

Detailed results from the principal survey in the Netherlands

M.C.M. Ehren – University of Twente

This section describes the results of the first year of data collection in the Netherlands. Descriptive information on the main variables will be presented, as well as the results of correlations, regressions and t-tests (comparing inspected and non-inspected schools and schools in different inspection treatments).

The analyses were executed and described separately for primary and secondary education, as inspection models for both types of schools may be different and background characteristics of both types of schools (e.g. size, school organization) are different. Results are also described separately for teachers, principals and school boards to analyze potential differences in impact on these different groups of respondents.

Note: some correlations and regressions (particularly measuring relations between the improvement and outcome variables) are at this point relatively meaningless as these variables were measured at one point in time, while relations can only occur subsequently in time (e.g. improvement actions only lead to changes in capacity-building after a period of time). Correlations in secondary education are also relatively meaningless due to low response rates.

Descriptives

Table 1 provides a description of the background characteristics of schools and principals in our study. The results in the table show that approximately half of both primary and secondary schools have been inspected in the previous academic year. The majority of the principals who responded to our survey have at least seven years' experience as a principal. They spend most of their time on administrative tasks, on the quality assurance of the school, or on other tasks. Discussing education with teachers or observing lessons is only a minor part of their task. Most of the schools in our survey are located in a rural area.

Table 1. Description of participating schools/principal survey (means, sd in brackets)

	P.S	S.S.
Number of schools:	73	15
Percentage of schools inspected in the previous year	61%	58%
Percentage of principals with 0-2 years' experience	9%	8%
Percentage of principals with 3-6 years' experience	19%	39%
Percentage of principals with >7 years' experience	72%	54%
Percentage of time spend on administrative tasks	31% (15)	24% (14)
Percentage of time spend on teaching	15% (12)	23% (21)
Percentage of time spend on discussing education with teachers	17% (9)	15% (7)
Percentage of time spend on observing lessons	10% (5)	6% (3)
Percentage of time spend on managing student behavior	8% (3)	11% (8)
Percentage of time spend on quality assurance/self-evaluation	18% (10)	20% (13)
Percentage of time spend on other tasks	19% (11)	30% (16)
Percentage of schools in area with <3000 inhabitants:	30%	15%
Percentage of schools in area with 3000-15.000 inhabitants:	36%	23%
Percentage of schools in area with 15.001-50.000 inhabitants:	21%	8%
Percentage of schools in area with 50.001-100.000 inhabitants:	5%	31%
Percentage of schools in area with 100.001-500.000 inhabitants:	9%	15%

	P.S	S.S.
Percentage of schools in area with >500.000 inhabitants:	0%	8%
Percentage of schools in urban area	12%	62%
Percentage of schools in suburban area	21%	0%
Percentage of suburban schools in metropolitan area	6%	0%
Percentage of schools in rural area	61%	39%
Average number of students in the schools	180 (104)	1403 (976)
Average number of fulltime teachers in the schools	10 (6)	58 (22)
Percentage of schools with majority of students (>50%) from low income groups	13%	0%
Percentage of schools with majority of students (>50%) from high income groups	33%	8%
Percentage of schools where majority of students (>50%) do not speak national language as first language	14%	8%

P.S.: Primary schools

S.S.: Secondary schools

Table 2 provides an overview of means and standard deviations of all the variables in our study. The results show that principals and teachers generally report relatively high satisfaction with school inspections. Principals also score high on the extent to which school inspections promote self-evaluations. Teachers perceive more actions of stakeholders in response to school inspections than principals. Both principals and teachers report of high levels and changes in the school's capacity to improve, particularly in the transformational leadership of principals. Teachers and principals in primary education also indicate relatively many changes in the achievement orientation in the school. Teachers and principals perceive relatively little unintended consequences of school inspections on the school level. Teachers in primary education seem to teach more to the inspection rubric than to the (Cito)-test, whereas in secondary this seems to be reversed and teachers indicate to teach more to the test than to the inspection rubric.

School boards in primary education indicate to primarily govern data use, achievement orientation and quality assurance and professional development of (teachers and principals in) schools and not so much the curriculum and instruction in their schools. They also perceive to have changed their governance of these aspects of schools in response to the introduction of the new inspection model (including meetings with school boards). They also indicate to have increased the amount of information they request from schools in response to this new inspection model.

Table 2. Description of variables in conceptual framework

	<i>Survey to teachers</i>		<i>Survey to principals</i>		<i>Survey to school boards</i>
	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>
	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>
General satisfaction with school inspections (5-point scale disagree-agree)	3.84 (0.60)	3.43 (0.90)	3.92 (0.53)	4.07 (0.33)	NA
Intermediate processes: setting of expectations (5-point scale disagree-agree)	3.61 (0.43)	3.24 (0.53)	3.87 (0.45)	3.64 (0.32)	NA
Intermediate processes: acceptance and use of feedback (5-point scale disagree-agree)	3.52 (0.48)	3.37 (0.64)	3.59 (0.39)	3.61 (0.19)	
Intermediate processes: promoting self-	3.81	3.84	4.07	4.17	NA

	<i>Survey to teachers</i>		<i>Survey to principals</i>		<i>Survey to school boards</i>
	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>
	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>
evaluations (5-point scale disagree-agree)	(0.75)	(1.00)	(0.70)	(0.46)	
Intermediate processes: actions of stakeholders (5-point scale disagree-agree)	3.76 (0.51)	3.76 (0.64)	2.94 (0.34)	2.91 (0.39)	NA
Outcome: changes in capacity to improve (5-point scale much less-much more)	3.65 (0.47)	3.59 (0.48)	3.70 (0.51)	3.63 (0.34)	NA
Sub outcome: changes in participation in decision-making	3.65 (0.47)	3.53 (0.48)	3.64 (0.63)	3.58 (0.59)	
Sub outcome: changes in cooperation between teachers	3.59 (0.52)	3.55 (0.54)	3.70 (0.57)	3.58 (0.59)	
Sub outcome: changes in transformational leadership	3.71 (0.60)	3.70 (0.84)	3.71 (0.58)	3.71 (0.44)	
Outcome: changes in effective school and teaching conditions (5-point scale much less-much more)	3.58 (0.35)	NA	3.70 (0.46)	NA	NA
Sub outcome: changes in opportunity to learn and learning time	3.38 (0.33)	NA	3.41 (0.34)	NA	
Sub outcome: changes in achievement orientation	3.85 (0.33)	3.68 (0.50)	3.75 (0.52)	3.52 (0.22)	
Sub outcome: changes in clear and structured teaching	3.56 (0.46)	3.48 (0.48)	3.71 (0.72)	3.40 (0.40)	
Sub outcome: changes in safe and stimulating learning climate	3.38 (0.58)	3.33 (0.55)	3.48 (0.58)	3.33 (0.49)	
Outcome: changes in governing instructional time in schools due to school inspections (5-point scale disagree-agree)	NA		NA		2.80 (0.81)
Outcome: changes in governing data use and achievement orientation in schools due to school inspections (5-point scale disagree-agree)	NA		NA		3.49 (0.58)
Outcome: changes in governing curriculum and instruction in schools due to school inspections (5-point scale disagree-agree)	NA		NA		2.88(0.84)
Outcome: changes in governing quality assurance in schools due to school inspections (5-point scale disagree-agree)	NA		NA		3.72 (0.37)
Outcome: changes in governing professional development of principals and teachers in schools due to school inspections (5-point scale disagree-agree)	NA		NA		3.37 (0.63)
Outcome: changes in governing school climate in schools due to school inspections (5-point scale disagree-agree)	NA		NA		3.17 (0.68)
Outcome: changes in data collection of the schools' functioning due to school inspections	NA		NA		3.76 (0.43)

	<i>Survey to teachers</i>		<i>Survey to principals</i>		<i>Survey to school boards</i>
	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>
	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>	<i>Mean (SD)</i>
(V12 and V15: (5-point scale disagree-agree; V14: 4 point scale: to no extent- to a large extent)					
Outcome: status in capacity to improve (5-point scale disagree-agree)	4.20 (0.48)	3.81 (0.54)	4.25 (0.43)	4.07 (0.37)	NA
Outcome: status in effective school and teaching conditions (5-point scale disagree-agree)	4.11 (0.50)	3.64 (0.50)	4.00 (0.49)	3.45 (0.48)	NA
Outcome: status in governing instructional time in schools (5-point scale disagree-agree; 6 'other')	NA		NA		2.78 (1.11)
Outcome: status in governing data use and achievement orientation in schools (5-point scale disagree-agree; 6 'other')	NA		NA		3.22 (1.16)
Outcome: status in governing curriculum and instruction in schools (5-point scale disagree-agree; 6 'other')	NA		NA		2.46 (0.99)
Outcome: status in governing quality assurance in schools (5-point scale disagree-agree; 6 'other')	NA		NA		3.65 (0.95)
Outcome: status in governing professional development of principals and teachers in schools (5-point scale disagree-agree; 6 'other')	NA		NA		3.91 (0.88)
Unintended consequences on the school level (5-point scale disagree-agree)	2.86 (0.49)	2.54 (0.59)	2.61 (0.55)	2.22 (0.40)	NA
Unintended consequences on the teaching level: teaching to the test (D23: 5 point scale never-every day; D24/25/26/28/29/30/31: 4 point scale never-to a large extent)	2.18 (0.62)	3.43 (0.57)	NA		NA
Unintended consequences on the teaching level: teaching to inspection (5-point scale disagree-agree)	3.41 (0.40)	2.90 (0.49)	NA		NA

The responses of teachers to items about teaching to the test and teaching to inspection were analyzed in more detail to gain insight into the specific strategies they use to teach to the test or to the inspection rubric, and to analyze differences between primary and secondary teachers. The results in table 10 indicate that teachers in primary education particularly perceive changes in the curriculum and teaching strategies, in the structuring of their lessons and in their pupil care due to school inspections. They seem to incorporate some practice for the Cito-test in their lessons to familiarize students with specific item formats (particularly in the month before testing) and they make sure to include tested topics in their lesson plan; teachers in primary schools do not seem to prepare students extensively and specifically for the test throughout the year.

Teachers in secondary education (in the final examination grade) also seem to particularly prepare students for testing in the month before the test, but they also have students practice old exams, teach students general test-taking skills and explain questions from old exams during regular instruction throughout the year. They make sure tested topics are covered in their lesson plan and use items from previous exams in their classroom assessments. Their curriculum and teaching strategies are (compared to primary school teachers) not so much aligned to the inspection rubric.

In secondary education, the Inspectorate also evaluates the (efficiency) of the throughput of the school. We therefore included a number of items in the teacher survey to ask teachers about the potential manipulation of these indicators. Teachers report of little actions to game these indicators; low performing students are, according to these teachers, generally not referred to lower educational tracks and students performing on the boundary of pass or fail are also not retained in a lower grade to improve the school's overall score in the inspection rubric.

Table 3. Descriptives of specific items on teaching to inspections/teaching to the test (teacher survey)

Note: first item between brackets refers to primary education, second item to secondary education

	Teachers	
	P.S.	S.S.
	Mean (SD)	Mean (SD)
I discourage teachers to experiment with new teaching methods that do not fit the scoring rubric of the Inspectorate (Q46), 5 points scale ranging from strongly disagree to strongly agree	2.17 (0.85)	2.10 (0.79)
School inspections have resulted in narrowing curriculum and instructional strategies in my school (Q47), 5 points scale ranging from strongly disagree to strongly agree	2.85 (0.95)	2.54 (0.92)
School inspections have resulted in refocusing curriculum and teaching and learning strategies in my school (Q48), 5 points scale ranging from strongly disagree to strongly agree	3.66 (0.82)	2.89 (1.00)
My lessons (e.g. activities in lessons, grouping of students, build-up of lessons) are structured to meet the evaluation criteria of the Inspectorate (D59), 5 points scale ranging from strongly disagree to strongly agree	3.79 (0.64)	3.00 (0.78)
The teaching strategies and methodologies I use (how I explain concepts to students) are structured to meet the evaluation criteria of the Inspectorate (D60), 5 points scale ranging from strongly disagree to strongly agree	3.83 (0.55)	2.85 (0.85)
How I provide care to students in need is structured to meet the evaluation criteria of the Inspectorate (bijvoorbeeld opstellen van handelingsplan/ bepalen ontwikkelingsperspectief) (D61-extra), 5 points scale ranging from strongly disagree to strongly agree	3.95 (0.51)	3.17 (0.83)
I feel under pressure to align my teaching to the evaluation criteria of the Inspectorate (D63), 5 points scale ranging from strongly disagree to strongly agree	3.31 (0.92)	3.25 (1.15)
Having students practice old tests/exams (D23a_rec en DS22a), 5 points scale ranging from never to almost every day	1.63 (0.91)	3.44 (1.24)
Teaching test-taking skills, such as pacing/timing, strategies for answering multiple-choice questions, eliminating wrong m.c. options, filling in answer sheets etc. (D23b_rec, DS22b), 5 points scale ranging from never to almost every day; SS 6 points scale	2.67 (1.16)	3.49 (1.06)
Providing practice on questions from the test/exam that was administered last year (D23c_rec), 5 points scale ranging from never to almost every day	1.56 (0.88)	-
Explaining questions from the test/exam that was administered last year (D23d_rec, DS22c), 5 points scale ranging from never to almost every day	1.83 (0.99)	3.54 (1.25)

Discussing responses to practice test/exam items (D23e_rec, DS22d), 5 points scale ranging from never to almost every day	1.68 (0.95)	3.56 (1.30)
I reviewed what was on last year's test/exam when designing my lesson plan/pacing calendar for this year (D24_rec, D24), 4 points scale	2.38 (0.97)	3.25 (0.73)
I emphasize particular styles and formats of problems in the test/exam in instruction (e.g. using particular styles of graphs; using specific key phrases) (D25_rec, D25), 4 points scale	2.50 (0.90)	3.52 (0.62)
I use m.c. questions from previous test/exam in my classroom assessments (D26_rec), 4 points scale	1.34 (0.66)	3.15 (0.97)
I rely on open-ended tests (essays, portfolios) in my own classroom assessment (D28_rec, D28), 4 points scale	2.62 (0.96)	2.72 (0.85)
I teach topics that are not on the test/exam after test administration (D29_rec;Q0024), 4 points scale	1.68 (0.85)	-
Within 1 month of testing, I use practice exercises/tests with multiple-choice questions and language similar to that found on the test/exam (D30_rec), 4 points scale	2.12 (0.98)	-
Within 1 month of testing, I provide a "refresher" on content and/or skill areas that specifically match those on the test/exam (D31_rec, D31), 4 points scale	2.24 (0.96)	3.85 (0.36)
Low performing students (not students in the categories LWOO or <3 years in the Netherlands) do not take the CITO-test (D31_extra_rec), 4 points scale	2.10 (1.04)	-
Low performing students who potentially lower our throughput indicators are referred to a lower educational track (DS30a_extra), 4 points scale	-	1.95 (0.86)
Students who potentially lower our average examination grades are referred to a lower educational track (DS30b_extra), 4 points scale	-	1.51 (0.65)
Students who are on the boundary of pass/fail are advised to repeat the grade before testing to make sure our average examination grades are not affected (DS30c_extra), 4 points scale	-	1.51 (0.69)

Correlations between variables

Correlations between the variables in the conceptual framework were calculated as a first test of the assumed relations between inspections, intermediate mechanisms and our outcome variables.

The results in table 4 show that teachers and principals who are overall satisfied with school inspections generally use inspection standards and feedback to a somewhat larger extent in their daily work and to improve the school's functioning. Principals and teachers (particularly in primary education) who indicate that inspection standards guide their daily work and who accept inspection feedback also indicate that school inspections promote self-evaluations and improvement actions in the school. Teachers who feel that inspections standards guide the daily work of the school perceive stakeholders to more frequently use the inspection standards and feedback in their actions towards the school. Teachers and principals (in primary education) who indicate that school inspections promote the implementation or improvement of self-evaluations in the school rate higher levels of innovation capacity in the school.

School boards that had an inspection meeting since 2007 indicate to have made some changes in their governance of quality assurance in their schools and in the amount of data they collect on the functioning of their schools. School boards that had an inspection meeting during the previous academic year score somewhat higher on the extent to which they govern professional development of teachers and principals in their schools. They also perceive to have implemented more changes in their governance of the data use, achievement orientation and professional development of teachers and principals in their schools, and in their data collection of their schools' functioning.

Table 4. Correlations between variables in conceptual framework

	<i>Teacher</i>		<i>Principal</i>		<i>School board</i>
	<i>P.S.</i>	<i>S.S</i>	<i>P.S.</i>	<i>S.S</i>	<i>P.S.</i>
<i>Relations school inspections and intermediate mechanisms</i>					
Satisfaction with school inspections - Setting expectations	0.08	0.27	0.43**	0.31	NA
Satisfaction with school inspections - Accepting feedback	0.77**	0.85*	0.77**	0.34	NA
Inspection school board meeting since 2007 – accepting feedback					-0.059
Satisfaction with school inspections - Promoting self-evaluations	0.13	-0.10	0.34*	-0.15	NA
Satisfaction with school inspections - Improvement actions	-	-	0.52**	NA	NA
<i>Relations between intermediate mechanisms</i>					
Setting expectations - Promoting self-evaluations	0.27*	-0.35	0.67**	0.07	NA
Setting expectations - Taking improvement actions	-	-	0.67**	NA	NA
Accepting feedback -Promoting self-evaluations	0.31*	-0.08	0.27	0.70	NA
Accepting feedback - Taking improvement actions	-	-	0.37*	NA	NA
Setting expectations - Actions of stakeholders	0.38**	-0.27	0.10	-0.21	NA
Accepting feedback - Actions of stakeholders	-0.02	-0.12	0.22	0.54	NA
Actions of stakeholders - Promoting self-evaluations	0.24*	0.17	0.24	-0.16	NA
Actions of stakeholders - Taking improvement actions	-	-	0.24	NA	NA
<i>Relations between intermediate mechanisms and outcome variables</i>					
Promoting self-evaluations - High improvement capacity	0.21*	0.33**	0.02	-0.16	NA
Promoting self-evaluations - Effective school and teaching conditions	0.31**	0.20	0.19	0.04	NA
Taking improvement actions - High improvement capacity	-	-	0.13	NA	NA
Taking improvement actions - Effective school and teaching conditions	-	-	0.21	NA	NA
<i>Relations between school inspections and outcome variables</i>					
Satisfaction with school inspections – high improvement capacity	0.11	-0.01	0.25	-0.06	NA
Satisfaction with school inspections – high effective conditions	0.27*	0.06	0.11	0.04	NA
Inspection school board meeting since 2007 – governing instructional time in schools	NA	NA	NA	NA	-,111
Inspection school board meeting since 2007 – governing data use and achievement orientation in schools	NA	NA	NA	NA	-,027
Inspection school board meeting since 2007 – governing curriculum and instruction in schools	NA	NA	NA	NA	-,133*
Inspection school board meeting since 2007 – governing quality assurance in schools	NA	NA	NA	NA	,002

	<i>Teacher</i>		<i>Principal</i>		<i>School board</i>
	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>
Inspection school board meeting since 2007 – governing professional development of teachers and principals	NA	NA	NA	NA	,132*
Inspection school board meeting since 2007 – changes in governing instructional time due to inspections	NA	NA	NA	NA	. ^a
Inspection school board meeting since 2007 – changes in governing data use and achievement orientation due to inspections	NA	NA	NA	NA	-,001
Inspection school board meeting since 2007 – changes in governing curriculum and instruction due to inspections	NA	NA	NA	NA	. ^a
Inspection school board meeting since 2007 – changes in governing quality assurance due to inspections	NA	NA	NA	NA	,315*
Inspection school board meeting since 2007 – changes in governing professional development of teachers and principals due to inspections	NA	NA	NA	NA	,048
Inspection school board meeting since 2007 - changes in governing school climate due to school inspections	NA	NA	NA	NA	,148
Inspection school board meeting since 2007 - changes in data collection of the schools' functioning due to school inspections	NA	NA	NA	NA	,255**
Inspection school board meeting 2010/2011 – governing instructional time in schools	NA	NA	NA	NA	-,025
Inspection school board meeting 2010/2011– governing data use and achievement orientation in schools	NA	NA	NA	NA	,019
Inspection school board meeting 2010/2011– governing curriculum and instruction in schools	NA	NA	NA	NA	-,108
Inspection school board meeting 2010/2011– governing quality assurance in schools	NA	NA	NA	NA	,044
Inspection school board meeting 2010/2011– governing professional development of teachers and principals	NA	NA	NA	NA	,241**
Inspection school board meeting 2010/2011– changes in governing instructional time due to inspections	NA	NA	NA	NA	,356
Inspection school board meeting 2010/2011– changes in governing data use and achievement orientation due to inspections	NA	NA	NA	NA	,479**
Inspection school board meeting 2010/2011 – changes in governing curriculum and instruction due to inspections	NA	NA	NA	NA	,180

	<i>Teacher</i>		<i>Principal</i>		<i>School board</i>
	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>
Inspection school board meeting 2010/2011 – changes in governing quality assurance due to inspections	NA	NA	NA	NA	,198
Inspection school board meeting 2010/2011– changes in governing professional development of teachers and principals due to inspections	NA	NA	NA	NA	,336**
Inspection school board meeting 2010/2011 - changes in governing school climate due to school inspections	NA	NA	NA	NA	,197
Inspection school board meeting 2010/2011 - changes in data collection of the schools' functioning due to school inspections	NA	NA	NA	NA	,272**

* $p < .05$, ** $p < .01$

Regression analyses

Regression analyses were executed to calculate the extent to which the predictor variables in our framework (occurrence of an inspection visit, general satisfaction with school inspections, setting expectations, actions of stakeholders, acceptance and use of feedback, promoting self-evaluations and improvement actions) explain variation in each of our outcome variables (capacity-building, school effectiveness). Unintended consequences were not included in the regression analyses due to low reliability of the scale. The response rates in secondary education in the Netherlands were also too low to execute a regression analysis. Overall response rates were also relatively small due to the fact that some of the variables are only measured in schools that were inspected in the previous year. The results in table 5 and 6 do not point to variables that convincingly explain levels of improvement capacity and school effectiveness in the schools in our study.

Table 5. Effects of intermediate mechanisms on improvement capacity of schools (regression analysis)

<i>Independent variable</i>	<i>Teachers</i>		<i>Principals</i>	
	<i>P.S.</i>	<i>S.S.</i>	<i>P.S.</i>	<i>S.S.</i>
Satisfaction with school inspections	-0.17 (0.23)	-1.15 (0.58)	0.19 (0.25)	-
Setting expectations	0.08 (0.20)	-0.11 (0.53)	0.28 (0.26)	-
Actions of stakeholders	-0.14 (0.19)	0.14 (0.28)	-0.14 (0.29)	-
Accepting feedback	0.26 (0.33)	2.38 (1.41)	-0.08 (0.31)	-
Promoting self-evaluations	0.09 (0.13)	-0.24 (0.40)	-0.04 (0.13)	-
Improvement actions	-	-	-0.10 (0.26)	-
Constant	3.87** (1.05)	0.59 (2.41)	3.64** (1.04)	-
R square	0.07	0.60	0.11	-
Sample size	34	9	34	-

Note: dependent variable: capacity-building; B values are reported with standard error in parentheses, * $p < .05$, ** $p < .01$

Note: B values tell the relationship between predictors (independent variable) and dependent variable. If the predictor increases by one unit, the dependent variable increases by B units (depending on the units used to measure the variables). The standard error indicates to what extent these values would vary across different samples and whether B differs significantly from zero.

Table 6. Effects of inspection and intermediate mechanisms on school effective conditions in schools

<i>Independent variable</i>	Teachers		Principals	
	P.S.	S.S	P.S.	S.S
Satisfaction with school inspections	0.22 (0.28)	-0.24 (0.27)	-0.24 (0.26)	-
Setting expectations	-0.08 (0.25)	-0.17 (0.25)	0.46 (0.31)	-
Actions of stakeholders	-0.04 (0.22)	0.16 (0.13)	-0.15 (0.28)	-
Promoting self-evaluations	0.11 (0.38)	-0.43 (0.19)	0.09 (0.16)	-
Accepting feedback	-0.07 (0.15)	0.72 (0.66)	0.13 (0.40)	-
Improvement actions	-	-	-0.18 (0.31)	-
Constant	3.69** (1.23)	3.69 (1.13)	3.52* (1.25)	-
R square	0.11	0.72	0.16	-
Sample size	33	9	35	-

Note: dependent variable: school effectiveness; B values are reported with standard error in parentheses, * $p < .05$, ** $p < .01$

Comparing schools (t-test)

A t-test was used to compare improvement capacity, school effectiveness and the time spent on changes in schools that received an inspection visit the previous year to schools that did not receive an inspection visit. The results in table 17 indicate small changes in the extent to which inspected schools in primary education implement more changes in their capacity to improve, compared to non-inspected schools. These changes were however not perceived by teachers. Teachers in inspected schools only perceive (a small amount of) more changes in the implementation of self-evaluations.

Table 17. Comparing inspected and not inspected schools

	Teachers		Principals	
	P.S.	S.S	P.S.	S.S
Capacity-building	-0.91 (118)	0.31 (77)	0.62 (64)	0.02 (10)
School effectiveness	1.72 (118)	-1.93 (77)	1.56 (63)	-0.73 (9)
Change in capacity building	1.68 (117)	-0.30 (74)	2.03* (64) Mean dif. 0.26	-0.36 (9)

	Teachers		Principals	
	P.S.	S.S	P.S.	S.S
Change in participation in decision making	1.26 (117)	-0.98 (76)	1.73 (64)	-0.31 (8)
Change in cooperation between teachers	0.93 (120)	0.35 (76)	1.89 (64)	-0.34 (7)
Change in transformational leadership	1.90 (117)	0.08 (77)	1.64 (64)	0.27 (10)
Change in school effectiveness	1.39 (101)	-	1.71 (61)	0.30 (8)
Changes in opportunity to learn	0.12 (103)	-	1.06 (61)	-0.10 (8)
Changes in achievement orientation	1.44 (117)	0.91 (76)	2.04* (64) Mean dif. 0.26	0.34 (10)
Changes in clear and structured teaching	1.59 (119)	0.97 (77)	1.40 (59)	0.59 (9)
Changes in safe and stimulating learning climate	-1.67 (120)	0.40 (78)	0.79 (55)	-0.78 (10)
Promoting self-evaluation	2.55* (118) Mean difference: 0.35	0.09 (77)	-0.58 (64)	0.20 (8)

Note: t-value, df between brackets, * $p < .05$, ** $p < .01$, mean differences reported for significant differences

Note: some variables (e.g. accepting feedback) not administered in non-inspected schools, therefore not included in the table.

Note: Levene's test shows if the variances in the two groups are equal. If this test is significant then we can conclude that the variances are significantly different.

In addition, we also compared schools that have been assigned to different inspection treatments. The table below shows differences between the groups of primary schools in the three inspection arrangements and also shows if these differences increase when schools are assigned to more intensive inspection arrangements. The results in the table indicate that both teachers and principals in primary schools in more intensive inspection arrangements report of a higher use of inspection standards to set expectations, more actions of stakeholders in response to the inspection report, more actions to improve the school's self-evaluation and more changes in the school's capacity to improve.

Table 7. Comparing primary schools in different inspection arrangements (planned comparisons/ANOVA)

	Teachers		Principals ¹	
	Significant differences between all 3 groups (was	Linear trend (was there a significant linear trend, indicating that as	Significant differences between all 3 groups (was	Linear trend (was there a significant linear trend, indicating that as

¹ Chscheff (changes in school effectiveness), chachorient (changes in achievement orientation), stakeact (actions of stakeholders), gsatinspect (general satisfaction with inspections) have significant different variances between the groups. F-value is incorrect.

	there a significant effect of inspection treatment on 'dependent variable'?)	inspection treatment increases, the dependent variable increases proportionately)	there a significant effect of inspection treatment on 'dependent variable'?)	inspection treatment increases, the dependent variable increases proportionately)
General Satisfaction with Inspection	F = 0.03 (2, 107)	F = 0.03 (1, 107)	F = 8.22 ** (2,37)	F = 5.92 (1,37)*
Setting Expectations	F = 6.96** (2,68)	F = 13.53** (1,68)	F = 4.63 (2,35)*	F = 9.19 (1,35)**
Accepting Feedback	F = 2.44 (2, 46)	F = 3.28 (1, 46)	F = 5.73 ** (2,36)	F = 4.66 (1,36)*
Actions of Stakeholders	F = 11.66** (2,93)	F = 22.70** (1,93)	F = 3.21 (2,36)	F = 6.30 (1,36)**
Promoting/ Improving Self-Evaluations	F = 3.76* (2, 129)	F = 5.85** (1,129)	F = 2.65 (2,63)	F = 4.36 (1,63)*
Improvement Actions	-	-	F = 6.72 ** (2,36)	F = 9.37 (1,36)**
Change in Capacity Building	F = 6.88** (2,128)	F = 9.53** (1,128)	F = 8.57 ** (2,63)	F = 17.15 (1,63)**
Change in Participation in Decision Making	F = 5.05** (2,131)	F = 2.97 (1,131)	F = 11.31** (2,63)	F = 21.29 ** (1, 63)
Change in Cooperation between teachers	F = 6.16** (2,131)	F = 7.56** (1,131)	F = 4.41* (2,63)	F = 8.46 (1,63)**
Change in Transformational Leadership	F = 3.84** (2,128)	F = 7.51** (1,128)	F = 5.74** (2,63)	F = 11.46 (1,63)**
Changes in School Effectiveness	F = 3.56* (2,103)	F = 5.12* (1,103)	F = 8.15** (2,60)	F = 15,68 (1,60)**
Changes in Opportunities to Learn and Learning Time	F = 0.04 (2,105)	F = 0.07 (1, 105)	F = 9.72** (2,60)	F = 18.89 (1,60)**
Changes in Achievement Orientation	F = 4.40** (2,119)	F = 6.95** (1,119)	F = 5.19** (2,63)	F = 9.83 (1,63)**
Changes in clear and structured teaching	F = 3.54** (2,121)	F = 1.92 (1, 121)	F = 5.1** (2,63)	F = 10.20** (1,63)
Changes in safe and stimulating learning climate	F = 2.00 (2, 122)	F = 0.07 (1, 122)	F = 0.56 (2,63)	F = 0.22 (1,63)
Capacity Building	F = 0.56 (2, 129)	F = 0.18 (1, 129)	F = 0.39 (2,63)	F = 0.48 (1,62)

School Effectiveness	F = 3.54* (2,120)	F = 4.54* (1,120)	F = 4.45* (2,64)	F = 1.34 (1,62)
Unintended responses school level	F = 2.33 (2, 120)	F = 3.44 (1, 120)	F = 1.83 (2,37)	F = 3.40 (1,37)
Teaching to the test	F = 1.74 (2,107)	F = 3.45 (1,107)	NA	NA
Teaching to inspection	F = 4.66* (2,72)	F = 7.14** (1,72)	NA	NA

Note: reported is (F(dfM, dfR) = F; * p < .05, ** p < .01

Additionally, we compared secondary schools who have been assigned to a basic or a weak inspection arrangement. The results only reflect responses of teachers, due to low response rates of principals, and also only reflect differences in two inspection treatments due to a lack of response of teachers in schools in a 'very weak inspection arrangement'. The results in table 8 indicate that teachers in the more intense inspection arrangement report of lower levels in the effectiveness of the school and more changes in the (transformational) leadership and climate of the school.

Table 8. Comparing secondary schools in two different inspection arrangements (basic/weak)

	Teachers	
	HAVO	VWO
Capbuilding: 'capacity-building'	2.22 (80)	-1.56 (80)
Sceffect: 'school effectiveness'	-2.61* (78)	-0.02 (78)
	Mean difference: -0.64	
Impact: 'Improvement actions'	-	-
Cbchange: 'Changes in capacity-building'	1.01 (6)	-1.16 (77)
Pdmchange: 'Changes in participation in decision-making'	0.85 (79)	0.17 (79)
Cbtchange: 'Changes in Cooperation between teachers'	0.17 (79)	-0.41 (79)
Ctlchange: 'Changes in Transformational leadership'	0.18 (80)	-2.11* (80)
Chscheff: 'Changes in school effectiveness'	-	-
Chlearn: 'Changes in opportunity to learn and learning time'	-	-
Chachorient: 'Changes in achievement orientation'	1.56 (77)	0.32 (77)
Chcst: 'Changes in clear and structured teaching'	0.75 (78)	-0.46 (78)
Chssl: 'Changes in safe and stimulating learning climate'	5.48** (76)	3.58** (64)
	Mean difference: 0.35	
Pise: 'Promoting/ improving self-evaluations'	0.36 (80)	-0.74 (80)
Setting expectations teachers	0.21 (26)	-0.24 (26)
accepting feedback teachers	-	-0.24 (8)
Actions of stakeholders teachers	-2.35 (47)	1.86 (47)
Unintended responses teachers	0.99 (69)	-0.25 (76)
General satisfaction with inspection teachers	-1.45 (17)	0.86 (17)
Teaching to the test	1.13 (43)	0.80 (43)
Teaching to inspection	-0.65 (37)	-1.56 (37)

Note: t-value, df between brackets, * p < .05, ** p < .01, mean differences reported for significant differences

Note: teaching to inspection not measured in non-inspected schools

Comparing school boards

We also compared school boards who had an inspection meeting since 2007 to school boards who have not had such a meeting yet. The results in table 9 indicate that school boards that had an inspection meeting since 2007 score higher in the extent to which they govern quality assurance in, and collect data on the functioning of their schools; they however score lower in their governance of the curriculum and instruction in their schools. The third column also shows a comparison of school boards who had an inspection meeting in 2010-2011 to school boards who didn't have such a meeting in that year. School boards that had an inspection meeting in 2010-2011 score higher in their governance of the professional development of principals and teachers in their schools; they also indicate to have implemented more changes in their governance of data use, achievement orientation and professional development in their schools, and in their collection of information on their schools.

Table 9. Comparing inspected to not inspected school boards

	School boards that had/did not have inspection meetings since 2007	School boards that had/did not have an inspection meeting in 2010-2011
Governing educational effectiveness of schools	0.81 (232)	-0.50 (232)
Changes in governing educational effectiveness of schools due to school inspections	0.10 (145)	NA
Changes in governance of schools due to inspection meetings	NA	-1.13 (1)
Acceptance and use of inspection feedback	0.74 (157)	NA
Status in governing instructional time in schools	1.72 (239)	0.38 (239)
Status in governing data use and achievement orientation in schools	0.41 (238)	-0.30 (238)
Status in governing curriculum and instruction in schools	2.06* (235) Mean difference: 0.39	1.67 (235)
Status in governing quality assurance in schools	-0.03 (239)	-0.68 (239)
Status in governing professional development of principals and teachers in schools	-1.72 (37)	-3.83** (238) Mean difference: -0.45
Changes in governing instructional time in schools due to school inspections	NA	-0.54 (1)
Changes in governing data use and achievement orientation in schools due to school inspections	0.01 (53)	-3.04** (16) Mean difference: -0.63
Changes in governing curriculum and instruction in schools due to school inspections	NA	-0.52 (8)
Changes in governing quality assurance in schools due to school inspections	-2.30* (48) Mean difference: -0.83	-1.00 (13)
Changes in governing professional development of principals and teachers in schools due to school inspections	-0.40 (69)	-2.97** (69) Mean difference: -0.50
Changes in governing school climate in schools due to school inspections	-0.90 (36)	-0.90 (11)
Changes in data collection of the schools' functioning due to school inspections	-2.80** (113) Mean difference: -0.45	-2.35** (31) Mean difference: -0.28

Note: t-value, df between brackets, * $p < .05$, ** $p < .01$, mean differences reported for significant differences

Comparing teachers in testing/non-testing grades

Finally we compared responses of primary school teachers in the (high stakes) testing grade 8 versus primary school teachers in the (low stakes) grades 3 and 5. In secondary education we compared teachers in the non-testing year 2 to teachers in the final examination grades. Teachers in the (high stakes) testing grade 8 in primary education indicate more changes in learning time, clear and structured teaching and higher levels of teaching to the test compared to their colleagues in the (low stakes) grades 3 and 5. Teachers in the final examination grades in secondary education perceive less actions of stakeholders in response to the inspection report compared to the teachers in the lower grades.

Table 10. Comparing teachers in (high stakes) testing and (low stakes/non) testing grades

	Primary school teachers in grades 3 and 5 compared to teachers in grade 8	Secondary school teachers in (low stakes/non-testing) year 2 compared to teachers in (high stakes) final examination grade
General Satisfaction with Inspection	-1.68 (54)	-0.36 (17)
Setting Expectations	-1.37 (69)	-0.52 (26)
Accepting Feedback	-0.31 (32)	1.11 (8)
Actions of Stakeholders	-1.80 (94)	-2.16* (47) Mean difference: -0.39
Promoting/ Improving Self-Evaluations	-1.80 (130)	-0.56 (70)
Improvement Actions	-	-
Change in Capacity Building	-0.34 (129)	-1.08 (63)
Change in Participation in Decision Making	0.11 (97)	-0.70 (77)
Change in Cooperation between teachers	-0.68 (132)	-1.36 (77)
Change in Transformational Leadership	-0.22 (129)	-1.06 (65)
Changes in School Effectiveness	-1.56 (104)	-0.52 (53)
Changes in Opportunities to Learn and Learning Time	-2.53** (106) Mean difference: -0.16	1.02 (76)
Changes in Achievement Orientation	-0.85 (120)	-1.49 (54)
Changes in clear and structured teaching	-2.47** (122) Mean difference: -0.21	0.25 (78)
Changes in safe and stimulating learning climate	-0.14 (86)	-1.81 (79)
Capacity Building	-1.12 (130)	0.72 (78)
School Effectiveness	-0.04 (121)	0.49 (78)
Unintended responses school level	-1.65 (84)	0.82 (76)
Teaching to the test	-5.34** (108) Mean difference: -0.62	NA
Teaching to inspection	-1.21 (73)	1.11 (37)

Note: items on teaching to the test not administered to teachers in lower grades in secondary education