Impact of school inspections on teaching and learning in primary and secondary education in Sweden
Technical report ISI-TL project year 1-3 data

Jan-Eric Gustafsson
Department of Education and Special Education, University of Gothenburg

1 Introduction

Sweden has since 2008 reestablished school inspections as a major instrument for controlling and promoting the quality of schools. As school inspections are increasingly used for the aim to improve schools, it is of great importance to gain more knowledge about which effects inspections have on school-level processes which may be expected to lead to improvement of student performance.

The study presented here is part of an EU-funded project that has the aim to expand this knowledge base by studying responses to survey questions about school inspections of principals in primary and secondary education in six European countries (England, Ireland, Sweden, Austria/Styria and the Czech Republic) during three consecutive years. This report describes the results of the three years of data collection and analyses in Sweden. The following research questions will be addressed:

1. What are effects and negative consequences of school inspections in terms of changes in school effectiveness conditions and innovation capacity of primary and secondary schools in Sweden?
2. What aspects of school inspections contribute to these changes?

Before presenting the conceptual framework of the project, there is, however, reason to present some contextual information about the development of school inspections in Sweden, and how the current inspection system is implemented.

1.1 Development of school inspections in Sweden

The description below is based on Gustafsson, Lander and Myrberg (2014), which reference provides more detail and references. A state school inspection was installed in Sweden in 1858 to influence municipalities to take more responsibility for the compulsory primary education. Inspectors could veto state subsidies to municipalities that did not follow state regulations and inspectors also gave advice to municipalities and teachers. In the beginning of the 20th century central bureaucracies were implemented for secondary and primary education. A tight financial and administrative state rule-governing was successively put into place, which caused state inspection to diminish in importance.

However, with the introduction of the comprehensive school in 1962 school inspections increased in importance again. Twenty-five regional state boards were added to the National Board of Education (NBE) in 1958. They supported schools during the experimentation phase before the reform, and inspected its later implementation. One task was to prevent informal ability grouping. The inspectorate's sanctions for rule-breaking were, however, very rarely used.

Another big change came in 1991, when municipalities were made the primary authorities of schools, which also meant the abolition of state inspection, together with the NBE and its regional boards. Instead the National Agency of Education (NAE) was established to uphold information governing within the decentralised system In 1992 a voucher system for both municipal and independent schools was introduced, and the previously highly regulated Swedish school system was deregulated and made strongly market oriented. No private fees are allowed in this system, but both small and big companies have found the school market profitable. About 13 per cent of the students at the compulsory level, and about 25 per cent at the secondary level, are enrolled in independent schools.

The ambition of the NAE was to support reflection and deliberation through rich and nuanced evaluations at local levels, but this system failed. Following this failure, and also prompted by declining achievement, a gradual re-centralisation of governing has taken place. The establishment of the Swedish Schools Inspectorate (SSI) in 2008 and introduction of large-scale school inspections is one important element of this re-centralisation. Thus, one main task of the SSI it to handle quality problems associated with dysfunctions of the decentralisation and marketization policy.

1.2 School inspections in Sweden

The Swedish Educational Act legally regulates school inspection. The government has issued an instruction, and as for all governmental bodies, there are annual "Letters of Regulation". Official descriptions of the SSI explain procedures and expected effects. The government decides on objectives, guidelines and allocation of resources for the SSI. The SSI works with interpretations of the national curricula, and other steering-documents, and research-based characteristics of good education and successful schools.

The inspectorate has the right to inspect municipal and independent schools whenever it decides, but usually it follows a four to five year cycle for the regular inspections. The Education Act from 2011 gave the inspectorate the right to induce sanctions. Some authorities have been threatened with economic penalties, but very few cases go to the necessary court trial. Between 2009 and 2012 about 20 independent schools annually have had their permission (approved by the SSI) to operate schools withdrawn.

During 2010 a system of differentiated supervisions was introduced. Before a visit, results are studied, and a survey is completed by the individual schools and the responsible authority. A so called "basic supervision" is done in schools which are judged to be well-functioning. If this is uncertain a "widened supervision" is made. Clear problems call for a "deepened supervision". During 2011 and 2012 about 70 per cent of the supervisions were of the widened type.

When inspecting an authority within the regular supervision all its schools are visited. School leaders, teachers, and school nurses are interviewed, as are pupils and the board members in charge. Some lesson observations are done. Two inspectors participate in the school visit, which lasts one or two days, but occasionally longer. Oral feedback is given to the head teacher, together with a preliminary report for the school and its authority to react to. The formal report is published at the internet, and also presented to media with a press release.

1.3 Conceptual framework of the study

The theoretical framework used to investigate the research questions of the project is based on analyses of assumptions on how school inspections are expected to lead to improvement of schools in each of the participating countries (Ehren, Altrichter, McNamara, & O'Hara, 2013). In constructing the framework, policy and inspection documents were analyzed within each country and interviews were held with inspection officials to describe the mechanisms through which each Inspectorate aims to affect school improvement. The assumptions describe the causal mechanisms of how school inspections are supposed to lead to the improvement of schools, linking school inspections to their intended outcomes of improved teaching and learning in each country. From these assumptions Ehren et al. (2013) derived the

intermediate processes and mechanisms and intended outcomes that are common to all six countries. These common processes, mechanisms and intended outcomes were included in the theoretical framework.

The framework is presented in Figure 1. The left-most part includes elements describing how school inspections, and in particular their criteria and procedures and the feedback given during inspection visits, are expected to influence schools to align their views and expectations of good education and good schools to the standards in the inspection framework. Schools are expected to act on these views and expectations and use the inspection feedback when conducting self-evaluations and when taking improvement actions. Self-evaluations by schools are expected to build their capacity to improve, and take improvement actions that will lead to more effective teaching and learning conditions. These conditions are expected to result in improved student achievement (Ehren et al., 2013).
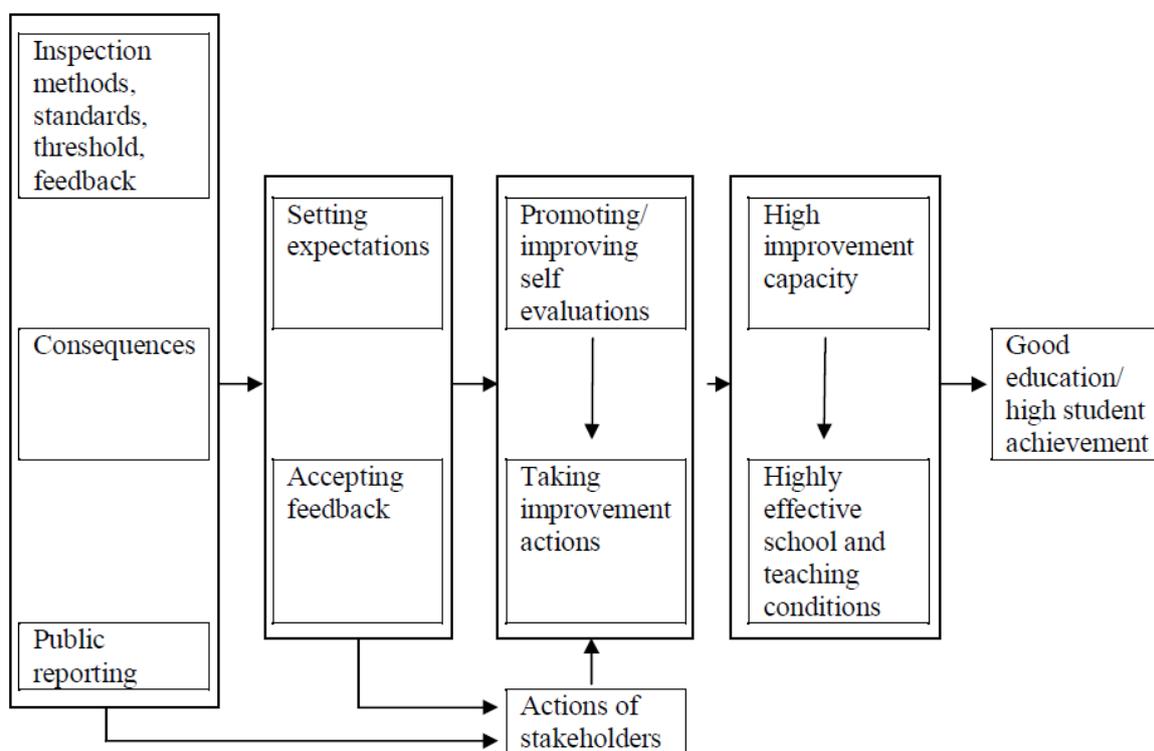

Figure 1. Conceptual model of intended effects of school inspections.

2 Research design

Below the basic idea of the research design is presented.

2.1 The longitudinal design

The research design used to investigate the research questions includes a three-year longitudinal design in which the variables in the model are measured using a survey to principals. The basic idea is to rely on a longitudinal design with switching replications as presented in Figure 2.

	Year 0	Year 1	Year 2	Year 3
Group of schools 1	X (inspection treatment)	O (data collection)	O (data collection)	O (data collection)
Group of schools 2		O (data collection) X (inspection treatment)	O (data collection)	O (data collection)
Group of schools 3		O (data collection)	O (data collection) X (inspection treatment)	O (data collection)
Group of schools 4		O (data collection)	O (data collection)	O (data collection) X (inspection treatment)

Figure 2. The basic principles of the research design

Given that the longitudinal design is such that the introduction of the treatment is done at a particular point in time, which differs between different groups, it is possible to take advantage of comparisons both over time for a given group of schools and across groups of schools. This design provides a strong basis for credible causal inference. However, because it was in Sweden not known at which point in time a particular set of schools was to be inspected, this design could not implemented in the economical way depicted in Figure 2. Instead a large random sample of schools was drawn for the first data collection which took place October-November 2011, and the schools have since been revisited twice for new data collections (October-November, 2012 and 2013). After the data collection was completed it has been possible to reconstruct when different schools were inspected, to emulate the design shown in Figure 2.

2.2 Sampling

Sweden does not have a clear distinction between primary and secondary schools, even though conventionally grades 1-6 (ages 7-13) are regarded as primary grades and grades 7-9 are regarded as (lower) secondary grades. In the decentralized Swedish school system, the municipalities are free to organize the schools as they like, and there is a wide range of combinations of grades in different schools. We define primary schools as schools which include grades in the 1-6 range, and secondary schools which include grades in the 7-9 range. Schools which include grades in both these ranges are treated as secondary schools, and the principal is asked to respond to the questionnaire with grades 7-9 in mind.

In the Swedish school system there are no national rules about the roles and responsibilities of principals of primary and secondary schools. However, there is a general rule which assigns responsibility to the principal for every aspect of the school, including budget, staffing, and instructional leadership. At larger schools the principal typically has one or more associate principals, which often have responsibility for a defined set of grades.

All types of inspection visits have been included in the data collection, and information from the principals, the Swedish Schools Inspectorate registers, and inspection reports was used to determine what type(s) of inspection took place in the different academic years

Simple random samples of primary and secondary schools were drawn. However, the 2013 survey was only sent to principals who responded in 2011 or 2012 or both 2011 and 2012 (N =1217, response rate = 62 %). Target populations and samples, along with obtained observations are reported in Table 1.

Table 1. Target populations and samples, along with obtained observations

	Target population	Target sample	Schools in year 1 data collection	Schools in year 2 data collection	Schools in year 3 data collection
Primary	3431	1167[1]	567	453	360
Secondary	1680	987[2]	464	388	318

Note 1: Out of the 1167 schools, 75 could not be reached, or were excluded because of recent change of principal
Note 2: Out of the 987 schools, 38 could not be reached, or were excluded because of recent change of principal

2.3 Data collection

Data collection included a survey to principals. The questionnaires included four sets of variables: background characteristics of schools, outcome variables, intermediate processes and inspection measures.

Background characteristics of schools
The questionnaire starts with a number of questions on background characteristics of schools and principals, such as the location of the school in a rural or urban area, the composition of the student population, the experience of the principal and his/her tasks, and the resources in the school. Questions also were included about the extent to which principals are responsible for administrative and pedagogical tasks.

Outcome variables
The second part of the questionnaire includes items intended to measure the outcome variables in the conceptual framework: capacity to improve and effective school and teaching conditions. Questions about these variables are framed in terms of the time principals have spent during the previous academic year to change the school's functioning in these areas (using a 5-point scale ranging from 'much less time' to 'much more time'), as well as the school's status and functioning on these variables (5-point scale ranging from 'strongly disagree' to 'strongly agree').Additionally questions were included about 'unintended consequences' of school inspections. These variables are described in more detail below.

Effective school and teaching conditions include conditions related to the school organization and management, such as educational leadership, a productive climate and culture and achievement-oriented school policy. These conditions are expected to contribute to and facilitate effective teaching and instruction and as a result lead to high student achievement. Teaching/instruction conditions include what a teacher does to create effective learning environments and to boost learning. The sub-variables for 'effective school and teaching conditions' included in the questionnaire are opportunity to learn and learning time,

achievement orientation, clear and structured teaching, and a safe and stimulating learning climate.

Capacity-building refers to the school's capacity to improve. A school with a high innovation capacity is one which is capable of implementing change, and is experienced at reflecting on its functioning and at changing and improving. Participation in decision-making, cooperation between teachers and transformational leadership are important factors in the school's capacity to improve and are therefore included as sub variables in the questionnaire.

Unintended consequences are potential side effects of school inspections. Potential unintended consequences include the extent to which school inspections lead to a narrowing of curricula and instructional processes in the school, the extent to which principals experience inspections as an administrative burden and manipulate documents and data they send to the Inspectorate.

Intermediate processes
The third part of the questionnaire includes questions about the intermediate processes that precede the outcome variables: setting of expectations, acceptance of feedback, promoting self-evaluations, and stakeholder sensitivity. Setting of expectations refers to the extent to which schools use the inspection standards to guide their work and define the school's goals and directions in working towards the inspection standards of a 'good school'. The acceptance of feedback is related to the assessment of the school and the suggestions for improvement provided during an inspection visit. Promoting self-evaluations refers to the effective implementation of internal systems of evaluation and self-review and the use of inspection standards and expectations of adequate self-evaluation to conduct self-evaluations and to implement self-evaluation systems. Stakeholder sensitivity includes parents using inspection findings to choose a school and stakeholders voicing necessary improvements to the school.

Inspection measures
The fourth part of the data collection includes information about inspection measures, such as the type of inspection visit in school, the methods of data collection used during inspection visits, the standards used to assess schools, the feedback provided to the school, the assessment of the school, sanctions and rewards to schools and how inspection findings are reported to stakeholders. The first year of data collection only included one item in the questionnaire on the occurrence of an inspection visit in the previous year, and data from the Inspectorate on the inspection arrangement to which schools were assigned.

3. Data analyses

The data has been subjected to a wide range of different analyses, most of which have been conducted with latent variable modelling techniques, such as confirmatory factor analysis (CFA) and structural equation modelling (SEM) (Muthén & Muthén, 1998 – 2012). However, scales have also been constructed from the items in the questionnaire to be used in analyses of observed variables, such as descriptive statistics. Information about the construction of scales is available at www.schoolinspections.eu.

The major focus of the current report is on estimation of causal effects using the longitudinal data, and the methods for doing that are described in Section 4, where also the results are reported. However, there is also a need to investigate the nature and quality of the data from the three waves of measurement, which is done through presenting cross-sectional descriptive results.

4 Cross-sectional results for years 1-3

Below descriptive results at the scale level are presented for all three years for Sweden.

4.1 Descriptives

The results presented below may be summarized as showing that the characteristics of the scales are generally quite satisfactory when it comes to central tendency and dispersion. While the internal consistency measures are acceptable to satisfactory for almost all scales, this is not the case for the Unintended consequences scale, which has alpha values around .40. The items intended to form this scale have therefore been analysed separately.

Year 1:

	Total			Secondary Schools			Primary Schools		
	M (SD)	N	Alpha (N)	M (SD)	N	Alpha	M (SD)	N	Alpha
Setting Expectations	3.55 (0.69)	357	.83	3.56 (0.69)	176	.83	3.54 (0.69)	181	.83
Stakeholder Sensitivity	3.56 (0.63)	351	.71	3.58 (0.60)	172	.71	3.54 (0.65)	179	.71
Accepting Feedback	3.93 (0.61)	359	.83	3.94 (0.61)	177	.83	3.93 (0.60)	182	.83
Capacity Building	4.13 (0.44)	1023	.73	4.09 (0.44)	460	.72	4.15 (0.43)	563	.74
Improvement in Capacity Building	3.73 (0.40)	1007	.77	3.72 (0.40)	457	.76	3.74 (0.40)	550	.79
School Effectiveness	3.86 (0.47)	1021	.66	3.82 (0.48)	461	.65	3.89 (0.46)	560	.66
Improvement School Effectiveness	3.53 (0.38)	1005	.81	3.52 (0.36)	456	.80	3.53 (0.39)	549	.82
Unintended consequences	2,73 (0.51)	356	.49	2.73 (0.53)	174	.55	2.73 (0.49)	182	.42

Year 2:

	Total			Secondary Schools			Primary Schools		
	M (SD)	N	Alpha (N)	M (SD)	N	Alpha	M (SD)	N	Alpha
Setting Expectations	3.77 (0.63)	822	.88	3.74 (0.65)	381	.88	3.81 (0.60)	441	.88
Stakeholder Sensitivity	3.56 (0.66)	821	.69	3.59 (0.67)	381	.70	3.52 (0.65)	440	.69
Accepting Feedback	3.87 (0.69)	813	.82	3.86 (0.73)	376	.83	3.88 (0.66)	437	.80
Capacity Building	4.17 (0.41)	840	.69	4.17 (0.41)	388	.67	4.17 (0.41)	452	.71
Improvement in Capacity Building	3.69 (0.42)	837	.80	3.68 (0.42)	388	.80	3.70 (0.42)	449	.80
School Effectiveness	3.89 (0.45)	838	.61	3.88 (0.44)	387	.58	3.90 (0.46)	451	.65
Improvement School Effectiveness	3.60 (0.40)	837	.81	3.56 (0.38)	388	.78	3.62 (0.41)	449	.83
Unintended consequences	2.86 (0.50)	822	.41	2.80 (0.53)	381	.46	2.91 (0.47)	441	.36
Inspection Measures	2.06 (0.64	773	.91	2.10 (0.64)	363	.90	2.03 (0.65)	410	.91

Year 3:

	Total			Secondary Schools			Primary Schools		
	M (SD)	N	Alpha (N)	M (SD)	N	Alpha	M (SD)	N	Alpha
Setting Expectations	3.67 (0.67)	663	.89	3.62 (0.63)	315	.86	3.72 (0.69)	348	.91
Stakeholder Sensitivity	3.49 (0.66)	661	.68	3.46 (0.66)	313	.67	3.52 (0.66)	348	.68
Accepting Feedback	3.81 (0.74)	652	.85	3.77 (0.75)	308	.85	3.84 (0.73)	344	.85
Capacity Building	4.19 (0.42)	677	.71	4.20 (0.41)	317	.69	4.18 (0.42)	360	.73
Improvement in Capacity Building	3.64 (0.42)	670	.83	3.63 (0.41)	312	.80	3.64 (0.44)	358	.84
School Effectiveness	3.94 (0.43)	677	.64	3.91 (0.43)	317	.63	3.97 (0.43)	360	.64
Improvement School Effectiveness	3.57 (0.39)	671	.80	3.56 (0.37)	314	.77	3.57 (0.40)	357	.84
Unintended consequences	2.80 (0.51)	664	.41	2.75 (0.51)	313	.45	2.75 (0.51)	351	.36
Inspection Measures	1.96 (0.68)	627	.93	2.04 (0.70)	295	.93	1.89 (0.65)	332	.92

4.2 Correlations

Year 1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Setting Expectations (1)	1						
Stakeholder Sensitivity (2)	.103	1					
Accepting Feedback (3)	.266	.236	1				
Capacity Building (4)	-.115	.281	.227	1			
Improvement in Capacity Building (5)	.225	.180	.198	.201	1		
School Effectiveness (6)	-.147	.184	.194	.532	.171	1	
Improvement School Effectiveness (7)	.189	.160	.173	.123	.591	.150	1

Year 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Setting Expectations (1)	1							
Stakeholder Sensitivity (2)	.228	1						
Accepting Feedback (3)	.454	.247	1					
Capacity Building (4)	.081	.254	.168	1				
Improvement in Capacity Building (5)	.226	.210	.158	.254	1			
School Effectiveness (6)	.085	.273	.200	.495	.207	1		
Improvement School Effectiveness (7)	.273	.160	.138	.227	.627	.181	1	
Inspection Measures (8)	.239	-.094	-.033	-.206	.130	-.306	.118	1

Year 3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Setting Expectations (1)	1							
Stakeholder Sensitivity (2)	.327	1						
Accepting Feedback (3)	.520	.361	1					
Capacity Building (4)	.034	.265	.174	1				
Improvement in Capacity Building (5)	.201	.178	.074	.194	1			
School Effectiveness (6)	-.008	.219	.130	.554	.120	1		
Improvement School Effectiveness (7)	.242	.153	.126	.140	.642	.143	1	
Inspection Measures (8)	.210	-.082	-.067	-.298	.141	-.334	.171	1

The correlations among the scales are generally positive, and almost all are highly significant. It is also interesting to note that corresponding correlations tend to be quite similar across years.

4.3 T-tests for differences between school types

The three tables below present results from comparisons of means of the scales across primary and secondary schools. There generally are few differences, the only exceptions being Capacity Building and School Effectiveness for Year 1, where primary schools tend to have somewhat higher means. These results suggest that there is little reason to conduct separate analyses of effects of school inspections for primary and secondary schools.

Year 1:

	Schooltype	M (SD)	t-value (df)	p-value
Setting Expectations	Sec	3.56 (0.69)	0.32 (355)	.75
	Prim	3.54 (0.69)		
Stakeholder Sensitivity	Sec	3.58 (0.60)	0.59 (349)	.56
	Prim	3.54 (0.65)		
Accepting Feedback	Sec	3.94 (0.61)	0.11 (357)	.91
	Prim	3.93 (0.60)		
Capacity Building	Sec	4.09 (0.44)	-2.14 (1021)	.03
	Prim	4.15 (0.43)		
Improvement in Capacity Building	Sec	3.72 (0.40)	-0.84 (1005)	.40
	Prim	3.74 (0.40)		
School Effectiveness	Sec	3.82 (0.48)	-2.26 (1019)	.02
	Prim	3.89 (0.46)		
Improvement School Effectiveness	Sec	3.52 (0.36)	-0.45 (1003)	.65
	Prim	3.53 (0.39)		

Year 2

	Schooltype	M (SD)	t-value (df)	p-value
Setting Expectations	Sec	3.74 (0.65)	-1.46 (820)	.14
	Prim	3.81 (0.60)		
Stakeholder Sensitivity	Sec	3.59 (0.67)	1.48 (819)	.14
	Prim	3.52 (0.65)		
Accepting Feedback	Sec	3.86 (0.73)	-0.51 (811)	.61
	Prim	3.88 (0.66)		
Capacity Building	Sec	4.17 (0.41)	-0.04 (838)	.97
	Prim	4.17 (0.41)		
Improvement in Capacity Building	Sec	3.68 (0.42)	-0.77 (835)	.45
	Prim	3.70 (0.42)		
School Effectiveness	Sec	3.88 (0.44)	-0.52 (836)	.61
	Prim	3.90 (0.46)		
Improvement School Effectiveness	Sec	3.58 (0.38)	-1.43 (835)	.15
	Prim	3.62 (0.41)		
Inspection Measures	Sec	2.10 (0.64)	1.44 (771)	.15
	Prim	2.03 (0.65)		

Year 3

	Schooltype	M (SD)	t-value (df)	p-value
Setting Expectations	Sec	3.62 (0.63)	-1.82 (661)	.07
	Prim	3.71 (0.69)		
Stakeholder Sensitivity	Sec	3.46 (0.66)	-1.70 (173)	.09
	Prim	3.52 (0.66)		
Accepting Feedback	Sec	3.77 (0.75)	-1.14 (659)	.26
	Prim	3.84 (0.73)		
Capacity Building	Sec	4.20 (0.41)	-1.10 (650)	.27
	Prim	4.18 (0.42)		
Improvement in Capacity Building	Sec	3.63 (0.41)	0.37 (675)	.71
	Prim	3.64 (0.44)		
School Effectiveness	Sec	3.91 (0.43)	-0.11 (668)	.91
	Prim	3.97 (0.43)		
Improvement School Effectiveness	Sec	3.56 (0.37)	-1.81 (675)	.07
	Prim	3.57 (0.40)		
Inspection Measures	Sec	2.04 (0.70)	-0.21 (669)	.83
	Prim	1.89 (0.65)		

4.4 T-Tests for differences between inspected and non-inspected schools

The three tables below investigate if there are differences in means of scales between inspected and non-inspected schools. For all three years, the Stakeholder Sensitivity scale shows a significant difference in favour of inspected schools, and the Accepting Feedback scale shows a significant difference in the same direction for Years 2 and 3. For some other variables there are significant differences or differences with borderline significance at some occasions. These results strongly suggest that there are effects of school inspections on at least some of the scales. Below these effects will be analysed with more powerful techniques than pair-wise t-tests.

Year 1:

	Inspected *	M (SD)	t-value (df)	p-value
Setting Expectations	Yes	3.51 (0.66)	1.05 (355)	.29
	No	3.59 (0.72)		
Stakeholder Sensitivity	Yes	3.65 (0.65)	-2.73 (349)	.01
	No	3.47 (0.59)		
Accepting Feedback	Yes	3.99 (0.64)	-1.73 (357)	.09
	No	3.87 (0.57)		
Capacity Building	Yes	4.15 (0.44)	-0.90 (1021)	.37
	No	4.12 (0.44)		
Improvement in Capacity Building	Yes	3.73 (0.39)	0.36 (1005)	.72
	No	3.72 (0.44)		
School Effectiveness	Yes	3.88 (0.47)	-0.74 (1019)	.46
	No	3.86 (0.47)		
Improvement School Effectiveness	Yes	3.57 (0.41)	-1.57 (1003)	.12
	No	3.52 0.37)		

* Inspected in the school year prior to answering the questionnaire.

Year 2:

	Inspected *	M (SD)	t-value (df)	p-value
Setting Expectations	Yes	3.78 (0.67)	-0.05 (820)	.96
	No	3.77 (0.65)		
Stakeholder Sensitivity	Yes	3.80 (0.62)	-4.91 (819)	.00
	No	3.50 (0.65)		
Accepting Feedback	Yes	3.97 (0.81)	-1.85 (811)	.07
	No	3.85 (0.66)		
Capacity Building	Yes	4.23 (0.39)	-1.82 (838)	.07
	No	4.16 (0.41)		
Improvement in Capacity Building	Yes	3.76 (0.42)	-2.03 (835)	.04
	No	3.68 (0.42)		
School Effectiveness	Yes	3.95 (0.43)	-1.64 (836)	.10
	No	3.88 (0.45)		
Improvement School Effectiveness	Yes	3.60 (0.43)	0.03 (835)	.97
	No	3.60 0.39)		

* Inspected in the school year prior to answering the questionnaire.

Year 3[1]:

	Inspected *	M (SD)	t-value (df)	p-value
Setting Expectations	Yes	3.67 (0.71)	-0.03 (661)	.98
	No	3.67 (0.65)		
Stakeholder Sensitivity	Yes	3.67 (0.58)	-3.84 (659)	.00
	No	3.44 (0.67)		
Accepting Feedback	Yes	3.97 (0.78)	-3.19 (650)	.00
	No	3.75 (0.72)		
Capacity Building	Yes	4.21 (0.41)	-0.84 (675)	.40
	No	4.18 (0.42)		
Improvement in Capacity Building	Yes	3.66 (0.43)	-0.70 (668)	.48
	No	3.63 (0.42)		
School Effectiveness	Yes	3.96 (0.42)	-0.62 (675)	.53
	No	3.94 (0.44)		
Improvement School Effectiveness	Yes	3.60 (0.39)	0.17 (669)	.24
	No	3.56 0.38)		

* Inspected in the school year prior to answering the questionnaire.

5 Causal effects of school inspections in Sweden

In this section analyses will be made of the Swedish longitudinal data with the aim to estimate causal effects of school inspections.

5.1 Method

Variables

Items were selected from the questionnaires described in section 2.3 to form scales and to be included in latent variable models. One set of questions asked about building teachers' capacity to develop the school, collaboration, and the instruction. There were five such items: "Teachers are involved in making decisions about educational matters such as teaching methods, curriculum and objectives", "Teachers collaborate in organizing and improving their teaching", "I use all possible opportunities to communicate the vision of the school to staff, parents and pupils"; "I support teachers in developing their career", and "I encourage teachers to improve their teaching practices". A Likert-type 5-point response scale ranging from Strongly disagree to Strongly agree was used, and the mean of the responses to these 5 items was computed to form a *Capacity Building* scale.

Another set of questions asked about to what extent the school adheres to principles of school effectiveness. There were five such items: "Students are provided with sufficient instruction time to reach their potential", "Teachers make good use of assessment results to inform their instruction", "I use assessment results to target areas for school improvement", "Teachers use clear, structured and challenging teaching approaches", and "The school overall has a safe and orderly social environment that is conducive to learning". Here too a 5-point response scale was used and the mean of the 5 items was computed to form a *School Effectiveness* scale.

Five questions on 'unintended consequences' asked whether school inspections lead to a narrowing of curricula and instructional processes, whether principals experience inspections as an administrative burden, and the extent to which school leaders manipulate documents and data they send to the Inspectorate. Principals could respond to these questions on a 5-point scale (strongly agree – strongly disagree). The five items did not form a scale with acceptable properties, so each item was analysed separately.

There also were questions aiming to get information about the sources of influence, if any, of school inspection activities. One was *Setting of expectations* (four items, e.g. "The inspection standards affect: (a) The evaluation and supervision of teachers, (b) The implementation of long term improvements, …"). Another one was *Accepting feedback* (four items, e.g., "The feedback provided to the school during the last inspection visit was insightful" and "The feedback received from the school inspectors was useful"). Yet another category of questions was whether the school's stakeholders are sensitive to the content of the inspection reports (*Stakeholders Sensitive to Reports*, three items, e.g., "The Parents' Representatives of the school are sensitive to the contents of the school inspection report"). These sets of items too used a 5-point response scale (Strongly agree – Strongly disagree).

Data

As has been described above (section 2.1) the sample was a random sample of the population of primary and secondary schools in Sweden in the autumn of 2011. The sample comprised 2154 schools, and 1217 principals responded to the questionnaire once, twice or three times. While it would have been more optimal to sample schools that were to be inspected one of the

three years during which the longitudinal study was conducted, this information was not available when the sample was drawn. Information was instead collected retrospectively about when the inspection was done.

The first round of measurement was conducted in the autumn (in most cases October) of the academic year 2011/2012, the second in the autumn of 2012/2013, and the third in the autumn of 2013/2014. We expect any effect of the inspection activities to show the year after the inspection was conducted, possibly also lasting the following couple of years. This implies that the 2011/2012 wave of measurement would reflect effects of inspections conducted in 2010/2011 (N = 228), the 2012/2013 wave effects of inspections conducted in 2011/2012 (N = 202), and the 2013/2014 wave effects of inspections conducted in 2012/2013 (N = 271; this number also includes some schools inspected early in the academic year 2013/2014). In addition, the sample included schools which had been inspected in 2009/2010 (N = 222).

For 294 schools no information was available about when the last school inspection took place. Data may be missing about when the inspection took place because the principal failed to respond to the question, but the questionnaire data may also be incorrect because the principal in the response included other forms of inspection than the regular one. For these reasons, dates for the regular inspections of the schools were also collected from the inspection protocols, which are made available on the internet by the SSI. Not even this information was complete, however, but combining the information from the protocols with that from the questionnaires improved the reliability.

All principals did not complete all three questionnaires, but an inclusion criterion was that the principal completed at least one questionnaire. Furthermore, in 2013 the questionnaire was only distributed to those principals who had responded at least once in 2011 and 2012.

Of the 1217 principals, 419 (34 %) responded at all three occasions, while 303 responded only once (224 in 2011 and 79 in 2012). 236 principals responded in 2011 and 2012, and 107 principals responded in 2012 and 2013.

If list-wise deletion was to be used, only 419 cases would remain for analysis. This would imply considerable loss of power, and unless the data is missing completely at random (MCAR) the cases included will not be representative of the population. For such reasons, list-wise deletion is generally considered to be the least appropriate method for dealing with missing data. An alternative method for handling missing-data is to use maximum likelihood (ML) missing-data modelling methods, as implemented for example in the Mplus program. This is generally considered to be one of the best methods for handling missing data, and it is based on the less restrictive assumption that the data is missing at random (MAR). The MAR assumption requires that the missingness is random given the information in the data. In a longitudinal design this assumption should generally be credible, given the substantial amount of correlation that is typically observed across waves of measurement. It may, however, be objected that this only applies to those cases for which information from at least one wave of measurement is available. Thus, the principals who elect to respond may differ from those who elect not to respond at all, with respect to for example, personality characteristics such as conscientiousness. With the data available in the current study, missing-data modelling based on MAR would not correct for such bias. It may thus be that the sample of principals included in the study is non-representative of the population in several different respects. However, this does not necessarily present a threat to the validity of the inferences drawn from the current study. The main aim of the study is to investigate causal effects of school inspections on beliefs and actions of the principals through use of a longitudinal design. Should there be such

effects, non-representativeness of the sample would be a problem only if there are interactions between the treatment and one or more of the characteristics with respect to which the sample differs from the population. If there are differences between the samples participating in the three waves of measurement, the MAR-based ML modelling should be able to control for these differences.

Approaches of analysis
The basic idea of the current study is that a longitudinal design is used, in which the cases are their own controls. Thus, the treatment effect is estimated by comparing the level of response after the treatment (i.e., the inspection) has taken place with the previous response or responses.

Technically, this estimation can be conducted in several different ways, such as for example with fixed-effects regression techniques, or with growth and path modelling techniques. . Here estimation procedures based on latent variable growth and path modelling approaches implemented in the Mplus 7 program (Muthén & Muthén, 1998 – 2012) were used because of their flexibility and versatility, the availability of tests of model fit, and access to powerful procedures for missing data modelling. Two different modelling approaches were used, one based on principles of growth modelling and the other based on principles of path modelling.

In the growth modelling approach an Intercept factor was estimated as a latent variable with a fixed relation of unity to all three measures, along with a Change factor as another latent variable with a fixed relation of unity to the variable observed the year after the inspection. The Change factor thus had the value of 1 for the relation to the 2011/2012 observation for those inspected in 2010/2011, to the 2012/13 observation for those inspected in 2011/2012, and to the 2013/14 observation for those inspected in 2012/2013. This specification makes the assumption that the effect of the inspection appears the year after the inspection, and then disappears. Another assumption, representing a more desirable outcome, would be that the inspection effect lasts, and preferably so for a longer period of time. A specification which captures this situation would be that the Change factor has a fixed value of unity for all the observations after the year of inspection (e.g., for the 2011/2012, 2012/2013 and 2013/2014 observations for those inspected in 2010/2011; for the 2012/2013 and 2013/2014 observations for those inspected in 2011/2012; and for the 2013/2014 observations for those inspected in 2012/2013.

The models were specified and estimated as multiple-group models with the Mplus 7 program (Muthén & Muthén, 2012), using the ML estimator. Data from the three groups who were inspected in 2010/2011, 2011/2012 or 2012/2013 were included (N = 656). Each and every parameter was constrained to be equal across groups. The treatment effect was evaluated with the estimated mean of the Change parameter. An effect size was computed by dividing the estimated mean with the standard deviation of the Intercept factor. A separate model was fitted for each of the two main outcomes, using observed scale scores for *Capacity Building* and *School Effectiveness*.

In the path modelling approach a one-group model was fitted, with treatment represented by three dummy variables, representing whether inspection took place one of the three years of the duration of the study. The reference group thus consisted of the schools which were inspected before the first wave of measurement of the study, or which had an unknown inspection date. This implies that in the reference group there must be schools which have

been inspected too, even though it is not known which particular year this happened. One likely consequence of this is that the treatment effect will be underestimated.

The logic of the modelling is most easily explained through demonstrating a hypothetical model (see Figure 3).
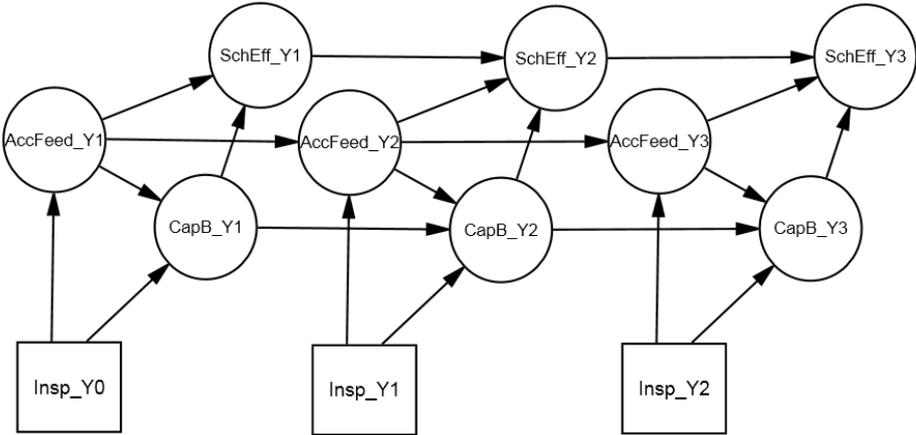

Figure 3. A hypothetical model for estimating effects of school inspections on *Capacity Building* (CapB) and *School Effectiveness* (SchEff) with *Accepting Feedback* (AccFeed) as a mediating variable

The main outcome variable is *School Effectiveness*, which is hypothesized to be influenced by *Capacity Building*. These two variables are also hypothesized to be influenced by *Accepting Feedback*, which in turn is influenced by whether the school was inspected the year before the measurement was made. The hypothesis thus is that the effect of Inspection on *Accepting Feedback* partially explains the effect of Inspection on *School Effectiveness* and *Capacity Building*. There also are autoregressive relations between corresponding variables over time. When estimating the model, constraints of equality have been imposed on the relations from Insp_Y0, Insp_Y1 and Insp_Y2, so that a common estimate of the effect of the inspection activities is obtained.

This model does not specify all possible relations. For example, the Inspection variables are taken to have direct effects on *Accepting Feedback* and on *Capacity Building* but not on *School Effectiveness*. This does not mean that Inspection is not influencing *School Effectiveness*, but only that there is no direct effect of Inspection on *School Effectiveness* over and above the indirect effects of Inspection via *Accepting Feedback* and *Capacity Building*. It could of course be that there is a direct effect of Inspection on *School Effectiveness* and should this be the case the model can easily be modified to estimate this effect too. It should be noted, however, that it is desirable not to have any direct effect of Inspection if there is a significant total effect, because this implies that the mediating effects via *Accepting Feedback* and *Capacity Building* explain the effects of Inspection on *School Effectiveness*.

A simpler version of this model does not include the *Accepting Feedback* variable. This model thus estimates the direct effect (which is also the total effect) of Inspection on *Capacity Building*, and the indirect effect of Inspection on *School Effectiveness* via *Capacity Building*.

However, the model that was actually used was more complex, and included all the variables that were hypothesized as explanatory variables on the basis of the conceptual model. This model is described in the Results section below.

5.2 Results

First the results from the growth modelling approach are reported, and then the results from the path modelling are presented.

Growth modelling

The analyses focus on the two main outcome variables *Capacity Building* and *School Effectiveness*. A model just including a latent variable representing the intercept was first fitted as a baseline model. The χ^2 goodness-of-fit test indicated a relatively good fit for this baseline model both for *Capacity Building* ($\chi^2 = 30.94$, df = 22, p < .10) and for *School Effectiveness* ($\chi^2 = 30.32$, df = 22, p < .11). To investigate effects of school inspections, models assuming both non-lasting and lasting effects were fitted. For *Capacity Building* the model assuming non-lasting effects had marginally better fit than the baseline model ($\chi^2 = 28.02$, df = 20, p < .11) while the model assuming lasting effects improved fit significantly as compared with the baseline model ($\chi^2 = 16.81$, df = 20, p < .66; $\Delta\chi^2 = 14.13$, Δdf = 2, p < .001). For *School Effectiveness* the fit of the model assuming non-lasting effects was about the same as the fit as the baseline model ($\chi^2 = 28.72$, df = 20, p < .09) while the model assuming lasting effects fitted somewhat better ($\chi^2 = 24.78$, df = 20, p < .22; $\Delta\chi^2 = 5.54$, Δdf = 2, p < .06). While the difference between the baseline model and this model did not quite reach significance, the parameter estimate for the change parameter was significant (t = 2.36, p < .02).

The unstandardized parameter estimate for the lasting effect of school-inspections on *Capacity Building* was 0.077 (t = 3. 69). The standard deviation of the intercept was 0.31, so an effect size estimate is 0.077/0.313 = 0.25. For *School Effectiveness* the unstandardized estimate was 0.054, and the effect size estimate was 0.17. We may thus conclude that there is a significant effect of school inspections on *Capacity Building*, and that there is a marginally significant effect of school inspections on *School Effectiveness*, in both cases with small effect sizes.

Analyses also were done at the item level to investigate homogeneity of effects across items. All estimates of the Change parameters were positive for both the *Capacity Building* and the *School Effectiveness* items. However, there was some variation in the parameter estimates, and all were not significant. For *Capacity Building* the highest estimate was observed for the item "I encourage teachers to improve their teaching practices" (t = 2.32) and for the item "Teachers collaborate in organizing and improving their teaching" (t = 1.97). For *School Effectiveness* the highest estimate for the Change parameter was observed for the item "I use assessment results to target areas for school improvement" (t = 2.74) and for the item "Teachers use clear, structured and challenging teaching approaches" (t = 2.00).

These results thus indicate that the regular inspections have an impact on principals' assessments of the schools' levels of capacity building and school effectiveness. But the mechanisms through which these effects come about are not revealed in these analyses. We therefore turn to the path modelling approach.

Path modelling

In the first step of analysis the total effects of inspections on *Capacity Building* and *School Effectiveness* were estimated, so no explanatory or mediating variable was included. First one model was estimated for *Capacity Building* and then *School Effectiveness* was added to the model. In addition to the relations indicated in Figure 3, the model allowed covariances among the residuals of the manifest variables over the three time points. The fit of the *Capacity Building* model was good (χ^2 = 311.31, df = 159, p < .00; RMSEA = 0.028, CI_{90} = 0.023 – 0.033; CFI = 0.959; SRMR = 0.039). The direct effect of inspection on *Capacity Building* was significant, with a small effect size (d = 0.14, t = 2.64). Furthermore, the inspection at t1 had a relatively strong indirect effect on *Capacity Building* at t2 (d = 0.10, t = 2.62) and at t3 (d = 0.08, t =2.61), indicating relatively large lasting effects of inspection over a three-year period. These lasting effects can be brought back to the high stability over time of *Capacity Building*, the correlation being 0.73 between t2 and t1, and 0.83 between t3 and t2.

In the next step of analysis the *School Effectiveness* latent variable was added to the model, assuming a unidirectional influence from *Capacity Building* to *School Effectiveness* at each time point, and autoregressive relations over time. The fit of the model was good (χ^2 = 1148.51, df = 540, p < .00; RMSEA = 0.030, CI_{90} = 0.028 – 0.033; CFI = 0.918; SRMR = 0.047). There was no direct effect of inspection on *School Effectiveness*, but there were significant indirect effects via *Capacity Building*. Thus, inspection at t0 had an indirect effect on *School Effectiveness* at t1 (d = 0.11, t = 2.27), and the effect was lasting to a large extent (t2: d = 0.09, t = 2.58; t3: d = 0.07, t = 2.51).

In the next step of analysis it was investigated to what extent the effects of inspection were mediated via variables representing *Setting Expectations*, *Stakeholder Sensitivity to Reports*, *Accepting Feedback* and *Improving Self-Evaluations*, as was hypothesized on the basis of the conceptual model. The model had acceptable fit (χ^2 = 6534.25, df = 3029, p < .00; RMSEA = 0.020, CI_{90} = 0.019 – 0.021; CFI = 0.925; SRMR = 0.061). The part of the model which represents the relations among the variables during a single year is presented in Figure 4.
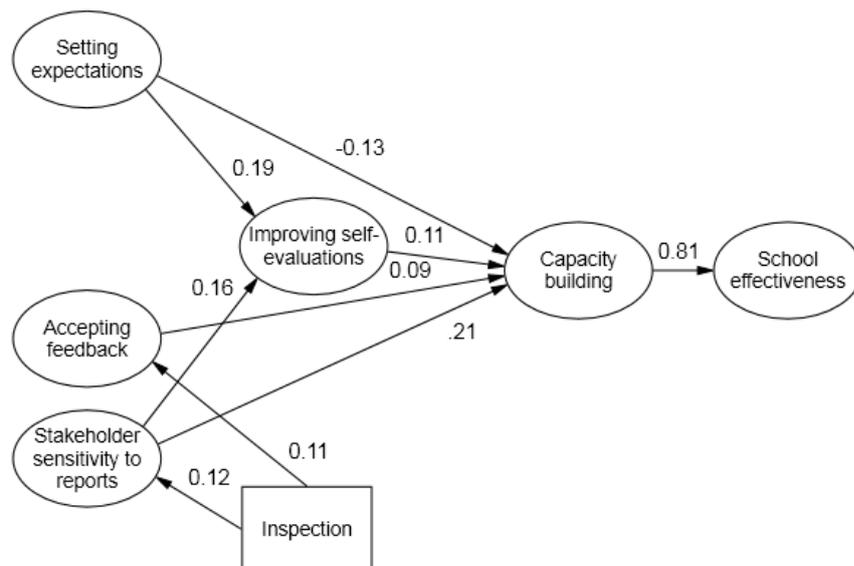

Figure 4. Estimates of effects of Inspection on mediators and outcomes in the path-model.

There was no significant effect of inspection on *Setting Expectations* (d =- 0.05, t = -0.94), but there were effects on *Stakeholder Sensitivity to Reports* (d = 0.30, t = 5.24) and on *Accepting*

Feedback (d = 0.28, t = 4.45). Given that these latent variables were related to *Improving Self-Evaluations* which in turn was related to *Capacity Building* and *School Effectiveness*, the effects of the inspections were mediated via these chains of relations between the variables of the model. It may also be noted that in this model there was no direct effect of inspection on *Capacity Building* (d = 0.02, t = 0.34) or on *School Effectiveness* (d = 0.02, t = 0.38), which implies that the effects of inspection on the outcome variables were completely mediated via effects on *Stakeholder Sensitivity to Reports* and on *Accepting Feedback*.

Table 1. Estimates of effects on the outcomes variables of inspection at year 0

	Total		Direct		Indirect	
	d	t-value	d	t-value	d	t-value
Capacity Building Year 1	0.12	2.24	0.02	0.34	0.10	5.45
Capacity Building Year 2	0.13	3.36			0.13	3.36
Capacity Building Year 3	0.12	3.99			0.12	3.99
School Effectiveness Year 1	0.12	2.32	0.02	0.38	0.10	2.23
School Effectiveness Year 2	0.11	3.34			0.11	3.34
School Effectiveness Year 3	0.11	3.93			0.11	3.93

Table 1 presents estimates of total and indirect effects of the inspection at t0 at each of the three waves of measurement. It is quite interesting to observe that even though the effects are small, they do remain stable over time for both *Capacity Building* and for *School Effectiveness*. This is a somewhat surprising finding, so it may be worthwhile to look more closely into how these effects come about. There were 58 specific indirect effects of inspection at t0 on *School Effectiveness* at t3, but only a few of these were significant and of substantial magnitude. The strongest indirect effect involved the first measure of *Stakeholder sensitivity to Reports* and the three measures of *Capacity Building*, and the second strongest effect involved the first two measures of *Stakeholder Sensitivity to Reports* and the two last measures of *Capacity Building*. Yet another path comprised the three measures of *Stakeholder Sensitivity to Reports* and the last measure of *Capacity building*. Three similar paths, which involved *Accepting feedback* instead of *Stakeholder Sensitivity to Reports* also were significant. Another set of paths, which were significant but with smaller effects, included *Improving Self-evaluations* along with either *Stakeholder Sensitivity* or *Accepting Feedback*. Thus, it seems that the inspection effect was propagated over time as a consequence of the substantial autoregressive stability of all the latent variables, and because both *Stakeholder Sensitivity to Reports* and *Accepting Feedback* influence development of *Capacity Building*.

Analyses also were made of effects of school inspections on the five items intended to assess unintended consequences. Table 2 presents estimates of the effects of inspection at t0 on the responses to the five items at the three measurement occasions.

Table 2. Estimates of unintended effects of the inspection at t0

	Total		Direct		Indirect	
	d	t-value	d	t-value	d	t-value
Discourage experimentation with teaching methods, t1	0.00	-0.07	-0.02	-0.36	0.02	1.21
Discourage experimentation with teaching methods, t2	0.01	0.43			0.01	0.43
Discourage experimentation with teaching methods, t3	0.01	1.16			0.01	1.16
Narrowing curriculum, t1	-0.19	-3.73	-0.13	-2.57	-0.06	-3.75
Narrowing curriculum, t2	-0.08	-5.02			-0.08	-5.02
Narrowing curriculum, t3	-0.04	-4.66			-0.04	-4.66
Refocusing curriculum and teaching, t1	-0.12	-2.23	-0.09	-1.63	-0.03	-1.61
Refocusing curriculum and teaching, t2	-0.05	-3.28			-0.05	-3.28
Refocusing curriculum and teaching, t3	-0.03	-3.74			-0.03	-3.74
Documents too positive, t1	-0.33	-7.60	-0.28	-6.56	-0.04	-3.07
Documents too positive, t2	-0.13	-6.77			-0.13	-6.77
Documents too positive, t3	-0.07	-5.44			-0.07	-5.44
Putting things into protocols and writing, t1	0.21	4.62	0.21	4.50	0.00	0.13
Putting things into protocols and writing, t2	0.05	3.46			0.05	3.46
Putting things into protocols and writing, t3	0.01	2.37			0.01	2.37

There was no effect for the item asking whether inspections discourage experimentation with new teaching methods. However, for the item concerning possible narrowing effects on the curriculum there was a relatively strong negative effect at t1, implying that the principals who had been inspected denied the existence of any negative effects. A similar outcome was observed for the item asking if inspections had caused a refocusing of curriculum and teaching, and an even stronger negative effect was observed for the item asking if documents present too positive a picture of the school. There was a positive effect of school inspection on the items asking to what extent principals agreed that the documents submitted before the inspection represent procedures that are in place at the school. For all the items with a significant direct effect, the effect vanished over time, and was generally close to zero at t3.

For these items there thus is a pattern of influence such that the year following an inspection the principals to a lower extent than before the inspection think that inspections have negative unintended consequences and that school leaders manipulate documents and data they send to the Inspectorate.

5.3 Discussion and Conclusions

The results of the different approaches to analysing the data from the longitudinal study of effects of school inspections in Sweden do indicate that inspections influence the principals' reports of activities that support learning and instruction. The estimated effect sizes are small, however, and they tend to vary somewhat over the different methods of analyses.

With the growth modelling approach the estimated effect size was 0.25 for *Capacity Building* and it was 0.17 for *School Effectiveness*, while with the path modelling approach the corresponding estimates were 0.14 and 0.11. The main reason for these differences is most likely that with the growth modelling approach, the inspected schools are compared with the schools which are not yet inspected during the three-year period, while with the path

modelling approach the comparison group also includes schools which may have been inspected some time before the three-year period. This causes an underestimation of the inspection effect with the path modelling approach, so the higher estimates obtained with the growth modelling approach are more trust-worthy.

It should also be noted that the growth modelling estimates were computed under the assumption that effects were lasting (i.e., for three years for the schools inspected in 2010/2011, and for two years for the schools inspected in 2011/2012). This thus implies an assumption that effects lasted for two years and there was empirical support for this assumption. It is reasonable to expect that the inspection effect vanishes successively, but the power of the empirical study was too small for the growth models to be able to demonstrate any such effect.

Even though the path modelling approach is likely to underestimate the effect of the inspection this approach has several advantages over the growth modelling approach. First, this approach is statistically more powerful than is the growth modelling approach, as is shown by the smaller standard errors of the estimates obtained with the path models. This higher power is due both to the larger sample used in the path modelling (N = 1217 vs N = 656), and to the fact that the path model takes advantage of the strong auto-regressive relations of the latent variables over the three time-points.

A second advantage of the path modelling approach is that it can explain how it comes that the effects of the school inspections are lasting beyond the first year. As has been described above, there is considerable stability of the outcome variables over time, and the mediating variables *Stakeholder Sensitivity to Reports* and *Accepting Feedback* also show strong autoregressive relations over time. These relations create indirect effects of inspection over time, which are equally strong as the direct effects.

A third advantage of the path modelling approach is that it can identify which hypothesized mechanisms can account for the effects of inspection on the outcome variables. The modelling results showed that inspection affected *Stakeholder Sensitivity to Reports* and *Accepting Feedback* and that these variables affected *Capacity Building* both directly and indirectly via *Improving Self evaluations*. *School Effectiveness* was in turn strongly affected by the *Capacity Building* variable.

The path model thus identifies the two factors *Stakeholder Sensitivity to Reports* and *Accepting Feedback* as the main mediators of the effects of inspections on the outcomes. These factors are in the conceptual model identified as two main drivers of the effects of inspections on activities to improve learning and instruction, so these empirical results are expected from the conceptual model, and provide support for the model.

6 Conclusions

The project of which this report is a part has generated a very considerable amount of descriptive information about the practices and impact of school inspections in Europe. However, the most important contribution of the project is that through the research design that was adopted it has made it possible to make credible inferences about causal effects of school inspections, and to account for the mechanisms behind those effects within a comprehensive theoretical framework. This progress, along with the expandable database constructed within the project, allows for further expansion of the knowledge base about

which effects school inspections have on school-level processes associated with improvement of student performance.

7 References

Ehren, M.C.M., Altrichter, H., McNamara, G. and O'Hara, J. (2013). Impact of school inspections on teaching and learning; Describing assumptions on causal mechanisms in seven European countries. *Educational Assessment, Evaluation and Accountability.* 25(1), 3-43.

Gustafsson, J.-E., Lander, R., & Myrberg, E. (2014). Inspections of Swedish schools: A critical reflection on intended effects, causal mechanisms and methods. Under review for publication in *Education Inquiry*.

Muthén, L.K., & Muthén, B.O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.