Studies in Educational Evaluation 74 (2022) 101181

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Use and impact of external evaluation feedback in schools

Björk Ólafsdóttir [a,b,*], Jón Torfi Jónasson [a,1], Anna Kristín Sigurðardóttir [a,2]

[a] University of Iceland, School of Education, Iceland
[b] Directorate of Education in Iceland, Iceland

ARTICLE INFO

Keywords:
Evaluation utilisation
School-based evaluation
School improvement
External evaluation

ABSTRACT

Past findings concerning whether and how feedback from external evaluations benefit the improvement of schools are inconsistent and sometimes even conflicting, which highlights the contextual nature of such evaluations and underscores the importance of exploring them in diverse contexts. Considering that broad international debate, we investigated the use and impact of feedback from external evaluations in compulsory schools in Iceland, particularly as perceived by principals and teachers in six such schools. A qualitative research design was adopted to examine changes in the schools made during a 4–6-year period following external evaluations by conducting interviews with principals and teachers, along with a document analysis of evaluation reports, improvement plans and progress reports. The findings reveal that feedback from external evaluations has been used for instrumental, conceptual, persuasive and reinforcement-oriented purposes in the schools, albeit to varying degrees. According to the principals and teachers, the improvement actions presented in the schools' improvement plans were generally implemented or continue to be implemented in some way, and the changes made have mostly been sustained.

1. Introduction

With the decentralisation of education systems in Europe in recent decades, decision-making regarding schools has largely been transferred from central governments to local authorities and the schools themselves (Hofer et al., 2020; Organisation for Economic Cooperation and Development (OECD), 2013). In Iceland the municipalities took over the operation of compulsory schools in 1996, and concurrently the professional responsibility of principals was increased (Ólafsdóttir, 2016). Although the growing autonomy of schools has afforded them some freedom to implement their own solutions and practices, decentralisation has also heightened the emphasis on the external evaluation of schools in order to hold them accountable for their decisions and to monitor whether they are operating in compliance with national legislation and policy. Aside from monitoring schools and ensuring their accountability, most evaluation systems aim at improving the quality of education in schools (Hofer et al., 2020; OECD, 2013; Penninckx, 2017), namely by issuing findings and recommendations for school staff to use as leverage for actions and measures to improve students' learning experiences (Van Gasse et al., 2018; Verhaeghe et al., 2010). However, such reactions from staff cannot be taken for granted. Several studies have indicated that receiving feedback from evaluations is not a sufficient condition for realising systematic reflection or improvement actions in schools (e.g., Ehren et al., 2013; Verhaeghe et al., 2010), and findings concerning how the results of external evaluation are used and impact improvement in schools have been inconsistent.[3] Whereas some studies have suggested that the results of external evaluation or inspections are helpful and used for learning and improvement in most schools (e.g., Dedering & Müller, 2011; Ehren & Visscher, 2008; McCrone et al., 2007), others have indicated that the use of such feedback and its impact are rather limited (e.g., Baughman et al., 2012; Chapman, 2002; Gärtner et al., 2014; Verhaeghe et al., 2010). Such inconsistent findings on the topic highlight the highly contextual nature of how schools use external evaluations (Ehren, 2016; Hofer et al., 2020). Likewise, a recent comparative study of six European inspectorates has drawn attention to the varying effects of external school

* Correspondence to: Daltún 6, 200 Kópavogur, Iceland.
E-mail addresses: bjo13@hi.is (B. Ólafsdóttir), jtj@hi.is (J.T. Jónasson), aks@hi.is (A.K. Sigurðardóttir).
[1] https://orcid.org/0000-0001-7580-3033.
[2] https://orcid.org/0000-0002-6713-1318.

[3] In this article, in line with usage in Iceland, the term *external school evaluation* is used instead of the term *inspection*. External evaluation is, however, considered to involve the same process as inspection and to have the same theoretical basis.

https://doi.org/10.1016/j.stueduc.2022.101181
Received 19 February 2022; Received in revised form 14 May 2022; Accepted 22 June 2022
Available online 6 July 2022

evaluations depending on pressure for accountability in the schools (Altrichter & Kemethofer, 2015; Ehren, Gustafsson et al., 2015).

The inconsistency also underscores the importance of investigating the use and impact of external evaluations in different models in diverse educational contexts (Ehren, Gustafsson et al., 2015). Most research on the impact of such evaluations has been conducted in countries where the pressure for accountability is greater than in Iceland, as discussed by Ólafsdóttir et al. (2022), which makes similar research in Iceland warranted. Moreover, most studies have been performed shortly after schools received the evaluation findings and therefore could not capture the (im)permanence of the measures for improvement taken by the schools (e.g., Behnke & Steins, 2017; Chapman, 2002; Ehren & Visscher, 2008; Verhaeghe et al., 2010). As an antidote, a longitudinal approach may be required to better determine the longer-term impact of external school evaluations. Because external school evaluations are a major component of ensuring the quality of Iceland's education system, identifying how their results are used and influence improvement can also afford school authorities critical insight into ways of redesigning or improving the evaluation process in order to increase its positive impact.

To partly fill those gaps in the literature, the purpose of our study was twofold. First, we aimed to contribute to current knowledge on the perceived use and impact of the feedback of external evaluations in compulsory schools in Iceland. Second, we sought to elucidate how well the improvements made, based on the feedback, have been sustained over time. To map the perceived use and long-term impact of the feedback, a qualitative research design was followed.

2. Conceptual framework

Ehren and Baxter (2021) have posited that three elements—trust, accountability and capacity—are the pillars of any education system and that their interaction affects the success of educational reforms. Their interaction can be complex, however, and vary across different education systems. For example, if the government introduces high-stakes external evaluations and if schools and teachers associate them with distrust, then accountability destroys trust. Fullan and Quinn (2016) and Six (2021) have highlighted the importance of approaching accountability as a strengthening, supporting process instead of as punishment for not meeting requirements. As such, accountability can contribute to building trust and capacity (Ehren & Baxter, 2021; Six, 2021). Evaluation feedback based on clear performance criteria is intended to hold schools accountable as well as to promote learning and thus develop schools' capacity to work towards improvement (Ehren et al., 2013; Ehren, Bachmann et al., 2021). To secure accountability, capacity has to be developed within schools so that they can incorporate the evaluation criteria and provide high-quality education (Ehren & Baxter, 2021; Fullan & Quinn, 2016).

Evaluation is a knowledge-generating undertaking (Vo, 2015) that assumes that the knowledge generated is useful (Alkin & Taut, 2003). Likewise, evaluations are worthwhile only if such knowledge is put to use. However, the term *use* can be understood in different ways (Rossi et al., 2004). Early studies employed a narrow definition of *use* focused on the decisions and changes prompted by evaluations, namely as "immediate, concrete, and observable influence on specific decisions and program activities resulting directly from evaluation findings" (Patton, 2008, p. 99). As such, that definition refers to *instrumental use*, which is the most commonly experienced, recognised and studied use of evaluations (Nunneley et al., 2015; Rossi et al., 2004; Vo, 2015; Weiss, 1998). Studies conducted on the instrumental use of external school evaluations have identified some products of such use, including changes in policy, teacher retraining, more distributed leadership and management, increased cooperation between teachers, improved self-evaluation and improvements in the quality of teaching, assessment, monitoring and pupil tracking (Dedering & Müller, 2011; Ehren & Visscher, 2008; Ehren, Perryman et al., 2015; Matthews & Sammons, 2004; McCrone et al., 2007; Ofsted, 2015; Van Gasse et al., 2018). However, other studies have documented the rather limited instrumental use of evaluations, especially in schools that have received positive evaluation judgements (Chapman, 2002; Gärtner et al., 2014; Penninckx et al., 2016a; Verhaeghe et al., 2010).

As research on the use of evaluations continued, scholars broadened the concept of use to include situations in which evaluations have affected an individual's thinking or understanding without immediately influencing decisions or actions (Alkin & Taut, 2003; Nunneley et al., 2015; Weiss, 1998). That kind of use is known as *conceptual use*, or *enlightenment*, and can impact individuals' actions in the long term (Nunneley et al., 2015; Rossi et al., 2004; Weiss, 1998). Several studies have identified the benefits of the conceptual use of external school evaluations, including a heightened awareness of the quality of schools and increased professional reflection and discussion amongst school staff (Chapman, 2002; Dedering & Müller, 2011; Gärtner et al., 2014; McCrone et al., 2007; Penninckx et al., 2016a; Schweinberger et al., 2017; Van Gasse et al., 2018; Verhaeghe et al., 2010).

A third kind of use is *persuasive use*, or when the evaluation results are used to convince others of an opinion or position already held by parties within the school about changes that they either consider to be necessary or are opposed to—that is, to either attack or safeguard the status quo (Rossi et al., 2004; Weiss, 1998). Research has revealed schools' persuasive use of the evaluation findings and other external feedback regarding school performance (Baughman et al., 2012; McCrone et al., 2007; Penninckx et al., 2016a; Van Gasse et al., 2018; Verhaeghe et al., 2010), and that such use is more widespread in schools that have received unfavourable evaluation judgements (Penninckx et al., 2016a).

A fourth type of use, *reinforcement*, added by Aderet-German and Ben-Peretz (2020), refers to "the use of positive data for reinforcing existing school strengths" (p. 7). The evaluation findings can give individuals and schools a sense of pride and confidence in what they do and thus reinforce good practices but do not directly prompt observable actions. Although the reinforcement-oriented use of the findings of external evaluations is seldom discussed in the literature, some studies have revealed the positive effects of favourable results on self-worth, self-efficacy (Behnke & Steins, 2017; McCrone et al., 2007; Penninckx et al., 2016a), and collective efficacy (Penninckx et al., 2016a).

Instead of *use*, some scholars prefer the term *utilisation* (Alkin & Taut, 2003; Nunneley et al., 2015; Patton, 2008). However, in this article we employ the term *use* based on the argument that *use* is a broader concept than *utilisation* and therefore more relevant when discussing use in a broad context (Kirkhart, 2000; Nunneley et al., 2015). In the context of evaluations, we define *use* "as the application of evaluation processes, products, or findings to produce an effect" (Johnson et al., 2009, p. 378). Following Rossi et al. (2004) and Aderet-German and Ben-Peretz (2020), we distinguish the instrumental, conceptual, persuasive and reinforcement-oriented use of external school evaluations and apply those uses to classify the outcomes of feedback published in evaluation reports. Based on that framework, two research questions guided the study, and both refer to the perceptions of principals and teachers in the schools:

1. How and to what extent do schools use the feedback presented in external evaluation reports?
2. To what extent do schools sustain the changes made after using the feedback from external evaluations instrumentally?

3. Research context

Representing both levels of governance in Iceland—that is, the state and municipalities—the Ministry of Education, Science and Culture and municipalities in Iceland are legally required to evaluate and assure the quality of individual schools ("Compulsory School Act, ", 2008, Articles 37 and 38). Whereas municipal authorities are responsible for following up on external evaluations and ensuring that they generate improvements in schools, the Ministry is responsible for ensuring that those

authorities fulfil their obligations. In 2013, when Iceland's education system adopted a new approach for conducting external evaluations in compulsory schools (Ólafsdóttir, 2016)—an approach developed collaboratively and jointly financed by the state and municipalities—the Educational Testing Institute, renamed the Directorate of Education in 2015, became tasked with performing the evaluations. Although only 10 schools were evaluated annually through 2017, the number was increased to 27 in 2018, and by late 2021, all compulsory schools in Iceland had been evaluated once (Ólafsdóttir et al., 2022). Designed to monitor whether schools are operating in compliance with laws and regulations and to promote improvement in schools, the approach is more oriented towards improvement than accountability and imposes few consequences for non-compliant and/or underperforming schools and can therefore be understood as a rather low-stakes approach.

Under the approach, external school evaluations are based on a set of criteria for school quality in three areas: the quality of learning and teaching, the quality of school leadership and management and the quality of internal evaluation (Sigurjónsdóttir et al., 2012). Involving document analysis, the analysis of students' performance and a school visit, the external evaluations focus on processes in schools instead of outcomes, and likewise, schools are not ranked based on the evaluation results. Schools are visited by two evaluators for 2–5 days or even longer, if required. During each visit, evaluators observe lessons, provide feedback to individual teachers and interview the school representatives (e.g., principals, middle management team members, teams of teachers, non-teaching staff, students, parents and members of the school council). The assessment of the school's strengths and recommendations for improvement are issued to both the school and the local authority in a written report. Regardless of the evaluation judgement (i.e., weak vs. strong), the school is required to develop an improvement plan in collaboration with the local school authority that addresses how it will implement the report's recommendations. The plan is delivered to the Ministry of Education, Science and Culture,[4] which analyses it before either approving or requesting revisions. To ensure the school's autonomy, the school and the local authority determine the improvement actions to pursue, whereas the Ministry endeavours to ensure that all recommendations are responded to in some way. Once the improvement plan is made public online along with the evaluation report, follow-up is undertaken in the form of communication between the Ministry, the municipality and the school. Every 6–12 months, until all improvements have been fully implemented, the Ministry requests progress reports from the local authority and the school. The follow-up process can thus last from one to several years depending on the improvement plan's timeline. Apart from the state's follow-up on the plan, however, the external evaluation imposes no consequences for the school or the municipality, neither of which the Ministry is authorised to sanction or reward.

4. Method

The research approach applied was a qualitative method (Creswell, 2014) involving interviews with principals and teachers and document analysis to obtain in-depth data from six compulsory schools in Iceland. The qualitative approach was appropriate given our aim to illuminate the perceived usefulness of external school evaluations and how it is woven into the complex fabric of each individual school.

To capture the long-term impact of the evaluations and how the schools have sustained their improvements and changes, interviews were conducted 4–6 years after the schools had received the evaluation reports. That strategy enabled us to examine the extent to which schools' goals for improvement actions were achieved according to the progress reports and interviews and how the improvements have been sustained, if at all.

4.1. Selection of schools and interviewees

Of the 22 schools first subjected to external evaluations in Iceland in 2013–2015, six were selected (see Table 1). To obtain a broad representation of schools with a wide range of contexts and variation in characteristics, the selection was informed by evaluation judgements, school size and geographical location (i.e., urban vs. rural). To protect the anonymity of the schools, all identifying information has been omitted in this article.

Schools A, B and C are relatively large schools that had 300–600 students each during the period investigated, whereas Schools D, E and F are much smaller schools that had 40–130 students each. Five of the schools serve students in Grades 1–10, while the other serves students in Grades 1–7. In Schools B, E and F, a new principal was appointed shortly after the evaluation and thus made responsible for processing the findings and developing as well as implementing the improvement plan.

Interviewees consisted of principals (i.e., one per school) and teachers (i.e., one or two per school), as detailed in Appendix A. The selection of teachers for the interviews was based on their active involvement in the evaluation and improvement process (see Appendix B: Selection criteria of teachers to interview). Although the intention was to interview one teacher in each school, in two cases the teacher requested to have another teacher with them in the interview, which was approved.

4.2. Data collection and analysis

The data consisted of official documents as well as of transcribed interviews. Evaluation reports and improvement plans were used to inform and prepare the interviews and to predetermine codes and themes. Annual progress reports from the schools to the Ministry of Education, Science and Culture regarding the implementation of the improvement plans were used to obtain information on the progress of improvements. In sum, the documents used in the study included six evaluation reports, six improvement plans and 17 progress reports.

The interviews were conducted with six principals and eight teachers in 2019. The first author arranged appointments at the interviewee's school except for one school where the interviews took place in connection with their participation at a conference. Absolute anonymity was promised to all participants and maintained, and all participants signed their written informed consent to participate. All interviews were semi-structured and based on the same generic questions but adapted to each school in light of the evaluation report and the school's improvement plan. To help each interviewee to review the improvement actions, the interviewer presented a copy of the school's improvement plan at each interview. The interviewees were asked about the actions taken and changes made in their school as a result of the external evaluations

Table 1
Information about the external evaluation of the six schools.

School	Location	Overall evaluation judgement	Duration of follow-up on improvement plan
A	Urban	Significant strengths	3 years
B	Urban	Strengths outweigh weaknesses	2 years
C	Urban	Strengths outweigh weaknesses	4 years
D	Rural	Significant strengths	3 years
E	Rural	Weaknesses outweigh strengths	4 years
F	Rural	Weaknesses outweigh strengths	2 years

[4] In this article, we discuss the arrangement for following up on external evaluations as it was when the studied schools underwent the process. Since 2019, the Directorate of Education has administered the follow-up process, not the Ministry of Education, Science and Culture.

and whether the improvements made had been sustained or were still in development. The interviews were recorded and lasted 48–90 min. They were transcribed, and selected citations were translated into English by the first author and reviewed by an English-language proofreader.

The software package NVivo R1 was used to store, organise and analyse both the interview transcripts and documentation and a thematic approach (Braun & Clarke, 2006) was followed. The segments of data relevant to the research focus were coded according to predefined coding structure in the three areas of the external evaluation: (1) leadership and management, (2) learning and teaching and (3) internal evaluation. Sub-codes for each of the three areas were developed (see Appendix D: Coding scheme). Predefined codes and themes referred only to the instrumental use and impact of the findings of external school evaluations. However, when additional themes were identified that did not represent instrumental use, we widened the scope of the analysis to encompass conceptual, persuasive and reinforcement-oriented uses as well. The analysis was guided by the research focus and was therefore more theoretical than inductive (Braun & Clarke, 2006). The data was first coded by each school and then assigned to the relevant theme. As the focus shifted to the themes, the themes were further analysed and refined.

The combination of different data sources, documents and interview data in each school was used for the purposes of triangulation. The analysis of the documents provided information on the initial status of the schools, planned changes and improvement actions in the 2–3 years after they received the evaluation feedback. That strategy allowed us to triangulate our thematic interview analysis and conclude that certain changes were indeed a result of external evaluations. In this article, our findings are presented primarily in excerpts from the interview transcripts, while document-based data were used at the stage of analysis.

Data collection was part of a formative assessment of the use of evaluation feedback by the evaluation agency, and the person conducting the interviews was affiliated with the body responsible for the external evaluation. Anonymity was promised and respected, and there were no risks involved; however, that may have biased the participants' responses towards presenting idealised interpretations of the schools' work and refraining from criticising the evaluations.

5. Results

The findings are presented according to the framework and the themes. The first findings concern the instrumental use of the feedback in external evaluation reports in terms of the quality of (1) leadership and management, (2) learning and teaching and (3) internal evaluation. Thereafter, the findings regarding the conceptual, persuasive and reinforcement-oriented use of the evaluation feedback are presented. Finally, findings on how the schools sustain the changes made, if at all, are discussed. However, before discussing the use of the feedback, we briefly outline the interviewees' perceptions of the evaluation results, because such perceptions affect their willingness to apply them as a means to make improvements.

5.1. Attitudes towards the external evaluation in the schools

In the interviews, both teachers and principals reported support for the external evaluations, which they generally characterised as being helpful and supportive of changes (i.e., instrumental impact) in the schools. According to the interviewees, it was helpful to receive concrete recommendations about which improvements to prioritise, especially in such constructive, positive wordings (i.e., "opportunities for improvement"). One principal commented:

> *[The report] comes with suggestions for improvements … and it's helped the school immensely because they're really good instructions about what needs to be done, [and are] structured in a positive way. There are few commands or big adjectives. They're just good, responsible recommendations. (P, School E)*

In the same vein, a teacher stated that despite their considerable anxiety leading up to the evaluation judgement, they found the evaluation feedback to be encouraging:

> *I think that we may have expected it [the evaluation] to be rather critical which, in retrospect, didn't happen at all. It was just about how we could go a step beyond where we are now, with what we have. (T, School F)*

Although the schools were generally satisfied and agreed with the recommendations, the principals in Schools C and E disagreed with some of them because they were perceived as being trivial or inconsistent with the school's policy. The principal in School C also expressed a certain resistance to the control exercised by external evaluations:

> *I think that schools always need the opportunity to step outside the framework being used. There will never be any development in schools unless someone doesn't quite follow all the rules. We may want to proceed in other ways.*

5.2. Instrumental use

5.2.1. Leadership and management

Recommendations for improvement in leadership and management, presented in the schools' evaluation reports and discussed here, are focused on the subthemes of professional collaboration amongst staff and the instructional leadership of school leaders.

Most of the schools received a recommendation to increase the professional collaboration amongst staff members. In Schools A and C, changes were made that consisted of clarifying the division of tasks on the management team and sharing leadership responsibilities with middle managers. In all schools but School E, external evaluations prompted increased professional collaboration and reflection amongst the teachers, and time for teachers' meetings was either increased or else meetings held more explicitly for collaboration. Teamwork on specific subjects across school levels was also increased. One teacher explained the changes as follows:

> *The collaboration between teachers–to help and work together–that's what I think is exactly the advantage of getting this kind of external evaluation. You know, it [the collaboration] became more holistic. We took everyone into the equation and worked much more together. It was more purposeful collaboration. (T, School D)*

Along similar lines, interviewees in Schools A and D also mentioned that increased classroom observations by principals recommended in the evaluation reports have increased their sense of the teachers' strengths, which has contributed to increased peer education, knowledge sharing and peer support. Other outcomes mentioned were more purposeful, results-oriented discussions about students' learning and more targeted professional development and learning programmes.

In the reports, leaders at all schools were advised to regularly evaluate teaching practices and provide feedback to teachers. By the time of their interviews, principals in Schools A and D had implemented systematic classroom observations and feedback for teaching staff yet were still developing their methods and focus. In Schools E and F, although the principals or other leaders had visited classrooms frequently, a formal, systematic process for observation and feedback was not apparent. In the others, Schools B and C, principals or other school leaders had made little or no effort to promote classroom observations or feedback for teachers.

5.2.2. Learning and teaching

The proposals for improving learning and teaching discussed here primarily concern differentiated strategies for instruction and the use of assessments to improve students' learning and democratic participation.

In evaluation reports, all schools were advised to improve differentiated instruction in order to meet students' diverse learning needs—for example, by strengthening the information technology used, emphasising collaboration and dialogue amongst students, considering students' fields of interest and strengthening their range of options. To address those recommendations, the schools took numerous steps, some even with various professional learning programmes. At each grade level, teachers' teamwork in planning and/or teaching also increased in most of the schools. Meanwhile, the availability of digital devices for students and staff to use increased as well, and tablets were implemented in learning and teaching. Indeed, in all six schools, substantial progress in information technology continued to be made, not necessarily due to the external evaluations, however, but owing to developments in the tech world or other external factors, including development projects in the municipality.

All schools but School C took actions to better meet the learning needs and interests of students and to expand their choice and collaboration in learning. Such actions included introducing a carousel strategy, group work, project-based learning, outdoor learning, a makerspace and art workshops. However, though the interviewees generally believed that professional development towards more differentiated instruction has occurred following the external evaluations, some stated that such instruction relies on the participation of every teacher, and despite productive discussions amongst the teaching staff and the joint decisions made, some teachers have continued to struggle to effect change towards realising differentiated instruction.

School C differed somewhat from the other schools, for its teacher and principal argued that the external evaluation has hardly impacted learning or teaching in the school even though the evaluation report had clearly recommended some changes. Few actions have been taken to increase differentiated instruction strategies apart from introducing teachers' teamwork at each grade level and increased collaborative learning amongst students.

All schools were advised to increase the democratic participation of students to enable them to express their views. Actions taken to that end in the schools included implementing class meetings and student discussion forums, increasing the activity of the student council and affording students opportunities to vote on topics and events. Although the planned reforms did not succeed in all cases, the interviewees generally stated that students' democratic participation in decision-making had intensified and become a more permanent part of school's daily life than before. However, work remains to be done. As one principal put it:

> *As for the democratic work of students–the evaluation report stated that it needs to be strengthened–and when I look back, we've been working on it, but it's not yet what we want it to be. (P, School B)*

Schools A, D, E and F received a recommendation to improve students' achievement. In Schools A and D, much emphasis has thus been placed on improving instruction, which their principals and teachers viewed as having improved assessment outcomes. Although goals were set in School F to promote achievement, the principal and the teacher stated that each teacher has been allowed to determine how they systematically worked towards those ends. Because the follow-up by the principal has been minimal, it is unclear whether any improvements have been made. In School E, no actions based on this recommendation were taken.

5.2.3. Internal evaluations

Most of the recommendations in the evaluation reports for improvement in internal evaluations concerned evaluation plans and methods, stakeholder participation and the improvement plans. In the progress reports of Schools D and F and in interviews with the principals, it was declared that the internal evaluations were systematically strengthened in accordance with recommendations to substantially improve the evaluations, and most of the changes made have been maintained or were still in development. Meanwhile, in School B, though the external evaluation report indicated a fairly mature internal evaluation, the recommendations were only partly met, and aspects of the internal evaluation in place when the external evaluation occurred have since declined, as stated in the interviews. In School E, almost no internal evaluation was performed at the time of the external evaluation; some improvements were made, but the principal admitted that not all recommendations have yet been met, even though the progress reports say otherwise. According to the principals in Schools A and C, no improvements were made to the internal evaluations despite recommendations; in both schools, the evaluators judged the internal evaluation as being rather mature. School C has shown a decline in its internal evaluation since its external evaluation; both a progress report and the principal during the interview attributed the decline to a lack of time and lack of perception of its importance. School A's situation has remained unchanged due to the evaluation team's lack of knowledge about making changes, as confirmed by the principal.

5.3. Conceptual use

Although the interviews did not focus on the conceptual use of the external evaluations, the findings suggest several ways in which the interviewees used the evaluation feedback in conceptual ways. They mentioned, for example, the usefulness of having an external view of the school's functions, which opened their eyes to existing practices and helped them to identify needed improvements and cultivate focus. On that subject, one teacher stated:

> *I thought in some aspects–"Yes, OK, we're not doing well enough there"–and that's why it was so good. You see, because sometimes you can just think, "Oh, we're on a really nice path here", but it's really lacking a lot. (T, School A)*

According to some interviewees, the external evaluations led to important, productive discussions and reflections in the schools and increased the scope of those discussions. Even in School C, where the instrumental influence of the external evaluation appeared rather slight, the evaluation has had a conceptual impact by stimulating discussion and teachers' reflections on their professionalism, at least according to the teacher:

> *Just those meetings, those discussions that started: it [the external evaluation] of course ignited interest and … a broader perspective on the school's work. I think that every teacher thought about their professionalism. It encouraged every teacher to think about their own performance.*

For the three principals who were appointed after the external evaluations, it had been useful to receive information about the school's status. On the one hand, the reports enabled them to familiarise themselves with their schools and gain a perspective on what needed to be done. On the other, it defined expectations for the principals in general and therefore afforded instructions for ones who had only recently assumed the role. The principal of School E captured the sentiment of all three of those principals by stating:

> *I got the best job description in the world. I just sat down and went over the external evaluation report and discovered little by little what was going on. … I wasn't an experienced principal when I started here, so it was very good to get it like this [in the evaluation report]. I simply got an introduction to how to be a school principal.*

Although the three principals shared that view, the principal of School B also reported struggling to immediately begin acting upon the findings of the evaluation upon entering their new school.

5.4. Persuasive use

As with the conceptual use of the external evaluations, their persuasive use by the principals and teachers was not specifically

addressed in the interviews. Nevertheless, three interviewees reported that the evaluation feedback was useful for such purposes. The teacher in School D stated that the evaluation report was a good instrument for supporting their existing opinions about changes needed in the school's operations and getting everyone involved in working to those ends:

> *I came here with new ideas and wanted to change a lot and wanted to do so many things, you know. So, I think it [the evaluation report] helped me a lot to introduce those new ideas and thoughts. Because then you can quote something like "As stated in the evaluation report, it's good for us to look at collaboration". And it's not just something that someone is saying, because there are professional arguments for it.*

In their interviews, the principals of Schools A and B also reported feeling that the evaluation reports have supported them in convincing others in the school to take certain actions. As the principal in School A said:

> *This [the evaluation report] is just one of the best tools I've ever received. Going into lessons and observing and giving feedback afterwards is very awkward for Icelandic teachers because they're not used to it, you know. I could just say, "Now the only thing we have left in this improvement plan is that I come, not only to visit, but to look at certain aspects". It's been really good to be able to refer to it.*

Those findings suggest that the persuasive use of external evaluation feedback strongly supports their instrumental use, especially when changes are needed that are likely to face resistance.

5.5. Reinforcement-oriented use

The last type of use identified in the interviews was the reinforcement-oriented use of the external evaluations. In the schools that had received favourable evaluation results (i.e., Schools A, C and D), the teachers and principals felt that obtaining feedback that the school was performing well and on the right track had been encouraging and empowering. The teacher in School D said:

> *Above all, I found it [the evaluation feedback] to be really encouraging. We could then quote the results, and we got the feeling that we were on the right track.*

In the same vein, the principal of School A stated:

> *It [the evaluation feedback] was really inspiring for us on the management team and in fact for the entire staff, [to learn] that we're doing a good job.*

5.6. Sustained changes

The interviews and progress reports suggest that most of the schools have implemented a range of strategies and actions owing to the external evaluations. Most of the improvement actions included in the schools' improvement plans have succeeded or else continued to be developed in some way, and the changes made have largely been sustained. However, School C was an exception, for only a few actions from their improvement plan had been implemented. When asked about the permanence of the improvements, principals in Schools A and F respectively said:

> *I think that most of the improvements that we made have been sustained. We've referred to it [the improvement plan] a lot. I went through it and X'ed many aspects, and I'm just really happy with how we have moved on them. So, maybe some [improvements] have not lasted but developed in another direction that's just as good. (P, School A)*

> *I think they [the improvements] are maintained in many aspects. Of course, there are some aspects that have fallen between the ship and the dock, and it will be necessary to come back [evaluate again] and point them out. But a lot of things have lasted. (P. School F)*

Although most changes made have been sustained, interviewees noted that some aspects, especially ones related to learning and teaching, needed a great deal of time to develop and were by no means complete. On that topic, the principal of School B said:

> *Of course, we continue to work according to those [the evaluation recommendations], but maybe we no longer think about that we've received a recommendation for this–such as students' responsibility for their own learning and democratic participation–we're working towards that end even if we're not always flipping through the report. It's simply become part of our culture.*

In Schools E and F, the need for improvements was substantial, the projects were extensive, and work remained to be done to realise planned improvements when the interviews were conducted, even though the improvement plans approved by the Ministry have been formally completed. As the principal of School E stated:

> *This [the evaluation report] was very useful for our organisation and will be used for a few more years and hopefully we'll have another external evaluation. We use our improvement plan, [but] I suppose it will be obsolete in a few years, so we'll need to make a new one.*

In general, the interviewees reported that the improvements have succeeded and that their schools have retained the knowledge for continuing such work. The teacher in School B was amazed by how much the school had accomplished when they reviewed the evaluation documents while preparing for the interview: "When I went through the report, I felt, 'Yes, we've done quite a lot'. It was just–*wow!*".

In Iceland, responsibility for improvements in schools is shared across the education system, and the municipalities play a significant role in supporting schools in that process, which is especially valuable for schools identified as having major weaknesses. In School F, improvements proved to be challenging for the principal even though awareness of the improvements within the school seemed high to them. The principal was retiring after some years in the position, partly due to burnout after striving to make various improvements at the school but receiving little to no support from the municipal administration:

> *The municipality was somehow—there was no support from it. You just become, when you're constantly facing adversity … As a principal I'd become slightly burned out, because the projects were just gigantic.*

Engagement is also needed at the national level in order to achieve greater improvement in schools. All principals reported that follow-up on the improvement plans by the Ministry of Education, Science and Culture mattered because it kept them focused on improvement, as the following quotation captures:

> *We'd made a progress report three times, then we naturally went over it to see whether we were making progress, and it was sometimes slightly like a checklist—"Are we definitely doing this?"—which is kind of good. It provides restraint. (P, School B)*

However, to the principal of School F, the Ministry's follow-up ended too early because the school still had far to go with its improvements and because the municipal administration was rather inactive. After the follow-up was completed, the school stopped working systematically on the improvement plan:

> *So, it just fell apart somehow–because there was no one to ask for information–then somehow, it's not as important, quite unconsciously. Because there are other factors that take priority. So, only if someone is always like. "How are you doing? You have three improvement actions left. How are you going to get them done?" Then you remember.*

6. Discussion

The aim of our study was to illuminate how principals and teachers in compulsory schools in Iceland have perceived the use and impact of

the external evaluation feedback given to their schools and how well improvements made at the schools based on the feedback have been sustained, if at all. The analysis drew from work by Rossi et al. (2004) and Aderet-German and Ben-Peretz (2020) that distinguishes the instrumental, conceptual, persuasive and reinforcement-oriented use of external school evaluations.

In their systematic synthesis of 30 years of research, Hofer et al. (2020) have identified several conditions that might increase the impact of external school evaluations. One of them concerned the importance for schools to accept external evaluations and the feedback that they offer. In our study, teachers and principals alike reported clear support for the external evaluations and had generally experienced the feedback as being helpful and as having contributed to changes in practice (i.e., instrumental impact) in their schools. Trust, which according to Ehren and Baxter (2021) affects the success of reforms, appears to be present when it comes to the external evaluations, even despite pressure for accountability in the form of publishing the evaluation reports and following-up on the improvement actions. Such a positive relationship between trust and accountability is more likely to facilitate the evaluations' positive impact and enhance education quality (Six, 2021).

As for the first research question, concerning how and to what extent schools use the feedback presented in the external evaluation reports, the findings confirm that the feedback has been used in the different types of ways identified by Rossi et al. (2004) and Aderet-German and Ben-Peretz (2020)—instrumental, conceptual, persuasive and reinforcement-oriented ways in the schools—albeit to varying degrees. First, concerning instrumental use, in the 4–6-year period after the external evaluations, substantial improvement actions have been implemented and developed in five of the six schools as a result of the evaluations, including actions to increase professional collaboration amongst staff, differentiate instruction strategies, integrate information technology in learning, stimulate students' democratic participation and enable them to express their views. In some cases, changes were also made regarding instructional leadership, internal evaluations and the use of assessment to improve students' learning. Second, interviewees from all schools also indicated that the evaluation feedback has been used in conceptual ways, especially for considering their schools from a broader perspective, for highlighting needed improvements and for cultivating a focus on action. Beyond that, the evaluations had prompted productive discussions and reflections in the schools and served as support for newly appointed principals. Third, without being asked about it, three interviewees indicated the persuasive use of the evaluation feedback, including that the evaluation results had supported them in implementing important changes. Fourth and finally, the reinforcement-oriented use of the feedback was also observed in three interviews. Consistent with research by Penninckx et al. (2016a) and Behnke and Steins (2017) and as stands to reason, such use has primarily occurred in schools that received positive evaluation judgements.

Concerning the instrumental use of the evaluation feedback, differences arose between the schools in how systematically they have worked to meet all of the recommendations for improvement in the evaluation reports. The two schools with the best evaluation results, Schools A and D, have worked systematically to meet all of those recommendations. Meanwhile, the two schools with the greatest need for improvement, Schools E and F, have sought to make improvements following most of the recommendations but not all, as has School B, which received a fairly positive external evaluation. School C, however, which also performed rather well according to the evaluation, differs from the other schools in having placed little emphasis on improvement based on the recommendations except in a few aspects, seemingly due to a certain opposition of the principal to the evaluations. That finding contradicts previous results from a study on school inspections in Flanders, which revealed the stronger instrumental and persuasive use of evaluation feedback in schools that received less favourable evaluation judgements (Penninckx et al., 2016a). Such inconsistency cannot be explained by a different evaluation model (e.g., low-stakes vs. high-stakes) because the inspection system in Flanders is rather low-stakes (Penninckx et al., 2016a), similar to the external evaluation system in Iceland. However, it may be explained by varying degrees of pressure for accountability placed on schools regulated by the follow-up process. In Iceland, follow-up on the behalf of the Ministry of Education, Science and Culture clearly sets the expectation that schools, regardless of their evaluation judgement, will use the evaluation feedback for improvement, even though no penalties exist for schools that do little to change their practices. Having to submit an annual progress report for 2 or more years following the evaluation creates (perceived) pressure and focuses the efforts of the school staff on improvements, which apparently increases the impact of a low-stakes evaluation model. That dynamic is important, given a recent, major European study on external school evaluation revealing that a low-stakes evaluation approach is not as effective as a high-stakes one, because pressure for accountability leads to more improvement actions (Altrichter & Kemethofer, 2015; Ehren, Gustafsson et al., 2015). Given the importance of accountability as one of the three chief elements for fruitful reform (Ehren & Baxter, 2021), our study speaks to the benefits of long-term follow-up in contexts in which external evaluation is a low-stakes affair.

Ehren and Visscher (2008) found that schools struggle to use the feedback that they receive from external evaluations as a basis for implementing complex improvement actions. Our study indicated that difficulties in making improvements may lie in certain areas, most of which relate to the purposeful use of internal evaluations and student assessments. Thus, some schools in our study have not made much progress in implementing aspects such as the instructional leadership of school leaders, the systematic use of assessments to improve students' learning and strengthened internal evaluations. Those results align with past findings showing that internal evaluations rank amongst the weakest areas in the management of schools and that school personnel often have limited skills in and experience with performing meaningful evaluations (Blok et al., 2008; Brown et al., 2017). Thus, the third chief element that affects successful reform (Ehren & Baxter, 2021), the capacity to implement recommended improvements, seems to be partly lacking, which has limited the impact of the external evaluations (Ehren, Bachmann et al., 2021).

With respect to the second research question, regarding the extent to which schools sustain the changes that they have made, the findings suggest that the improvement actions presented in the schools' improvement plans were generally implemented or continued to develop in some way. However, we also acknowledge that the issue of sustainability is particularly complex for several reasons. The spectrum of actions to be considered is quite wide, and the judgement of whether something is sustained or not is a complex one that depends on the point of departure—that is, whether much or little change is needed. Even so, the comments made by the interviewees indicate that they were aware of such complications, and we can infer that many of the changes made have been sustained, according to examples mentioned in the interviews and documents and the reference made to using the evaluation feedback to continue encouraging change and acknowledge that, in some cases, sustainability had not been achieved. However, it remains to be seen whether the changes judged to have been sustained persist.

Based on their synthesis of the literature, Hofer et al. (2020) have recommended nuanced feedback instead of judgement in evaluation reports in order to prevent the negative effects of critical judgement. Our

findings support that recommendation in showing that the constructive wording of evaluation reports when pinpointing areas for improvement is indeed important to teachers and principals, largely because it prevents the impression that improvements are being forced on the staff, which can critically limit the sustainability of improvement actions (Penninckx et al., 2016b). Constructive feedback also increases trust in the evaluation and thus its impact (Six, 2021). Altogether, we conclude that many of the recommended changes were implemented and have been sustained, at least from the perspectives of the teachers and principals interviewed.

6.1. Strengths and limitations

A strength of our study was its longitudinal design, which enabled us to study the different uses of external evaluation feedback and how well the changes made over the years have been sustained, if at all. That perspective may be pivotal given research showing that many changes and improvement actions in schools seem to peter out (Ehren, 2016). Furthermore, whereas studies on the use of school evaluations have often been based solely on the views of principals, our study benefited by including the perceptions of both principals and teachers as well as building on documents generated during the improvement process for the purposes of triangulation.

Despite those strengths, the study's limitations also warrant attention. First, the findings are largely based on participants' perceptions and reports in only six schools in Iceland. Therefore, the extent to which generalisations can be made to the wider population of schools in Iceland is limited. Nevertheless, the most important part of the findings is the rich content of the material obtained, which is quite clear even from only those schools. Second, the validity of the findings is restricted to a specific educational context that involves the use of external evaluations, which offer a relatively low-stakes system of accountability, albeit one with a fairly transparent, substantive follow-up system in the form of progress reports. That restriction should be taken into account when using the findings from Iceland's education accountability system to reflect on other accountability systems. Even so, the analysis sheds light on how schools use evaluation feedback in such a setting and, as such, may offer important insights. Third, self-report, which the study relied upon, may be biased and thus overemphasise improvements made and the use of evaluation feedback, not least because the interviewer came from the agency responsible for the evaluations. However, anonymity was clear, as was the fact that the purpose was to evaluate processes, not the schools or their individual responses. Last, other stakeholders, including students and municipalities, were not included in the study, which would have given more weight to the results, especially regarding how they have been affected by the changes (e.g., in teaching and learning). On that note, research in the future should take into account the views of more stakeholders in the context of external school evaluations.

Furthermore, we recognise that the findings are based on the teachers' and principals' perceptions of the changes and improvements made as a result of the external evaluations as well as on the progress reports, which also reflect the schools' interpretation of their progress. In the study, no position was taken on the nature of the changes or whether the measures were adequate responses to the recommendations of the evaluation reports. It can therefore not be stated that the changes made to practice in the schools have been equal to improvements in line with the evaluation agency's expectations. Further research is therefore needed that examines the nature and depth of such changes and improvements and to what extent they align with the expectations for the schools.

7. Conclusions

We consider the data sources to have been valuable for answering the research questions. Our qualitative study has clearly shown that external school evaluations can have various uses, for the data revealed clear examples of the instrumental, conceptual, persuasive and reinforcement-oriented uses of the feedback in the external evaluation reports. It has also illustrated that schools seem to sustain many of their improvements or continue to develop them in some way, at least in the few years following the evaluations. The results moreover show clear evidence that in a system based on low-stakes accountability and trust between schools and authorities, the improvement-oriented evaluation approach works well, provided that schools receive support to increase their capacity in areas in which they are facing difficulties. In that light, the findings can inform policymakers as they attempt to understand and shape the future use of the feedback of external school evaluations.

Although our study focused on Iceland, its findings tentatively suggest that policymakers in other countries may find the results and suggestions interesting, given the apparent positive impact of a low-stakes but thorough evaluation procedure, and thus indicative for the development of such external evaluation and its follow-up process. In that light, the results of the study can be used to improve the role of external evaluations in national and local school governance. To that end, we make three suggestions, all of which assume that the basic ingredients of the system are retained. First, the length of follow-up needs to be adjusted according to the school's status so that schools in great need of improvement are monitored for longer periods. Second, external support for schools regarding internal evaluation, the appraisal of teachers and the purposeful use of assessments to enhance student achievement needs to be developed. Third, responsibility for school improvement needs to be shared across the education system.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-for-profit sectors.

CRediT authorship contribution statement

Björk Ólafsdóttir: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Jón Torfi Jónasson:** Validation, Writing – review & editing. **Anna Kristín Sigurðardóttir:** Validation, Writing – review & editing.

Declaration of interest statement

The first author is employed at the Directorate of Education, which is responsible for conducting the external school evaluations in Iceland. She has been a proponent of and has played a considerable role in the external school evaluation programme from its inception. Although she currently does not participate in its implementation, she works with those who are responsible for its implementation. The two additional authors have not participated in any stage of the programme being discussed. The authors have not received any financial or other support for the study.

Acknowledgements

None.

Appendix A. Interviewees' demographic information

Table A1 shows the demographic information of the participants interviewed for the study. To ensure confidentiality, the school where each interviewee works is not identified.

Table A1
Demographic information of the interviewees.

Occupation	Gender	Age group	Seniority in the school
Principal	Woman	30–39	1–3 years
Principal	Man	30–39	1–3 years
Principal	Woman	50–59	4–8 years
Principal	Man	60–69	>15 years
Principal	Woman	50–59	1–3 years
Principal	Woman	50–59	9–14 years
Teacher	Woman	50–59	>15 years
Teacher	Woman	50–59	9–14 years
Teacher	Woman	40–49	4–8 years
Teacher	Woman	40–49	4–8 years
Teacher	Man	30–39	4–8 years
Teacher	Woman	40–49	4–8 years
Teacher	Woman	40–49	9–14 years
Teacher	Woman	40–49	4–8 years

Appendix B. Selection criteria of teachers to interview

It was assumed that teachers, who were members of improvement or internal evaluation teams, had access to information that would qualify them to answer questions about the implementation of improvement actions following external evaluations. The selection criteria for teachers to interview were thus:

(1) If the school had assembled a team to work on the improvement plan, as was the case in three schools, then one teacher from the team was selected to be interviewed; and
(2) If no team was working on the improvement plan, as was the case in one school, then a member of the internal evaluation team was selected to be interviewed.

Two of the schools did not have a dedicated team to handle the improvement plan or the internal evaluations. In those cases, a third selection criterion was used:

(3) A teacher was selected from the group of teachers published on the school's website. The participation of a teacher who taught at the school level that had received the most recommendations for improvement, especially regarding student achievement, was requested.

Appendix C. Framework for interviews with principals and teachers

At the beginning of the interview, the purpose of the interview, how we would use the data and the length of the interview were communicated to the interviewee, their permission to record the interview was obtained, and their full confidentiality was ensured. Each participant signed a form stating that they were informed about the subject of the study.

Each interviewee then received the school's improvement plan for their review, after which the following script was followed:

For both principals and teachers:

1. In your opinion, how useful were the results of the external evaluation to the school?
2. To what extent did the results of the external evaluation reflect the strengths and weaknesses of the school?
3. How did you work the results of the external evaluation?
4. What was the process of making the improvement plan like?
5. Did you find it easy or complicated to decide on the improvement actions?
6. Did the school have the resources that it needed to work on the improvements?
7. What resources (e.g., time, training and staffing) were allocated to work on the improvements?
8. Based on your experience, how open to innovation and change are the school's principals, teachers and other staff? Are they open to doing things differently? What is the attitude of teachers towards professional development and changes in their teaching practices?
9. How did you monitor the progress and/or success of the improvements?

For principals only:

10. Did the representatives of the municipality take part in the process of making an improvement plan?
11. Did the municipality provide any support? What resources (e.g., time, education and staffing) did the municipality allocate to the school so that it could work on the improvements?
12. In your opinion, were enough resources allocated so that you could work on the improvements? Is there a need for more external support?
13. What do you think about the follow-up process of the Ministry of Education, Science and Culture?

Now we turn to the questions about change and development in the school after the external evaluation regarding the three aspects of the

evaluation: management and leadership, learning and teaching and internal evaluation. We will focus on leadership and management first.

For both principals and teachers:

14. In general, how did the external evaluation contribute to changes in the leadership and management of the school? What recommendations regarding the aspects of leadership and management did you work with in particular?
    a. Let's review the school's improvement plan and selected improvement actions related to leadership and management (e.g., strengthen professional leadership, distribute leadership, increase parental involvement and information to parents, strengthen staff cooperation, form a clearer vision and school policies and appraise teachers better).
    b. Have the changes or improvements that you made lasted? Are they being sustained?
    c. How important do you think that those improvements and changes are? Why?
    d. Would you have liked to have done something different? In what way, and why?

For principals only:

15. Do you or other leaders in the school attend lessons in order to appraise teaching and give feedback to teachers?
    a. If yes:
        i. How often?
        ii. How prepared are you to appraise teaching and give feedback to teachers?
        iii. What do you think that teachers think about the appraisal and feedback?
    b. If no:
        i. How do you monitor teachers' performance?

For teachers only:

1. How are teachers appraised in the school? Do the principals visit classrooms and/or do teachers evaluate other teachers' lessons?

Let's now turn to the aspects of learning and teaching.

For both principals and teachers:

16. In general, how did the external evaluation contribute to changes in learning and teaching in the school? What recommendations regarding the aspects of learning and teaching did you work with in particular?
    a. Let's review the school's improvement plan and selected improvement actions related to learning and teaching (e.g., promote results in Icelandic and mathematics, analyse what causes poor scores on standardised tests, increase integration in learning, increase students' choice, increase dialogue and collaboration in learning, host student meetings, promote the democratic participation of students and better meet students' interests).
    b. Have the changes or improvements that you made lasted? Are they being sustained?
    c. How important do you think that those improvements and changes are? Why?
    d. Would you have liked to have done something different? In what way, and why?
    e. How much knowledge does the school's staff have about making improvements in learning and teaching? Is there enough knowledge amongst the teachers to work on the improvement actions, or is more knowledge needed?

Last, let's discuss the internal evaluation.

17. In general, how did the external evaluation contribute to changes in internal evaluation in the school? What recommendations regarding the aspect of internal evaluation did you work with in particular?
    a. Let's review the school's improvement plan and selected improvement actions related to internal evaluation (e.g., organize the implementation of and responsibility for internal evaluation, evaluate learning and teaching, increase stakeholder participation, diversify data collection, use the results of standardised tests and make internal evaluation reports and improvement plans).
    b. Have the changes or improvements that you made lasted? Are they being sustained?
    c. How important do you think that those improvements and changes are? Why?
    d. Would you have liked to have done something different? In what way, and why?
    e. How much knowledge does the school's staff have about making improvements in internal evaluation? Is there enough knowledge, or is more knowledge needed?
18. Do you use the quality criteria used in the external evaluation in the internal evaluation? Why or why not?

We've arrived at the final question:

19. Would you like to see external evaluations continue or not?

I have gone over all of the questions. Is there anything that you would like to add before we end the interview?

Thank you for your participation.

Appendix D. Coding scheme

	Level 1	Level 2	Level 3	Ref.	Example (s)
Instrumental use	Leadership and management	Distributed leadership		5	We've changed the management. There are more people involved now, more who share responsibility (P, School C).
		Professional collaboration amongst staff	Teachers' meetings and collaboration	21	We've started having team meetings and school-level meetings where a professional discussion takes place and an agenda for all staff meetings is set (Progress report, School F).
					Since the evaluation, we've deepened our collaboration. We do a lot of talking about what we're doing in meetings and

(*continued on next page*)

(*continued*)

	Level 1	Level 2	Level 3	Ref.	Example (s)
					sharing with others—we're bringing other teachers into the classroom—and then we're working holistically with projects between [students'] age levels (T, School D).
			Teachers' subject-focused teamwork	9	We started engaging in closer teamwork and more collaboration (P, School B) I think we're all on at least two teams (T, School F).
		Instructional leadership - Classroom observation and feedback	Implementation	37	The principals don't visit classrooms. They're in fact on the ground very little, which can perhaps be criticised. (T, School C) Right after it [ex. ev.], X [the principal] started to get more into lessons and spend some time in lessons with teachers (T, School A) This [visiting classrooms] was dropped, and instead more emphasis was placed on teamwork and collaboration between teachers (Progress report, School C) Principals regularly visit lessons, but the visits need to be more organised, and the feedback needs to be more purposeful (Progress report, School F) Both principals have come in sometimes but not to evaluate learning or teaching (T, School F).
			Benefits	5	When we're talking about, for example, disciplinary management, there's a teacher whom the school principal has noticed has a good grasp on it, who is now sharing his knowledge with other teachers (T, School A) I think this [ex.ev.] has given us the opportunity to strengthen ourselves in being instructional leaders (P, School A).
			Feedback	8	They've always been very diligent in looking into lessons, but they haven't given any feedback. Maybe they should. (T, School E)
			Barriers	3	Going into lessons and observing and giving feedback afterwards is very awkward for Icelandic teachers because they're not used to it (P, School A).
			Teachers' attitude	7	It's just *great* to get X [the principal] into a lesson, absolutely great. You just meet X, and X might give you some points. X comes in at all times, and she experiences all kinds of things. Sometimes there's chaos, and sometimes everything goes as planned (T, School A) I want more of it. … I think it's really good … when X [the principal] comes, X can see what's going on and. I just feel … the more often the better. … I just experience it as very positive (T, School D).
	Level 1	**Level 2**	**Level 3**	**Ref.**	**Example(s)**
Instrumental use	Leadership and management	Instructional leadership: Classroom observation and feedback	Principals' attitude	3	I know some principals have started giving feedback to teachers after sitting in class and even filling out a checklist. I don't like it. I could never do that (P, School C).
			Procedures in development	6	I've been trying this [new method of classroom observation] out slightly this winter, and next fall I'm going to start doing it purposefully (P, School A).
	Learning and teaching	Differentiated instruction	Diversity in teaching	18	I think that we're more moving towards what the new curriculum is entrusting us to do, such as using more diverse teaching methods (T, School D). This winter, teachers were aware of a variety of teaching methods: interest-based lessons, carousel forms in mathematics, outdoor teaching, theme days and weeks (Progress report, School B). Teachers were encouraged to use a variety of teaching methods, and ideas for ways how were discussed at meetings. Some have succeeded, but others need to improve. We'll work systematically on it next school year (Progress report, School F). Following this [the ex. ev.], we slightly systematically looked into our teaching methods and ways of changing them, to some extent (T, School F).
			Information technology	20	There were several courses for teachers on the use of tablet computers during the school year. iPad lessons are at all levels (Progress report, School B). We've purchased tablet computers for the school and participated in two Erasmus+ projects on the introduction of tablet computers in learning and teaching. Courses have been held for teachers (Progress report, School A). Information technology has increased diversity. We have, for example, become very advanced in programming (T, School E).
			Collaboration and dialogue	11	In teaching programmes, emphasis was placed on students' collaborative learning (Progress report, School A).

(*continued on next page*)

(*continued*)

	Level 1	Level 2	Level 3	Ref.	Example (s)
					We're talking about using every single shot. group here, group there. you see it when you walk around (T, School B). We broke down a wall and put the teenagers in one class, and we bought a round table instead of a rectangular one to facilitate cooperation (P, School E).
			Choice of optional subjects	19	The choice of subjects at the adolescent level has been increased, and students were allowed to influence the optional subjects that were made available. (Progress report, School B).
			Students' areas of interest	10	Teachers revised their syllabi and added more choice for students and individualised learning objectives (Progress report, School D). We have workshops where they select a project based on students' interests (P, School A). Linking learning to interests: we're doing it really well (P, School B).
	Level 1	**Level 2**	**Level 3**	**Ref.**	**Example(s)**
Instrumental use	Learning and teaching	Student achievement and use of assessment		20	We worked systematically with reading during the school year and presented a school reading policy (Progress report, School F). This [increase student achievement in mathematics] is a long-term goal. We started by increasing the number of lessons in mathematics at the youngest level (Progress report, School D). It [increase student achievement in mathematics] has been an ongoing process. Teaching methods were reviewed. A mathematics course was held for teachers (Progress report, School A). I don't think that we've changed much about it [systematically work with test results to improve achievement] (P, School E). It [identify the causes of poor performance in standardized examinations] was only discussed at the intermediate level and not at the adolescent level (T, School F).
		Students' democratic participation		33	I would actually say that today we have a more powerful student council (T, School C). We tried democratic meetings with students, but it didn't work (P, School B). It came as a result of this [the ev. report] … this democracy… to allow them to have a choice about their studies and the study material (T, School D).
		Team teaching and planning		17	We introduced team teaching to strengthen cooperation and restraint and feedback (T, School C). We've become an entirely team-teaching school, almost entirely … and we've developed it, I think, in a very successful way (P, School A).
	Internal evaluation	Changes		36	We didn't work systematically according to the improvement plan regarding the internal evaluation because it was just another project that took a lot of time (P, School C). We completely revised the internal evaluation, and we linked it to the curriculum as we were advised. There's always a four-year plan for what we're going to evaluate (P, School D).
		Evaluation team		29	Look, the evaluation team just hasn't been active for a long time … or since we were in all this work [making an improvement plan] (T, School C). No, we don't have an evaluation team this winter, but we had it then [when ex.ev. took place] (T, School B). The evaluation team isn't active now (P, School E).
		Knowledge to perform internal evaluation		11	I want to see a change in our internal evaluation but X, who's the assistant principal and leads the internal evaluation, hasn't wanted to make any changes. I think it's because of his lack of professional knowledge about internal evaluation. But now, a new assistant principal is coming, so there's an opportunity to change the way that we work, and I want to change it (P, School A). I outsource the internal evaluation. I use Skólapúlsinn, and they bring the results back in a user-friendly way. There's no one here in house who needs to have knowledge. We get those numerical results, and we can come up with improvements. I did it myself once, created a questionnaire and something. but the result was not useful (P, School E).
	Level 1	**Level 2**	**Level 3**	**Ref.**	**Example(s)**
Instrumental use	Internal evaluation	Publication and improvements		21	We made an improvement plan in 2017. I think it was our last improvement plan. [*Laughs.*] It's high time for another, I think. (T, School C). We always publish it [internal evaluation report], it's always accessible to everyone, and there need to be improvements

(*continued on next page*)

(*continued*)

	Level 1	Level 2	Level 3	Ref.	Example (s)
					based on the results–not just report on the situation–so we present an action plan (P, School D). Look, we've compiled reports and put them on the website, but we don't do that now. We just don't think that it matters now somehow (P, School C). We're supposed to publish it [internal evaluation results] on the website, but we don't. And if we were completely professional, then we'd make an improvement plan, write it down, but we haven't done it, but I will do it this winter (P, School E).
Conceptual use	Usefulness of external view			10	You went like this. "Yes—A-ha". and it was incredibly beneficial. (T, School B). There were of course certain factors that were very good to get such an external view of. getting an outside party to come up with suggestions on what could be done better (T, School F).
	Discussions and reflections			4	We had the opportunity to go deeper into things and what it is that we could improve (T, School D). I think this [ev. report] has created a professional discussion, and we've benefited a lot from it (P, School D).
	Professionalism			8	I think this [ex.ev. feedback] is good to keep us professional. … It isn't an attack, you know. It's just a reminder, really (T, School C).
	Support for new principals			6	I learned a lot from it [ex. ev.]. if I can speak completely personally, and I think it's a great tool. I learned a lot just about being a principal (P, School F). As a new principal, I could go in there and see exactly what needs to be improved (P, School B).
	Create focus			2	It [ex.ev.] helped us to focus on what needed to be done (P, School B). This [ex.ev.] has helped the school quite enormously, because it's a really good guide (P, School E).
Persuasive use				6	In fact, it's just a good tool, because I can say, "This is reflected in the external evaluation report, and we need to work on it" (P, School B). What I found helpful. because this [ev. results] were in line with my views. The practice that was being asked for. diversity and integration and all that. a great interest of mine. so it was such a good tool for me to get people to join me, you know. because I hadn't been here for that long and was still creating a niche for myself. and it helped. It was a professional document that I could use and quote to get people more oriented towards what I was aiming for. [.] That way you can better lead people in the same direction. (T, School D).
	Level 1	**Level 2**	**Level 3**	**Ref.**	**Example(s)**
Reinforcement-oriented use				6	We were just very proud because the school came out really well (T, School C). Of course, it's a pleasure to be able to show others what we're doing good things. You feel good about having the opportunity to do so (P, School D).
Sustained improvements	Sustained or progressing			28	The improvements aren't over. It's in development, and the improvements remain in progress (P, School C). We constantly have to keep working on it [the improvements] because if we don't, then everything will go the same way again (P, School E). Of course, a lot has changed since this report was made, and we remain in progress (T, School C). We've done a lot, and it [the changes] has simply become part of our daily work (T, School B). I think that the improvements have mostly been sustained, at least regarding the aspects that we've been discussing (T, School E).
	Professional knowledge in school			11	I think that our professional knowledge is good when you add it all up. We're a very active group in lifelong learning, which is of course part of being able to deal with this [ev. feedback] (T. School F). We had really good knowledge about how to do that kind of work [decide on improvements], and I think that people were active in it (P, School C). The group of teachers here—and the professional group as a whole—the majority have a very strong professional vision and are always striving to do things better (P, School A). I think that knowledge about working on improvements is available in a lot of people here … but not in everyone (P, School B).
	Restraint			7	Just wow—what great progress we've made. We did this and that and that … and we probably wouldn't have ever done it if

(*continued on next page*)

(*continued*)

	Level 1	Level 2	Level 3	Ref.	Example (s)
					it hadn't been for the external evaluation. … It wouldn't have been like this (T, School B). I think that this [ex.ev.] is a certain restraint on the schools. knowing that school practices are being monitored (T, School F).
	Ministry follow up			10	I'm happy with the follow-up. It's formal, and it's clear. it's just simple, not complicated. it's just asking for information on the progress of the improvements. It helped (P, School E).
	Municipality support			17	The municipality could, for example, come and sit with us in meetings and work on it and not only be some kind of regulator (P, School C). Project manager at the school office helped me to decide the focus and set up the plan (P, School A). The education committee was completely inactive, and once we'd finished our work, when they were supposed to discuss the progress. they never discussed it, and it was just a mess at the end with the Ministry. endless correspondence (P, School F).
	Level 1	**Level 2**	**Level 3**	**Ref.**	**Example(s)**
Attitude towards external evaluation	Attitudes			45	That external parties come to evaluate the school's activities is very important. Evaluation of the school system is necessary, and it should be done far more often (T, School C). I found the evaluation very positive and professional, and I'd like it much more often (T, School A). I think that our school benefited from the external evaluation. quite directly (T, School F). I thought that it [ex.ev.] was very beneficial for us. I was very satisfied (P, School D). It [ex.ev.] is just one of the best tools that I've ever received. I was incredibly grateful for it, and I thought that it was really good, and I just want to see it every five years. (P, School A). This [ex.ev.] benefited all of the professional work. I want to say that it's what all schools have to go through. just completely regularly and purposefully (P, School F).
	Stress prior to the external evaluation			8	When this [ex. ev.] came, there was such anxiety; people were slightly stressed at school. But I found it just fun above all (T, School D). I remember a teachers' meeting where it [the ex. ev.] was announced, and it was like "Okay, we're just lucky. Not everyone gets it, and it's an opportunity to make a good school better". I think that it set the tone (T, School B). It was stressful, yes, I remember that. But I think that everyone thought that it was okay once it started and we just kept working like we used to (T, School E). I told the staff, "I'm not going to beautify anything because we just want to be seen as we are so that we can see our situation and where we need to improve" (P, School A).
	Recommendations acceptance/ resistance			4	There were responsible, good recommendations, and then there were small things that I didn't agree with (P, School E). There were issues that we were extremely happy to get recommendations on. Other issues we may not have found important and maybe not even in line with our policy (P, School C).

References

Aderet-German, T., & Ben-Peretz, M. (2020). Using data on school strengths and weaknesses for school improvement. *Studies in Educational Evaluation, 64*, Article 100831. https://doi.org/10.1016/j.stueduc.2019.100831

Alkin, M. C., & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation, 29*(1), 1–12. https://doi.org/10.1016/S0191-491X(03)90001-0

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

Baughman, S., Boyd, H. H., & Franz, N. K. (2012). Non-formal educator use of evaluation results. *Evaluation and Program Planning, 35*(3), 329–336. https://doi.org/10.1016/j.evalprogplan.2011.11.008

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

Blok, H., Sleegers, P., & Karsten, S. (2008). Looking for a balance between internal and external evaluation of school quality: Evaluation of the SVI model. *Journal of Education Policy, 23*(4), 379–395. https://doi.org/10.1080/02680930801923773

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Brown, M., McNamara, G., O'Hara, J., & O'Brien, S. (2017). Inspectors and the process of self-evaluation in Ireland. In J. Baxter (Ed.), *School inspectors: Policy implementers, policy shapers in national policy contexts* (pp. 71–96). Springer.

Chapman, C. (2002). Ofsted and school improvement: Teachers' perceptions of the inspection process in schools facing challenging circumstances. *School Leadership & Management, 22*(3), 257–272. https://doi.org/10.1080/1363243022000020390

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

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage.

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

Ehren, M. C. M. (2016). School inspections and school improvement: The current evidence base. In M. Ehren (Ed.), *Methods and modalities of effective school inspections* (pp. 69–85). Springer. https://doi.org/10.1007/978-3-319-31003-9.

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*, 3–43. https://doi.org/10.1007/s11092-012-9156-4

Ehren, M. C. M., & Bachmann, R. (2021). Accountability to build school and system improvement capacity. In M. C. M. Ehren, & J. Baxter (Eds.), *Trust, accountability and capacity in education system reform* (pp. 102–123). Routledge.

Ehren, M. C. M., & Baxter, J. (2021). Trust, accountability and capacity: Three building blocks of education system reform. In M. Ehren, & J. Baxter (Eds.), *Trust, accoubntability and C capacity in education system reform* (pp. 1–29). Routledge.

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., & Visscher, A. J. (2008). The relationships between school inspections, school characteristics and school improvement. *British Journal of Educational Studies, 56*(2), 205–227. https://doi.org/10.1111/j.1467-8527.2008.00400.x

Fullan, M., & Quinn, J. (2016). *Coherence: The right drivers in action for schools, districts, and systems*. Sage.

Gärtner, H., Wurster, S., & Pant, H. A. (2014). The effect of school inspections on school improvement. *School Effectiveness and School Improvement, 25*(4), 489–508. https://doi.org/10.1080/09243453.2013.811089

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*, Article 100864. https://doi.org/10.1016/j.stueduc.2020.100864

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation, 30*(3), 377–410. https://doi.org/10.1177/1098214009341660

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation, 88*, 5–23.

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

McCrone, T., Rudd, P., Blenkinsop, S., Wade, P., Rutt, S., & Yeshanew, T. (2007). *Evaluation of the impact of Section 5 inspections*. National Foundation for Educational Research. ⟨https://www.nfer.ac.uk/publications/ISH01/ISH01.pdf⟩.

Nunneley, R. D., King, J. A., Johnson, K., & Pejsa, L. (2015). The value of clear thinking about evaluation theory: The example of use and influence. In C. A. Christie, & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin* (pp. 53–71). Information Age Publishing.

Ofsted. (2015). *School leaders' views on the impact of inspection*. Retrieved from ⟨https://www.gov.uk/government/publications/school-leaders-views-on-the-impact-of-inspection⟩.

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

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. Retrieved from ⟨http://netla.hi.is/greinar/2016/ryn/14_ryn_arsrit_2016.pdf⟩.

Ólafsdóttir, B., Jónasson, J. T., Sigurðardóttir, A. K., & Aspelund, T. (2022). The mechanisms by which external school evaluation in Iceland influences internal evaluation and school professionals' practices. *Nordic Journal of Studies in Educational Policy, 8*. https://doi.org/10.1080/20020317.2022.2076376

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Sage.

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

Penninckx, M., Vanhoof, J., De Maeyer, S., & Van Petegem, P. (2016aa). Effects and side effects of Flemish school inspection. *Educational Management Administration & Leadership, 44*(5), 728–744. https://doi.org/10.1177/1741143215570305

Penninckx, M., Vanhoof, J., De Maeyer, S., & Van Petegem, P. (2016bb). Explaining effects and side effects of school inspections: A path analysis. *School Effectiveness and School Improvement, 27*(3), 333–347. https://doi.org/10.1080/09243453.2015.1085421

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Sage.

Schweinberger, K., Quesel, C., Mahler, S., & Hochli, 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]. Retrieved from ⟨http://www.samband.is/media/mat-og-rannsoknir-a-skolastarfi/Vidmid-um-gaedi-i-skolastarfi.pdf⟩.

Six, F. (2021). Trust-based accountability in education. The role of intrinsic motivation. In M. C. M. Ehren, & J. Baxter (Eds.), *Trust, accountability and capacity in education system reform* (pp. 55–77). Routledge.

Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2018). Instrumental, conceptual and symbolic effects of data use: The impact of collaboration and expectations. *Educational Studies, 44*(5), 521–534. https://doi.org/10.1080/03055698.2017.1382328

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

Vo, A. T. (2015). Foreword. In C. A. Christie, & A. T. Vo (Eds.), *Evaluation use and decision making in society: A tribute to Marvin C. Alkin (pp. vii–xix)*. Information Age Publishing.

Weiss, C. (1998). Have we learned anything new about the use of evaluation. *The American Journal of Evaluation, 19*(1), 21–33. https://doi.org/10.1016/s1098-2140(99)80178-7