cluster 1, hardly any school (self) evaluation, cluster 2 with average school (self) evaluation, cluster 3 typified as advanced in school (self) evaluation, and cluster 4 as variable in school (self) evaluation.

Accountability and SSE types of schools
In general, it is obvious that schools that are hardly working on the development of school (self) evaluation system especially seem to lack a clear focus or vision on school (self) evaluation. Interestingly, these schools feel to be least encouraged by the Inspectorate of Education to improve their SSE-sytem and are also less influenced by external organizations or the community around their school. In this line of thinking, it is even more interesting that that schools with *variable SSE* (cluster 4) are indeed positively encouraged by the Inspectorate and more strongly influenced by external organizations and other schools support. This arises the question whether the cluster 4 group of schools (variable SSE) could have been encouraged in an earlier stage to work on their quality because of their lagging behind and showing insufficient school quality according to the Inspectorates assessment. It seems that the external focus on SSE could stimulate the schools that lag behind. On the other hand schools that have already accomplished a high level of SSE seem to possess certain internal characteristics that are of importance to SSE (a learning organization and high reliability approach). Moreover, it is helpful that of all schools in our (939) sample at least 83% of them prefer support of the Inspectorate to a certain extent (Hofman et al., 2004a).

In terms of school (self) evaluation and the quality of schools our research confirmed our *hypothesis*: schools that implemented hardly any school self evaluation measures (type 1) score significantly lower on mathematics. Furthermore, a significant relation has been found between the quality of the teaching-learning process and one of the types of school (self) evaluation. Schools with an advanced SSE score highest on the scale teaching and learning. The advanced schools show a significant higher quality of the teaching-learning process. This indicates that schools with an advanced school (self) evaluation system show a higher quality (according to the Inspectorate) regarding the curriculum, the use the available learning time, the pedagogical and didactic performances of teachers, the school climate, harmonization with the educational needs of pupils, an active and independent role of pupils and finally a higher quality of support and guidance for pupils, in comparison to the rest of the schools in our study.

SSE and schools as learning organisations
The multilevel analyses showed the relevance of the theory of schools as learning organizations. A definition of a 'learning organization' used by many authors is the one Leithwood & Aitken (1995) introduced: "*A group of people pursuing common purposes (and individual purposes as well) with a collective commitment to regularly weighing the value of those purposes, modifying them when that makes sense, and continuously developing more effective and efficient ways of accomplishing those purposes*". Those primary schools typified as 'learning organizations' optimize the talents of the staff so they can contribute maximally to the quality of the school. Further these are schools with (high) innovative capacities able and willing to respondoptimally to contextual changes.

These analyses once again seem to support the theory that teachers or more strongly the team of teachers are at the are central to the processes through which schools are stimulated to school improvement. Further research should keep in mind that attention has to be paid to the schools' interpretations of an accountability policy and to the processes in school and classroom constituting the implementation of such a policy. Although accountability policies are constructed at the central government level, they are in the end school specifically shaped by the different actors in the school. Teachers' construction of understandings and interpretations are influenced by social interaction with colleagues and by characteristics of the school environment. A context specific analysis of pressure to improve and the schools improvement efforts must not be forgotten (Coburn, 2005).

References

Arts, J., Kok, J., Verbiest, E., Sleegers, P. & de Wit, C. (Eds.) (2003). *Professionele ontwikkeling en schoolontwikkeling*. [Professional and school development.] Den Haag: Q*Primair.

Bosker, R.J. (Ed.) (2001). *Kwaliteitszorg Onderwijskundig Lexicon, Editie III*. [School (self) evaluation Educational Dictionary.] Alphen aan den Rijn: Kluwer.

Creemers, B.P.M. (1994). *The effective classroom.* London: Cassell.

Dalin, P. (1993). *Changing the school culture.* London: Cassell.

Deming, W.E. (1989). *Out of the crisis.* Cambridge, MA: MIT Press.

Earl, L. (2004). *Accountability is not paint by numbers*. Key-note paper at ICSEI, Rotterdam 2004.

Hofman, R.H. (1993). *Effectief schoolbestuur. De bijdrage van schoolbesturen aan de effectiviteit van basisscholen*. [Effective school governing boards. The contributions of school governing boards to the effectiveness of primary schools.] Groningen: RION/RUG.

Hofman, R.H. & Lugthart, E. (1997). *Identiteit, schoolkenmerken en schooleffectiviteit*. [Identity, school characteristics and school effectiveness.] Groningen: ION/RUG.

Hofman, R.H., Hofman, W.H.A., & Guldemond, H. (2000). *Effectieve sociale contexten van leren.* [Effective social contexts of learning.] Groningen/Rotterdam: GION/RUG, RISBO/EUR.

Hofman, R.H. & Hofman, W.H.A., (2003). *Ontwerp van een beoordelingskader voor zelfevaluatie-instrumenten voor scholen*. [Design of an assessment framework for school self-evaluation.] Groningen/Rotterdam: GION/RUG, RISBO/EUR.

Hofman, R.H., Hofman, W.H.A., Gray, J., & Daly, P. (2003). *Institutional contexts of education systems in Europe. A cross-country comparison of quality and equity.* Kluwer Academic Press.

Hofman, R.H., Dijkstra, N.J., Hofman, W.H.A., & de Boom, J. (2004a). *Deelstudie 1: Kwaliteitszorg in het primair onderwijs, Peiling 2003/2004*. [Study 1: School (self) evaluation in primary education, Survey 2003/2004]. Groningen/Rotterdam: GION/RUG, RISBO/EUR.

Hofman, R.H., Dijkstra, N.J. & de Boom, J. (2004b). *Deelstudie 2: Typologie-analyse en discrepantie analyse interne en externe kwaliteitszorg*. [Study 2: Typology analysis and discrepancy analysis internal and external school (self) evaluation] Groningen/Rotterdam: GION/RUG, RISBO/EUR.

Hofman, R.H., Dijkstra, N.J. & Hofman, W.H.A. (2005). School self-evaluation instruments: an assessment framework. Article in review for *International Journal for Leadership in Education*.

Horsman, K. (2002). Samen doordacht leren: Drie onmisbare bouwstenen voor goede kwaliteitszorg. *Meso magazine*, 123, 14-18.

National Inspectorate of Education (2001). *Onderwijsverslag over het jaar 2000*. [Education Report for 2001.] Den Haag: SDU.

National Inspectorate of Education (2003a). *Onderwijsverslag over het jaar 2002*. [Education Report for 2002.] Den Haag: SDU.

National Inspectorate of Education (2003a). *Veranderend toezicht. De inspectie van het onderwijs en de Wet op het onderwijstoezicht*. [Changing supervision. The inspectorate of education and the law on supervision of education.] Utrecht: Inspectie van het onderwijs.

National Inspectorate of Education (2003b). *Toezichtkader Primair Onderwijs. Inhoud en werkwijze van het inspectietoezicht conform de WOT*. [Supervision framework Primary Education: Content and method of the inspection supervision according to the Law on Educational Supervision.] Utrecht: Inspectie van het onderwijs.

National Inspectorate of Education (2003c). *Onderwijsverslag over het jaar 2002*. [Education Report for 2002.] Den Haag: SDU.

National Inspectorate of Education (2004). *Onderwijsverslag over het jaar 2003* (Onderwijsverslag 2002/2003) [Education Report for 2003.]. Utrecht: Inspectie van het onderwijs.

Leithwood, K., Edge, K. & Jantzi, D. (1999). *Educational Accountability: The State of the Art*. International Network for Innovative School Systems (INIS) Gütersloh: Bertelsmann Foundation Publishers.

Leithwood, K., Aitken, R., & Jantzi, D. (2001). *Making schools smarter: a system for monitoring school and district progress*. Thousand Oaks, CA: Corwin Press.

MacBeath, J., Meuretz, D., Schratz, M. & Jakobssen, L.B. (1999). *Evaluating quality in school education. A European pilot project. Final report*. European Commission Educating Training Youth.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, NJ: Prentice Hall.

Mintzberg, H. (1989). *Mintzberg on management*. Englewood Cliffs, NJ: Prentice Hall.

Newmann, F.M., King, M.B. & Rigdon, M. (1997). Accountability and School Performance: Implications from restructuring schools. *Harvard Educational Review*, 67, 41-74.

Scheerens, J. (1989). *Wat maakt scholen effectief?* [What makes schools effective?] 's-Gravenhage: SVO.

Stoll, L. & Wikeley, F. (1998). Issues on linking school effectiveness and school improvement. In W.T.

Hoeben (1998). (Ed.). *Effective School Improvement: State of the Art Contribution to a Discussion*. pp. 29-58. Groningen: GION/RUG.

Stringfield, S. & Slavin, R. (2001). *Title I: Compensatory education at the crossroads, socio-cultural, political, and historical studies in education* .Mahwah, NJ: Lawrence Erlbaum Association.

Stringfield, S., Reynolds, D. & Schaffer, E.C. (2001**). *The High Reliability Schools Project'*.** Paper presented at the ICSEI, Toronto, 2001.

Reezigt, G.J. (Ed.). (2001). *A framework for effective school improvement. Final report of the SI*. Leeuwarden: Gemeenschappelijk Centrum voor Onderwijsbegeleiding.

Reynolds, D. & Teddlie, C. (2000). The process of school effectiveness. In Reynolds, D. & Teddlie, C. (2000). *The International Handbook of School Effectiveness Research.* London/New York: Falmer Press.

Romesburg, H.C. (1984). *Cluster analysis for researchers.* Belmont CA: Lifetime Learning Publications.

Websites: http://www.minocw.nl; http://www.owinsp.nl; http://www.europa.eu.int; www.eurydice.nl;

Wastiau-Schlüter, P. (Ed.). (2004). *A close-up on the evaluation of schools.* Brussels: Eurydice.

Wilcox, B. & Gray, J. (1996). *Inspecting schools: holding schools to account and helping schools to Improve.* Buckingham Philadelphia: Open University Press.

Referenties

Bosker, R.J. (red) (2001). *Kwaliteitszorg. Onderwijskundig Lexicon. Editie III.* Aphen aan den Rijn: Kluwer.

Driessen, G., Langen, A. van, & Oudenhoven, D. (1994). *De toetsen voor het cohort Primair Onderwijs: Verantwoording.* Nijmegen: ITS.

Hofman, R.H. (1993). *Effectief schoolbestuur. De bijdrage van schoolbesturen aan de effectiviteit van basisscholen.* (Academisch proefschrift). Groningen: Rijksuniversiteit: Groningen/RION.

Hofman, W.H.A., Boom, de J., Hofman, R.H., & van den berg, M.A. (2005). *Variatie in en effectiviteit van de gemeentelijke regiefunctie in het onderwijsachterstandenbeleid. Eindrapportage.* Rotterdam/Groningen: RISBO/GION.

Hofman, R.H., Dijkstra, N.J., de Boom, J., Hofman, W.H.A. (2005). *Kwaliteitszorg in het primair onderwijs. Eindrapport.* GION/RUG, Groningen. 237 p.

Hofman,R.H., Hofman, W.H.A. & Guldemond, H. (2000). *Effectieve sociale contexten van leren.* Groningen/Rotterdam: GION/RISBO.

Hofman R.H, W.H.A. Hofman & H. Guldemond (2001). Social context effects on pupils' perception of school. *Learning and Instruction.* 11, Pp.171-194.

Hofman R.H., Hofman W.H.A. & H. Guldemond (2003). Effective families, peers and schools. A configurational approach. *Educational Research and Evaluation. Vol 9*, No 3. Pp. 213-237.

Hofman, R. Hofman, A, Gray, J. Daly, P. (2004). *Institutional contexts of education systems in Europe. A cross-country comparison of quality and equity.* Kluwer Academic Press.

Hofman, R.H. & W.H.A. Hofman (2003). *Ontwerp van een beoordelingskader voor zelfevaluatie-instrumenten voor scholen.* Groningen/Rotterdam: GION/RISBO.

Inspectie van het onderwijs (2004). *Onderwijsverslag over het jaar 2002/2003* (Den Haag: SDU).

Inspectie van het onderwijs (2005). *Onderwijsverslag over het jaar 2003/2004* (Den Haag: SDU).

Inspectie van het onderwijs (2006). *Onderwijsverslag over het jaar 2004/2005* (Den Haag: SDU).

Inspectie van het onderwijs (2003a). *De kern van de zaak. Werken aan kwaliteitszorg in het primair onderwijs.* Inspectie van het onderwijs, Utrecht.

Inspectie van het onderwijs (2003b). *Toezichtkader Primair Onderwijs. Inhoud en werkwijze van het inspectietoezicht conform de WOT.* Brochure. Inspectie van het onderwijs, Utrecht.

Inspectie van het onderwijs (2003c). *Veranderd toezicht.* De inspectie van het onderwijs en de Wet op het onderwijstoezicht. Inspectie van het onderwijs, Utrecht.

Inspectie van het onderwijs (2006). *Toezichtkader Primair Onderwijs 2005. Inhoud en werkwijze van het inspectietoezich.* Brochure ook op (www.owinsp.nl). Inspectie van het onderwijs, Utrecht.

Leithwood, K., K. Edge & D. Jantzi (1999). *Educational accountability: The state of the art*. International Network for Innovative School Systems (INIS). Gutersloh: Bertelsmann Foundation Publishers.

Leithwood, K. & L. Earl (2000). *Educational accountability effects: an international perspective*. *Peabody Journal of Education*, Vol. 75, number 4. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

MacBeath, J., Meuret, D., Schratz, M. & Jakobssen, L.B. (1999). *Evaluating quality in school education.* A European pilot project. Final report. European Commission (Educating Training Youth).

Ministerie van Onderwijs, Cultuur & Wetenschappen (1994). *Schevenings Beraad Bestuurlijke Vernieuwing. Gezamenlijke richtinggevende uitspraken.* Den Haag: SDU.

Newmann, Fred M. & King, M. Bruce (1997). *Accountability and school performance: Implications from restructuring schools.* In Harvard Educational Review, 67 (1) 41-74.

Van Petegum, P. (1997). *Scholen op zoek naar hun kwaliteit. Effectieve scholenonderzoek als inspiratiebron voor de zelfevaluatie van scholen*. Gent: Universiteit Gent.

Veen, I. van der, Meijden, A. van der & Ledoux, G. (2004). *School- en klaskenmerken Speciaal Basisonderwijs.* Basisrapportage PRIMA-cohort onderzoek. Vierde meting 2002-2003. Amsterdam: SCO-Kohnstamminstituut.

Websites: [www.owinsp.nl;](www.owinsp.nl) [www.qprimair.nl;](www.qprimair.nl)

Wilcox, B. & Gray, J. (1996). *Inspecting school: holding schools to account and helping schools to improve*. Open University Press, Buckingham, Philadelphia.

Willms (1992). Monitoring school performance: A guide for educators. London: The Falmer Press.

Appendix 1 *Psychometric characteristics of quality assurance indicators*

	Mean M	SD	Range	Nr. items	Cronbach's alpha
Vision quality assurance (QA)					
• QA as learning organization	3.33	.51	1-4	5	.70
• QA as high reliability organization	3.11	.43	1-4	9	.70
• QA by external pressuredruk	3.11	.70	1-4	2	.60
• QA system inspectorate accepted	3.69	.26	1-4	10	.79
• Support of QA inspectorate	3.61	.35	1-4	7	.81
Rol e/influence actors on quality policy					
• Influence interaction director/team	5.33	.82	1-6	2	.62
• Influence interaction external organisations/schools	2.50	.59	1-6	6	.60
• Vision/time director QA	2.81	.35	1-3	2	.66
• Vision/expertsie teachers QA	2.64	.48	1-3	4	.88
Characteristics QA system					
• Aplication of PDCA-cyclus	2.93	.40	1-4	14	.86
• Integral en systematic QA system	2.57	.29	1-3	12	.72
• Autonomy QA system	2.54	1.04	1-5	11	nvt