The mechanisms by which external school evaluation in Iceland influences internal evaluation and school professionals' practices

Björk Ólafsdóttir, Jón Torfi Jónasson, Anna Kristin Sigurðardóttir & Thor Aspelund

Published online: 19 May 2022.

Submit your article to this journal

Article views: 377

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

The mechanisms by which external school evaluation in Iceland influences internal evaluation and school professionals' practices

Björk Ólafsdóttir[a], Jón Torfi Jónasson [b], Anna Kristin Sigurðardóttir [b] and Thor Aspelund [c]

[a]School of Education, University of Iceland, Reykjavík, Iceland; [b]School of Education, University of Iceland, Reykjavik, Iceland; [c]School of Health Sciences, University of Iceland, Reykjavik, Iceland

ABSTRACT

The main purpose of this research is to analyse school principals' and teachers' attitudes towards external school evaluation in Iceland, in particular, the ways in which they consider the evaluation affects their schools' internal evaluation and drives changes in their own practices. The study uses a quantitative method and is based on a survey conducted among principals and teachers in 22 schools that were externally evaluated during the years 2013 to 2015. The results indicate a positive attitude towards external school evaluation among both teachers and principals. *Acceptance, setting expectations*, and *teacher participation* were found to be significant predictors of perceived changes in internal evaluation in the teachers' data. However, only *acceptance* significantly explained perceived changes in teaching practices. In the principals' data, the only variable that had a significant association with perceived changes in internal evaluation was *setting expectations*, and only *acceptance* had a significant association with perceived changes in leadership practices. In accordance with the hypothesis of this study, the results underpin the importance of acceptance of the evaluation feedback and setting expectations through quality standards. However, contrary to the hypothesis, external stakeholder involvement did not prove to be a strong determinant of change as perceived by principals.

Introduction

Governments and education policy-makers have increasingly emphasized the internal and external evaluation of schools and education systems to obtain information about their educational performance and to improve school practices (Eurydice, 2015; Organisation for Economic Cooperation and Development [OECD], 2013). In European countries, the external evaluation is conducted in a variety of ways and has different functions as it is embedded in different political priorities, national circumstances and educational traditions. Some national governments use a 'soft' governance approach ('low-stake'), and consequently, their external evaluation tends to have low accountability pressure, while others use a 'hard' governance approach ('high-stake') with greater accountability pressure (Altrichter & Kemethofer, 2015; Ehren, Gustafsson et al., 2015). The Ehren et al. (2013) study distinguished four main dimensions of external school evaluations: (1) types of inspection and frequency of visits; (2) standards and the threshold for identifying failing schools; (3) consequences: sanctions, rewards and interventions; and (4) the presence or absence of reporting on individual schools to the general public. Based on this classification, Ehren et al. (2015) and

Altrichter and Kemethofer (2015) pointed out that countries with a 'high-stake' external evaluation approach use differentiated evaluation models, outcome-oriented evaluation, sanctions for failing schools, and reports on individual schools to inform the public. By contrast, countries with a 'low-stake' approach to their external school evaluations use a cyclical evaluation of all schools that is process-oriented, impose no sanctions for failing schools, and do not inform the public with reports on individual schools.

The current state of knowledge on school evaluations in Europe is based mainly on studies from countries with a long history of external school evaluation that use a somewhat 'high-stake' evaluation approach, such as England and the Netherlands (Altrichter & Kemethofer, 2015; Ehren, Gustafsson et al., 2015). Research in countries using a rather 'low-stake' approach, such as Germany, Austria, and Switzerland, has increased in recent years (see e.g. Behnke & Steins, 2017; Dedering & Müller, 2011; Schweinberger et al., 2017). To gauge the influence of context, studies from an even wider spectrum of systems are needed to add to this corpus of European research, which thus clarifies the context sensitivity of the issue. No research on external school evaluations

has been undertaken in Iceland; hence, no results are available on the effect of evaluations on school improvement. The results of this study contribute in this respect and can both enhance the theoretical understanding of the contribution of external evaluation to school development and serve to support policy- and decision-making on the future of external evaluations. The main purpose of this research is to identify and analyse school principals' and teachers' attitudes towards external evaluation in Iceland, looking at, in particular, the ways in which they consider that such evaluations affect their schools' internal evaluation and drives changes in their own practices.

In Iceland, the term 'external school evaluation' is used instead of the term 'inspection'. Therefore, external evaluation is used throughout this article to avoid confusion. External evaluation is, nevertheless, considered to involve the same process as inspection and to be based on the same theoretical foundation. Similarly, the term 'internal evaluation' is used for evaluation that takes place within each school. This article does not differentiate between internal evaluation and self-evaluation, and both terms are used interchangeably, the former in keeping with the Icelandic practice and the latter in relation to some of the literature.

The next section outlines the main characteristics of the Icelandic external evaluation system that may influence school improvement according to the four main dimensions of external school evaluation listed above.

The Icelandic context and process of external school evaluation

The Icelandic compulsory school system has two administrative levels of governance: the state and the municipalities. Compulsory education is under the jurisdiction of the Ministry of Education, Science and Culture. Local authorities (the municipal councils) are liable for the operation and expense of compulsory school and each municipality is required to operate a school board, which acts on behalf of the municipal council, and manages the compulsory school's affairs as prescribed by law and regulations ("Compulsory School Act," 2008). The school's principal is responsible to the municipal council for the school's practice, directing the school, and providing professional leadership. Both the state and the municipalities are obliged to evaluate schools. Local authorities (the municipal council and the school board) are liable for following up external evaluations to ensure that they lead to improvements in the schools, and the ministry is responsible for supervising local authorities to ensure that they fulfil their obligations in that context. The principal, in collaboration with the school's personnel, is responsible for the quality of school activities and

each school is required to carry out a systematic internal evaluation (The Icelandic National Curriculum Guide for Compulsory Schools, 2011). Schools choose their own evaluation methods, focus areas and evaluation criteria.

In 2013, an approach to external school evaluation that is a cooperative project between state and municipalities and jointly financed was implemented in the Icelandic education system (Ólafsdóttir, 2016). This project is the subject of the present study. The Educational Testing Institute from 2013 and its successor from 2015, the Directorate of Education, were assigned to implement the external evaluation on behalf of the state and municipalities. The directorate is an administrative institution, and its task is to improve quality and progress in education by, for example, monitoring and evaluating school progress and outcomes ("Act on the Directorate of Education," 91/2015). In addition to the external evaluation, the directorate conducts, among other things, the implementation of national coordinated examinations and international studies such as PISA. The main objectives of the external evaluation are to provide information, ensure that schools are operated in accordance with the law, stimulate school improvement, increase the quality of education, and ensure that students' rights are respected ("Compulsory School Act," 2008).

Types and frequency of evaluation visits: The overall program for external school evaluations is in the form of a cyclical evaluation. For the first five years, 10 schools were evaluated annually, and then this number was increased to 27 schools in 2018. The first evaluation cycle was completed in 2021, nine years after the external evaluation was first implemented. The annual selection of schools is based on their distribution across the country and among the municipalities. The evaluation is carried out primarily through a qualitative methodology and involves document analysis, assessment data and a school visit, where classroom observations are made, and interviews conducted with key stakeholders (principal, management team, teachers, other staff, students and parents).

Standard and threshold: The external evaluation is based on three main areas of school quality: (1) *quality of learning and teaching*, (2) *quality of school leadership and management* and (3) *quality of internal evaluation* (Sigurjónsdóttir et al., 2012). The quality criteria focus on school activities and processes and are expected to stimulate improvements, although schools are not required to use them in their internal evaluation. No threshold levels are used to rank schools as failing or satisfactory.

Reporting to the public: Based on the evaluation criteria, recommendations for improvement are

presented in a written report published on the website of the Directorate of Education.

Consequences – sanctions, rewards and interventions: In collaboration with the local education authority, each school is required to develop an improvement plan and deliver it to the Ministry of Education, Science and Culture (beginning in 2019, to the Directorate of Education). Further follow-up is undertaken in the form of communication between the ministry/directorate, the municipality and the school until all improvements have been implemented. The follow-up process can last from one to several years depending on the plan's timelines. The local education authorities are expected to use the evaluation data to inform the school of their support and challenge them to improve. The external evaluation carries no specified consequences either for schools that are judged to have weaknesses or schools that fail to implement their improvement targets. Neither the ministry nor the directorate is in a position to impose sanctions on schools or reward them.

To summarize, the external school evaluation in Iceland can be understood as rather 'low-stake' since it is based only on regular but infrequent cyclical evaluation with process-oriented quality criteria and carries no specified consequences except for the improvement recommendations and the ministry's/directorate's follow-up with the improvement plan in all schools. The results are published, but comparisons are not expected to be based on them.

Theoretical framework and previous research

This study is inspired by the theoretical framework developed by Ehren et al. (2013), which was based on programme theories (i.e. assumptions on the causal mechanisms underlying the program) of six European countries' school-inspection systems (England, the Netherlands, Ireland, Sweden, Austria and the Czech Republic). Although these six countries have different assumptions about the effectiveness of inspection, their common ideas were incorporated into the framework. As shown in Figure 1, the framework describes the intermediate processes of how inspection/external school evaluation is expected to impact the improvement of schools and promote quality education. Key characteristics of school inspection that may influence improvement are highlighted in the first column, such as evaluation methods, standards and criteria, feedback, and the degree of accountability pressure including consequences and public disclosure of evaluation results. The second column defines two mediating mechanisms that are expected to be causal in bringing about school improvement: *setting expectations* and *accepting feedback*. *Stakeholder involvement* (i.e. actions) is the third mediating causal factor. These three mechanisms are interconnected and are expected to trigger a process sequence that links external evaluation to its intended impact on *promoting/improving self-evaluation* and stimulating *improvement actions*, as shown in the third column. A high-quality *self-evaluation* and *improvement actions*, in turn, are expected to promote increased
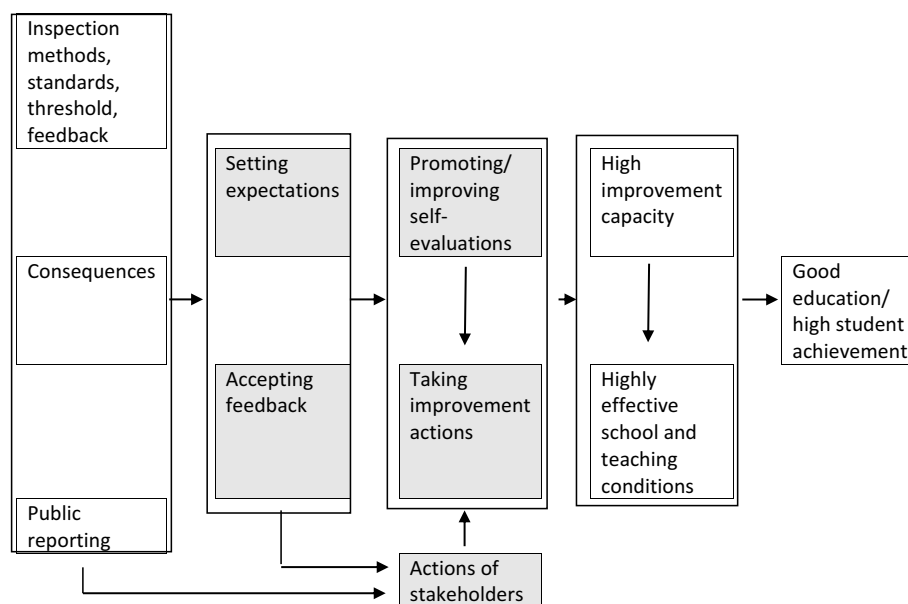

Figure 1. Framework of causal mechanisms of school inspection (Ehren et al., 2013, p. 14, with permission).

improvement capacity and more *effective school and teaching conditions* (column four), which will lead to the ultimate goal of the inspection – *higher-quality education and improved student results* (fifth column; Ehren et al., 2013).

The framework assumptions were tested by means of a school principal survey in the six European countries on which the framework was based (Ehren et al., 2013). The results of those studies are discussed below and used as a basis for discussing the results of this study, which only deals with the Icelandic situation.

Although inspired by Ehren et al. (2013), the analysis in this study is based on a simpler and reduced model than shown in Figure 1 – more specifically, on the part of the framework marked with shaded boxes. Accordingly, the focus is not on the boxes to the right in the figure that point to various important aspects of the school operation.

Setting expectations

The anticipated first causal mechanism of the framework involves setting standards and criteria that define school quality and on which external school evaluation and school improvement are based. The criteria and standards are intended to influence schools and their stakeholders to align their views with those expressed in the criteria in regard to the factors that constitute quality education and good schools (Ehren et al., 2013; Ehren, Gustafsson et al., 2015). The idea is that schools will seek to adapt to these norms by shaping their goals, work structure, daily practices, and internal evaluation in accordance with them. Therefore, the criteria are intended primarily to contribute to a school's progress in relation to the expectations they set, and the external evaluation can be seen as an effective way of communicating those expectations and norms (Ehren, Eddy-Spicer et al., 2016; Kemethofer et al., 2017).

'High-stake' accountability systems are among other factors based on the assumption that rational actors strive to meet the standards because of the threat of sanctions if they do not (Ehren, Gustafsson et al., 2015). In this way, a high level of external pressure increases the likelihood that schools will make use of the standards. The studies by Ehren et al. (2015) and Altrichter and Kemethofer (2015), based on the survey responses of principals in the above-mentioned European countries, largely confirm this. They found that principals who felt more 'accountability pressure' were more alert to the expectations issued by the inspectorate; however, unintended consequences also increased with greater pressure (Altrichter & Kemethofer, 2015; Jones et al., 2017). Such unintended side effects are, for example, strategic behaviour (intended or unintended) on

behalf of the school, such as narrowing the curriculum, misrepresenting data, teaching to the test and discouraging teachers from experimenting in teaching (Ehren, Gustafsson et al., 2015; Ehren, Jones et al., 2016; Jones et al., 2017). In their study of Dutch schools, where accountability pressure is high (Altrichter & Kemethofer, 2015), Ehren, Perryman et al. (2015) revealed that setting clear expectations is a strong determinant of improvements and changes made in the capacity of the school. Yet a comparative study by Kemethofer et al. (2017) between Austria and Sweden, where Austria represents a 'low-stake' system and Sweden a 'medium- to high-stake' system (Altrichter & Kemethofer, 2015), showed a significant effect of *setting expectations* in Austria but not in Sweden and, thus, does not comply with the assumption of 'high-stake' systems.

Accepting feedback

The second causal mechanism in the framework is through the feedback that schools receive during the evaluation visit and in the evaluation report (Ehren et al., 2013). The feedback is based on the quality standards and criteria, and it is expected that schools will use the feedback to address the areas in need of development that emerged when the quality of the school's work was compared to the criteria. They do so by realizing, accepting, deciding on and implementing improvement actions that meet the requirements of the standards, thereby, improving school quality (Ehren et al., 2013; Ehren, Gustafsson et al., 2015). Based on various studies, the acceptance of the feedback is essential if decisions about school improvement are to be made based on them (Dedering & Müller, 2011; HMIE, 2010; Penninckx, 2017; Schildkamp & Ehren, 2013). If the feedback received is considered compelling, relevant, valuable and supportive, it is more likely to be accepted and used for improvement (Altrichter & Kemethofer, 2015; Behnke & Steins, 2017; HMIE, 2010; Verhaeghe et al., 2010). In a recent research summary and analysis, Hofer et al. (2020) concluded that the better the school staff's impression of the quality of the evaluation is, the more positive the consequences of the evaluation will be, meaning it will have an impact on process and outcome variables. However, they and other researchers have pointed out that principals and teachers show partly different reactions to external evaluation, with in most cases the principals being more positive than the teachers (Ehren, Perryman et al., 2015; Hofer et al., 2020; Matthews & Sammons, 2004).

Studies conducted to verify the hypotheses of the theoretical framework of Ehren et al. (2013) have indicated, contrary to the hypotheses, that *acceptance of feedback* is not necessarily a key factor in the

impact of external evaluation on school improvement (Altrichter & Kemethofer, 2015; Ehren, Perryman et al., 2015; Gustafsson et al., 2015). It was concluded that schools that accept feedback do so because of the other two influencing factors, namely *setting expectations* and *stakeholder involvement*. In the study of Gustafsson et al. (2015), no improvement actions were found as a result of *accepting feedback*, and in the study of Ehren, Perryman et al. (2015) no improvement in *self-evaluation* or *capacity to improve* was found because of *accepting feedback*, only changes on conditions of school effectiveness (i.e. opportunity to learn, quality of teaching and use of assessment to monitor students' progress and the quality of the school). Contrary to those results, Kemethofer et al. (2017) study revealed the effect of inspection for *accepting feedback* positive and significant in both Austria and Sweden, albeit it was much higher in Austria. A similar conclusion was reached by Altrichter and Kemethofer (2015), who found that *accepting feedback* did significantly influence *improving self-evaluation* in Austria and Sweden but not in England. They concluded that *accepting feedback* is more important in 'middle' and 'low-stake' evaluation approaches than in 'high-stake' approaches, due to lower accountability pressure (Altrichter & Kemethofer, 2015).

Stakeholder involvement

Stakeholder involvement is the third causal mechanism (Ehren et al., 2013). Feedback from external evaluation is not only between schools and evaluators; it also takes place in a complex, multilevel system where different stakeholders external to the school (e.g. authorities, school boards, community members and parents) have access to information about the evaluation results through presentations and/or the publication of evaluation reports (Ehren, Perryman et al., 2015). External stakeholders are expected to review the results, support the schools in their development and pressure them to respond to the feedback with improvement actions to fulfil the evaluation criteria (Ehren et al., 2013; Ehren, Eddy-Spicer et al., 2016; Kemethofer et al., 2017). Thus, *stakeholder involvement* refers to their awareness of the school evaluation report and the identified improvement needs and to their use of their 'voices' to bring about change (Ehren et al., 2013; Ehren, Gustafsson et al., 2015).

Ehren, Perryman et al. (2015) suggested that the mechanism of *stakeholder involvement* was an important driver for the implementation of a *school's self-evaluation* in the Netherlands and that it was also related to increased *feedback acceptance*. By contrast, in their study, Kemethofer et al. (2017) did not identify any

effect of *stakeholder involvement* on improvement in either Austria or Sweden. Similarly, Gustafsson et al. (2015) concluded that *stakeholder involvement* did not directly impact improvement actions but rather had an effect on school responses early in the improvement process by motivating them to accept feedback and inspection standards.

The inconsistencies in research findings regarding the three causal mechanisms of the framework – *setting expectations, accepting feedback*, and *stakeholder involvement* – indicate that the effects are not at all harmonious and depend on a variety of other conditions, such as the national context in which they are implemented (Kemethofer et al., 2017). However, in this study and in accordance with the assumptions of the framework, it is hypothesized that there is a positive correlation between the three causal mechanisms of the model and changes in internal evaluation, teaching and leadership practices in Icelandic schools. It is also assumed that other influencing factors – discussed below under the heading capacity to implement improvements – will associate positively or negatively with changes, given the variables in question.

Capacity to implement improvements

Building capacity refers to a school's capacity to address weaknesses, improve the professional learning of teachers and take actions to make teaching practices student-oriented, which may ultimately result in higher student achievement (Ehren, 2016). In the theoretical framework, improvement in capacity-building is expected to be the result of improved self-evaluation. The framework does not address the pre-existing capacity of the school as an influencing factor for improvement (Ehren et al., 2013; Hofer et al., 2020). However, in her writings, Ehren (2016) has highlighted capacity-building as a fourth mechanism of change that explains how external evaluation can lead to improvements in school quality. Capacity to improve, therefore, plays a dual role in explaining external evaluation effects; it acts as a condition for schools to respond to expectations with improvements, and at the same time external evaluation is expected to increase the school's capacity to improve (Ehren, 2016; Ehren, Eddy-Spicer et al., 2016). Influencing organizational factors that have been shown to be relevant for external evaluation success are, for example, favourable attitudes towards internal evaluation, participation of teachers in decision-making and the existence of resources and knowledge in schools that support improvement (Ehren, 2016;

Ehren, Eddy-Spicer et al., 2016; Hofer et al., 2020; Schildkamp & Lai, 2013).

Improving self-evaluation and taking improvement actions to promote professional development

Through the causal mechanism described above, external evaluation aims to stimulate and drive improvement and self-evaluative actions by schools (Altrichter & Kemethofer, 2015; Ehren et al., 2013). *Improving self-evaluation* and *taking improvement actions* are described as 'intermediate mechanisms' that contribute to a school's ultimate success (Gustafsson et al., 2015). This is based on the hypothesis that internal evaluation plays an important role in enhancing improvement processes (Ehren, 2016; Ehren et al., 2013; Hanberger et al., 2016). Studies based on the theoretical framework (Figure 1) have shown that external evaluation has a positive impact on school self-evaluation, and that schools that improved their self-evaluation process also improved their capacity-building, thus leading in turn to improvement in the conditions for an effective school (Altrichter & Kemethofer, 2015; Ehren, Perryman et al., 2015; Gustafsson et al., 2015; Kemethofer et al., 2017). However, a comparison between the 'low-stake' Austrian approach and the 'high-stake' British model revealed that the external evaluation seemed to have less impact on changes in self-evaluation in Austria than in England (Altrichter & Kemethofer, 2015).

Research focus and questions

The literature review above discussed research carried out in countries outside Iceland. In the present study, the focus shifts towards Iceland. Part of the mechanisms of the theoretical framework by Ehren et al. (2013) are used to understand the reception and potential influence of external evaluation in Iceland on internal evaluation and school professionals' practices. This reflects the shaded boxes in Figure 1.

Regarding this focus, the following research questions will be investigated in this article:

RQ1a – How can the scales constructed to delineate reception and attitude (*accepting feedback* and *teacher participation*) of compulsory school teachers and principals to the evaluation findings be characterized, compared and related?

RQ1b – With reference to Hofer et al. (2020) and Matthews and Sammons (2004) are principals more positive than teachers in their attitudes and reactions?

RQ2 – On the basis of Ehren et al. (2013), are the mechanisms *accepting feedback, setting expectations,*

stakeholder involvement, and *capacity to implement improvements* positively correlated with *changes in internal evaluation, teaching* and *leadership practices*?

Method

Sampling and data collection

This section reports the results of an online survey of the perception of principals and teachers regarding external school evaluation and its effect. As previously stated, 2013 marked the beginning of the cyclical external school evaluation covered by this research. In May 2016, surveys were sent to teachers (N = 550) and principals (N = 22) in all the schools that were evaluated in 2013, 2014 and the first half of 2015, a total of 22 schools representing 13% of all compulsory schools in Iceland at that time. The online software Survey Monkey was used to collect data.

The response rate among principals was 100% and among teachers 56% (n = 309). Of these 309 respondents, 17% (n = 51) were not employed as teachers in the school at the time the external evaluation took place. This group answered only a small portion of the questionnaire that is not relevant to the focus of this paper. Therefore, the following analysis of the 22 schools is based solely on responses from those teachers (n = 258) who were employed at the time of the external evaluation.

Instrument

This study is based on questions from the principal and teacher surveys; see the question overview in Tables A1-A3 in Appendix 1. As the overview shows, some of the questions were the same for both teachers and principals, while some questions were designed to be answered by only one of the two groups. The teachers were asked 28 questions and the principals 30 questions. All questions utilized a Likert-type scale for responses.

Context

The surveys were conducted to evaluate the implementation and effect of the external evaluation project on behalf of the Ministry of Education, Science and Culture, the Icelandic Association of Local Authorities and the Directorate of Education. Consequently, there is a risk of social desirability bias because those who answered the questions were responding to a questionnaire sponsored by education authorities, which may have led them to express overly positive statements. It is difficult to assess the impact this may have had on the results.

Table 1. The scales, their reliability and inter-item correlations derived from both the teachers' and principals' surveys.

Scales		Number of items	Question numeral	Cronbach's alpha	Average inter-item correlation
(1) Acceptance of evaluation feedback	Teachers	4	(Q1–Q4)	0.76	0.44
	Principals	4		0.66	0.35
(2) Setting expectations – quality criteria	Teachers	3	(Q5–Q7)	0.7	0.44
	Principals	3		0.76	0.52
(3) Stakeholder involvement	Principals	7	(Q8–Q14)	0.83	0.41
(4) Teacher participation in working with feedback	Teachers	2	(Q21–Q22)	0.76	0.62
(5) Attitudes towards internal evaluation	Teachers	5	(Q32–Q36)	0.82	0.48
	Principals	5		0.77	0.41
(6) Experienced hindrances that prevented improvements	Teachers	9	(Q23–Q31)	0.79	0.3
	Principals	9		0.81	0.34
(7) External evaluation leading to change in teaching practices	Teachers	4	(Q16–Q19)	0.7	0.37

Data analysis

Based on the self-report surveys, a descriptive, exploratory and multilevel correlational analysis was conducted in four steps:

(1) To construct scales based on composite variables for the analysis, exploratory factor analysis was performed on the teacher data. Table 1 shows the main factors that teachers and principals were asked about and the number of questions/items belonging to each factor. The 'question numeral' column shows each question's number in Tables A1-A3 (Appendix 1). In addition to what is shown in Table 1, teachers and principals were asked if the external evaluation has led to changes in internal evaluation, and principals were also asked about changes in leadership practices following the evaluation. Cut-off values of 0.3 were used as a minimum for significant factor loadings (Kline, 1994). The principals' data were based on too few subjects for us to be able to trust reliable correlation coefficients in factor analysis (Kline, 1994). However, we considered it important to compare the answers of teachers and principals; therefore, we combined variables in the principals' data in the same way as in the teachers' data but removed variables in both datasets when their removal led to an increase in Cronbach's alpha. We calculated Cronbach's alpha for each factor variable to evaluate the reliability of the instrument. Further information on the factor-loading of questions can be found in Tables A1-A3 in Appendix 1.

Table 1 shows that 10 of the 11 proposed scales are reliable according to Cronbach's alpha's consensus threshold ($\alpha \geq 0.7$) (Hinton, 2014). The scale of the variable *acceptance of evaluation feedback* in the principal's survey has a marginal reliability coefficient ($\alpha = 0.66$). However, as the average inter-item correlation is acceptable, we decided to keep the scale in the analyses to make the teacher and principal scales comparable; as can be inferred, the small number of respondents in the principals' survey is the reason. All variables in the study are latent variables except

the two variables *change in internal evaluation* and *change in leadership practices*, each of which consists of one item (see Appendix 1, Q15 and Q20). Descriptive statistics (mean and SD) were calculated for all measures.

(2) The second step in the analysis was to compare the responses of teachers and principals to four variables. To take into account the multilevel structure of the data, we used a linear mixed model to analyse the difference by including a random effect for school.

(3) Correlation coefficients were calculated to describe the relationship between the variables. To account for the clustering in the data from teachers within a school we used a mixed model with school as a random effect. Then the regression slope between standardized variables was used as an estimator for Pearson's correlation coefficient. We used the method described in Rosner and Glynn (2017) for maximum likelihood estimation of the Pearson correlation in the clustered data setting. Spearman's Rho rank was calculated for the principal's data. For interpreting the magnitude of the correlation, the following criteria were used: correlations with magnitudes 0.0–0.1 were considered non-existent, 0.1–0.3 were small, 0.3–0.5 moderate, 0.5–0.7 large and 0.7–0.9 very large (Hopkins, 2002).

(4) Finally, to deepen the correlation analysis, a multivariable regression analysis was performed on the teachers' data. We used a linear mixed model to investigate whether two of the three assumed mechanisms of school inspection (for which information was available), according to the theoretical model, could significantly explain together *changes in internal evaluation* (model 1) and *teaching practices* (model 2), more specifically, the explanatory variables a*ccepting feedback* and s*etting expectations*. In addition, the explanatory variable *teacher participation* was added to the model. Normality of residuals was checked by visual inspection of the QQ plots, which revealed that the observations approximately followed a straight line, indicating that the

assumption of normally distributed residuals was satisfactory.

All results were obtained by multilevel analyses, where teachers' and principals' answers were modelled at level I and schools were modelled at level II. Exceptions were descriptive analyses and correlation analyses based on principal's data. The variables were standardized, and only standard variables were used in calculations other than descriptive statistics. Analyses were carried out in SPSS27, except for linear mixed model analysis, which was performed in R.

Findings

The findings are presented in line with the key factors discussed in the theoretical framework and guided by the model in Figure 1, as follows: *acceptance of the feedback, setting expectations, stakeholder involvement, schools' capacity to implement improvements,*

research questions are answered in parallel. Descriptive results and correlations for the teachers' and principals' surveys, which are noted below, are summarized in Tables 2 and 3. The results of the linear mixed model regression analysis are discussed and summarized in Table 4.

Acceptance of evaluation feedback. The analysis is based on scale 1 in Table 1. The answers of teachers (M = 3.7, SD = 0.68) and principals (M = 4.1, SD = 0.59) indicate a positive attitude amongst both groups towards the external evaluation and its usefulness for the school (see, Tables 2 and 3). However, the t-test of the fixed effect, from the linear mixed model with a random effect for school, comparing principals to teachers, indicated that principals were significantly more positive than teachers ($t = -2.14$, $p < 0.05$). The highest significant relationship in the teachers' correlation analysis (see, Table 2) was observed for *acceptance of evaluation feedback* and

Table 2. Teachers' survey: mean, standard deviation and estimator for the Pearson's correlation coefficient between the different scales.

	Acceptan	SetExpe	TeacPartip	AttituIE	ExpHindra	ChangeIE	ChangeTe
M	3.7	3.4	3.8	4.1	2.8	3.5	3.5
SD	0.68	0.74	0.92	0.59	0.8	1.02	0.81
Acceptan – Acceptance of evaluation feedback (4)	1						
SetExpe – Setting expectations: criteria knowledge (3)	0.32**	1					
TeacPartip – Teacher participation in working with feedback (2)	0.72**	0.37**	1				
AttituIE – Attitudes towards internal evaluation (5)	0.48**	0.23**	0.34**	1			
ExpHindra – Experienced hindrances preventing improvements (9)	−0.34**	−0.11	−0.28**	−0.18**	1		
ChangeIE – External evaluation changes internal evaluation (1)	0.52**	0.35**	0.56**	0.46**	−0.22**	1	
ChangeTe – External evaluation changes teaching practices (4)	0.55**	0.23**	0.45**	0.29**	−0.20**	0.37**	1

Note: In parentheses following variable labels are the number of Likert variables that comprise the scale. Answer categories: 1 = strongly disagree; 2 = disagree; 3 = cannot answer; 4 = agree; 5 = strongly agree. Note: *significant at p < .05 level; **significant at p < .01 level. N = 258. The regression slope from a mixed model between standardized variables was estimator for the Pearson's correlation coefficient.

Table 3. Principals' survey: mean, standard deviation and Spearman's Rho rank order correlation coefficient between the different scales.

	Acceptan	SetExpe	StakSuPr	AttituIE	ExpHindra	ChangeIE	ChangeLe
M	4.1	3.8	3.25	4.4	2.3	3.4	3.7
SD	0.59	0.72	0.93	0.4	0.81	1.37	0.94
Acceptan – Acceptance of evaluation feedback (4)	1						
SetExpe – Setting expectations: criteria knowledge (3)	0.11	1					
StakSuPr – Stakeholder support and pressure after evaluation (7)	0.49*	0.29	1				
AttituIE – Attitudes towards internal evaluation (5)	0.14	0.10	0.20	1			
ExpHindra – Experienced hindrances preventing improvements (9)	−0.54*	−0.02	−0.69**	−0.10	1		
ChangeIE – External evaluation changes internal evaluation (1)	0.00	0.50*	−0.02	0.09	0.36	1	
ChangeLe – External evaluation changes leadership practices (1)	0.61**	−0.05	0.15	0.12	−0.14	−0.27	1

Note: In parentheses following variable labels are the number of Likert variables that comprise the scale. Answer categories: 1 = strongly disagree; 2 = disagree; 3 = cannot answer; 4 = agree; 5 = strongly agree. Note: *significant at p < .05 level; **significant at p < .01 level. N = 22.

and *changes in internal evaluation, teaching and leadership practices.* Slight changes in the order of topics in the results section compared to the theoretical framework section are due to the focus in the first research question. Furthermore, in the theoretical framework, the focus is on improvement, but as the perceived changes in this study may not be equivalent to school improvements, the term 'change' will be used instead of improvement. In the following, the

teacher participation in working with evaluation feedback ($r = 0.72$, $p < 0.01$). Furthermore, a strong significant correlation was found between teachers' *acceptance* and *change in teaching practices* ($r = 0.55$, $p < 0.01$) and *change in internal evaluation* ($r = 0.52$, $p < 0.01$). The same is true for principals (see, Table 3), where a strong significant correlation between *acceptance* and *changes in leadership practices* ($\rho = 0.61$, $p < 0.01$) was found. However, no

correlation of *acceptance* with *change in internal evaluation* was identified, which is undeniably an unexpected result given the strong correlation between the variables in the teachers' data.

Setting expectations – knowledge and use of quality criteria. The analysis is based on scale 2 in Table 1. A relatively low mean and some variance (M = 3.4, SD = 0.74) in teachers' answers (see, Table 2) indicate that, although the majority of them know and use the quality criteria from the external evaluation to some extent, this does not apply to all. The t-test of the fixed effect showed that principals reported significantly (t = −2.76, $p < 0.01$) more knowledge about and use of the criteria (M = 3.8, SD = 0.72). A significant correlation – moderate in the teachers' answers ($r = 0.35$, $p < 0.01$) while moderate to strong in the principals' answers ($\rho = 0.50$, $p < 0.05$) – was found between *setting expectations* and *change in internal evaluation* (see, Tables 2 and 3). *Setting expectations* has, according to the teachers' answers, a small but significant correlation with *changes in teaching practices* ($r = 0.23$, $p < 0.01$), but, according to the principals' answers, it has no correlation with *changes in leadership practices*.

Stakeholder support and pressure. Scale 3 in Table 1 is the basis of the analysis. Questions about *stakeholder involvement* were included only in the principals' questionnaire and concerned only the support and pressure applied by educational authorities. The mean, being non-significantly different from the midpoint in the principals' answers (M = 3.25, SD = 0.93), implies that, on average, they do not consider *stakeholder involvement* to be high (see, Table 3). Nevertheless, there was a notable variance in principals' answers regarding this aspect, which indicates that their experiences differ in this respect. This inference is supported by the highly significant negative correlation observed between *stakeholder involvement* and the principals' *experiences of hindrances that prevented improvements* ($\rho = -0.69$, $p < 0.01$), which indicates that there is an important relationship between the principals' experiences of *stakeholder involvement* and their feeling of support or pressure from their authorities. *Stakeholder involvement*, however, seems to have no correlation with *change in internal evaluation* or *change in leadership practices*. Thus, despite the very clear correlation noted above, this support does not seem to have been clearly translated into action.

School's capacity to implement improvements. Three scales in Table 1 were analysed to measure this aspect: scale 4, scale 5 and scale 6. Questions about *teacher participation* were included only in the teachers' questionnaire. Teachers generally seem to believe that they have been involved in working with the external evaluation feedback and making decisions regarding improvements (M = 3.8,

SD = 0.92), but the variability in their answers indicates significant differences between respondents, with perhaps the most notable difference in the ranking of the scales (see, Table 2). Moderate-to-strong significant correlations were found between *teacher participation* and other variables measured. It is clear that for both principals and teachers, the most positive attitude emerges from the scale focusing on the *attitude towards internal evaluation*, which also has the smallest variance. Nevertheless, the t-test of the fixed effect indicated that principals (M = 4.4, SD = 0.4) were significantly (t = −2.38. $p < 0.05$) more positive than teachers (M = 4.1, SD = 0.59). Furthermore, both teachers (M = 2.8, SD = 0.8) and principals (M = 2.3, SD = 0.81) were least concerned about *experienced hindrances* (i.e. lack of time, support, knowledge and resources) that might have made improvements less feasible, although there is some variability in their responses (see, Tables 2 and 3). Teachers, however, generally experienced significantly more hindrances than principals did (t = 3.12, $p < 0.01$).

External evaluation changing teaching and leadership practices and internal evaluation. This analysis is based on scale 7 in Table 1 and two single-item variables: one for principals about *changes in leadership practices* and one for each group of teachers and principals about *changes in internal evaluation*. Regarding *changes in teaching practices*, the average for the teachers' responses and the high standard deviation (M = 3.5, SD = 0.81) imply that, although some of them have made changes to their teaching methods, this does not apply to all teachers (see, Table 2). The results are similar for *changes in internal evaluation* although with a higher standard deviation (M = 3.5, SD = 1.02). Similar results were observed with principals (see, Table 3) regarding the effect of the evaluation on *changes in leadership practices* (M = 3.7, SD = 0.94) and *changes in internal evaluation* (M = 3.4, SD = 1.37), and the standard deviations are also high.

To study further the second research question regarding the contribution of key variables, as guided by the model shown in Figure 1, on both *changes in internal evaluation* and *changes in teaching practices*, according to the teachers' data, we used a linear mixed model analysis, where teacher's answers were modelled at level I and schools were modelled at level II (see, Table 4). The two available mediating variables from the theoretical model were added to the analysis – *acceptance* and *setting expectations* – as well as the *teacher participation* variable. Due to a lack of a significant relationship between variables in the principals' data, it was decided not to go any further with the analysis for the principals. Furthermore, there was no basis for combining the teacher and principal data for this analysis as the

Table 4. Summary of linear mixed model analysis based on teacher's data for changes in 1) internal evaluation and 2) teaching practices.

Fixed effects variables	Model 1: Change in internal evaluation				Model 2: Change in teaching practices			
	β (SE)	95% CI	t-value	Cond. R²	B (SE)	95% CI	t-value	Cond. R²
				0.38				0.36
Intercept	−0.003 (0.074)	[−0.149, 0.142]	−0.042		−0.008 (0.052)	[−0.112, 0.094]	−0.163	
Acceptance	0.235 (0.080)	[0.076, 0.393]	2.914**		0.575 (0.079)	[0.419, 0.731]	7.236***	
Setting expectations	0.177 (0.058)	[0.063, 0.292]	3.044**		0.017 (0.057)	[−0.095, 0.130]	0.309	
Teacher participation	0.289 (0.082)	[0.128, 0.450]	3.518***		0.034 (0.080)	[−0.123, 0.192]	0.426	

Notes: *p < 0.05; **p < 0.01; ***p < 0.001. Schools (N = 22) were treated as random effects subject groupings. Only standard variables were used in the analyses.

principal data did not include data on all key variables in the model.

Table 4 presents the results of the linear mixed model for *changes in internal evaluation* (model 1) and *changes in teaching practices* (model 2). In accordance with the assumptions of the theoretical model in Figure 1 and the hypothesis of this study, the analysis showed that *change in internal evaluation* was significantly increased as a function of *acceptance* and *setting expectations*. *Teacher participation* also significantly impacted *changes in internal evaluation* (model 1, Table 4). In model 2 (Table 4), *change in teaching practices*, significantly increases as a function of *acceptance*. However, contrary to the hypothesis, neither *setting expectations* nor *teacher participation* proved to have significant impact on *changes in teaching practices* when *acceptance was* accounted for. The proportion of total variance explained through both fixed and random effects was 38% for model 1 and 36% for model 2. In neither model was the random effect for school statistically significant, indicating a small amount of clustering within schools relative to the between variability among all teachers.

Discussion

Guided by the model shown in Figure 1, the aim of the present study was to describe and understand how external evaluation might influence certain aspects of school activities as perceived by both teachers and principals. This meant first examining the reception and attitude among professional school staff towards the external evaluation findings and determining whether there was a difference between the answers provided by teachers and those provided by principals in that context (RQ1a and RQ1b). Secondly, there was an exploration of which mechanisms of external school evaluation the school professionals perceived as contributing to changes in internal evaluation and to their own practices—in particular, to assess whether external evaluation is seen to contribute to changes through (1) the expectations expressed in the quality criteria, (2) the feedback the school was

receiving in the evaluation report, (3) the support and pressure the school was receiving from its external stakeholders, and (4) the capacity to implement the improvements recommended (i.e. participation of teachers, the attitude in schools towards internal evaluation, and the hindrances experienced in implementing improvements) (RQ2). We used survey data to collect the views of the teachers and principals in 22 recently evaluated compulsory schools.

To answer the first research question (RQ1a), we infer participants' receptivity and attitudes towards external evaluation findings. According to the theoretical framework (Ehren et al., 2013), it is expected that *accepting* feedback is an important mechanism for driving changes at schools. Both principals and teachers who responded in this study expressed a positive attitude towards the evaluation feedback, and they considered it useful for school improvement. Consistent with studies from Austria, Germany and Switzerland (Dedering & Müller, 2011; Kemethofer et al., 2017; Schweinberger et al., 2017), this positive attitude might reflect the relatively 'low-stake' and improvement-focused approach used in external evaluation in Iceland. Even though the results are publicly available – considered one of the characteristics of 'high-stakes' inspection systems (Altrichter & Kemethofer, 2015) – there is no ranking of schools based on the results or any threats of punitive action for failure to address identified needs. As Ehren, Jones et al. (2016) points out, an important condition for effective school evaluation, is an environment based on trust, where school staff are open about their performance and improvements. Such an environment is also less likely to lead to undesired side effects such as stress, window dressing and narrowing the curriculum (Ehren, Jones et al., 2016; Jones et al., 2017). Furthermore, this study indicates that teachers who were involved in working with the evaluation feedback were more accepting of and motivated to use the feedback, reported greater knowledge and use of the quality criteria and experienced fewer hindrances (i.e. those

characterized by lack of time, support, knowledge and resources) in regard to implementing the improvements than teachers who were less involved. They also reported more changes in their teaching practices and in their schools' internal evaluations. This finding supports the importance of teacher participation and collaboration and might encourage school leaders, policy-makers and administrators to emphasize a collaborative atmosphere within schools, which in turn may affect the developmental culture of the school.

The second part (RQ1b) of the initial research question was about differences between the reactions of teachers and principals. In line with the results of Hofer et al. (2020) and Matthews and Sammons (2004), the results of this study indicated that principals were generally more positive than teachers—they reported significantly higher scores than the teachers for *acceptance of evaluation feedback, setting expectations* and *attitude towards internal evaluation*, and significantly lower scores for *experienced hindrances* that prevented improvements. However, no difference was found in the position of these two groups as to whether the external evaluation had led to *changes in the internal evaluation*. We also noted that both groups showed a positive attitude towards internal evaluation (the highest for both groups and higher for principals), but the relationship to other variables was very different. For the principals, this variable did not correlate significantly with any other variable, whereas for the teachers, it correlated significantly to all the other variables. Both the clear positive attitude of the principals and this lack of correlation may indicate an inflated positive assessment by the principals on this score due possibly to social desirability bias as noted in the method section. Many studies on the effect of external evaluation are based solely on the attitudes of principals (see e.g. Dedering & Müller, 2011; Ehren, Gustafsson et al., 2015; Kemethofer et al., 2017). Our results highlight the importance of gathering information from diverse sources within the school to gain a more holistic picture of the actual evaluation effect.

With respect to the second main research question (RQ2) regarding the mechanisms that school professionals perceive as contributing to changes in internal evaluation, teaching and leadership practices following external evaluation, the mechanism of *feedback acceptance* will be discussed first. The findings revealed a strong significant positive correlation between *feedback acceptance* and *changes in teaching* and *leadership practices*. When it comes to *changes in internal evaluation*, a strong significant positive correlation was found in teachers' answers, but no correlation was revealed in principals' answers. Linear

mixed model analysis based on the teachers' data showed that *feedback acceptance* could be seen as a significant predictor of improvements in both *teaching practices* and *internal evaluation*, taking into account *settings expectation* and *teacher participation*. This is in accordance with the hypothesized effect according to Ehren et al. (2013) but contrary to the results reported by Gustafsson et al. (2015) and Ehren, Perryman et al. (2015). As pointed out by Altrichter and Kemethofer (2015), feedback acceptance is more important in 'low-stake' evaluation approaches than in countries with 'higher-stake' approaches, which may explain this difference.

Setting expectations through standards and criteria for quality education is another causal mechanism of the theoretical framework (Ehren et al., 2013). In line with the study by Ehren, Perryman et al. (2015), which suggested that *setting expectations* is an essential factor for promoting improvements, our findings revealed that quality standards contribute to *changes in internal evaluation* but not to *changes in teaching* or *leadership practices*. Linear mixed model analysis indicated that, even though the effect of *setting expectations* was not as great as the effect of *feedback acceptance* on *change in internal evaluation*, it was persistent when controlling for other variables. *Setting expectations*, furthermore, had a moderately significant correlation with *feedback acceptance* in teachers' answers but an insignificant correlation in principals' answers. However, given the rather low means regarding knowledge and use of the quality criteria in our results, especially among teachers, this suggests that more should be done to promote the criteria and emphasize that schools should make use of them, e.g. in their internal evaluation. Increasing knowledge of the quality standards could increase the desired development in schools.

Stakeholder involvement is the third causal mechanism (Ehren et al., 2013). Information on this aspect was available only from the principals' questionnaire. Results revealed that *stakeholder involvement* had a strong significant correlation with *feedback acceptance* but no statistically significant correlation with *setting expectations*. Furthermore, no correlation was found between *stakeholder involvement* and *changes in internal evaluation* or *leadership practices*. Kemethofer et al. (2017) and Gustafsson et al. (2015) came to a similar conclusion in their research; they did not identify any effect of *stakeholder involvement* on improvement. Gustafsson et al. (2015) concluded that the effect of *stakeholder involvement* was indirect and consisted mainly of encouraging schools to accept evaluation feedback and quality criteria. There is evidence of this in our study as well in terms of *feedback acceptance* but not in terms of *setting expectations*. It is important,

however, to stress that, based on our data, it is not possible to establish any causal relationship between *stakeholder involvement* and other variables. Our results can, therefore, be considered only suggestive in this respect. Nevertheless, there is a reason to draw attention to the strong negatively significant association between *stakeholder involvement* and *experienced hindrances* that prevent improvements. This highlights the important role of the support of education authorities when schools are implementing improvements. By contrast, principals' answers imply that they do not consider *stakeholders involvement* to be high, which gives a reason to encourage the education authorities to do better in that regard. Schools need support for using feedback, adapting to expectations and addressing their weaknesses. Thus, the authorities should make an effort to create a more collaborative atmosphere around the schools as well as within them.

Schools' capacity to implement improvements is, in this study, suggested as the fourth mechanism of change after external evaluation (Ehren, 2016). A number of scholars have made strong argument for the importance of teacher participation in decision-making (see e.g. Ehren, Perryman et al., 2015; Kools & Stoll, 2016; Schildkamp & Lai, 2013). The results of this study support this argument since the *participation of teachers* seems to be a crucial factor with moderate-to-strong statistically significant associations with other variables in the study. More aspects regarding schools' capacity were under consideration, including *attitude towards internal evaluation*, which researchers have pointed out as essential if the internal evaluation is to be improved (O'Brien et al., 2015; Schildkamp & Ehren, 2013). Analysis of the teachers' data in this study supports this. However, the principals' data shows a total lack of an association between *attitudes towards internal evaluation* and *changes in internal evaluation*, which was unexpected and not in accordance with the studies noted above. This study does not give a reason to draw any conclusions about these findings and therefore calls for further research. The final factor reflecting schools' capacity in this study is *experienced hindrances*. Neither teachers nor principals seem to experience major hindrances to the improvement process, and correlations with *changes in internal evaluation* and *teaching and leadership practices* are either small or insignificant.

Summarizing the findings from the perspective of the research questions posed, the conclusion for RQ1a and RQ1b is very clear, but for RQ2 the answer is more complex. As for RQ1a and RQ1b, there is a positive attitude towards external school evaluation in Iceland among both teachers and principals, although principals are more positive towards it than teachers are. The evaluation is based on a 'low-stake' and improvement-oriented approach, which may influence this positive attitude. Furthermore, teachers who were involved in working with the evaluation feedback were more accepting and motivated to use the feedback than teachers who took little or no part in the decision-making. As for RQ2, different answers were obtained from the teachers and principals related to influence on internal evaluation. The scales a*cceptance, setting expectations* and *teacher participation* were all found in linear mixed model analysis to be significant predictors for perceived *changes in internal evaluation* in the teachers' data. Furthermore, *changes in internal evaluation* were positively correlated to *attitude towards internal evaluation* and negatively correlated to *experienced hindrances*. In the principals' data, the only scale that had a significant correlation with perceived *changes in internal evaluation* was *setting expectations*. No relationship was measured in relation to *acceptance, stakeholder involvement, attitude towards internal evaluation* or *experienced hindrances*. Regarding influence on *change in teaching and leadership practice*, only *acceptance* significantly explained perceived *changes in teaching practices* in linear mixed model analysis, although the scales *setting expectations* and *teacher participation* had small-to-moderate significant correlations with *changes in teaching practices*. However, only *acceptance* was found to be significantly related to perceived *changes in leadership practices*; all other relationships were found to be small and not significant. The different and sometimes paradoxical answers of teachers and principals were unexpected, but it should be kept in mind that the results were based on answers from 258 teachers but only 22 principals, which may have had an effect.

In accordance with the theoretical framework in Figure 1 and the hypothesis of this study, the results suggest the importance of *acceptance of the evaluation feedback* and *setting expectations* through quality standards. However, contrary to the hypothesis, *stakeholder involvement* did not prove to be a strong determinant of change following the external evaluation as perceived by principals, given the form the evaluation has in Iceland. *School's capacity* to implement improvements is positively correlated with the changes in the teacher's data, however not in the principal's data.

The study has several limitations that need to be addressed. First, it set out to explore what professional personnel thought about the implications of the external evaluation, primarily on the conduct of the internal evaluation but also on teaching and leadership practices. This limits the inferences that can be made about actual changes. The effort to

triangulate data, including both principals' and teachers' insights, is our way of reducing this limitation. A second limitation concerns the generalizability of the findings, which is limited by the fact that only 22 schools were included in the study. Furthermore, the validity of our findings may be restricted to the Icelandic school system, with its particular school-evaluation context. However, the aim of this study was to examine how mechanisms of external evaluation influence improvement and change in schools in the Icelandic system, and the data sources are highly relevant for a discussion of our questions to consider its unique characteristics and to provide a specific case of the influence of an evaluation process on school development.

Suggestions for policies, practices, and future research

This study highlights important issues and implications for policy development and future research. Regarding future research, the following suggestions are made. First, in view of the educational system in which the study is conducted, it clearly shows the crucial role of teachers' and principals' acceptance of evaluation feedback in order for them to accept its use and implementation to drive school improvement. Hence, since school improvement is one of the main aims of the external school evaluation, it is important to further investigate how and to what extent improvement actually takes place and to what extent it can be attributed to the evaluation effort. The perceived changes, clearly shown in the results, may not be equivalent to school improvement as envisaged by policy-makers and school authorities. A fuller understanding of the nature and extent of the changes actually taking place and how they may be facilitated is needed. The second guideline for future research relates to Iceland's participation over the years in Nordic cooperative efforts, and in many ways, the Icelandic education system has developed in a way that is similar to the education systems of other Nordic countries. Based on the theoretical framework used in this study, research has indicated that the Swedish government takes a 'medium'-to-'high-stake' approach in their external school evaluation (Altrichter & Kemethofer, 2015), while the Icelandic government takes a 'soft' governance approach. The theoretical framework by Ehren et al. (2013), inspiring the questions asked in this study, and future studies in other Nordic countries can offer important insights into the extent to which the external evaluation systems of these countries have developed in similar or different ways. The third call for future exploration is that this study could draw conclusions regarding only part of the underlying theoretical framework. If and how external evaluation affects school-improvement capacity and students' educational outcomes remains unexplored, thereby calling for further study. Lastly, future research should aim to be more independent of the system and should also include information from other groups participating in the process, as the model implies.

In turning attention to the present study's implications for policy and practice, the findings give rise to suggestions that could increase the impact of the external evaluation on school development. Given the formal status of the evaluation process, it would seem a key challenge to, first, better inform schools about what is expected of them in terms of the quality standards and, second, provide them with resources and support in the process of meeting these expectations and monitoring their progress through their internal evaluations. Our analysis suggests that collaboration, both within and around schools, is a central factor in the success of the external evaluation; hence, school leaders, policy-makers and administrators should consider strategies for its promotion.

Disclosure statement

ORCID

Jón Torfi Jónasson 🔘 http://orcid.org/0000-0001-7580-3033
Anna Kristin Sigurðardóttir 🔘 http://orcid.org/0000-0002-6713-1318
Thor Aspelund 🔘 http://orcid.org/0000-0002-7998-5433

References

Act on the directorate of education, (91/2015)

Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, *26*(1), 32–56. https://doi.org/10.1080/09243453.2014.927369

Behnke, K., & Steins, G. (2017). Principals' reactions to feedback received by school inspection: A longitudinal study. *Journal of Educational Change*, *18*(1), 77–106. https://doi.org/10.1007/s10833-016-9275-7

Compulsory School Act, (2008). https://www.government.is/media/menntamalaraduneyti-media/media/law-and-regulations/Compulsory-School-Act-No.-91-2008.pdf

Dedering, K., & Müller, S. (2011). School improvement through inspections? First empirical insights from Germany. *Journal of Educational Change*, *12*(3), 301–322. https://doi.org/10.1007/s10833-010-9151-9

Ehren, M. C. M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools: Describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, *25*(1), 3–43. https://doi.org/10.1007/s11092-012-9156-4

Ehren, M. C. M., Gustafsson, J. E., Altrichter, H., Skedsmo, G., Kemethofer, D., & Huber, S. G. (2015). Comparing effects and side effects of different school inspection systems across Europe. *Comparative Education*, 51(3), 375–400. https://doi.org/10.1080/03050068.2015.1045769

Ehren, M. C. M., Perryman, J., & Shackleton, N. (2015). Setting expectations for good education: How Dutch school inspections drive improvement. *School Effectiveness and School Improvement*, 26(2), 296–327. https://doi.org/10.1080/09243453.2014.936472

Ehren, M. C. M. (2016). Mechanism and contexts of effective inspections. In M. Ehren (Ed.), *Methods and modalities of effective school inspections* (pp. 111–139). Springer International Publishing.

Ehren, M. C. M., Eddy-Spicer, D., Bangpan, M., & Reid, A. (2016). School inspections in low- and middle-income countries: Explaining impact and mechanisms of impact. *Compare: A Journal of Comparative and International Education*, 47(4), 468–482 doi:10.1080/03057925.2016.1239188.

Ehren, M. C. M., Jones, K., & Perryman, J. (2016). Side effects of school inspection: Motivations, and contexts for strategic responses. In M. C. M. Ehren (Ed.), *Methods and modalities of effective school inspections* (pp. 87–106). Springer International Publishing.

Eurydice. (2015). *Assuring quality in education: Policies and approaches to school evaluation in Europe*. https://eacea.ec.europa.eu/national-policies/eurydice/content/assuring-quality-education-policies-and-approaches-school-evaluation-europe_en

Gustafsson, J.-E., Ehren, M. C. M., Conyngham, G., McNamara, G., Altrichter, H., & O'Hara, J. (2015). From inspection to quality: Ways in which school inspection influences change in schools. *Studies in Educational Evaluation*, 47, 47–57. https://doi.org/10.1016/j.stueduc.2015.07.002

Hanberger, A., Carlbaum, S., Hult, A., Lindgren, L., & Lundström, U. (2016). School evaluation in Sweden in a local perspective: A synthesis. Education Inquiry. 7(3), 349–371. https://doi.org/10.3402/edui.v7.30115

Hinton, P. R. (2014). *Statistics explained* (3rd ed.). Routledge.

HMIE. (2010). *Learning together: Lessons about school improvement. An HMIE report on how schools get better*. https://education.gov.scot/Documents/LearningTogetherLessonsSchoolImprovement.pdf

Hofer, S. I., Holzberger, D., & Reiss, K. (2020). Evaluating school inspection effectiveness: A systematic research synthesis on 30 years of international research. *Studies in Educational Evaluation*, 65. https://doi.org/10.1016/j.stueduc.2020.100864

Hopkins, W. G. (2002). *A scale of magnitudes for effect statistics*. New View of Statistics Web site. Retrieved October 18 from http://www.sportsci.org/resource/stats/effectmag.html

The Icelandic National Curriculum Guide for Compulsory Schools. (2011). *Mennta- og menningarmálaráðuneyti*. http://www.menntamalaraduneyti.is/utgefid-efni/namskrar/adalnamskra-grunnskola/

Jones, K. L., Tymms, P., Kemethofer, D., O'Hara, J., McNamara, G., Huber, S., Myrberg, E., Skedsmo, G., & Greger, D. (2017). The unintended consequences of school inspection: The prevalence of inspection side-effects in Austria, the Czech Republic, England Ireland the Netherlands Sweden and Switzerland. *Oxford Review of Education*, 43(6), 805–822 doi:10.1080/03054985.2017.1352499

Kemethofer, D., Gustafsson, J. E., & Altrichter, H. (2017). Comparing effects of school inspections in Sweden and Austria. *Educational Assessment, Evaluation and Accountability*, 29(4), 319–337. https://doi.org/10.1007/s11092-017-9265-1

Kline, P. (1994). *An easy guide to factor analysis*. Routledge.

Kools, M., & Stoll, L. (2016). *What makes a school a learning organisation?* (OECD Education Working Paper, Issue. O. Publishing. http://www.oecd-ilibrary.org/docserver/download/5jlwm62b3bvh-en.pdf?expires=1505943700&id=id&accname=guest&checksum=40BFFD79E0E160911BF991D77832C792

Matthews, P., & Sammons, P. (2004). *Improvement through inspection: An evaluation of the impact of Ofsted's word*. Ofsted and the Institute of Education University of London. http://dera.ioe.ac.uk/4969/3/3696.pdf

O'Brien, S., McNamara, G., & O´Hara, J. (2015). Supporting the consistent implementation of self-evaluation in Irish post-primary schools. *Educational Assessment, Evaluation and Accountability*, 27(4), 377–393. https://doi.org/10.1007/s11092-015-9218-5

Ólafsdóttir, B. (2016). Tilurð og þróun ytra mats á Íslandi frá 1991 til 2016 [The origin and development of external evaluation in Iceland, 1990–2016]. *Netla: Veftímarit um uppeldi og menntun*. http://netla.hi.is/greinar/2016/ryn/14_ryn_arsrit_2016.pdf

Organisation for Economic Cooperation and Development [OECD]. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. https://doi.org/10.1787/9789264190658-en

Penninckx, M. (2017). Effects and side effects of school inspections: A general framework. *Studies in Educational Evaluation*, 52, 1–11. https://doi.org/10.1016/j.stueduc.2016.06.006

Rosner, B., & Glynn, R. J. (2017). Estimation of rank correlation for clustered data. *Statistics in Medicine*, 36 (14), 2163–2186 doi:10.1002/sim.7257.

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education. Challenges and opportunities* (pp. 177–191). Springer. https://doi.org/10.1007/978-94-007-4816-3

Schildkamp, K., & Ehren, M. (2013). From "intuition"- to "data"-based decision making in Dutsch secondary schools? In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education. Challenges and opportunities* (pp. 49–67). Springer. https://doi.org/10.1007/978-94-007-4816-3

Schweinberger, K., Quesel, C., Mahler, S., & Höchli, A. (2017). Effects of feedback on process features of school quality: A longitudinal study on teachers' reception of school inspection of Swiss compulsory schools. *Studies in Educational Evaluation*, 55, 75–82. https://doi.org/10.1016/j.stueduc.2017.07.004

Sigurjónsdóttir, B., Ólafsdóttir, B., & Jónsdóttir, Þ. B. (2012). Viðmið um gæði í skólastarfi fyrir ytra mat á grunnskólum [Quality criteria for external evaluation of compulsory schools]. http://www.samband.is/media/mat-og-rannsoknir-a-skolastarfi/Vidmid-um-gaedi-i-skolastarfi.pdf

Verhaeghe, G., Vanhoof, J., Valcke, M., & Van Petegem, P. (2010). Using school performance feedback: Perceptions of primary school principals. *School Effectiveness and School Improvement*, 21(2), 167–188. https://doi.org/10.1080/09243450903396005

Appendix. Overview of questionnaire items, latent and single-item variables, and factor loadings

Table A1. Factor loadings for the mediating variables according to the theoretical framework: Acceptance, setting expectations, stakeholder involvement.

Questions/Items:	Acceptance	Setting expectations	Stakeholder involvement
Q1. The findings reflect the strengths and weaknesses of the school.	0.595 (T) 0.766 (P)		
Q2. The findings were useful for school development.	0.844 (T) 0.598 (P)		
Q3. The findings led to actions aimed to promote professional development.	0.775 (T) 0.747 (P)		
Q4. Improvement actions had a positive effect on practices in the school.	0.825 (T) 0.746 (P)		
Q5. I know the criteria and indicators used for the external evaluation of schools.		0.813 (T) 0.808 (P)	
Q6. The criteria and indicators are taken into consideration in the school's internal evaluation.		0.761 (T) 0.858 (P)	
Q7. As a teacher/principal, I utilize the criteria and indicators to improve my own practice.		0.799 (T) 0.805 (P)	
Q8. The local education authorities supported the preparation of the improvement plan following the external evaluation.			0.723 (P)
Q9. The local authorities monitor the implementation of the improvement plan.			0.618 (P)
Q10. We have formally announced to the local authorities our progress in implementing the improvement plan.			0.792 (P)
Q11. The school council supported the preparation of the improvement plan following the external evaluation.			0.624 (P)
Q12. The school council monitors the implementation of the improvement plan.			0.726 (P)
Q13. We have formally announced to the school council our progress in implementing the improvement plan.			0.890 (P)
Q14. The Ministry follow-up with the implementation of the improvement plan has facilitated its implementation.			0.542 (P)

Note: Between parentheses: T = teachers and P = principals

Table A2. Factor loadings on factors classified as output variables in this study: External evaluation changes internal evaluation, teaching practices and leadership practices.

Questions/Items:	External evaluation changes internal evaluation	External evaluation changes teaching practices	External evaluation changes leadership practices
Q15. The findings of the external evaluation led to improvements in the school's internal evaluation.	Single item Answered by principals and teachers		
Q16. The external evaluation had the effect that I now consider my own teaching practices.		0.754 (T)	
Q17. The findings of the external evaluation led to a revision of teaching methods at the school.		0.620 (T)	
Q18. Being visited by an evaluator and receiving his/her feedback on the lesson was useful for my professional development.		0.692 (T)	
Q19. I did utilize the feedback from the evaluator to improve my teaching practices.		0.833 (T)	
Q20. The findings of the external evaluation led to improved leadership in the school.			Single item Answered only by principals

Note: Between parentheses: T = teachers and P = principals

Table A3. Factor loadings on variables that reflect schools' capacity for improvement at the time of external evaluation.

Questions/Items:	Teacher participation	Experienced hindrances	Attitude towards internal evaluation
Q21. Teachers participated in formal discussions on improvement following the external evaluation.	0.899 (T)		
Q22. The principal encouraged teachers to utilize the external evaluation findings for improvement.	0.899 (T)		
Q23. It was difficult to find suitable ways to make improvements.		0.512 (T) 0.548 (P)	
Q24. The daily workloads of teachers/principal are extensive.		0.496 (T) 0.386 (P)	
Q25. Teachers' collective agreement imposes restrictions.		0.561 (T) 0.573 (P)	
Q26. Teachers have limited opportunities for professional development.		0.687 (T) 0.666 (P)	
Q27. There is a lack of support by the municipality for principals' and teachers'professional development.		0.749 (T) 0.895 (P)	
Q28. There is a lack of support by the municipality for improvements.		0.756 (T) 0.851 (P)	
Q29. Teachers lack knowledge to utilize the evaluation findings for improvements.		0.619 (T) 0.789 (P)	
Q30. The principal lacks knowledge to utilize the evaluation findings for improvements.		0.599 (T) 0.670 (P)	
Q31. There is a lack of time to work with the evaluation findings.		0.500 (T) 0.307 (P)	
Q32. I believe internal evaluation is important for school development.			0.792 (T) 0.849 (P)
Q33. I believe internal evaluation is necessary to monitor the quality of the school's work.			0.784 (T) 0.864 (P)
Q34. There is a positive attitude at the school towards conducting an internal evaluation.			0.739 (T) 0.445 (P)
Q35. The general view of the school staff is that internal evaluation is important for school development.			0.757 (T) 0.595 (P)
Q36. Knowledge of the strengths and weaknesses of the school has increased through the school's internal evaluation.			0.740 (T) 0.845 (P)

Note: Between parentheses: T = teachers and P = principals