The origin, use, and perceived impact of the external evaluation of compulsory schools in Iceland

Björk Ólafsdóttir

Dissertation towards the degree of Doctor of Philosophy

October 2023

School of Education

UNIVERSITY OF ICELAND

The origin, use, and perceived impact of the external evaluation of compulsory schools in Iceland

Björk Ólafsdóttir

Dissertation towards the degree of Doctor of Philosophy

Supervisor(s)

Professor emeritus Jón Torfi Jónasson, University of Iceland
Professor Anna Kristín Sigurðardóttir, University of Iceland

Doctoral committee

Professor emeritus Jón Torfi Jónasson, University of Iceland
Professor Anna Kristín Sigurðardóttir, University of Iceland
Professor emeritus Rúnar Sigþórsson, University of Akureyri

Opponents

Professor Melanie Ehren, Vrije Universiteit Amsterdam
Professor Elisabet Nihlfors, Uppsala University

Dissertation towards the degree of Doctor of Philosophy

October 2023

School of Education

UNIVERSITY OF ICELAND

Uppruni, notkun og áhrif ytra mats á skólastarf grunnskóla á Íslandi

Björk Ólafsdóttir

Leiðbeinendur

Prófessor emeritus Jón Torfi Jónasson, Háskóli Íslands
Prófessor Anna Kristín Sigurðardóttir, Háskóli Íslands

Doktorsnefnd

Prófessor emeritus Jón Torfi Jónasson, Háskóli Íslands
Prófessor Anna Kristín Sigurðardóttir, Háskóli Íslands
Prófessor emeritus Rúnar Sigþórsson, Háskólinn á Akureyri

Andmælendur

Prófessor Melanie Ehren, Háskólinn Vrije í Amsterdam
Prófessor Elisabet Nihlfors, Háskólinn í Uppsölum

Ritgerð til doktorsgráðu

Október 2023

Menntavísindasvið

HÁSKÓLI ÍSLANDS

Ágrip

Viðfangsefni þessarar doktorsrannsóknar er ytra mat á grunnskólum á Íslandi. Rannsóknin er tvíþætt og miðar í fyrsta lagi að því að auka skilning á hvernig ytra mat á grunnskólastarfi er tilkomið og hver þróun þess hefur verið frá því að það kom inn í opinbera menntastefnu á tíunda áratug 20. aldar. Í öðru lagi að varpa ljósi á viðhorf kennara og skólastjóra til endurgjafar í skýrslum um ytra matið og hvaða áhrif endurgjöfin hefur til umbóta og breytinga á námi og kennslu, stjórnun og innra mati að þeirra áliti. Mikilvægi rannsóknarinnar felst fyrst og fremst í því að niðurstöður geta aukið skilning á því hvernig endurgjöf ytra mats er nýtt af skólum og hvort matið hafi tilætluð áhrif. Rannsóknin getur í því samhengi gefið hagnýtar upplýsingar til að nota við frekari þróun á ytra mati og eftirfylgd með því. Til þessa hafa engar rannsóknir verið gerðar á ávinningi af ytra mati á skólum á Íslandi og því mikilvægt að afla skilnings á ferlinu og nýta hann við frekari þróun.

Rannsóknin var unnin með blönduðu rannsóknarsniði (e. mixed method design) og stuðst við viðhorfakannanir, viðtöl og greiningu fyrirliggjandi gagna. Rannsókninni var skipt í þrjá rannsóknaráfanga og var mismunandi rannsóknaraðferð notuð í hverjum áfanga. Í fyrsta áfanga rannsóknarinnar var gerð greining á lögum, reglugerðum, stefnuskjölum, skýrslum og öðrum opinberum gögnum ásamt því að taka viðtöl við ellefu lykilaðila sem höfðu tekið þátt í stefnumótun og/eða þróun ytra mats. Í öðrum hluta var lagður spurningalisti fyrir skólastjóra og kennara í þeim 22 grunnskólum sem fyrstir tóku þátt í reglubundnu ytra mati á árunum 2013 til 2015. Í þriðja hluta voru tekin viðtöl við sex skólastjóra og átta kennara í sex af þeim 22 grunnskólum sem svöruðu viðhorfakönnuninni ásamt því að greina matsskýrslur, umbótaáætlanir og framvinduskýrslur skólanna.

Með birtingu greinanna þriggja er leitast við að ná meginmarkmiðum rannsóknarinnar sem eru að varpa ljósi á annars vegar hvernig ytra mat á grunnskólastarfi er tilkomið og hvernig það hefur þróast til dagsins í dag. Hins vegar að hvaða marki væntingar til skólanna um að nýta niðurstöður matsins til umbóta hafa orðið að veruleika og hvaða áhrif matsendurgjöfin hefur haft á breytingar á námi og kennslu, stjórnun og innra mati að mati kennara og skólastjóra. Niðurstöður sem lúta að fyrra markmiðinu sýna að upphaf hugmynda og mótun stefnu um ytra mat á grunnskólum megi rekja til umbótaaðgerða í opinberri stjórnsýslu sem áttu sér stað víða um heim frá níunda áratug síðustu aldar undir heitinu nýskipan í ríkisrekstri (e. New Public Management). Ytra mat náði þó ekki fótfestu á Íslandi á þeim tíma. Það var ekki fyrr en í byrjun annars áratugar tuttugustu og fyrstu aldar sem farið var af stað með reglubundið ytra mat á grunnskólum í þeim tilgangi að stuðla að skólaumbótum. Allir skólar á landinu utan Reykjavíkur voru

metnir einu sinni á níu ára tímabili frá 2013 til 2021. Niðurstöður sem tengjast seinna markmiðinu gefa til kynna að ytra mat hafi stuðlað að umbótum á starfsháttum kennara og skólastjóra og einnig á innra mati í flestum skólanna, þó svo að umbætur tengdar innra mati hafi reynst sumum skólanna sem tóku þátt í rannsókninni erfiðar. Endurgjöf í skýrslu með niðurstöðum ytra matsins nýttist skólunum í ýmsum tilgangi: til að stuðla að breytingum, vekja starfsfólkið til vitundar og umhugsunar, réttlæta breytingar sem voru líklegar til að mæta fyrirstöðu og styrkja og valdefla skóla. Niðurstöður sýndu að viðhorf meðal kennara og skólastjóra til endurgjafar ytra mats var að jafnaði gott og þeir voru almennt samþykkir henni—en samþykkt endurgjafar reyndist vera sá þáttur sem hafði mest áhrif á hvort niðurstöður væru nýttar til umbóta innan skólanna. Fleiri þættir höfðu einnig jákvæð áhrif á breytingar, svo sem þátttaka kennara í að ákveða umbótaðgerðir í kjölfar ytra matsins og þekking og notkun skólanna á viðmiðum um gæði í skólastarfi.

Lykilorð:

ytra mat á skólastarfi, endurgjöf ytra mats, umbætur á skólastarfi, notkun matsniðurstaðna, áhrif ytra mats

Abstract

This thesis focuses on the external evaluation of compulsory schools in Iceland. Aligned with that focus, the aim of the research conducted for the thesis was twofold: first, to shed light on the origin of the external evaluation of compulsory schools in Iceland and its development since becoming part of official education policy in 1991; and second, to identify and analyse school principals' and teachers' attitudes towards a recent external evaluation, how and to what extent they have used the feedback from the evaluation, how such use has affected internal evaluation at their schools and driven change in their own practices, and how well the changes implemented have been sustained over time. To that aim, longitudinal, mixed-methods research was conducted involving document analysis, questionnaires, and interviews. The primary data were collected in three stages. First, policy documents and interviews with 11 key informants who had participated in the policymaking process or were familiar with the process were examined. That part of data collection was undertaken in 2015–2016, and an article presenting the results was published in 2016 (i.e. Paper I). Second, the 22 compulsory schools that were the first to participate in the external evaluation in 2013–2015 were identified, and a survey of the principals and teachers of those schools was conducted in 2016. Those results were published in another article in 2022 (i.e. Paper II). Third, six of the 22 schools were selected for further analysis, and the principal and one to two teachers in each school were interviewed, followed by a document analysis of each school's improvement process. That final part of data collection was undertaken in 2019, 4–6 years after the external evaluation took place, and the findings were published in 2022 (i.e. Paper III).

The overall results are discussed in Chapter 8 of this thesis. In line with the two aims of the research, that discussion is divided into two parts. The first sheds light on how external evaluation at compulsory schools in Iceland came about and how it developed after becoming part of the country's official education policy in the early 1990s. That part focuses on the historical background and expected benefits of the external evaluation of schools in Iceland. After that, the second part discusses the extent to which the anticipated benefits of the evaluation, especially regarding the use of the feedback provided, and its impact have been realised. Several key factors expected to promote the improvement of schools are highlighted, and their effects in the Icelandic context are assessed. Based on the most significant findings, the thesis concludes that external evaluation can play a role in changing not only teachers' and principals' practices but also the internal evaluation of schools. The feedback from evaluation is shown to be used for various purposes, and the factor with the greatest overall impact on such use is shown to be the school staff's acceptance of the feedback.

Because this thesis is inspired by pragmatism, it aims to contribute to the existing knowledge base and to expand understandings of how schools use feedback from external evaluation and whether such use impacts changes in practice. Beyond that, it seeks to highlight important ways to improve the role of external evaluation in national and local school governance.

Keywords

external school evaluation, evaluation feedback, school improvement, evaluation use, evaluation impact

Acknowledgements

This thesis is the result of the efforts of many individuals who have supported me through the various stages of my research and to whom I owe a debt of gratitude. Indeed, without their support, I would not have completed my journey towards becoming a researcher.

First, I would like to thank my primary supervisor, Dr Jón Torfi Jónasson, for his valuable support, encouragement, and guidance throughout my whole doctoral study. His enthusiasm and academic interest in my research topic, his vast knowledge and wisdom about research in general, and his knack for giving constructive feedback instead of criticism have all been invaluable resources during my time as a doctoral student. To my co-supervisor, Dr Anna Kristín Sigurðardóttir, I also extend my deepest gratitude, especially for her assistance and insightful comments and suggestions. Both Dr Jónasson and Dr Sigurðardóttir have been exceptional supervisors, role models, and teachers. Their profound patience and countless comments throughout my research process have enabled me to develop as a researcher as well as a writer. I also want to thank Dr Thor Aspelund for his valuable contribution to Paper II; Dr Rúnar Sigþórsson and Dr Gerry MacNamara for their helpful feedback on the research proposal at the interim evaluation; and Dr Sigþórsson in particular for his constructive comments on a draft of the thesis. I would additionally like to express my gratitude to Dr Melanie Ehren and Dr Elisabeth Nihlfors for providing feedback on the thesis before and during its defence.

To the former administrator at the Directorate of Education, Dr Arnór Guðmundsson, and to my manager at the Directorate, Sverrir Óskarsson, I wish to express my appreciation for both giving me the opportunity to work part-time and being flexible with my schedule as I completed my studies. I would also like to thank my colleagues and friends within the Directorate, namely Þóra Björk Jónsdóttir, Hanna Hjartardóttir, Gunnhildur Harðardóttir, and Svanhildur María Ólafsdóttir, for their patience and support throughout my research. I additionally wish to thank the principals and teachers who took time out of their busy workdays to participate in my research. I am indebted to them for their time and generosity in sharing their experiences and attitudes.

I am extremely grateful to my husband, Kristján Zophaníasson, who has been especially encouraging and supportive throughout my academic journey, which I never could have undertaken without his emotional and financial support. He has endured my ceaseless refrain of "I can't—I'm working on my doctorate" with understanding, especially during the intensity of the final stages of writing. He always showed faith in me and let me know how proud he was of me. Last, I would like to thank my amazing three sons—Ingi Þór, Arnar Breki and Hugi Snær—my daughter-in-law, Ewelina, and my granddaughter, Díana Líf, for their patience, hugs, and steadfast support of me during my work—and for making my life fantastic.

Contents

List of Figures

List of Tables

List of Original Papers

This thesis is based on the following three original publications, which are referred to in the text as "Paper I", "Paper II", and "Paper III", respectively:

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

II. 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*(3), 209–224.
    https://doi.org/10.1080/20020317.2022.2076376

III. Ólafsdóttir, B., Jónasson, J. T., & Sigurðardóttir, A. K. (2022). Use and impact of external evaluation feedback in schools. *Studies in Educational Evaluation*, *74*. Article 101181.
    https://doi.org/10.1016/j.stueduc.2022.101181

All three papers were published with open access and with the authors' right to redistribute and republish them. Obtaining special permission to republish them in this thesis was therefore unnecessary.

Declaration of contributions

This thesis is the result of my own work. All persons who meet the criteria for authorship are listed as its authors, and all authors certify that they have participated in the work, including in the conceptualisation, design, analysis, writing, and/or revision of the manuscript, to an extent that justifies their taking public responsibility for its content.

CRediT author statement:

Björk Ólafsdóttir: Conceptualisation, Methodology, Formal analysis, Investigation, Data Curation, Writing — Original draft preparation, Review & Editing (Papers I, II, and III)

Jón Torfi Jónasson: Conceptualisation, Validation, Writing — Review & Editing (Papers II and III)

Anna Kristín Sigurðardóttir: Validation, Writing — Review & Editing (Papers II and III)

Thor Aspelund: Formal analysis (Paper II)

I did not receive any financial support for my research. However, because I worked throughout doctoral study, I undertook only a 50% study load during the period, which extended my total time of study.

1 Introduction

Today, most education systems in Europe engage in the external evaluation of schools—hereafter, "external school evaluation"—which they consider to be an important tool for managing and promoting the quality of schools (Ehren et al., 2013). In general, external evaluation has been viewed as a potential stimulus for change and support for decision-making about improvement in schools (Organisation for Economic Co-operation and Development [OECD], 2013). In Europe in particular, external evaluation has a long history in some countries and, since the 1990s, has enjoyed increased attention in increasingly more countries (OECD, 2013). In Iceland, the trend of external school evaluation began rather slowly. The municipality of Reykjavík, Iceland's capital, first evaluated its compulsory schools using so-called comprehensive evaluation in 2007 (Sigurjónsdóttir, 2010). In 2010, the Ministry of Education and Children began to evaluate compulsory schools at the national level,[1] albeit on a small scale and in a project that lasted only three years. In 2013, the external evaluation of compulsory schools was launched, in a separate project jointly funded by the state and municipalities outside Reykjavík. Both the comprehensive evaluation in Reykjavík and the joint evaluation project of the state and municipalities were primarily aimed at supporting schools and promoting improvements. However, when the external evaluation was launched in 2013, no studies had been performed on the extent to which comprehensive evaluation in Reykjavík had benefited the schools or contributed to their development, while international studies had shown inconsistent findings on the topic. That lack of empirical research and evidence to inform decision-making about external school evaluation in Iceland was the impetus for the research conducted for this thesis.

This thesis reports research on the external evaluation of compulsory schools in Iceland, particularly ones outside Reykjavík, viewed in the international context of research on external school evaluation and Iceland's education policy. In the thesis, I first examine the origin of the recent external evaluation of compulsory schools in Iceland and its development since appearing in official policy documents in 1991. Second, I analyse school principals' and teachers' attitudes towards the external evaluation, especially the ways in which they consider that it has affected their schools' internal evaluation and driven changes in their own practices. Third, I investigate the use and impact of feedback from the external evaluation as perceived by principals and teachers.

---

[1] The current name of the Ministry is used throughout the thesis, although until 2021 it was called the "Ministry of Education, Science and Culture".

Therefore, when discussing the findings of the research regarding changes resulting from the use of such feedback in schools, I am always referring to the perspectives of teachers and principals.

In 2018, the proposal for the doctoral project that culminated in this thesis was subjected to an interim evaluation. In accordance with the comments provided during the evaluation, the research's agenda changed, especially regarding the focus of the research. Initially, my intention was to chiefly examine the impact of the external evaluation on changes in internal evaluation; however, the evaluation committee suggested that I should examine additional changes occurring within the schools in the wake of the evaluation. That suggestion prompted me to shift the research's focus to the impact of external evaluation on the three key aspects that such evaluation in Iceland addresses: learning and teaching, leadership and management, and internal evaluation. Another suggestion was to pay systematic attention to the realisation of planned improvements, which later became the subject of one of the three papers written for this thesis. Other comments were aimed at explaining the underlying philosophy of the research and clarifying certain points, which I hope to have done in the thesis to the satisfaction of readers. I believe that all of those suggestions greatly improved the research and, in turn, have improved the thesis.

This thesis is intended to contribute to the body of knowledge concerning external school evaluation. I hope that it will assist governments, policymakers, and other interested parties in formulating and implementing future mechanisms, concepts, and resources needed to successfully implement external school evaluation.

1.1   Definitions of key concepts

Following is a brief clarification of some of the central terms used in the thesis: *evaluation, programme evaluation, school evaluation, external evaluation, internal evaluation, external school evaluation, school improvement, evaluation judgement, evaluation feedback,* and *evaluation use.*

Many definitions of *evaluation* have been offered over the years. One of the earliest was proposed by Scriven (1991), who defined *evaluation* as "the process of determining the merit, worth, or value of things" and *evaluations* as "the products of that process" (p. 1). Worthen et al. (1997) added to Scriven's definition by stating that *evaluation* is "the identification, clarification, and application of defensible criteria to determine an evaluation object's value (worth or merit), quality, utility, effectiveness, or significance in relation to those criteria" (p. 5).

Although evaluation can cover numerous evaluation objects, including programmes, products, people, process, and policies, this thesis focuses on the evaluation of programmes, namely external school evaluation, as an ongoing, planned intervention that seeks to achieve certain outcomes (Worthen et al., 1997). On that count, Rossi et

al. (2004) have defined *programme evaluation* as a sort of social science research "aimed at collecting, analysing, interpreting and communicating information about activity and effectiveness of social intervention programs" (p. 2). Scriven (1991), by contrast, has argued that that perspective is too narrow and that programme evaluation is a transdisciplinary process. In any case, evaluation involves a careful, systematic, or methodological approach, one that is planned and purposeful (Dahler-Larsen, 2018; Russ-Eft & Preskill, 2009). Patton (2002), in another definition of *programme evaluation*, has emphasised that such evaluation has the purpose of prompting action: "Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming" (p. 10).

Also according to Patton (2002), evaluation's purpose of prompting action is what distinguishes it from research, which has the primary purpose to "contribute to knowledge for the sake of knowledge" (p. 10). By comparison, evaluation may be used to improve the programme being evaluated, to develop organisational capacity, to empower people, to increase the understanding of a project, to inform decision-making, and/or to promote greater justice and equity (Greene, 2007; Worthen et al., 1997). Greene (2007) has pointed out that evaluation also differs from research in terms of the audience. Where the audience of research is primarily other researchers (Russ-Eft & Preskill, 2009), the audiences of evaluation can encompass policy- and other decision-makers, managers, board members, programme staff members, programme participants or clients, families, and/or communities (Greene, 2007; Weiss, 1998). In each of the two forms of inquiry, the communication and presentation of the results therefore take place in different ways and are aimed at different groups (Russ-Eft & Preskill, 2009). Another distinction between evaluation and research is that evaluations are used to serve various political functions and therefore reflect political and policy debates (Greene, 2007; Worthen et al., 1997). However, most relevant to the research conducted for this thesis is whether external evaluation is used to support decision-making and changes in schools.

Along other lines, formative evaluation has been distinguished from summative evaluation (Davíðsdóttir, 2008; Worthen et al., 1997). Evaluation is considered to be formative when its principal purpose is to provide information chiefly to parties who are working on the programme or responsible for it—in this thesis's case, the staff of the school and the municipality—so that it can be improved (Davíðsdóttir, 2008). Summative evaluation, by contrast, is meant to provide decision-makers and stakeholders—in the case of evaluating schools, the school's authorities and parents—with an overall judgement about a programme's performance and effectiveness on which to base their decisions about, for instance, actions, rewards and sanctions, the programme's continuation or discontinuation, or whether to continue or discontinue its use (Patton, 2002; Rossi et al., 2004; Worthen et al., 1997). The purpose of

summative evaluation primarily relates to accountability (Rossi et al., 2004). Despite those differences, formative evaluation and summative evaluation are often intertwined and difficult to distinguish in practice, and the results of an evaluation can be used for both summative and formative purposes (Worthen et al., 1997). However, as various scholars have indicated, tension can arise between those purposes when they are coupled together (Ehren, 2016a; Landwehr, 2011; Schweinberger et al., 2017).

On the topic of evaluating schools, the European network Eurydice (2015) has stated that school evaluation focuses on the school's activities and "seeks to monitor or improve school quality and/or student results" (p. 55). A report issued by the OECD (2013) has further explained that school evaluation "concentrates on key processes such as teaching and learning, school leadership, educational administration, school environment, and the management of human resources" (p. 384). The evaluation is performed with attention to learning outcomes and students' progress as well as considers other factors influencing the school.

School evaluation can be either external or internal. On the one hand, external school evaluation is conducted by evaluators who "are not directly involved in the activities of the school being evaluated" (Eurydice, 2015, p. 54). In such cases, the external evaluators report to the education authority that oversees the school. Internal school evaluation, on the other hand, is "undertaken by persons or groups of persons who are directly involved with the school" (Eurydice, 2015, p. 55). As in external school evaluation, various tasks can be evaluated in internal school evaluation, including teaching and management. All evaluations of a school conducted by the school itself are regarded as internal school evaluations.

According to Altrichter and Kemethofer (2015, p. 33), external school evaluation is a part of "evidence-based governance" in which governments set expectations in the form of quality criteria and assess the quality of education in individual schools in light of evidence collected during visits to schools and from evaluation and assessment instruments. Schools are held "accountable for a broad range of goals related to student achievement, teaching, organization, and leadership", with the aim to "stimulate school and system improvement by producing reports which point to strengths and weaknesses of individuals schools or the authorities in charge of them".

In this thesis, external school evaluation is used as a synonym and includes terms as *school inspection, school review*, and *whole school evaluation*. In Iceland, although the term *external school evaluation* is used instead of the most common term *inspection*, it is nevertheless considered to involve a process identical or similar to inspection and to be based on the same theoretical foundation. Likewise, in this thesis, I do not differentiate *internal school evaluation* from *self-evaluation*, and both terms are used interchangeably: the former in keeping with Icelandic practice, the latter in keeping with some of the literature.

At base, external school evaluation is a tool used to support school improvement. This thesis's definition of *school improvement* comes from Lander and Ekholm (2005) who have described it as "a process of deliberate change in structures, rules, norms, conceptions, habits and working patterns, which immediately, or over a longer period, help students to improve their learning and development according to requirements of school and society" (p. 86). Because the results of the research presented in this thesis reflect teachers' and principals' perceptions of the changes and improvements made as a result of external evaluation, it cannot be stated that the changes made qualify as improvements per the mentioned definition. To draw conclusions with caution, the term *change* is therefore used instead of *improvement* when discussing the results of the research.

In this thesis, external school evaluation is more specifically viewed as feedback instrument for schools (Behnke & Steins, 2017). Judgements produced in light of evaluation, or *evaluation judgements*, are generally presented in evaluation reports and can include an aggregate score for the school and/or an overview of strengths as well as aspects that requires improvement (Ehren, 2016a). The way in which schools use that feedback, or *evaluation feedback*, plays a key role in its potential impact (Verhaeghe et al., 2010). Feedback is generally considered to be an important element of learning (Ehren, 2016b), and to prompt a learning process in schools that expands knowledge and brings about school improvement in the wake of external evaluation, feedback on the quality of schools is essential (Schweinberger et al., 2017). In the context of evaluation, the term *evaluation use* is defined "as the application of evaluation processes, products, or findings to produce an effect" (Johnson et al., 2009, p. 378).

1.2 Personal motivation for the research and the positioning of the researcher

Given my involvement in the research conducted for this thesis, I dedicate this section to explaining my professional background and my relationship to the topic studied in the research.

My educational and professional background has played a significant role in shaping my interest as a researcher in internal and external school evaluation. My interest first arose in 2003, when I began pursuing a master's degree in the Faculty of Social and Human Sciences at the University of Iceland, in studies focused on the evaluation and development of schools. After earning my master's degree, I worked freelance for the Ministry of Education and Children, among others, on school evaluation and in advising compulsory schools on internal evaluation.

From 2009 to 2012, I worked full-time as a project leader on a temporary cooperative project between the Ministry of Education and Children and the Icelandic Association of Local Authorities. The project, supported by the Municipal Equalisation Fund,

involved implementing new legislation concerning compulsory schools that came into effect in 2008 with the Compulsory School Act. My task was to manage the implementation of new requirements for local governments stated in the act, which consisted of performing external evaluations of schools and establishing a general policy on the operation of the schools. Part of the project involved leading an ad hoc team of representatives of the Ministry and local authorities, whose role was to formulate an approach to external evaluation for compulsory schools (see Section 2.4). After the team presented its proposals in 2011, I was assigned to execute the work on the implementation of the proposed actions—that is, to formulate a framework for evaluation and to provide materials for guiding and supporting the evaluation. I also led the formulation of quality criteria that informed the work of evaluators, criteria that were formulated with representatives of both the Ministry and municipalities and sent to representatives of all interest groups for comment before publication. Following that work, I directed a pilot project on the evaluation of six schools. When my temporary appointment as project leader ended in 2012, I continued on the project development team (i.e. for the external evaluation) for another year. The external evaluation of compulsory schools across Iceland, except for Reykjavík, was launched in 2013, and that project was the subject of my doctoral research. Additional details about the project appear in Section 2.4.

In the fall of 2012, I was hired as an evaluation specialist by the City of Reykjavík's Department of Education and Youth, where I worked until 2015. My work there primarily consisted of supervising a comprehensive evaluation of preschools and both after-school and youth centres administered by the Department, leading the formulation of quality criteria, creating a framework for the external evaluation to be performed, providing training and support for evaluators, and supporting the implementation of internal evaluation at the same institutions. Another project involved engaging in the comprehensive evaluation of compulsory schools in Reykjavík and evaluating the status of their internal evaluation.

During that time, my colleagues and I in the Department of Education and Youth began to wonder whether the comprehensive evaluation of compulsory schools in Reykjavík, as well as the joint project between state and local associations on external evaluation, was impacting the schools as school authorities had expected. Because the evaluation demanded considerable human resources and time, I considered it to be important to answer questions about its usefulness and contribution to developing schools, especially in relation to how the evaluation could be modified to promote the intended impact. Given my interest and because the external and internal evaluation of schools in Iceland remains a largely unexplored topic, I decided to apply to a doctoral programme to investigate the matter and subsequently commenced doctoral study in late 2014. Because I was no longer directly involved in the joint project between the state and municipalities, I decided to focus my attention as a doctoral student on that project. In 2015, I began working for the Directorate of Education, which is responsible for,

among other things, the operation of the joint project on the external evaluation of compulsory schools on the behalf of the state and municipalities. There, my chief role is to participate in developing policy and leading the external evaluation of pre-primary schools. Since 2019, I have also been in charge of following up with external evaluations in preschools and compulsory schools.

According to Creswell (2014), a strategy for validating findings from research is to interrogate the bias that researchers bring to their studies and how their positions might shape the findings. Given my involvement in the external evaluation project, I have been particularly mindful of the effects of my subjectivity. Due to its interpretative nature, research always involves the risk of researcher bias during the analysis of findings. My particular relationship to my research as the researcher presented two major challenges. On the one hand, I was working as an evaluation specialist at the Directorate of Education, and my feelings towards evaluation and my knowledge about it might have influenced the processes of data collection and data analysis. On the other, I was examining a programme that I had helped to develop and had participated in since its inception. Having played a significant role in introducing the external school evaluation, I was obviously at risk of being biased towards many of the practices of evaluation and their indented effects. However, my deep knowledge of the project, sincere interest in it, and aim to offer insights and information for the project's further development as a means to benefit the education system and the schools may also be considered to be a strength of the research. I believe that my knowledge and experience have enhanced my awareness and understanding of the object of study. Although that argument is not strong enough to eliminate the possibility of bias, it does partly explain why I decided to study the project in the first place. I discussed that challenge at considerable length with my principal supervisor before I began the doctoral programme. He was adamant that the strengths far outweighed the obvious weaknesses, but we also agreed on the importance of being transparent about and cognizant of it.

Although I have not led the project about the external evaluation of compulsory schools since 2012 and have never participated in evaluating such schools, as an employee at the Directorate of Education, I nevertheless work closely with the evaluators of compulsory schools. I also actively participate in strategic planning in the field of external evaluation at all levels of education. That circumstance also risked conflicts of interest in my study that I, as well as my supervisors, had to constantly be mindful of throughout the research process. Indeed, throughout that period, I actively sought to acknowledge and confront my own opinions and subjectivity and to reconcile them with my role as the researcher.

1.3   Structure of the thesis

This thesis consists of two parts. Part 1 contains the main text of the thesis, after which Part 2 consists of the three papers (i.e. Papers I–III) and appendices. Part 1 is divided into nine chapters. Chapter 1 is this introduction, in which I have given an overview of the research, defined the key concepts, and explained my motivation for conducting the research. In Chapter 2, I discuss the background of the research and contextualise it in Iceland. Next, in Chapter 3, I review the analytical frameworks on which the research is based and offer an overview of other research in the field. In Chapter 4, I present apparent gaps in such studies and, in Chapter 5, present the chief aim and purpose of the research conducted for the thesis. Later in that chapter, I also provide an overview of the three papers, their aims, and their research questions. After that, in Chapter 6, I provide a detailed account of the methodological framework structuring the research. After introducing the perspective of pragmatism, drawing specifically on Deweyan pragmatism, I justify the pragmatic orientation of the research and explain the mixed-methods research design followed. I also give an overview of the methods used to collect and analyse the data and elaborate on ethical considerations made during the research process. Last, I discuss some advantages and limitations of the research. In Chapter 7, I summarise each of the three papers reported in this thesis and presented in Part 2. Next, in Chapter 8, I discuss the research's overall findings in light of its aim and purpose. The research's contributions and the implications of its findings are presented as well, along with a discussion of threats to validity of the research and suggestions for future studies. Chapter 9 concludes the thesis by summarising the major findings and offering some concluding remarks. Thereafter, Part 2 of the thesis presents Papers I–III and appendices.

2 Contextual background: Iceland

This section introduces the context of the research conducted for this thesis: schools and systems of school government in Iceland. After that, it describes the characteristics and aims of the recent external evaluation of schools in the country. It should be noted that, earlier in Iceland's history, inspection was performed as a type of school evaluation, in both the 1740s and the 1930s, as discussed by Jónasson, Bjarnadóttir, et al. (2021). Thus, from a certain perspective, the more recent evaluation—the subject of the thesis—can be regarded as the "third wave" of education or school evaluation in the country.

Iceland, with a population exceeding 387,000 (Statistics Iceland, 2023b), has two administrative levels of government: the state and municipalities (Eurydice, 2018). All children 6–16 years old are required to receive compulsory education, which generally lasts 10 years and is organised as a single structure system with primary and lower secondary education at the same school level (Eurydice, 2018). Compulsory education is provided in compulsory schools, and if a child's family lives in the same neighbourhood throughout their compulsory education, then the child usually attends the same compulsory school for the entire period. In 2022, there were 47,115 students in compulsory schools in Iceland (Statistics Iceland, 2023a), 33% of whom attend compulsory schools in Reykjavík, the capital and only city in the country.

Ratified in 2008, the Compulsory School Act regulates the goals, demands, and tasks for the Ministry of Education and Children, municipalities, and other school owners, as well as school committees, principals, school councils, school staff, students, and parents. The National Curriculum Guide for Compulsory Schools, with a legal status equivalent to regulations, is based on the Act and provides details of its implementation (Eurydice, 2018).

2.1 The Ministry of Education and Children and the state's obligations

Compulsory education in Iceland falls under the jurisdiction of the Ministry of Education and Children, which is responsible for, among other things, implementing the Compulsory School Act, developing and implementing regulations and the National Curriculum Guide for Compulsory Schools, and monitoring such implementation (Eurydice, 2018). The Ministry is also responsible for ensuring that local authorities fulfil their obligations pursuant to the Compulsory School Act and to regulations and rules based it, including the National Curriculum Guide for Compulsory Schools, and for

evaluating and monitoring the school system and individual schools (Compulsory School Act, 2008). As written in the legislation, "The Compulsory School Act obligates the Ministry, or by delegation the Directorate of Education, to undertake analysis and dissemination of information about compulsory schools based on information from the municipalities and their own data collection".

2.2 Obligations of municipalities and local governments

Iceland's municipalities are responsible for covering the costs of operating compulsory schools (Eurydice, 2018). The country currently has 64 municipalities, ranging in size from Reykjavík, with 135,688 inhabitants, to the smallest, Árneshreppur, with 42 inhabitants (Statistics Iceland, 2022). For comparison with Reykjavík, Iceland's second-largest municipality, Kópavogsbær, has 38,998 inhabitants. Approximately two-thirds of the country's population lives around the capital city, in Greater Reykjavík, which includes seven municipalities. Most municipalities are quite small; nearly half of them (45%) have fewer than 1,000 inhabitants, and only eight (12.5%) have more than 10,000 (Icelandic Association of Local Authorities, 2023).

Table 1 shows how small the populations in most Icelandic municipalities are. In fact, the country's 10 largest municipalities are home to slightly more than 80% of all compulsory school students (Statistics Iceland, 2022).

Table 1. The size of municipalities in Iceland

Population size	Number	Percentage
<500	16	25
500–1,000	13	20
1,001–2,500	15	23
2,501–5,000	9	14
5,001–10,000	3	5
10,001–40,000	7	11
>40,000	1	2
Total	**64**	**100**

According to the Compulsory School Act (2008), municipalities are responsible for the general organisation of school operations within their jurisdiction, as well as for developing individual schools, evaluating and monitoring them, collecting and disseminating information about schools, and implementing activities at the schools, among other things. Moreover, each municipality is required to maintain a school

committee, which acts on the municipal council's behalf and manages the affairs of compulsory schools as prescribed by law and regulations. One of the committee's chief tasks is to ensure that school operations accord with laws, regulations, and the National Curriculum Guide; another is to submit proposals to the municipal council about improvements. The municipalities are additionally responsible for providing school support services for the full range of activities and practices in compulsory schools and for their staff members (Regulation on Municipal Services to Preschools and Primary Schools and Student Support Councils in Primary Schools, 2019). According to the cited regulation, support for a school's staff should include counselling on the school's practices, innovation, and development-focused activities.

Owing to their small size, some of Iceland's municipalities do not run their own schools, and their students attend the school in the neighbouring municipality. Such municipalities do not have their own school committees but usually have a representative on the school committee of the receiving municipality. In all cases, compulsory education is financed primarily by local taxes; however, due to significant variation in the size of municipalities, the Municipality Equalisation Fund has the task of evening out the cost of operating compulsory schools (Icelandic Association of Local Authorities, n.d.). The Fund makes allocations to municipalities for use in balancing their potential revenue and necessary costs based on rules set for the Fund.

About half of the country's municipalities operating schools have a superintendent, who serves on the municipal school committee and manages the provision of local education. Normally, the superintendent is seniormost principal of the compulsory schools. In some cases, a few municipalities maintain a forum for cooperation and hire one superintendent to serve them all. In other cases, municipalities employ no superintendent, and the principal is a direct subordinate of the municipal director. Some smaller municipalities therefore lack the resources available to their larger counterparts (Jónasson, Ragnarsdóttir, et al., 2021). Local authorities are responsible for performing evaluation and quality assurance assessments in schools, as well as for providing the Ministry of Education and Children with information about the operation of schools, internal school evaluation, external evaluation performed by local authorities, the implementation of school policy, and plans for improvement (Compulsory School Act, 2008). They are also obliged to follow up on internal and external evaluations in order to ensure that they have brought about improvements in the schools.

A recent study on the degree to which Iceland's municipalities support compulsory schools revealed that participating municipalities do not fully comply with the provisions of the mentioned regulation on municipal services to schools (Sigurðardóttir et al., 2022; Sigþórsson et al., 2022; Svanbjörnsdóttir et al., 2021), particularly the part of the school service aimed at strengthening schools as professional institutions. Municipalities were also generally found to have not fulfilled their leadership duties and

to be more reactive than proactive (Sigurðardóttir et al., 2022). As highlighted by Jónasson, Ragnarsdóttir, et al. (2021), owing to each municipality's autonomy, their influence on schools tends to be fragmented and often fragile.

2.3 Compulsory schools and the obligations principals

The Compulsory School Act (2008) defines the roles of the principal of each compulsory school as involving the direction of the school, the provision professional leadership, and being responsible to the municipal council for the school's work. The principal, in collaboration with the school staff, is also responsible for the quality of the school's activities (National Curriculum Guide for Compulsory Schools, 2011). Each school is obliged to have a school council containing representative parents, students, teachers, and other school staff as well as the principal, who chairs the council (Compulsory School Act, 2008). The council is a forum for consultation between the principal and the school community regarding the operation and activities of the school.

Per the Compulsory School Act (2008), each compulsory school is required to carry out a systematic internal evaluation of the achievement and quality of the school's activities, with the active participation of the school's staff, students, and parents. Schools choose their own methods, focal areas, and criteria for the evaluation. Information about their internal evaluation, its relation to the National Curriculum Guide, and plans for improvements are required to be publicly disclosed. The findings of the evaluation should be used to improve the school's activities in consultation with the school council (National Curriculum Guide for Compulsory Schools, 2011).

2.4 Characteristics and aims of external school evaluation

In Iceland, a systematic, comprehensive evaluation of compulsory schools commenced in 2007 in Reykjavík, carried out by the municipality's school authorities. In 2008, new legislation required both the Ministry of Education and Children and the municipalities to evaluate compulsory schools (Compulsory School Act, 2008), and, in 2010, the Ministry began evaluating six schools annually and continued to do so until 2012 (Ólafsdóttir, 2016a). During that time, an ad hoc team was founded with representatives of the Ministry and local authorities, whose role was to formulate an approach to external evaluation and discuss the idea of cooperation between the Ministry and local authorities on regular external evaluation of compulsory schools (Ad Hoc Team on the External Evaluation of Compulsory Schools, 2011). The ad hoc team presented its proposal in a report in 2011, recommending that evaluation of compulsory schools to be performed jointly by municipalities and the state. In the team's proposal, it was further recommended that the comprehensive evaluation of compulsory schools in Reykjavík would be considered as a model when formulating the external evaluation. In 2013, following a pilot project in 2012, an approach to external school evaluation

developed in a cooperative project between municipalities and the state and jointly financed by the Ministry and the Municipal Equalisation Fund was implemented in the Icelandic education system (Ólafsdóttir, 2016a). Reykjavík's school authorities, however, waived participation in favour of continuing to evaluate the municipality's compulsory schools on their own. Even so, the same evaluation process and criteria were used in Reykjavík as in the joint evaluation project. Iceland has 173 compulsory schools, 44 of which are in the municipality of Reykjavík. Therefore, the joint project—which is the subject of the research conducted for this thesis—covered the evaluation of 129 compulsory schools.

In 2013, the Educational Testing Institute, renamed the "Directorate of Education" in 2015, was tasked with performing the evaluation of schools on behalf of the state and municipalities (Ólafsdóttir, 2016a). The overall programme for the external evaluation took the form of a cyclical evaluation. The annual selection of schools was based on their distribution across the country and among the municipalities, and urban and rural areas as well as small and large schools had to be represented. For the first five years, 10 schools were evaluated annually, and that number increased to 27 in 2018 (Ólafsdóttir et al., 2022). By late 2021—nine years after the external evaluation was first implemented—all compulsory schools in Iceland had been evaluated once. In 2022 and in the first half of 2023, no external evaluation of compulsory schools was performed. It has been decided to close down the Directorate of Education in 2023, and it remains uncertain what will happen to the continuation of external school evaluation in the country.

The external evaluation of compulsory schools had four primary objectives: (1) providing information about schooling, its outcomes, and the development of schools and making that information available to interested parties; (2) ensuring that schools operate in accordance with the law, regulations, and the National Curriculum Guide for Compulsory Schools; (3) stimulating school improvement and increasing the quality of education; and (4) ensuring that students' rights are respected and that students receive the services entitled to them by law (Compulsory School Act, 2008). Furthermore, emphasis was to be placed on enhancing and supporting schools' internal evaluation and quality assurance, motivating teachers to collaborate on school improvement and stimulating the development of schools, and supporting schools, principals, and teachers in improving their practices. The central aspiration was that the external evaluation would support the development of school processes and improve outcomes among students (Ad Hoc Team on the External Evaluation of Compulsory Schools, 2011).

The evaluation focused on three major areas representing the quality of schools: the quality of learning and teaching, the quality of school leadership and management, and the quality of internal evaluation (Sigurjónsdóttir et al., 2012b). The adopted criteria elaborated on legal requirements and set further expectations for the performance of

schools and the quality of education. Beyond that, the three areas were subdivided into 16 criteria of quality. The area of learning and teaching included quality criteria regarding the curriculum, student achievement, teaching, study arrangements, students' active learning, and students' responsibility and participation. The area of leadership, by contrast, included criteria regarding school policy, professional cooperation, parent–school relationships, improvements and changes, procedures and plans, personnel management, and leadership. Last, the area of internal evaluation consisted of planning, implementation, and improvement (Directorate of Education, 2018). Each quality criterion had indicators in the form of statements and was intended to facilitate the evaluators' decision-making and improve the precision of evaluation. The continuum used to evaluate the indicators had four levels: (1) "Significant strengths", (2) "More strengths than weaknesses", (3) "More weaknesses than strengths", and (d) "Significant major weaknesses". In addition to the fixed evaluation themes, local education authorities, in consultation with the school, could request a more detailed evaluation of certain aspects of the school.

No threshold levels were used to rank schools as failing or being satisfactory; instead, the evaluation identified only how schools were functioning in terms of the evaluation criteria. The schools faced no consequences such as sanctions or interventions for failing to meet the standards. However, if a school failed to comply with the law, then the Ministry of Education and Children issued serious comments on the violation(s). Because the evaluation focused on processes in schools instead of outcomes, its results were not intended for use to compare schools. Emphasis was placed on making the evaluation process transparent by publicly publishing information about the process and the documents involved therein.

Each external evaluation was conducted by a pair of evaluators: a formal leader, typically employed by the Directorate of Education, and an external contractor. Both individuals had to meet certain criteria, primarily regarding their competence and independence (Ólafsdóttir, 2012). Concerning competence, they had to have education in and experience with teaching as well as knowledge and skills in methods of evaluating a school's practices. Concerning independence, they were not allowed to be personally related to the schools, to have direct interests in the results, or to evaluate schools that they had provided services to in the past. Those safeguards were applied to increase the reliability and validity of evaluation. Beyond that, mutual respect and trust between the evaluators and schools were considered foundational for the development of improvement actions in schools. The external contractors were mainly former employees of the school system, including former principals, teachers, or professionals in municipal school offices.

The evaluation process involved three stages: pre-evaluation, in-school evaluation, and post-evaluation (Sigurjónsdóttir et al., 2012a). First, pre-evaluation involved communication with the school to be evaluated and the analysis of documents provided

by the school. Evaluators requested a range of information from the schools, including about its curriculum, operational plan, results from internal evaluations, improvement plans, lesson plans, plans for retraining staff, and class timetables. Moreover, from the Directorate of Education, they received data on students' performance and progress on standardised national tests conducted by the Directorate. Evaluators received a list of teachers, non-teaching staff, students, and parents, and from those groups, individuals were chosen at random to participate in focus group interviews.

Second, during in-school evaluation the pair of evaluators visited the school and performed the evaluation. Each visit took 2–5 days but sometimes longer depending on the size of the school. A large proportion of the evaluators' time during the visit was spent on classroom observation and providing oral feedback to individual teachers based on observations. Overall, approximately 60%–70% of the teachers received classroom visits from the evaluators. The school visit also included interviews and focus groups with the principal, middle management, members of the school council, teachers, non-teaching staff, students, and parents.

Third and last, post-evaluation entailed the analysis of documents, interviews and observations, along with reporting the results. Based on the evaluation criteria, judgements on strengths and recommendations for improvements were issued to the school and the local authority in a written report. Before the final version of the report was issued, a draft of the report was given to the principal for comment. The post-evaluation phase ended with a meeting with the staff, the members of the school council, and the local school authorities at which the major findings were presented. All parents received a letter summarising the evaluation results, and the full report was made public on the Directorate of Education's website.

Regardless of the evaluation judgement (i.e. weak vs. strong), the school was required to develop an improvement plan in collaboration with the local school authority that addressed how it was going to implement the report's recommendations. The improvement plan had to include agreed-upon targets along with a plan and timeline to implement measures to achieve the targets. The municipality delivered the improvement plan to the Ministry of Education and Children, which analysed it and either approved or requested revision. To ensure the school's autonomy, the school and the local authority determined the improvement actions to pursue, whereas the Ministry endeavoured to ensure that all recommendations were responded to in some way. The final improvement plan was made public along with the evaluation report. Further follow-up was undertaken in the form of communication between the Ministry, the municipality, and the school. Every 6–12 months, until all improvements had been implemented, the Ministry requested a progress report from the local authority and the school. The follow-up process thus lasted 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 imposed no consequences for any school, or its municipality, that

failed to implement the measures for targeted improvements. Indeed, neither the Ministry nor the Directorate was in a position to impose sanctions on schools or municipalities or to reward them (Ólafsdóttir et al., 2022).

Although the evaluation and follow-up process changed slightly during the evaluation period, especially in 2018 and 2019, the above description applies to the implementation of the evaluation and follow-up process as it was from 2013–2018 when the schools examined in the research for the thesis were evaluated and their improvement plans activated.

3 Review of the literature and analytical frameworks

In this section, the key concepts underpinning the conceptual and analytical frameworks in the research conducted for this thesis are defined and presented together with relevant evidence from studies on external school evaluation conducted in other educational contexts.

In this section, I first describe the analytical framework underlying Paper I that concerns the origin and development of the external evaluation of compulsory schools in Iceland from 1990 to 2016. The prehistory of the external evaluation is examined in light of theory on how issues end up on the government's agenda, new public management (NPM), and theory on the adoption of policy and change. Second, I briefly discuss how external school evaluation is conducted in various ways and with different approaches across several countries in Europe. Third, I introduce Ehren et al.'s (2013) theoretical framework of the causal mechanisms of school inspection, which is the basis of the analysis in Paper II. The results of studies conducted to test the assumptions of the model are presented as well. Fourth, I present additional findings from research on the impact of external evaluations and discuss factors that can facilitate or hinder an evaluation's influence on a school's improvement. Last, I describe the conceptual framework utilised in Paper III regarding the different types of evaluation use, distinguished by Rossi et al. (2004) and Aderet-German and Ben-Peretz (2020). Findings from research on the different uses of evaluation feedback are discussed as well.

3.1 Decentralisation and external evaluation

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; OECD, 2013). 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 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. In response, in the past few decades, the comprehensive evaluation of school operations in various parts of the world has expanded significantly (OECD, 2013).

In Iceland, that development manifested when the country's municipalities took responsibility for the operation of compulsory schools in 1996. The purpose of transferring the operation from the central government to local authorities was to ensure equality as well as quality in education (Jónasson, Ragnarsdóttir, et al., 2021). On that topic, Paper I seeks to shed light on the sequence of events that ensued, with special focus on the external evaluation of compulsory schools. Such developments were examined in the research for this thesis in light of (1) theory on how issues end up on the government's agenda, (2) the period of reforms in public administration attributed to NPM, (3) the increased delegation of power and responsibility to lower-level administrations and the accompanying rise in the control and monitoring of school operations, and (4) theory on the adoption of policy and change.

3.1.1   How issues end up on the government's agenda

U.S. political analyst Kingdon (2014) has advanced a theory about how certain issues find their way into political debate and subsequently become part of a government's agenda. Although Kingdon's theory, based on research addressing official health and communications policy in the U.S. government from 1976 to 1979, does not encompass education policy, it could be expanded to include that dimension of government administration. In his book *Agendas, Alternatives, and Public Policies*, Kingdon discusses three streams of agenda-setting processes that have to exist in order to bring about significant changes in government policy. First, there has to be problem recognition, which thrusts the matter into political debate in which potential solutions are discussed. Second, viable policy proposals have to have been put forward by experts and policymakers. Third, the political environment has to be conducive to change—for example, a newly formed government that has opened up an opportunity for new proposals to enter the political agenda. When those three streams flow together, a *policy window* opens, meaning favourable timing for policy changes that enable advocates of innovative proposals to present their ideas or to direct administrative attention to a specific problem. At the same time, Kingdon writes that problems are not necessarily the only reason why certain issues find their way to the government table, for the accumulation of knowledge and expert opinions in a particular field can also catalyse innovative discussion and proposals. The influence of the public and interest groups in agenda setting is also considerable, via both positive promotion and negative blocking. Proposals that do not meet the criteria of being technically feasible, aligning with the values of policymakers, involving tolerable costs, and being likely to receive public acceptance and be supported by elected representatives are not likely to be taken seriously as viable proposals.

Another political scientist, Schick (2002), has identified factors that impact administrative amendments, and there is much accord between his ideas and Kingdon's (2014) theories. Schick's work stresses that, in some instances, the time for change had simply arrived, often due to trends and policies in other countries. In such cases,

elections and changes in government have sometimes opened up opportunities for change, often because major administrative and policy innovations are launched shortly after a new government takes power. Schick argues that entrepreneurs can also be highly effective mediators of new ideas but need to have enough power to be able to persuade politicians to assume the risks of new approaches. He adds that politicians themselves also have to be convinced that they are making the right choice.

3.1.2 New public management (NPM)

NPM is a concept encompassing an assortment of ideas and theories aimed at stimulating reform in public administration (Hood, 1991). As a policy of innovation, NPM dominated administrative reform in developed countries around the world, beginning in the 1980s and continuing well into the 2000s, albeit to varying degrees and with different areas of emphasis in each country (Gunter et al., 2016b; Margetts & Dunleavy, 2013).

During the period in which NPM was adopted, public organisations were often regarded as inefficient and ineffective—that is, marred by lagging performance (Moynihan, 2008). In response, the thinkers who laid the foundation of NPM put forth special effort to provide practical solutions to that problem (Dunn & Miller, 2007). A common core of their ideas was to utilise management practices and principles applied in the private sector to develop a success-oriented, performance-driven official administration system (Dunleavy et al., 2005; Eurydice, 2007; Gunter et al., 2016a; Moynihan, 2008).

Hood (1991) has identified seven factors most frequently referred to in relation to NPM during its initial period. The first was professional management, meaning that management came to be increasingly regarded as a profession in its own right. CEOs were to be given the mandate and freedom to manage and be plainly accountable for their activities. Second, explicit standards and measurements for performance were to be established, which involved defining measurable targets for success in order to determine whether the CEO and the organisational entity could sustain their responsibilities. Furthermore, clear objectives served the purpose of increasing efficiency. Third, focus on controlling output increased. The allocation of resources and rewards was to be directly linked to performance, while centralised, bureaucratic personnel management was to be dismantled. Fourth, formal organisational units were to be divided into smaller, more manageable independent units. Fifth, competition in the public sector was heightened by introducing temporary contracts and public tendering procedures as means to lower costs and increase quality. Sixth, emphasis was placed on using management practices from the private sector in public institutions, including greater flexibility in recruitment and reward and incentive systems. Seventh and last, discipline and parsimony in the use of resources were to be given greater weight.

Also considering the concept of NPM, Koliba et al. (2010) have highlighted associated areas of emphasis. Among other areas, devolution due to transferring tasks and powers to the government at lower administrative levels was assumed to result in improved organisational efficacy and more flexibility in the provision of services. Those authors also underscore that another consequence of decentralisation is the increased necessity to practise management via legislation and regulations. In that vein, Dunn and Miller (2007) have discussed several laws that followed NPM policy and noted that the government's role was to strengthen the power of the population and municipalities to self-govern. To that end, the authors argue, the government should maximise the participation of as many stakeholders and institutions in policy- and decision-making as possible. In that way, residents were often seen as customers of public services.

Taking a different perspective, Gunter et al. (2016a) have considered how NPM affected educational reform in Europe. Among the key features highlighted were the practices of generating, analysing, and using outcome data, including grades on inspections and the results of examination, to make judgements about a school's performance. Another feature was using processes of self-evaluation and strategic management to plan and monitor the implementation of a school's strategy. Along those lines, emphasis was placed on inspection and assessment to judge the performance of individuals, teams, and schools and to make benchmarking comparisons. The publication of data and league tables was supposed to better enable consumers to choose schools and service providers. Gunter et al. (2016b) have also pointed out that those NPM reforms have had different effects in European countries depending on the institutional context and historical, political, and administrative nuances in each country. They note, for example, that in the Nordic countries, including Iceland, the historical legacy and tradition of welfare has moderated the effects of NPM. For instance, Nordic governments have generally not taken the view that the market will ensure quality education, albeit with some exceptions (Jónasson, Bjarnadóttir, et al., 2021).

NPM was a declared policy of the Icelandic government from 1991 to 2007 (Iceland Government Office, 2015). In 1993, a policy report published by the Ministry of Finance titled "Reforms and NPM" stated that the essence of the policy was "distributing power, enhancing responsibility and shifting decision-making as close as possible to the relevant setting, thus attaining a more economical performance and improved services" (Ministry of Finance, 1993, p. 4). Regarding objectives and the assessment of results, the report stressed the importance of having public institutions evaluated on the basis of their objectives and subsequent results and stated that, in some cases, institutional quality systems might facilitate successful performance.

3.1.3 Delegation of power and responsibility: The increased control and monitoring of school operations

The practice of devolution is based on the delegation of power and responsibility. In a democratic society, the activities of the state should reflect the wishes of the electorate as far as possible (Kristinsson, 2007). Thus, it is a fundamental principle that decisions made at various administrative levels be led and monitored by democratically elected representatives and government ministers. Education ranks among the domains of society in which elected representatives and the legislature are obliged to serve the public interest (Simkins, 2003). By means of devolution, the political responsibility for education is divided between the central and local governments, municipal councils, headmasters, and teachers.

As tasks are transferred to lower-level administrations, government departments and other central institutions retain the crucial functions of determining policies, setting targets, and overseeing performance (Ministry of Finance, 2000; Schick, 2002). Jónsson's (2014) analysis of changes made to legislation regarding compulsory schools in Iceland in 1995 (i.e. when the administration of compulsory schools was transferred from the state to municipalities), as well as to the National Curriculum Guide in 1999, revealed that the changes from previous legislation and curricula were primarily of two sorts. On the one hand, objectives became considerably more detailed and comprehensive than before; on the other, those objectives were more closely linked to the school's activities and therefore could serve as a basis for evaluation. In that context, Jónsson observed that as the central administration was revoked by delegating the administration of compulsory schools to municipalities, the central government's control was strengthened by the use of stricter criteria regarding the objectives and evaluation of a school's activities.

In 1997, the OECD reported that trends towards devolution require more thorough performance management and possibly a special supervisory authority for the purpose of monitoring the implementation of legislation (OECD, 1997). That ideal, however, does not appear to have been fully actualised. As Schick (2002) has observed, in many cases centralised institutions and government departments had been so weakened due to devolution that they lacked the capacity to implement essential coordinating functions. From his perspective, in the eagerness to grant increased freedoms to local authorities and administrators, insufficient care had been taken to ensure the continued maintenance of necessary resources and centralised authority, both to formulate cross-institutional plans and policies and to maintain strict accountability. Thus, responsibilities remained relatively unformulated and embryonic in many countries. In 2000, a report by Iceland's Ministry of Finance indicated that such was also the case in Iceland, where devolution and the enhanced freedom of institutions had been prioritised over developing ways of determining institutional responsibilities, and the report emphasised the need for change in that respect (Ministry of Finance, 2000).

In 2007, a report from Eurydice indicated that, in most countries, the central government is responsible for matters concerning education and inspectorates for monitoring and evaluating schools. In some locations, especially in the Nordic countries, the independence of schools and their delegation to municipal councils seems to have sprung from the idea that municipalities should play a pivotal role in managing schools (Eurydice, 2007). In such circumstances, schools are primarily accountable to the local authorities that, in turn, serve the crucial function of supervising and evaluating schools. In general, however, there is a trend towards multi-accountability, meaning that schools answer to a range of bodies, including departments of education, municipalities, and parents (Eurydice, 2007).

3.1.4 The adoption of policy and changes

When confronted with change in public governance, the authorities face countless decisions about introducing the changes in question, including who is to implement them and how (Schick, 2002). Hogwood and Gunn (1984) have identified several factors that influence whether a policy is viable, including external circumstances, adequate time, the requisite combination of resources, the policy's basis on a valid theory of cause and effect, and objectives that are clearly understood and agreed upon. Beyond that, time needs to be devoted to developing an understanding of the changes in question and encouraging interest in their implementation. For several years, Kotter (2012) studied the process of change in companies and institutions, in which he observed the need to demonstrate a sense of urgency so that people experience change as a solution to a certain challenge. No less crucial are top-level management's support and commitment to change as well as leadership of a powerful guiding coalition. In cases in which introducing change has been most successful, the guiding coalition has been endowed with leadership skills in combination with authority, knowledge, and professional connections. Kotter also identified key factors of achieving change, including having a clear vision and strategy, pursuing well-defined avenues for reaching objectives, and seizing every opportunity to highlight the vision of change. On that count, Fernandez and Rainey (2006) have cautioned that a vague strategy may prompt confusion and give official management the opportunity to reinterpret objectives and arbitrarily implement change such that only part of the programme envisioned by policymakers is ever accomplished. In line with Kotter (2012), they argue that changes in the public sector depend on the support of government ministers and key external stakeholders whose impact stemmed from their ability to impose statutory changes and control the flow of resources to public institutions.

According to a report from the OECD (2013), many countries have encountered problems in introducing the evaluation of schools' operations. The report suggested various reasons for the difficulty, including policy design, education authorities without the muscle to put the evaluation into practice, the absence of a culture of evaluation, and the insufficient use of evaluation results. The report also stated that the

implementation of evaluation is indeed complex and requires the inclusion of numerous participants with different sorts of interests. It is therefore important, according to the report, to strengthen the cooperation of stakeholders via consultation, to harness the services of experts, to clearly articulate the objectives of the evaluation, to gather data and assess implementation, to cultivate skills, to provide resources, and to allow enough time for completing the task. It was additionally emphasised, however, that one size does not fit all; the process always needs to correspond to the education system and conventions of the country in question.

In the next section, I briefly discuss how external school evaluation is conducted in various ways and following different approaches in certain European countries and how it is based on a certain theory of action. I explain how the theory of action of the Icelandic model is based on models from other European countries and thus why Ehren et al.'s (2013) framework of causal mechanisms, based on programme theories of six European countries' school inspection systems, was chosen as an analytical tool for the study presented in Paper II.

3.2 External school evaluation in Europe

Governments and education policymakers have increasingly emphasised the evaluation of schools and education systems as a means to gather information about the educational performance and improve school practices (Eurydice, 2015; Gärtner & Pant, 2011). The external evaluation is conducted in various ways and has different functions, for it occurs in light of different political priorities, national circumstances, and educational traditions. In countries whose governments take the approach of soft governance, external evaluation tends to have little pressure for accountability compared with governments that practise hard governance (Altrichter & Kemethofer, 2015; Ehren, Gustafsson, et al., 2015). In that context, Altrichter and Kemethofer (2015) have delineated what they call "low-stake" and "high-stake" inspection systems. Ehren et al. (2013) have distinguished four major dimensions of external school evaluation: (1) types of inspections and the frequency of visits; (2) standards and thresholds for identifying failing schools; (3) consequences such as sanctions, rewards, and interventions; and (4) the presence or absence of reporting on individual schools to the general public. Based on those classifications, Ehren, Gustafsson, et al. (2015) and Altrichter and Kemethofer (2015) have demonstrated that countries that use differentiated evaluation models (e.g. cyclical evaluation of all schools and more frequent evaluation if schools are suspected of poor quality), outcome-oriented evaluation, sanctions for failing schools, and reports on individual schools to inform the public can be deemed to use a high-stakes approach to external evaluation. By contrast, countries that use the cyclical evaluation of all schools, are process-oriented, impose no sanctions for failing schools, and do not inform the public with reports on individual schools can be regarded as using a low-stakes approach. In particular, Altrichter and Kemethofer (2015) study on several countries in Europe revealed that two of them—

England and the Netherlands—can be considered to use a high-stakes approach to external evaluation, whereas Austria and Switzerland use a low-stakes approach. By comparison, Sweden, Ireland, and the Czech Republic take an approach that can be located between being high-stakes and low-stakes. Germany was also found to take rather low-stakes approaches to external school evaluation (Dedering & Müller, 2011). Although all of those countries take different approaches, they share certain basic ideas about the assumptions of the causal mechanisms of external evaluation (Ehren et al., 2013).

Behind any evaluation process is a theory of how external evaluation is intended to make an impact (Ehren et al., 2013; Ehren et al., 2005). The *theory of action* defines the purpose of a programme and the expectations behind interventions, as well as what is likely and what is desired to happen as a result (Kools & Stoll, 2016). The theory also explains how interventions are intended to effect changes and why certain outcomes occur. As introduced in Section 2.4, the manifold purpose of the external evaluation of compulsory schools in Iceland is to enhance and support schools' internal evaluation and quality assurance; to support schools, principals, and teachers in improving their practices; to motivate teachers to collaborate on improvement; and stimulate the development of schools (Ad Hoc Team on the External Evaluation of Compulsory Schools, 2011). Moreover, it is expected that the feedback in the evaluation report, in the form of recommendations for improvement, provide the basis for an improvement plan, later followed up by local education authorities and the Ministry of Education and Children. The fact that external school evaluation in Iceland and inspections in other countries in Europe show significant similarities in their assumptions about the causal mechanisms underlying evaluation is no coincidence. After all, when formulating external evaluations, developers in Iceland have attempted to learn from the experience and models of other countries in Europe (Ad Hoc Team on the External Evaluation of Compulsory Schools, 2011).

The following section introduces the theoretical framework of causal mechanisms of external school evaluation, used in the research conducted for this thesis and in Paper II as a basis for analysis and discussion.

3.3 Framework of causal mechanisms of external school evaluation

The theoretical framework developed by Ehren et al. (2013) is based on programme theories (i.e. theory of action) behind systems for school inspection in six European countries: England, the Netherlands, Ireland, Sweden, Austria and the Czech Republic. Although those six countries have different assumptions about the effectiveness of inspection, their common ideas are incorporated into the framework.

As shown in Figure 1, the framework describes the intermediate processes of how inspection and external school evaluation are expected to impact the improvement of schools and promote quality education. Key characteristics of school inspection that may influence improvement are highlighted in the first column; they include evaluation methods, standards and criteria, feedback, and the degree of pressure for accountability (e.g. consequences and the public disclosure of evaluation results). The second column defines two mediating mechanisms expected to bring about school improvement: setting expectations and accepting feedback. Stakeholder involvement (i.e. actions) is the third mediating causal factor. Those three mechanisms are interconnected and expected to trigger a sequence of processes that link external evaluation to the intended promotion and/or improvement of self-evaluation and stimulation of improvement actions, as shown in the third column. In turn, high-quality self-evaluation and improvement actions are expected to promote increased improvement capacity and more effective school and teaching conditions, as shown in the fourth column, which will lead to the ultimate goal of the inspection—that is, higher-quality education and improved student results—as shown in the fifth column (Ehren et al., 2013).

Figure 1. Framework of causal mechanisms of school inspection (Ehren et al., 2013, pp. 14, with permission)
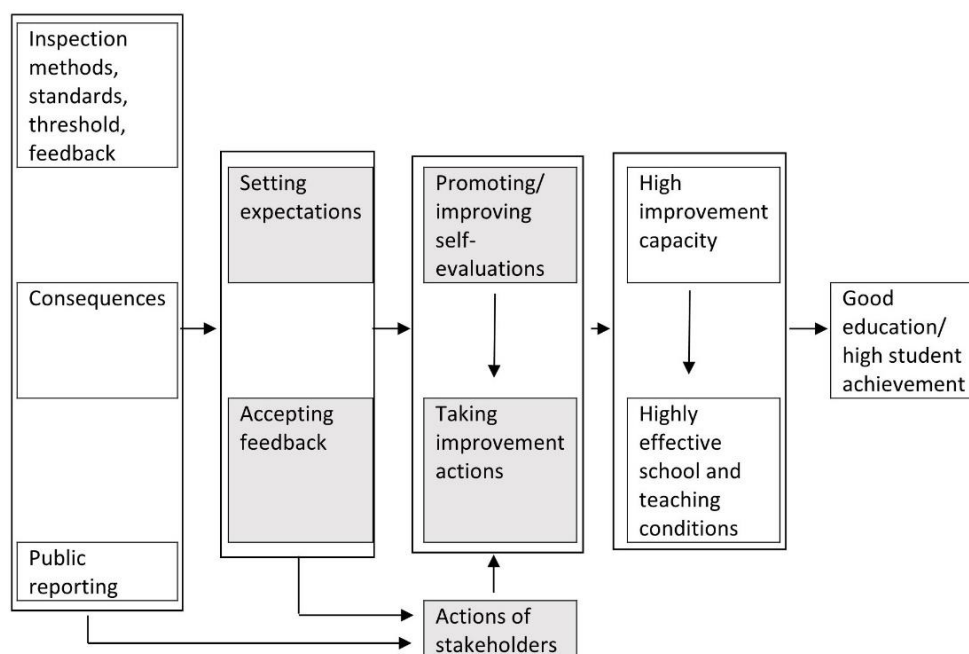
Ehren et al. (2013) tested the framework's assumptions by surveying the principals of schools in the six European countries represented in the framework. The results of their research are discussed below and used as a basis for discussing the results of the research conducted for this thesis.

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

3.3.1 Setting expectations

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

High-stakes accountability systems are based, among other factors, on the assumption that rational actors strive to meet the standards of the systems due to the threat of sanctions if they do not (Ehren, Gustafsson, et al., 2015). In that way, a high degree of external pressure increases the likelihood that schools will use the standards. Studies by Ehren, Gustafsson, et al. (2015) and Altrichter and Kemethofer (2015), both based on the survey responses of principals in the six mentioned European countries, largely confirm that trend. They found that principals who felt more pressure for accountability were more alert to the expectations issued by the inspectorate and implemented more improvement actions; however, unintended consequences also increased with greater pressure (Altrichter & Kemethofer, 2015; Jones et al., 2017). In that context, Fullan (2016) has pointed out that excessive pressure can spur changes but only ones made primarily on the surface. In their study of schools in the Netherlands, where pressure for accountability is rather high (Altrichter & Kemethofer, 2015), Ehren, Perryman, et al. (2015) found that setting clear expectations is a strong determinant of improvements and changes made in the capacity of schools. Even so, Kemethofer et al.'s (2017) study comparing Austria and Sweden, with Austria representing a low-stakes system and Sweden a medium- to high-stakes system (Altrichter & Kemethofer, 2015), showed the significant effect of setting expectations in Austria but not in Sweden. Thus, those results do not meet the assumptions of high-stakes systems.

3.3.2 Accepting feedback

The second causal mechanism in the framework occurs via the feedback that schools receive during evaluation visits and in their evaluation reports (Ehren et al., 2013). The feedback is based on quality-focused standards and criteria, and schools are expected to use it to address areas in need of development that emerged when the quality of the school's performance was assessed in light of the criteria. They are expected to do so by realising, accepting, deciding on, and implementing improvement actions that meet the requirements of the standards and, thereby, improve their quality as schools (Ehren et al., 2013; Ehren, Gustafsson, et al., 2015). As various studies have illustrated, the acceptance of feedback is essential if decisions about school improvement are to be made based on them (Dedering & Müller, 2011; Her Majesty's Inspectorate of Education, 2010; Penninckx, 2017; Schildkamp & Ehren, 2013). If the feedback received is considered to be compelling, relevant, valuable, and supportive, then it is more likely to be accepted and used for improvement (Altrichter & Kemethofer, 2015; Behnke & Steins, 2017; Her Majesty's Inspectorate of Education, 2010; Verhaeghe et al., 2010). Furthermore, several researchers have highlighted that principals and teachers show partly different reactions to external evaluation (Ehren, Perryman, et al., 2015; Hofer et al., 2020; Matthews & Sammons, 2004; Penninckx et al., 2016a) with principals being more positive and accepting of the evaluation feedback than teachers (Matthews & Sammons, 2004) or vice versa (Ehren, Perryman, et al., 2015).

Studies conducted to verify the hypotheses of Ehren et al.'s (2013) framework have indicated, contrary to those hypotheses, that the acceptance of feedback is not necessarily a key factor of the impact of external evaluation on school improvement (Altrichter & Kemethofer, 2015; Ehren, Perryman, et al., 2015; Gustafsson et al., 2015; Gustafsson et al., 2014). In those studies, it was concluded that schools that accept feedback do so because of the other two factors—that is, setting expectations and stakeholder involvement. In Gustafsson et al.'s (2015) study, no improvement actions were detected as a result of the acceptance of feedback, while Ehren, Perryman, et al.'s (2015) study showed no improvement in self-evaluation or the capacity to improve due to accepting feedback, but only changes in the conditions of a school's effectiveness (i.e. opportunity to learn, quality of teaching, and use of assessment to monitor students' progress and the quality of the school). Contrary to those results, Kemethofer et al. (2017) observed that the effect of inspection on the acceptance of feedback was positive and significant in both Austria and Sweden, albeit much higher in Austria. A similar conclusion was reached by Altrichter and Kemethofer (2015), who found that accepting feedback significantly improved self-evaluation in Austria and Sweden but not in England. They thus concluded that accepting feedback is more important in middle- and low-stakes evaluation approaches than in high-stakes ones due to less pressure for accountability (Altrichter & Kemethofer, 2015).

3.3.3 Stakeholder involvement

Stakeholder involvement is the third causal mechanism in Ehren et al.'s (2013) framework. Feedback from external evaluations emerges not only between schools and evaluators but also stems from a complex, multilevel system in which different stakeholders external to the school (e.g. authorities, school boards, community members, and parents) have access to information about the evaluation results through presentations and/or the publication of evaluation reports (Ehren, Perryman, et al., 2015). External stakeholders are expected to review the results, support the schools in their development and pressure them to respond to the feedback with improvement actions to fulfil the evaluation criteria (Ehren et al., 2013; Ehren, Eddy-Spicer, et al., 2016; Kemethofer et al., 2017). Thus, stakeholder involvement refers to their awareness of the school evaluation report and the identified needs for improvement and to their use of their voices to bring about change (Ehren, 2016b; Ehren et al., 2013; Ehren, Gustafsson, et al., 2015). Parents are also expected to put pressure on schools by using the evaluation reports to inform the choice of school for their children—for example, to exit a failing school and transfer their children to a higher-performing one (Ehren, 2016b). The inclusion of stakeholders in the process of issuing feedback is expected to raise their awareness of standards of quality and, in turn, to support the expectations established in the criteria and increase the likelihood that schools will respond to the results (Ehren, Perryman, et al., 2015).

Ehren, Perryman, et al. (2015) have suggested that the mechanism of stakeholder involvement was not only an important driver for the implementation of a school's self-evaluation in the Netherlands but also related to the increased acceptance of evaluation feedback. A similar conclusion was reached by Ehren, Gustafsson, et al. (2015). By contrast, Kemethofer et al. (2017) did not identify any effect of stakeholder involvement on improvement in either Austria or Sweden. Along similar lines, Gustafsson et al. (2015) concluded that stakeholder involvement did not directly impact improvement actions but instead affected schools' responses early in the improvement process by motivating them to accept feedback as well as inspection standards.

Inconsistencies in the findings of research regarding the three causal mechanisms of the Ehren et al.'s (2013) framework—setting expectations, accepting feedback, and involving stakeholders—indicate that the effects of external evaluations are far from uniform and depend on a variety of other conditions, including the national context in which they are implemented (Kemethofer et al., 2017). However, the research conducted for this thesis and in accordance with the assumptions of the framework, I hypothesised a positive correlation between the framework's three causal mechanisms and changes in practices of internal evaluation, teaching, and leadership in Icelandic schools. I also assumed that other factors, discussed in the next section, are positively or negatively associated with changes given the variables in question.

3.3.4 Capacity to implement improvements

Building capacity refers to a school's capacity to address weaknesses, improve the professional learning of teachers, and take actions to make teaching practices student-oriented, which may ultimately boost the achievement of students (Ehren, 2016b). In Ehren et al.'s (2013) theoretical framework, improvement in capacity-building is expected to result from improved self-evaluation. Although the framework does not address a school's pre-existing capacity as a factor of improvement (Ehren et al., 2013; Hofer et al., 2020), Ehren (2016b) has highlighted capacity building as a fourth mechanism of change that explains how external evaluation can lead to improvements in the quality of schools. Capacity to improve therefore plays a dual role in explaining the effects of external evaluation; it acts as a condition for schools to respond to expectations with improvements, while external evaluation is simultaneously expected to increase the school's capacity to improve (Ehren, 2016b; Ehren, Eddy-Spicer, et al., 2016). Beyond that, organisational factors shown to influence the success of external evaluation include favourable attitudes towards internal evaluation, teachers' participation in decision-making, and the existence of resources and knowledge in schools that support improvement (Ehren, 2016b; Ehren, Eddy-Spicer, et al., 2016; Hofer et al., 2020; Schildkamp & Lai, 2013). Section 3.4.1 elaborates on both organisational and contextual factors that affect the impact of external evaluation.

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

Via the causal mechanisms described above, external evaluation aims to stimulate and drive improvement and self-evaluative actions by schools (Altrichter & Kemethofer, 2015; Ehren et al., 2013). Improving self-evaluation and taking improvement actions are described as "intermediate mechanisms" that contribute to a school's ultimate success (Gustafsson et al., 2015). That is based on the hypothesis that internal evaluation plays an important role in enhancing improvement-oriented processes (Ehren, 2016b; Ehren et al., 2013; Hanberger et al., 2016). Studies based on the theoretical framework (Figure 1) have shown that external evaluation positively impacts schools' self-evaluation and that schools that improved their process of self-evaluation also improved their capacity building, thereby prompting improvement in the conditions needed for an effective school (Altrichter & Kemethofer, 2015; Ehren, Perryman, et al., 2015; Gustafsson et al., 2015; Kemethofer et al., 2017). However, a comparison between the low-stakes Austrian approach and the high-stakes model used in England revealed that external evaluation seemed to have less impact on changes in self-evaluation in Austria than in England (Altrichter & Kemethofer, 2015).

In sum, Ehren et al.'s (2013) theoretical framework evinces assumptions shared among various European countries about the intermediate mechanisms required for external evaluation to be effective. However, different national evaluation systems do not use

identical means to achieve that effect, as discussed in Section 3.2. For that reason, research conducted to test the framework's assumptions has not shown a uniform picture of each mechanism's effect. Along those lines, the next section discusses additional research on the impact and use of the findings of external evaluation and spotlights factors that can encourage or discourage evaluation feedback from having the desired influence on school improvement following external evaluation. It further discusses how such feedback can be used in different ways.

3.4 Impact and use of the findings of external evaluation

Most evaluation systems aim at both monitoring schools and ensuring their accountability as well as improving the quality of education in schools (Hofer et al., 2020; OECD, 2013; Penninckx, 2017). Those aims are pursued by giving evaluation feedback—that is, 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 evaluation feedback is not a sufficient condition for realising systematic reflection or improvement actions in schools (e.g. Ehren et al., 2013; Jónasson, 2019; Verhaeghe et al., 2010), and findings concerning how the results of external evaluation are used and impact improvement in schools have been inconsistent, as discussed above. 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. AlKutich & Abukari, 2018; 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; Wenger et al., 2022). In particular, it has proven difficult to demonstrate a significant relationship between external evaluation and students' achievement (Ehren et al., 2013; Wenger et al., 2022). It has been shown, however, that evaluation feedback may not have the desired effect, because the influence of an evaluation arises in the lead-up to and preparation for the evaluation, not in reaction to its feedback (Gärtner, 2021; Gärtner et al., 2014; Wurster & Gärtner, 2013). Other causes have also been highlighted and are discussed in greater detail below, including a lack of capacity, support and external pressure, bad experiences with the evaluation process, and resistance to results (Ehren, 2016b; Hofer et al., 2020; Landwehr, 2011). Whatever the case, inconsistent findings on the topic highlight the highly contextual nature of how schools use the results of external evaluation (Ehren, 2016c; Hofer et al., 2020).

Studies on the impact of external school evaluation show not only a positive or lack of effect, however. On the contrary, external evaluation sometimes appears to lead to unintended side effects that are often negative and tend to undermine the intended positive effects (Jones et al., 2017). Side effects, which can occur during pre-evaluation, in-school evaluation, or post-evaluation (Ehren, Jones, et al., 2016; Ehren &

Visscher, 2006), include stress, anxiety, and strain, along with increased workload in schools in anticipation of and during evaluation (Drinck et al., 2013; Ehren, Jones, et al., 2016; Matthews & Sammons, 2004); the misrepresentation of data in school documents, strategic behaviour during evaluation, and the rejection of the evaluators' recommendations (Ehren & Visscher, 2006; Jones et al., 2017); the narrowing of the school's curriculum and teachers' hesitation to experiment in teaching (Ehren, Gustafsson, et al., 2015; Ehren, Jones, et al., 2016; Jones et al., 2017); and strategic behaviour regarding the measures of students' achievement and teaching to the test at the expense of curricular aspects not represented on tests (Ehren & Honingh, 2011). For internal evaluation, the side effects of external evaluation appear to include focusing exclusively on factors that occur in the external evaluation framework (i.e. "tunnel vision"), aiming at short-term targets at the expense of long-term objectives (i.e. "myopia"), using evaluation indicators rigidly and refraining from innovation (i.e. "ossification") (Ehren & Visscher, 2006), and the increased possibility that the primary motivation for schools to engage in internal evaluation is external instead of internal (O'Brien et al., 2015).

Various enablers and barriers influence evaluation use in schools and, in turn, affect the evaluation's impact on the development of schools and the outcomes of education (Schildkamp & Lai, 2013). Those enablers and barriers can be broadly categorised as representing the perceived quality of the external evaluation performed, the features of the school itself, and external pressure and support. The impact of external evaluation depends on the interaction of those factors (Ehren & Baxter, 2021; Landwehr, 2011), as detailed in the next section.

3.4.1 Impact of external evaluation: Enablers and barriers

The characteristics of external school evaluation and its quality as perceived by school staff, the features of the school itself, and pressure and support from external sources all affect what happens in schools following external evaluation (Ehren et al., 2013; Ehren & Visscher, 2006; Hofer et al., 2020; Landwehr, 2011; Penninckx, 2017; Visscher & Coe, 2003). Landwehr (2011) has identified four evaluation-related factors that influence how external evaluation affects the development of quality in schools: (1) the trustworthiness of the evaluation, (2) the transparency and plausibility of the criteria and indicators used, (3) the communication style of the evaluators during the evaluation, and (4) the quality of the feedback from the evaluation given to schools. Other studies have also shown that the extent to which the results of external evaluation align with the school's understanding of its strengths and weaknesses affects the extent to which schools use the results to achieve improvement (Drinck et al., 2013; Ehren et al., 2005; Ehren & Visscher, 2006). More recently, Penninckx et al. (2016c) have revealed that the perceived quality of external evaluation played a crucial role in explaining its effects. They highlighted that if external evaluation is aimed at supporting the development of schools, then ensuring its perceived quality is pivotal, especially in

terms of reliability, validity, the transparency of the process and criteria, trustworthiness, and respect shown to school staff. Other scholars have also emphasised the importance of the credibility of external evaluation (e.g. Ehren, 2016b; Ehren & Visscher, 2008; Hofer et al., 2020; Penninckx, 2017; Penzer, 2011; Verhaeghe et al., 2010). Beyond that, in a recent summary and analysis of research from 1990 to 2020, Hofer et al. (2020) concluded that the better the school staff's impression of the quality of evaluation, the more positive the evaluation's effect, particularly concerning process and outcome variables.

On the topic of organisational and contextual factors, Landwehr (2011) has also identified eight school-related features that are important for external evaluation to have the desired effect on the development of quality in schools: (1) the extent to which the school is a learning organisation and its leadership and teachers have a positive attitude towards change; (2) interest in the results of evaluation; (3) competence in developing lessons; (4) experience with the productive use of the results of internal evaluation; (5) the staff's involvement in a collaborative discussion and analysis of the results; (6) the creation of targets for improvement based on that analysis and the evaluation report's recommendations; (7) external pressure; and (8) the availability of resources. As mentioned in Section 3.3.4, studies as well as literature reviews have stressed the importance of such organisational factors on how much improvement can be expected as a result of external evaluation (e.g. Ehren, 2016b; Moynihan & Pandey, 2010; Penninckx, 2017; Schildkamp & Ehren, 2013; Verhaeghe et al., 2010). The mentioned study by Penninckx et al. (2016c), for instance, revealed that a school's characteristics may influence the effect of external evaluation but not as much as the quality of evaluation. Similar to Landwehr, Ehren and Visscher (2006) found that factors in and around a school—for example, external pressure and support—along with the school's characteristics affect how the school responds to the results of external evaluation. In response, they advised adapting the follow-up to evaluation to suit the characteristics of the school to stimulate improvement and bring about the desired effects. Beyond that, they discovered that schools with a low level of innovation capacity needed greater external pressure and support to improve, whereas schools with high innovation capacity were capable of improving on their own. In the same vein, Matthews and Sammons (2004) have demonstrated that a lack of follow-up in schools whose administrators are not proactive can reduce the impact of external evaluation on the school.

More recently, Ehren and Baxter (2021) have linked those three influencing factors (i.e. characteristics of external evaluation, school characteristics, and external pressure) and posited that three elements—trust, capacity, and accountability—are not only the pillars of any education system but also 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 such a level

of accountability will undermine trust. On that topic, Fullan and Quinn (2016) and Six (2021) have highlighted the importance of approaching accountability as a strengthening, supporting process, not as punishment for not meeting requirements. As such, accountability can contribute to building trust and capacity—for instance, by creating performance criteria for schools that clearly define the expectations placed on staff and inform their work—as well as increase the trust of parents and the public (Ehren et al., 2013; Ehren & Baxter, 2021; Six, 2021). In an environment of trust, staff can be open about their performance and reflect on improvements, which is an important condition for effective evaluation, especially because the evaluation will have side effects if staff are under significant pressure to do well during external evaluation (Ehren, Jones, et al., 2016; Ehren & Visscher, 2006). As discussed, evaluation feedback based on clear performance-focused 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, 2021). Such feedback is expected to do so by impacting internal evaluation, promoting the professional development of teachers and principals, encouraging cooperation in the school, and creating incentives for seeking out external support (Ehren, 2016b). To secure accountability, capacity—meaning skill, competencies, knowledge, and decision-making power—has to be developed within schools so that they can incorporate the evaluation criteria and provide high-quality education. The interaction of the three major elements of trust, capacity, and accountability varies from education system to education system, based on varying degrees of pressure for accountability, cultural traditions of trust or distrust in authority, and the knowledge and resources available in the system (Ehren & Baxter, 2020).

The following section introduces four ways of using the results of evaluation that were applied in the research conducted for this thesis and as a basis for the analysis and discussion presented in Paper III: instrumental, conceptual, persuasive and reinforcement-oriented use. A review of literature on evaluation use is included as well.

3.4.2 The uses of external evaluation

Evaluation, as a knowledge-generating undertaking (Vo, 2015), assumes that the knowledge generated is useful (Alkin & Taut, 2003). Likewise, evaluation is worthwhile only if such knowledge is put to use. However, the term *use* can be understood in different ways (Rossi et al., 2004). In early studies, scholars employed a narrow definition of *use* focused on the decisions and changes prompted by evaluation, namely as "immediate, concrete, and observable influence on specific decisions and program activities resulting directly from evaluation findings" (Patton, 2008, p. 99). That definition of *use* now more precisely refers to *instrumental use*, the most commonly experienced, recognised, and studied application of evaluation (Nunneley et al., 2015; Rossi et al., 2004; Vo, 2015; Weiss, 1998). As research on evaluation use continued, scholars broadened the concept of use to include situations in which evaluation has

affected an individual's thinking or understanding but without immediately influencing their decisions or actions (Alkin & Taut, 2003; Nunneley et al., 2015; Weiss, 1998). That kind of use, known as *conceptual use,* can impact individuals' actions in the long term (Nunneley et al., 2015; Rossi et al., 2004; Weiss, 1998). Among other ways in which evaluation may be used, two more that formed the basis for the research for this thesis are *persuasive use* and *reinforcement-oriented* use. The following sub-sections describe those four categories of evaluation use in greater detail and report the results of studies on such use and its effects.

3.4.2.1 Instrumental use

Instrumental use refers to types of evaluation use in which decisions, actions, and changes are based on the results and recommendations of evaluation. Instrumental use, as an effect of evaluation, is observable; it can be seen, heard, or felt in some way (Russ-Eft & Preskill, 2009). Weiss (1979) has called that conception of use a "problem-solving model", in which a problem exists and a decision has to be made to generate a solution for it. She later added that, in the case of instrumental use, the results of evaluation become the basis for decisions in the short term (Weiss et al., 2005).

The findings and recommendations of evaluation have the potential to support and increase the instrumental use of evaluation (Weiss, 1998). It is therefore not surprising that most studies on the effects of external school evaluation have primarily examined the instrumental use of the findings. On that count, many have also identified outcomes of such use, including changes in policy, teacher retraining, more distributed leadership and management, increased cooperation between teachers, better teacher–student relationships, improved self-evaluation, and improvements in the quality of teaching, assessment, monitoring, and student tracking (Dedering & Müller, 2011; Ehren, Perryman, et al., 2015; Ehren & Visscher, 2008; Gärtner et al., 2009; Matthews & Sammons, 2004; McCrone et al., 2009; McCrone et al., 2007; Ofsted, 2015; Van Gasse et al., 2018). However, other studies have shown 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., 2014; Penninckx et al., 2016a; Verhaeghe et al., 2010). Otherwise, stronger instrumental use in schools given less favourable evaluation judgements might be due to an imposed need to act and/or more pressure for accountability (Altrichter & Kemethofer, 2015; Penninckx et al., 2016a). Among other findings, teachers appear to be less likely than principals to use the results of evaluation instrumentally in order to improve their teaching practice (Chapman, 2002; Hanberger et al., 2016).

3.4.2.2 Conceptual use

Less direct than instrumental use, the *conceptual use* of the results of external school evaluation refers to the impact of the evaluations on how principals, teachers, and other school staff understand and reflect on the school's practices (Alkin & Taut, 2003;

Penninckx, 2017; Rossi et al., 2004). From evaluation, they gain new ideas, awareness, and insight into their and/or the school's strengths and weaknesses, and although concrete changes or decisions may not immediately materialise following evaluation, evaluation nevertheless impacts their actions in the long term (Russ-Eft & Preskill, 2009; Weiss, 1998). Conceptual use, referred to by Weiss (1979) as "enlightenment use", can be viewed as the first step towards more instrumental use, both in future decision-making and the development of schools, even if it is often difficult to trace (Rossi et al., 2004; Weiss, 1998). Conceptual use involves using the findings of evaluation to initiate discussion within the school and sometimes even criticising their validity and/or usefulness (Aderet-German & Ben-Peretz, 2020).

Few scholars interested in external school evaluation have distinguished instrumental from other types of effects of evaluation, including conceptual effects (Penninckx et al., 2016a). Even so, several have identified the benefits of the conceptual use of the results of external school evaluation, including a better understanding and heightened awareness of the school's quality, along with increased professional reflection and discussion among the 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). By some contrast, Penninckx et al. (2014) found that the conceptual effect occurred primarily before evaluation and arose from a reaction to the announcement of the evaluation (i.e. indirect conceptual effect), whereas the evaluation report increased reflection to only a limited extent. Most often, the results of evaluation do not provide new insights into the schools' strengths and weaknesses but confirm the staff's own experience (Drinck et al., 2013; McCrone et al., 2007; Penninckx et al., 2014; Penninckx et al., 2016a).

3.4.2.3 *Persuasive use*

Persuasive use refers to when the results of external evaluation 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). Principals and teachers often know where action is needed and what they need to do to improve, and they use the results of evaluation to convince, inform, and educate others in order to gain supporters for their cause (Penninckx et al., 2016a; Weiss, 1998). In the category of persuasive use, the use of findings can be positive—for instance, to advocate for important improvements—or negative—for instance, used selectively or to deceive (Baughman et al., 2012; Patton, 2008; Weiss et al., 2005).

Research has revealed schools' persuasive use of the findings and feedback from external evaluation regarding school performance (Baughman et al., 2012; McCrone et al., 2007; Penninckx et al., 2016a; Van Gasse et al., 2018; Verhaeghe et al., 2010). In McCrone et al.'s (2007) case study, some principals expressed the view that the evaluation findings gave them "added authority to introduce innovations or changes"

(p. 57). Later, in other interview study, principals indicated that external feedback on performance supported them in clarifying and presenting their ideas within their school (Verhaeghe et al., 2010). Penninckx et al.'s (2016a) survey also revealed that the persuasive use of the findings of external evaluation is not as common as their conceptual or instrumental use; however, such use is more widespread in schools that have received unfavourable evaluation judgements.

3.4.2.4 Reinforcement-oriented use

Reinforcement-oriented use, first conceptualised by Aderet-German and Ben-Peretz (2020) refers to "the use of positive data for reinforcing existing school strengths" (p. 7). Such use is similar to what Verhaeghe et al. (2010) have called "motivating use" and what Cain (2015) has called "confirmatory use". Penninckx et al. (2016a) have discussed the effects of evaluation on self-efficacy and collective efficacy, both of which also relate to reinforcement-oriented use. Behind all of those concepts is the idea that the findings of evaluation 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 (Aderet-German & Ben-Peretz, 2020; Cain, 2015). In the same way, negative judgement can reduce self- and collective efficacy and thus have a discouraging effect on school staff (Hofer et al., 2020; Penninckx et al., 2016a). Reinforcement-oriented use is similar to conceptual use in that it affects people's thoughts and beliefs but does not directly lead to visible actions (Aderet-German & Ben-Peretz, 2020).

Although the reinforcement-oriented use of the findings of external evaluation 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). Those effects have been greater in schools with more positive evaluation judgements (Penninckx et al., 2016a). A similar connection has been found with negative evaluation judgements, which result in more negative feelings than positive judgements do (Hofer et al., 2020; Penninckx et al., 2016a, 2016b).

4 Gaps in research on external school evaluation

The summary of findings from research on the use and impact of external evaluation has offered an overview of the field of such research in different European settings. It also highlights certain gaps in such research to date.

The above summary of studies on external school evaluation reveals few to no citations to research conducted in Iceland. Thus, the significance of the research conducted for this thesis mostly lies in the fact that external school evaluation in Iceland has hardly been investigated. That oversight is a gap in the current state of knowledge about the Icelandic context, which the research for the thesis was designed to address. I considered it to be necessary to shed light on that phenomenon, ideally by analysing key policy documents, legal statutes, questionnaires, and interviews. Thus, my research was aimed at gaining insight into the origin of external evaluation and the use and impact of evaluation feedback with reference to (1) theory on how issues end up on the government's agenda, NPM, and theory on the adoption of policy and change; (2) Ehren et al.'s (2013) theoretical framework of the causal mechanisms of school inspection; and (3) Rossi et al.'s (2004) and Aderet-German and Ben-Peretz's (2020) conceptual frameworks of different ways of using feedback.

In the international context, the current state of knowledge about external school evaluation has mostly been informed by studies in countries with a long history of such evaluation (OECD, 2013) and where the pressure for accountability is greater than in Iceland. In response, studies from a wider spectrum of systems with different school evaluation policies, models, and pressure for accountability are needed. Such research is important especially given inconsistency in published findings regarding the use and impact of evaluation feedback, which has highlighted the highly contextual nature of how schools use feedback from external evaluation (Ehren, 2016c; Hofer et al., 2020). Furthermore, most studies have been performed shortly after schools received the results of evaluation and therefore could not capture the (im)permanence of the measures for improvement taken by the schools. As an antidote, a longitudinal approach may be required to better determine the longer-term impact of external school evaluation.

Against that background, the research conducted for this thesis was unique due to its mixed-methods approach. Therein, survey data on the mechanisms by which external school evaluation influences internal evaluation and school professionals' practices were statistically analysed, while interview and document data into the different uses of evaluation feedback in the schools were thematically analysed.

5 Aim, purpose, and research questions

The focus of this thesis, influenced by a pragmatic orientation, is the external evaluation of compulsory schools in Iceland, based on the mentioned joint project on external evaluation between the state and municipalities across Iceland, with the exception of the capital Reykjavík (see Section 2.4).

The aim of the research conducted for the thesis was twofold:

1. The first purpose was to shed light on the origin of the recent external evaluation of compulsory schools in Iceland and how it developed as part of official education policy beginning in 1991 until 2016. Doing so was intended to highlight the context in which external evaluation in Iceland was created, how it developed, and what benefits were expected as a result. That purpose was addressed in Phase I of the research, which built an important foundation for the two phases that followed.

2. The second purpose was to identify and analyse the attitudes of school principals and teachers towards the recent external evaluation, how they view their use of the evaluation feedback, how they think that the feedback has affected their schools' internal evaluation and driven change in their own practices (i.e. teaching and leadership), and how well the changes made appear to have been sustained over time. The second purpose was dual in itself: (a) to add to the current knowledge base about external school evaluation and expand current understandings of how schools use evaluation feedback and whether its use impacts improvement and changes; and (b) to gain insight into the extent to which the use and impact of evaluation feedback aligns with its expected benefits and highlight important issues that need to be addressed in order to improve the role of external evaluation in national and local school governance.

In sum, the overarching aim of the research was to illuminate how external school evaluation came about in Iceland and how it has developed, as well as the extent to which the expected benefits, regarding the use of its feedback and perceived impact, have been realised. To achieve that aim, seven research questions, specified in the next section, were used to guide the formulation of the three research papers based on the data collected.

5.1 Research questions and project design

The research consisted of three phases, each with a different focus and each of which formed the basis of a published journal article. The research questions presented below were developed in those papers on the basis of the relevant research and theoretical considerations.

5.1.1 Phase I

The aim of Phase I was to contribute to an improved understanding about the origin of the external evaluation of compulsory schools in Iceland and how it developed after becoming part of official education policy in 1991. The study for Phase I was launched by articulating two research questions:

1. How did the external evaluation of compulsory schools in Iceland come about, and what explains its origin?

2. How has the external evaluation of compulsory schools developed to date, and what explains its development?

5.1.2 Phase II

The aim of Phase II was to identify and analyse principals' and teachers' attitudes towards external evaluation in Iceland, with special attention paid to the ways in which they consider that the evaluation has affected internal evaluation at their schools and driven changes in their own practices (i.e. teaching and leadership). To that aim, the following three research questions were investigated:

3. How can the scales constructed to delineate compulsory school teachers' and principals' reception of and attitudes towards the findings of evaluation (i.e. Accepting Feedback and Teacher Participation) be characterised, compared, and related?
4. With reference to Hofer et al. (2020) and Matthews and Sammons (2004), are principals' attitudes and reactions more positive than teachers' attitudes and reactions?
5. With reference to Ehren et al. (2013), are the mechanisms of accepting feedback, setting expectations, involving stakeholders, and building the capacity to implement improvements positively correlated with changes in practices of internal evaluation, teaching, and leadership?

5.1.3 Phase III

The aim of Phase III of the research was twofold. First, it sought to contribute to current knowledge on the perceived use and impact of the feedback of external evaluation in compulsory schools in Iceland. Second, it sought to elucidate how well the

improvements made, based on the feedback given, have been sustained over time. Two research questions guided that phase of the research, both of which refer to the perceptions of principals and teachers in the schools:

6. How and to what extent do schools use the feedback presented in external evaluation reports?

7. To what extent do schools sustain the changes made after using the feedback from external evaluations instrumentally?

5.1.4 Overview of data collection

To address the research questions, I employed mixed methods, namely document analysis, questionnaires, and interviews. The overall research was longitudinal, and data were collected in three stages. First, I relied on policy documents and interviews with 11 key informants who had participated in the policymaking process or were familiar with it. That part of data collection was undertaken in 2015 and 2016. Second, for participants, I chose the 22 compulsory schools that were the first to participate in the external evaluation in 2013–2015 and conducted a survey in 2016 among the principals and teachers in those schools. Third, for further analysis, I selected six of the 22 schools and interviewed the principal and one or two teachers in each school, as well as performed a document analysis of the school's improvement process. That final part of data collection was undertaken in 2019—that is, 4–6 years after the external evaluation.

This thesis integrates the findings of the three phases of research and explores them in light of relevant literature and conceptual frameworks. Chapter 8 discusses the findings presented in the three papers. In the next chapter, however, the methodology is discussed in detail, in terms of the research design, data collection, and data analysis.

6 Materials and methods

This section describes the methodology followed in conducting the research for this thesis. After providing an overview of historical paradigms in social science, I introduce the perspective of pragmatism and how it informed the research before outlining the research design. Next, I characterise the participants in the three phases of the study, followed by the methods of data collection and data analysis used, first in the two qualitative phases of the research (i.e. Phases I and III) and, after that, in the quantitative phase (i.e. Phase II). In parallel, I discuss the validity and reliability of the methods applied in each phase. Last, ethical concerns are considered, and the advantages and limitations of the research are explained.

6.1 Philosophies and paradigms

The starting point of any research project is to establish the direction of the research strategy—in other words, its *philosophy of science* or *paradigmatic belief*. Because every researcher conducts their study "from within a particular understanding of the nature, texture, and role of social inquiry in society" (Greene & Hall, 2010, p. 121), the researcher's philosophy of science matters.

According to Greene (2007), paradigms "comprise sets of various philosophical assumptions regarding reality, knowledge, methodology, and values" (p. 69), assumptions that form a consensual set of beliefs and practices (Morgan, 2007), that guide and direct thinking and action (Mertens, 2005). Such assumptions about the social world and the nature of knowledge that can be acquired about it differ widely (Greene, 2007), as shown in debates over realism versus constructivism as ways of knowing and whether inquirers should take objectivity or subjectivity as their stance (Greene, 2008). Discussions about the philosophy of social science in the literature (e.g. Greene & Hall, 2010; Morgan, 2007; Teddlie & Tashakkori, 2010) often highlight three broad concepts used to define and compare paradigms: ontology, epistemology, and methodology. *Ontology* refers to the "nature of the social world we study" (Greene & Hall, 2010, p. 121), *epistemology* refers to the theory of knowledge and what counts as warranted knowledge, and *methodology* refers to ways of producing and justifying knowledge (Greene & Hall, 2010; Morgan, 2007). The linkage of ontology, epistemology, and methodology is what Morgan (2007) calls the "metaphysical paradigm".

Of the various research paradigms, post-positivism and constructivism arguably stand opposed and, at any rate, are widely discussed in the literature (e.g. Creswell, 2014; Greene, 2007; Johnson & Onwuegbuzie, 2004; Mertens, 2005; Patton, 2002). In

terms of ontology, post-positivism assumes that the social world is real but knowable only within the limits of probability due to the limitations of the researcher (Mertens, 2005). In constructivism, by contrast, the world is assumed to be multiple and socially constructed (Mertens, 2005), such that no single objective reality exists. Meanwhile, in terms of epistemology, post-positivist research aims to generate, test, and validate theories about human behaviour (Greene, 2007) in order to explain the social world and identify causes, factors, and/or correlations that can be generalised to a larger group (Biesta, 2010; Morgan, 2007). In that paradigm, objectivity is the standard that researchers pursue, given the need for neutrality to prevent their values and biases from influencing their methods and findings (Mertens, 2005). To that end, they require and follow prescribed procedures. By contrast, constructivist researchers seek a contextualised understanding of the varied, multiple meanings that participants make of their experiences and realities (Creswell, 2014; Mertens, 2005). In their studies, data are normally collected in the participants' environments, in an interactive process in which researchers and participants mutually influence each other. Although the researchers indeed rely on the participants' subjective views, they also recognise that their values influence their interpretation and therefore account for them as explicitly as possible. Regarding methodology, research conducted under the post-positivist paradigm mostly follows quantitative methodologies, for the researchers are admittedly external to what they are studying (Creswell, 2014; Mertens, 2005). Objective theories are tested by examining the relationships between measurable variables (Landman, 2006), and numerical data are analysed following statistical procedures (Creswell, 2014). However, qualitative methodologies dominate in research conducted under the constructivist paradigm, because the researchers gather data about their research objects (e.g. participants) in the field. The nature of such methods requires focusing on only a few units of analysis at a time (Landman, 2006).

According to Biesta (2010), the above view on paradigms, which he calls the "mind-world scheme", affords only two options: objectivity and subjectivity. The view also consistently assumes the incompatibility of opposed paradigms owing to different assumptions about the nature of reality and truth (Morgan, 2007; Teddlie & Tashakkori, 2010). Although proponents of mixed-methods research have sought to integrate quantitative and qualitative research methods, that approach does not fall comfortably within one or the other metaphysical paradigm or the mind-world scheme. Biesta (2010) thus asks whether the "scheme is itself inevitable or whether it is possible to think about knowledge and reality in a different way, starting from different assumption" (p. 106). To that purpose, he argues that John Dewey's pragmatism furnishes a good foundation, one that provides an understanding of knowledge that does not require the dualistic mind-world scheme. On the contrary, Deweyan pragmatism affords a different view on knowledge and a different understanding of how humans acquire knowledge (Biesta & Burbules, 2003). In a similar vein, Morgan (2007) calls into question the metaphysical paradigm's basic attempt "to 'impose order' on the practices in social

science research through an externally defined, a priori system from the philosophy of knowledge" (p. 61). By extension, he deconstructs the term *paradigm* and proposes alternative versions, including one that envisions a paradigm "as shared beliefs within a community of researchers who share a consensus about which questions are most meaningful and which procedures are most appropriate for answering those questions" (p. 53).

Different conceptual stances are associated with mixed-methods research, including pragmatism. Morgan (2007) proposes a pragmatic approach as an alternative to the metaphysical paradigm and dualistic mind-world scheme discussed above. Basing his pragmatic approach on Dewey's brand of pragmatism, Morgan advocates the approach as a new guiding paradigm in social science research. In the approach, he rejects the "top-down privileging of ontological assumptions" and focuses on "methodology as an area that connects issues at the abstract level of epistemology and the mechanical level of actual methods" (p. 68). In a diagram of the approach, he places methodology at the centre and connects it to both epistemology and methods with bidirectional links: epistemology ↔ methodology ↔ methods. Morgan argues that the approach's strength is "its emphasis on the connection between epistemological concerns about the nature of the knowledge that we produce and technical concerns about the methods that we use to generate that knowledge" (p. 73)—in other words, that researchers have to reflexively consider what they choose to study and how they choose to study it. Beyond that, methods of inquiry have to fit the problem and questions posed, as argued by Dewey (1938):

> The problem is not epistemological [...] nor is it metaphysical or ontological. In saying that it is logical, it is affirmed that the question at issue is that of the relation to each other of different kinds of *problems*, since difference in the type of problem demands different emphases in inquiry. It is because of this fact that different logical forms accrue to common sense and scientific objects. From this point of view, the question, summarily stated, is that of the relation to each other of the subject-matters of practical uses and concrete enjoyments and of scientific conclusions; not the subject matters of two different domains whether epistemological or ontological. (p. 383)

Dewey's emphasis on the practical worth and actionable value of inquiry is that the process focuses on solving an identified problem (Greene & Hall, 2010).

Morgan (2007) argues that pragmatism provides new options for resolving methodological conflicts and proposes a framework that contrasts his pragmatic approach with the two dominant methodologies: quantitative and qualitative. To relate theory to data, he presents *abduction* as a counterpart to both qualitative induction and quantitative deduction, one guided by abductive reasoning that "moves back and forth between induction and deduction" (p. 71). By extension, considering the relationship

between the researcher and the research process and arguing that being completely subjective or objective is impossible, Morgan recommends a pragmatic *intersubjective* approach to replace the subjectivity of the qualitative approach and the objectivity of the quantitative one. It was Dewey who introduced the pragmatic perspective of intersubjectivity, from which humans are assumed to construct objects of knowledge first individually and then co-reconstruct them in interaction with their social environments (Biesta & Burbules, 2003). Central concepts in that approach are thus communication and shared meaning (Morgan, 2007). Regarding what can be inferred from data, Morgan uses the term *transferability*, meaning a "distinction between knowledge that is either specific and context-dependent or universal and generalized" (p. 72). The pragmatic approach rejects the dichotomy of those two extremes and instead opts to inquire into the extent to which knowledge gained from research conducted in certain situations can be applied in other circumstances. To that end, Morgan (2007) states that researchers "always need to ask how much of our existing knowledge might be usable in a new set of circumstances, as well as what our warrant is for making any such claims" (p. 72). Added to that, Biesta (2010) points out Dewey's preference to refer to results or outcomes of inquiry and research as '"warranted assertions"' not truths (p. 111). *Warranted assertion* is a term that represents potentiality, not actuality, and implies that the result of an inquiry stems from activity that occurs in a particular situation. Thus, although whatever is possible in one situation may be possible in other situations as well, such is not necessarily the case, for the context and transactional situation may differ. For that reason, researchers have to be careful in "putting the label *true* on them [the assertions that they make]—of thinking that they will be warranted for all time and all similar situations" (Biesta, 2010, p. 111). Along similar lines, *warranted inferences* "represent actionable knowledge, that is, knowledge that can be acted upon, in this context and others, or knowledge that is directly actionable for improving the important practical problem being studied" (Greene & Hall, 2010, p. 139). Morgan (2007) argues that those three key concepts in his framework—abduction, intersubjectivity, and transferability—create new opportunities for thinking about methodological issues in social science.

Before introducing the methodology of the research conducted for this thesis, I offer additional insights into Deweyan pragmatism in the following section as the basis for the philosophy behind the research. Thereafter, I elaborate on how those ideas underpinned the entire research project.

6.1.1 Deweyan pragmatism

John Dewey's ideas concerning social inquiry derive from his transactional or interactional conceptualisation of knowledge and understanding (Biesta, 2010; Greene, 2007). The term *experience* is a fundamental concept in Dewey's theory, one that "refers to the transactions of living organisms and their environments" (Biesta, 2010, p. 106), which is a continuous, ever-changing process mediated by culture. Dewey (1917) emphasises that experience "doesn't occur in a vacuum" (p. 50), for "knowledge manifests itself first of all in the way in which organisms transact with and respond to

changes in their environment" (Biesta & Burbules, 2003, p. 11). In that context, Dewey (1938) writes, "We never experience nor form judgements about objects and events in isolation, but only in connection with a contextual whole" (p. 383). However, because experience itself does not ensure knowledge, the only way that humans can acquire knowledge is by combining action and reflection or thinking (Biesta, 2010; Biesta & Burbules, 2003; Dewey, 1917). Furthermore, knowledge stems from the relationships between actions and consequences: "The organism has to endure, to undergo, the consequences of its own actions. Experience is no slipping along in a path fixed by inner consciousness" (Dewey, 1917, p. 49). According to Biesta and Burbules (2003) that dynamic means "that knowledge is not a passive registration of reality 'out there'" but is "always a human construction" (p. 51), one that is contextual, temporal, and related to action.

Dewey's theory maintains that no sort of knowledge affords a more real or truer picture of the world than another (Biesta, 2010). In fact, Dewey (1917) writes that "the chief characteristic trait of the pragmatic notion of reality is precisely that no theory of Reality in general, *überhaupt*, is possible or needed" (p. 64). From that perspective, "Different knowledges are simply the result of different ways in which we engage with the world. They are, in other words, the consequences of different actions" (Biesta, 2010, p. 113). Biesta (2010) adds that the described stance is important for mixed-methods research because it eliminates the hierarchies claimed between different approaches. It also implies that different approaches reveal different sorts of knowledge, which needs to be pragmatically judged in relation to how it was generated.

Dewey's transactional view "forms the theoretical basis for his pragmatist approach that goes beyond the traditional distinction between objectivism and relativism" (Biesta & Burbules, 2003, p. 11). According to Biesta and Burbules (2003), the point of his philosophy was "to overcome the dilemma of inhuman rationality versus human irrationality" (p. 23), as captured in a comment about pragmatism from Dewey himself:

> Pragmatism is content to take its stand with science; for science finds all such events to be subject-matter of description and inquiry [...]. It also takes its stand with daily life, which finds that such things really have to be reckoned with as they occur interwoven in the texture of events. (Dewey, 1917, p. 64)

As that quotation indicates, Deweyan pragmatism honours both realism and constructionism and is therefore not devoted to any one paradigm of philosophy or reality (Biesta, 2010; Biesta & Burbules, 2003; Greene, 2008). Biesta and Burbules (2003) point out that Dewey's transactional approach arguably derives from realist assumptions for his philosophy "does not deny or doubt the existence of a world 'outside'" (p. 10). However, his brand of transactional realism can also be repositioned as transactional constructivism, because knowledge, according to Dewey, "is at the very same time a construction and based on reality" (Biesta & Burbules, 2003, p. 11).

6.1.2 Pragmatism and mixed methods

Many scholars have advocated mixing methods (e.g. Creswell, 2014; Greene, 2007; Johnson & Onwuegbuzie, 2004; Mertens, 2005; Patton, 2002; Teddlie & Tashakkori, 2010). Some have additionally argued that the traditional paradigms of post-positivism and constructionism are reconciled in paradigms such as pragmatism (Greene, 2007; Johnson & Onwuegbuzie, 2004), which fully respects the wisdom of both paradigmatic traditions (Greene, 2007; Johnson et al., 2007). Johnson and Onwuegbuzie (2004), Johnson et al. (2007), and Creswell (2014) all emphasise that pragmatism provides a philosophical basis for mixed methods. By contrast, other scholars, including Biesta (2010) and Biesta and Burbules (2003), maintain that Dewey's brand of pragmatism does not clearly provide a philosophical foundation for combining research. In particular, Biesta and Burbules (2003) argue that Dewey's pragmatist theory of knowledge is not an epistemology in the traditional sense because it relies on "completely different metaphysics and completely different theory of knowledge" (p. 72). They even argue that Dewey's approach should be considered an "anti-epistemology" because his understanding of knowledge is not based on the "dualism of mind and matter" (p. 10). Instead, his approach shifts the focus from traditional questions of epistemology to the transaction between humans and their environments (Greene & Hall, 2010). For that reason, Biesta (2010) claims that "pragmatism should not be understood as a philosophical framework among others, but rather as a set of philosophical tools that can be used to address problems" (p. 97). In a similar vein, Greene and Hall (2010) write that though pragmatism encompasses philosophical attributes, the "pragmatist attends to context, practicality, and instrumentality—not to philosophy—in service of this overall commitment to problem solutions" (p. 138). The following section explains how pragmatism as a stance underpinned the research project on which this thesis is based as well as my practice as a researcher therein.

6.1.3 Pragmatic orientation: An argument for the methodology

Greene and Hall (2010) claim that it "is simply not possible to conduct social inquiry without some self-understanding of what it means to be an inquirer, what the purpose and role of such activity is in society, and what a competent study looks like" (p. 121). In the research conducted for this thesis, I adopted a pragmatic stance and viewed human behaviour and knowledge "as both constructed and as a function of organism–environment transactions" (Greene & Hall, 2010, p. 131). Although the research involved both testing correlations and generating an understanding of a phenomenon, constructionism served as the philosophical grounding because the pragmatic stance was taken to produce constructive knowledge that can be used for action. Nevertheless, respect was demonstrated for the different ways of knowing, which was manifested in the sequential design of the research by using different methods with substantial fidelity to their own assumptive framework. Each method was implemented separately and with integrity to the respective procedures. To further inform and guide the research, I

relied on Morgan's (2007) pragmatic approach for resolving methodological conflicts. Thus, when the results of different methods were compiled and discussed, I had in mind the three key concepts in his framework: abduction, intersubjectivity, and transferability.

Dewey's theory aligned well with my ideas about what it means to be a researcher, especially in prioritising the practical value of the research, addressing important questions, and emphasising that the results lead to action. In a sense I sought to capture what Greene and Hall (2010) call "actionable knowledge". Pragmatically inspired research means performing studies in a way that generates fruitful results, and, in that light my purpose was dual. On the one hand, I sought to contribute to current understandings of how schools use evaluation feedback and whether such use impacts improvements and changes made in schools. On the other, I sought to highlight important ways to improve the role of external evaluation in national and local school governance. In this thesis, I hope to have found a practical way to communicate the results of the research so that they can support policymakers and education authorities in developing effective means of external school evaluation that can serve schools and help them to continuously improve with students' interests in mind. In line with Deweyan pragmatism, the findings presented in this thesis are not meant to be understood as truths but as warranted assertions that can be used as a resource for improvement and decision-making regarding the external evaluation. The findings describe what has been observed in a certain context under certain circumstances and reveal possible connections between actions and consequences. It is my hope that, as such, they can be transferred and used for improvement in other situations.

The methodology of mixed-methods research has been described as "the broad inquiry logic that guides the selection of specific methods and that is informed by conceptual positions common to mixed methods practitioners" (Teddlie & Tashakkori, 2010, p. 5). From that perspective, I rejected the either–or of quantitative and qualitative approaches at all levels of the research's process. I sought to find practical ways to conduct the research and chose methods, techniques, and procedures, both quantitative and qualitative, able to provide the most robust understanding of the problem being studied (Creswell, 2014). As pointed out by Greene (2007), mixed-methods researchers respect multiple ways of knowing and both understand and respect the contradictions between different traditions without feeling forced to choose sides and thus invite "multiple ways of knowing into the same study" (Greene, 2007, p. 27).

6.2 Research design: Mixed-methods approach

According to Johnson et al. (2007), *mixed-methods research* is a "type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches [...] for the broad purposes of breadth and depth of

understanding and corroboration" (p. 123). As those authors explain, qualitative and quantitative approaches exist on a continuum; pure qualitative research methods lie on the one end, while pure quantitative research methods lie on the other, and both may be adopted to collect and analyse data. Between those ends of the continuum are various mixed-methods options available to researchers depending on their research question or questions (Johnson et al., 2007). The core assumption of mixing methods is that combining qualitative and quantitative approaches in research generates a better understanding of the subject than either approach alone (Creswell, 2014; Greene, 2007). In that context, Poteete et al. (2010) highlight that the two dominant research methodologies (i.e. qualitative and quantitative) offer different strengths and weaknesses, and to overcome the limitations of each, researchers need to draw on both of them.

Mixed methods take a variety of forms. In the research conducted for this thesis (i.e. in Phase II and III), I applied an explanatory sequential mixed method, in which a quantitative (i.e. numerical) data are collected and analysed, followed by qualitative (i.e. textual) data, in two consecutive phases (Creswell, 2014). Explanatory sequential mixed methods are "considered explanatory because the initial quantitative data results are explained further with the qualitative data. It is considered sequential because the initial quantitative phase is followed by the qualitative phase" (Creswell, 2014, pp. 15–16). The sequential explanatory design offers the opportunity to explain, or elaborate on, quantitative findings with qualitative results and use the latter to fill gaps that exist in the former. In the research for this thesis, the quantitative and qualitative phases (i.e. Phases II and III) were connected by selecting the case for the qualitative phase and developing an interview guide that was grounded, among other things, in the results of the survey in the quantitative phase. In addition, quantitative and qualitative data were integrated when discussing the results and pinpointing their implications. The qualitative study in Phase I provided an important starting point for Phases II and III but was not directly included in the sequential explanatory mixed method design.

There were three major reasons for mixing methods in the research conducted for this thesis. First of all, to achieve complementarity: to form a broader, deeper, more comprehensive understanding of the subject (Greene, 2007). By combining the descriptive narratives from in-depth interviews with survey data and document analysis, I strove to engage different ways of acquiring knowledge. For example, in Phase I, the interviews provided more detailed information than could be obtained through document analysis alone. Furthermore, by using the survey data from Phase II, I gained a fairly comprehensive picture of the views of principals and teachers on the feedback use, and impact of external evaluation. In turn, the interview data from Phase III afforded a deeper understanding of the survey data by giving the principals and teachers an opportunity to explain their professional practices and decisions, as well as to articulate their opinions, when it comes to evaluation. By collecting data from interviews, I was able to fill in some of the gaps detected when analysing the survey

data. The second major reason for employing mixed methods was to promote development: to use the results of the survey in Phase II to inform the development of Phase III (i.e. identify cases and develop the interview guide). The third and final reason was to achieve triangulation and thereby enhance the validity and credibility of inferences drawn from the evidence, for each method used captured a different perspective on the subject (Greene, 2007).

According to Ivankova et al. (2006), researchers who conduct mixed-methods sequential research have to consider issues such as the weight or priority given to the collection and analysis of quantitative versus qualitative data. Because the focus of the research conducted for this thesis was both to identify correlations and to generate an understanding of phenomenon, quantitative and qualitative data were given equal status during the collection, analysis and interpretation of data. To clarify the mixed-methods procedures used in the Phase II and III, Figure 2 presents the research process graphically.

Figure 2. Visual model of the sequential explanatory mixed-methods design followed in the research, inspired and based on Ivankova et al.'s (2006) visual model

Phase	Procedure	Product
Quantitative Data Collection	Questionnaire: • Principals (n = 22) • Teachers (n = 258)	Numerical data
Quantitative Data Analysis	Data screening, factor analysis, comparison, correlation, and multivariable regression analysis SPSS and R software	Descriptive statistics, t–test, Pearson's and Spearman's correlations, and summary of linear mixed-model analysis
Connecting quantitative and qualitative phases	Purposeful selection of 6 of 22 schools based on evaluation judgement, size and location Development of interview questions	Schools (n = 6) Interview guide
Qualitative Data Collection	Individual in-depth interviews with 6 principals and 8 teachers Documents: Evaluation reports, improvement plans, and annual progress reports Answers to open-ended questions from surveys	Textual data (i.e. interview transcripts and documents)
Qualitative Data Analysis	Coding and thematic analysis Within-school and across-school development of themes Cross-thematic analysis Data analysis using NVivo	Codes and themes Similar and different themes and categories Cross-theme matrix
Integration of the quantitative and qualitative results	Interpretation and explanation of the quantitative and qualitative results to generate inferences	Two published articles: • Discussion of results • Implications PhD thesis

The next section provides an overview of the three papers produced.

6.2.1 Overview of the research

In Table 2, the three papers included in the thesis are summarised in terms of their titles, research questions, and method(s) of data collection and analysis. As noted in Section 5, the overarching aim of the doctoral project was to illuminate (1) how external school evaluation came about in Iceland and how it has developed and (2) the extent to which the expected benefits regarding the use of its feedback and perceived impact have been realised.

Table 2. Focus, research questions, and method(s) in the three papers

	Paper I	Paper II	Paper III
Focus	The origin and development of external evaluation in Iceland from 1990 to 2016	The mechanisms by which external school evaluation in Iceland influences internal evaluation and school professionals' practices	Use and impact of external evaluation feedback in schools
Research questions	RQ I. 1: How did the external evaluation of compulsory schools in Iceland come about, and what explains its origin? RQ I. 2: How has the external evaluation of compulsory schools developed to date, and what explains its development?	RQ II. 1: How can the scales constructed to delineate compulsory school teachers' and principals' reception of and attitudes towards the findings of evaluation be characterised, compared, and related? RQ II. 2: Are principals' attitudes and reactions more positive than teachers? RQ II. 3: Are the mechanisms of accepting feedback, setting expectations, involving stakeholders, and building the capacity to implement improvements positively correlated with changes in practices of internal evaluation, teaching, and leadership?	RQ III. 1: How and to what extent do schools use the feedback presented in external evaluation reports? RQ III. 2: To what extent do schools sustain the changes made after using the feedback from external evaluations instrumentally?
Method(s) of data collection	*Methods*: Interviews with 11 individuals from government and academia Policy document analysis	*Method:* Survey conducted among principals and teachers in 22 schools that were externally evaluated in 2013–2015	*Methods*: Interviews with teachers and principals in six schools evaluated in 2013–2015 Document analysis
Method(s) of data analysis	Descriptive analysis	Descriptive, exploratory, and multilevel correlational analysis	Descriptive analysis.

In the next sections, I characterise the participants in the three phases of the research.

6.3 Participants

Data collection took place in three phases. In Phase I, data from interviews and documents were collected in 2015 regarding the origin and development of the current effort to evaluate schools externally in Iceland. In Phase II, survey data were gathered in 2016 regarding the contribution of the recent external evaluation to improvements and changes in schools. Last, in Phase III, data from interviews and documents were collected in 2019 to gain insight into how schools have used external evaluation feedback. Each phase involved different participants, as described in the following sections.

6.3.1 Participants: Phase I

Participants in Phase I were selected via purposeful sampling, with a view to gather information from parties whom the researcher considered to have the best overview of the topic and insight into its individual elements (Patton, 2002). A total of 10 interviews were conducted with 11 individuals, including five former or current specialists or administrators in the Ministry of Education and Children, two former ministerial advisors or assistants, two academics from universities, and two specialists from the Icelandic Association of Local Authorities. Seven of the 11 interviewees had participated in the formation of education policy in the early 1990s and were thus able to provide information related to the context described in Research Question I.1 (see Table 2). Of the 11 interviewees, 10 had been associated in some way with the development of external evaluation in Iceland during the period covered by the research and were therefore able to provide insights in relation to Research Question I.2 (see Table 2).

6.3.2 Participants: Phase II

As stated, 2013 marked the beginning of the cyclical external school evaluation covered by the research. In May 2016, surveys were sent to teachers ($N = 550$) and principals ($N = 22$) in all of the schools that were evaluated in 2013, 2014, and the first half of 2015, for a total of 22 schools, or 13% of all compulsory schools in Iceland at the time. Schools evaluated from the second half of 2015 until May 2016, when the surveys were conducted, were not included because asking them about the external evaluation's contribution to school improvement so soon after the evaluation was unrealistic.

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

employed at the time of the evaluation. The low response rate among teachers was at least partly due to the timing of the administration of the survey in late spring, as detailed in Section 6.4.2.1.

6.3.3 Participants: Phase III

Of the 22 schools subjected to external evaluations in Iceland in 2013−2015, six were selected (see Table 3). To obtain a broad representation of schools with a wide range of contexts and variation in characteristics, the selection of schools was informed by evaluation judgements, size, and location (i.e. urban vs. rural). To protect the anonymity of the schools, all identifying information was omitted in Paper III.

Table 3. 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

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, with only 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). The selection of teachers for interviews was based on certain criteria. It was assumed that teachers, all of whom 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 the recent external evaluation.

The selection criteria 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.

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.

In the following sections, I explain the procedures of data collection and analysis. First, I outline data collection and analysis in the two qualitative phases of the research (i.e. Phases I and III), after which I turn to discuss data collection and analysis in the quantitative phase (i.e. Phase II). I also discuss the methods used to increase the validity and reliability of the study in each phase of the research.

6.4 Data collection and analysis

In Phases I and III of the research conducted for this thesis, data were collected via individual interviews and document analysis. By contrast, in Phase II, data were gathered using questionnaires. The three phases and timeline of the research are illustrated in Figure 3.

Figure 3. Outline of data collected in the research

Phase 1: May to September 2015		Phase 2: May 2016		Phase 3: May to October 2019
• Public policy document analysis over the period 1974 to 2016 • Interviews with 11 selected participants in policy development and implementation	→	• Surveys to teachers and principals in 22 schools evaluated in 2013, 2014 and 2015	→	• Interviews with 6 principals and 8 teachers in six of the 22 schools evaluated in 2013, 2014 and 2015 • Document analysis: Evaluation reports, improvement plans and progress reports.

In the following sections, I first discuss the interview and document data collection and analysis in Phases I and III and, after that, discuss the survey data collection and analysis in Phase II.

6.4.1 Qualitative methods: Interviews and document analysis in Phases I and III

Phases I and III of the research focused on collecting qualitative data, namely using in-depth semi-structured interviews and document analysis. In general, qualitative data describe, tell a story, and capture and/or communicate someone's experience (Patton, 2002).

An in-depth interview is a method of data collection that includes one interviewer and at least one interviewee discussing a specific topic in a detailed, intensive manner (Hennink et al., 2011). Each such interview has a certain purpose, and the researcher seeks to gain insight into the topic being studied with the goal of producing knowledge about it (Brinkmann, 2018). In-depth interviews can range from being relatively unstructured to relatively structured, although semi-structured interviews are the most common (Brinkmann, 2018). In semi-structured interviews, the interviewer identifies a series of themes that the interview needs to include as well as suggested questions (Kvale, 1996). At the same time, it is possible to change the order and form of the questions used to follow up on the interviewee's answers (Brinkmann, 2018). According to Creswell (2014), interviews are useful to acquire historical information, as collected in Phase I of the research for this thesis. In the research, I sought to hold the interviews at each interviewee's workplace, where they could be expected to feel most at ease and speak freely (Hennink et al., 2011).

Document analysis can be executed both quantitatively and qualitatively and used as a stand-alone method or in complement to other methods (Flick, 2014). Documentary evidence may serve as the basis for further data collection (Bickman et al., 1998) or play a subsidiary or complementary role (Peräkylä & Ruusuvuory, 2018). In Phase I of the research conducted for this thesis, an analysis of policy documents was the mainstay, and the interview data were analysed to gain a better understanding and answer questions that arose during the analysis of policy documents. Through the analysis of the documents, interviewees who could provide detailed information were also identified. That process was the opposite of the process in Phase III, in which the interview data were the chief pillar. Document analysis served two purposes: to guide preparation for the interviews and to support the analysis of data from them.

6.4.1.1 Phase I: Data collection

In Phase I, the research was implemented based on the acquisition, examination, and analysis of documents, articles of law, regulations, reports, and other official documents, as well as interviews. The interviews were carried out from May to September 2015. I conducted the interviews myself, always at the participants' workplace, except the two conducted at my workplace because the interviewers could not provide a suitable facility for the interview and requested me to provide one.

The first step in the process was to gather relevant written materials, undertaken following a systematic procedure that evolved from the topic of the study. A wide range of documents were closely read, interpreted, and analysed, including policy documents, legal statues, conference papers, national curriculum guides, reports, commentaries on draft laws, and explanatory memoranda accompanying a parliamentary bill. Those documents were accessed primarily by searching official websites at the central governmental level and were critically read in order to understand the origin and development of external school evaluation in Iceland.

Reading the documents revealed which parties had been involved in making policy or implementing it during the various periods covered by the study and who were thus able to provide more detailed information. All parties requested to be interviewed responded affirmatively, for 11 interviewees in total. The interviews were semi-structured and supported by predetermined open-ended questions, which provided the basis for the specific development of each interview in light of the information that the interviewee provided (Lichtman, 2006). Thus, the interviewees did not always answer the same questions, and no one interviewee answered all of the questions. The interview guide addressed five areas: (1) the origin of the external evaluation of compulsory schools, (2) the implementation of external evaluation after legislation in 1995, (3) the lead-up to and preparation of the legislation in 2008, (4) the development of external evaluation after the law took effect in 2008, and (5) the purpose of the external evaluation and its impact. The interview guide appears in the Appendix to Paper I.

The interviews lasted 40−93 minutes, were recorded with the interviewees' permission, and were later transcribed. Absolute anonymity was guaranteed to all participants.

6.4.1.2 Phase I: Data analysis

Data analysis covered the progression of external evaluation in Iceland from the time when the concept first appeared in official policy documents in 1991 until the latter half of 2016. Processual analysis was applied, in which researchers are generally "purposefully looking at action/interaction and noting movement, sequence, and change as well as how it evolves (changes or remains the same) in response to changes in context or conditions" (Strauss & Corbin, 1998, p. 167). Time and history are central to process analysis (Pettigrew, 1997).

Initially, available public documents were gathered and analysed by reading and rereading the materials to pinpoint key themes, sequences of events, and transition points. Next, the process was divided into timeline-based sub-processes representing the periods 1974−1990, 1991−1994, 1995−2008, and 2008−2016; Table 4 presents the sub-processes and their associated themes and coding categories. In sum, the analysis of the topic was based on 38 written sources and documents.

An interview framework based on document analysis was created to gain a deeper, more comprehensive perspective of the case in question and thereby integrate information about the what, why, and how of the process being studied (Pettigrew, 1997). Interview data were gathered until saturation was considered to have been reached. The interviews were analysed in the same way as the documents (see Table 4).

Last, the data were examined and explained in light of Kingdon's (2014) theory on how matters come to be included in a government's agenda and Kotter's (2012) theory on the introduction of policy and change. Ideas and concepts originating from NPM were also used when interpreting the data.

Table 4. Themes, sequence of events, and transition points regarding the implementation of external school evaluation in Iceland

1974—1990 ■ External evaluation in the Compulsory School Act (1974) ■ External evaluation in the National Curriculum Guide for Compulsory Schools (1976) ■ The situation circa 1990
1991—1994 ■ External evaluation in the Compulsory School Act (1991) ■ External evaluation in the "Towards a New Century" policy (1991) ■ External evaluation in new education policy (1994) ■ Emphasis on internal evaluation audits ■ Perceived problems in compulsory schools ■ Response to proposals for external evaluation

1995—2008
■ External evaluation in the Compulsory School Act (1995)
■ Internal evaluation audits
■ Ministers and their interest in external evaluation
■ Evaluation of compulsory schools
■ In anticipation of the 2008 legislation
    ○ What drove changes in school evaluation
    ○ Which stakeholders were involved in preparing the new law
    ○ Stakeholders' responses to proposals for external evaluation

2008—2016 ■ External evaluation in the Compulsory School Act (2008) ■ Cooperation with local authorities on external evaluation ■ Funding for evaluation ■ Expectations of external evaluation ■ Perceived usefulness of the evaluation for education authorities ■ Attitudes towards external evaluation

6.4.1.3 Phase III: Data collection

The data in Phase III came from official documents as well as transcribed interviews. Evaluation reports and improvement plans were used both to inform and prepare the interviews and to predetermine codes and themes. Annual progress reports from the schools made to the Ministry of Education and Children regarding the implementation of the improvement plans were used to obtain information about the progress of improvements and changes. 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. I arranged all appointments at the interviewees' schools except for one school where the interviews took place in connection with their participation at a conference. Because the given information might have been sensitive, absolute anonymity was promised to all participants and maintained. All participants signed their written informed consent to participate (see Appendix A). 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. Presented in the Appendix to Paper III, the interview guide addressed five areas: (1) attitudes towards external evaluation and how the feedback was handled in the school, (2) changes in leadership and management, (3) changes in learning and teaching, (4) changes in internal evaluation, and (5) whether changes made in those three areas had been sustained or further developed. To help each interviewee to review the improvement actions, I 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 evaluation and whether the improvements and changes made had been sustained or remained in development. The interviews lasted 48–90 minutes, were recorded, and were later transcribed.

6.4.1.4 Phase III: Data analysis

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 first step in the thematic approach was familiarising myself with the data. Because I both conducted and transcribed all of the interviews myself, I knew the data well when it came to analysing them. While transcribing the interviews, I began to reflect on what I observed and took notes. After sufficiently familiarising myself with the data, I formed initial ideas for codes.

I coded the segment of data that was relevant to the research's focus 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. During the coding process, I also looked at the attitude towards the evaluation and how sustained the improvements and changes the schools made were. Sub-codes for each area were

developed, and the codes were modified as I proceeded through coding, beginning with coarse coding and followed by more detailed coding. Ideas about the main themes were also created in parallel. Predefined codes and themes referred only to the instrumental use and impact of the findings of external evaluation, how sustained the improvements and changes made were, and attitudes towards the evaluation and its feedback. However, when additional themes were identified that did not represent instrumental use, I 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 were first coded by each school and subsequently assigned to the relevant theme. As the focus shifted to the themes, the themes were further analysed and refined. Table 5 presents an overview of the themes and codes.

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 about the initial status of the schools, planned changes, and improvement actions in the 2–3 years after evaluation feedback was received. That strategy allowed me to triangulate my thematic interview analysis and conclude that certain measures were indeed a result of external evaluation.

The data were examined and explained in light of Rossi et al.'s (2004) and Aderet-German and Ben-Peretz's (2020) distinction of ways of using feedback—that is, for instrumental, conceptual, persuasive, and reinforcement-oriented use.

Table 5. Themes and codes, with sub-sub-codes shown in parentheses

Instrumental use:
• Leadership and management
   o Distributed leadership
   o Professional collaboration among staff (e.g. teachers' meetings, collaboration, and subject-focused teamwork)
   o Instructional leadership, classroom observation and feedback (e.g. implementation, benefits, feedback, barriers, teachers' attitude, principals' attitude, and procedures in development)
• Learning and teaching
   o Differentiated instruction (e.g. diversity in teaching, information technology, collaboration and dialogue, choice of optional subjects, and students' areas of interest)
   o Students' achievement and use of assessment
   o Students' democratic participation
   o Team teaching and planning
• Internal evaluation
   o Changes
   o Evaluation team
   o Knowledge to perform internal evaluation
   o Publication and changes

Conceptual use: • Usefulness of external view • Discussions and reflections • Professionalism • Support for new principals • Creation of focus Persuasive use Reinforcement-oriented use
Sustained improvements and changes: • Sustained changes and progress • Professional knowledge in the school • Restraint • Follow-up from the Ministry of Education and Children • Support from the municipality
Attitudes towards external evaluation: • Attitudes towards and experience of the external evaluation • Attitudes towards the continuation of the external evaluation • Stress prior to the external evaluation • Acceptance of or resistance towards recommendations

6.4.1.5 Validity and reliability in Phases I and III

Conducting pragmatically oriented research interwoven with natural conditions affords researchers less control over conditions than in experimental designed research, which presents challenges when ensuring the validity of research. Nevertheless, it is important for professionals, practitioners, and policymakers who want to use the results to trust them, at least to a certain degree (Merriam, 1998; Morse, 2018). Scholars who have written about methodology have highlighted that increasing confidence in the results requires accounting for their validity and reliability (e.g. Bickman et al., 1998; Creswell, 1998; Merriam, 1998; Mertens, 2005). The terms *validity* and *reliability* have their origins in positivist, quantitative research and include, among other things, the generation of rigorous data, of replicable research, and of generalisable results (Creswell, 2014; Lincoln et al., 2018; Merriam, 1998). By contrast, the nature of qualitative research means that such accounting takes different forms from ones used in quantitative research (Creswell, 2014; Merriam, 1998). Some scholars have suggested using the concept of validation instead of validity in qualitative research because the idea of validation emphasises that qualitative research is a legitimate research method in its own right (Creswell, 1998; Flick, 2014). Additional terms used to address validity in the literature reporting qualitative research include *trustworthiness, credibility,* and *transferability* (Bickman et al., 1998; Creswell, 2014; Mertens, 2005). Because the research conducted for this thesis involved using mixed methods, concepts coming from the two paradigmatic positions of post-positivism and constructivism were applied in parallel.

In qualitative research, validity requires the researcher to apply several strategies in order to ensure that results are credible and dependable (Creswell, 2014). To demonstrate trustworthiness in the research for the thesis, I employed data and methodological triangulation, clarified my bias as a researcher, sought feedback from my supervisors on the data and interpretation, used longitudinal design and rich data, and consciously sought to minimise my influence as a researcher on the interviewees. Those precautions are explained in the following paragraphs, although the last-mentioned point is discussed in Section 6.5.1, which addresses ethical aspects of the interviews.

Studies that use only one method of data collection are at higher risk of error than studies that use multiple methods (Patton, 2002). Triangulation is one of the strategies for ensuring the validity of research (Flick, 2018). According to Flick (2018), *triangulation* refers to when "an issue of research is considered [...] from (at least) two points or perspectives" (p. 445). There are several types of triangulation, including *data triangulation*, or "the use of different data sources" (Flick, 2014, p. 183), and *methodological triangulation*, or the use of different research methods. Both data triangulation and methodological triangulation were used in Phases I and III of the research in order to validate the findings. Data triangulation involved (1) comparing

data from interviews and documents and (2) comparing principals' and teachers' answers to the same questions. Meanwhile, methodological triangulation involved using both interviews and documentary analysis.

According to Patton (2002), validity depends on the careful construction of an instrument, such that it measures what it intends to measure. Because the researcher is the instrument in qualitative research, a strategy for validation is to clarify the bias that the researcher brings to the study (Creswell, 2014). Section 1.2 addresses researcher bias, a topic also addressed in Section 8.5. Furthermore, I sought to take a critical view during the analysis in order to avoid biased interpretation, and throughout the research period, ongoing discussions between my instructors and me about the data and the interpretation of the data took place. Co-authorship (i.e. Paper III) also provided us with an opportunity for discussion, enhanced the collective understanding of the data, and reduced researcher bias.

An advantage of longitudinal studies is greater validation because data are gathered over a period of time and become richer (Merriam, 1998). The longitudinal design of the research conducted for this thesis allowed me to examine the subject over a longer period, which increases the credibility of the results. Furthermore, the research was based on rich data; interviews with 25 participants (i.e. 11 in Phase I and 14 in Phase III) transcribed verbatim and 67 written sources and documents (i.e. 38 in Phase I and 29 in Phase III). Thus, the data are detailed and suitable for giving a holistic picture of the topic studied.

The reliability of qualitative research indicates that the researcher's approach is consistent (Creswell, 2014), which in the research for this thesis was achieved by having the same researcher—me, that is—collect all of the data. Furthermore, predetermined questions, albeit only semi-structured ones, were used to make the interviews comparable (Merriam, 1998; Morse, 2018), and efforts were made to avoid leading questions (Kvale, 1996). Another effort was made to continuously document the research process, because detailed documentation may increase reliability in qualitative research (Flick, 2014). A rigorous, trustworthy process analysis (i.e. Phase I) and thematic analysis (i.e. Phase III) were conducted, as described in the sections above. Last, to enhance reliability, the assumptions and theory underlying the research (Merriam, 1998) were explained, and multiple methods of data collection used, as also discussed above.

6.4.2 Quantitative: Questionnaires in Phase II

The quantitative phase of the research conducted for this thesis consisted of online questionnaires completed by principals and teachers. A questionnaire survey provides a numerical description of trends, attitudes, and/or opinions in a population or sample of that population (Creswell, 2014). The purpose of using a survey in the research was to obtain answers from more respondents than possible using interview data alone.

Furthermore, gathering principals' and teachers' answers to the same questions allowed me to compare their answers. Descriptive and correlational statistics played an important role in answering the research questions and thus justify a higher level of confidence in the findings.

6.4.2.1 Data collection

Data collection was based on questions from the surveys for principals and teachers, as summarised in the Appendix to Paper II. Some of the questions were the same for both teachers and principals, whereas other were designed to be answered by only one of the two groups. The teachers were asked 28 questions and the principals 30 questions. All questions utilised a Likert-type scale for responses.

The surveys were conducted to evaluate the implementation and impact of the external evaluation project on behalf of the Ministry of Education and Children, the Icelandic Association of Local Authorities, and the Directorate of Education. Because the subject of the survey was directly related to the research, I offered to conduct the evaluation free of charge in return for being able to collect more information than needed for the evaluation itself and to utilise all of the data in my doctoral study. Results of the evaluation were published in a report in the fall of 2016 (Ólafsdóttir, 2016b). For the purposes of the research, questions were added to the surveys; although not processed for the evaluation report, they became the subject of the research. The online platform SurveyMonkey was used to collect data.

Before issuing the questionnaire, I called all of the principals, presented the study, and requested the school's participation. All of the principals agreed to participate and to encourage teachers to answer the questionnaires as well. Following the call, the principals sent me the email addresses of all teachers teaching in the school at the time.

The survey began in May 2016. In the second week of May, the initial questionnaire and cover letter explaining the purpose of the research were emailed to the principals and teachers (see Appendix B for details). A reminder email was sent two weeks later to the entire sample. In the reminder email sent to the principals, the principals were also asked to encourage teachers to answer the survey. In the first week of June, an email was sent to all principals, who were again asked to encourage teachers to respond to the survey and, if possible, allocate time for them to answer the survey on a specific day at the end of the school year. If the teachers' response rate was low (i.e. <40%), a call was made to the principals in addition to the email for the same purpose. Ultimately, it was impossible to reach all of the principals. At the end of May, I also called the principals who did not respond to the survey and encouraged them to respond.

6.4.2.2 Data analysis

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

First, to construct scales based on composite variables for the analysis, an exploratory factor analysis of the teachers' data was performed. Cut-off values of 0.3 were used as a minimum for significant factor loadings (Kline, 1994). The principals' data were based on too few participants to be able to trust correlation coefficients in factor analysis (Kline, 1994). However, given the importance of comparing the answers of teachers and principals, variables in the principals' data were combined in the same way as in the teachers' data, although variables were removed from both datasets when their removal increased Cronbach's alpha. Cronbach's alpha was calculated for each factor variable to evaluate the reliability of the instrument. Descriptive statistics (*M* and *SD*) were calculated for all measures.

Second, the responses of teachers and principals to four variables were compared. To account for the multilevel structure of the data, we used a linear mixed model to analyse the difference by including a random effect for school.

Third, correlation coefficients were calculated to describe the relationship between the variables. To account for the clustering in the data from teachers within a school, we used a mixed model with school as a random effect. Next, the regression slope between standardised variables was used as an estimator for Pearson's correlation coefficient. After that, the method described in Rosner and Glynn (2017) was applied for the maximum likelihood estimation of the Pearson correlation in the clustered data setting, and Spearman's rho rank was calculated for the principals' data. The interpretation of the magnitude of the correlation relied on the following criteria: 0.0–0.1 was considered to be non-existent, 0.1–0.3 to be small, 0.3–0.5 to be moderate, 0.5–0.7 to be large, and 0.7–0.9 to be very large (Hopkins, 2002).

Last, to deepen the correlation analysis, a multivariable regression analysis was performed on the teachers' data. A linear mixed model was used to investigate, according to the theoretical model, whether two of the three assumed mechanisms of school inspection for which information was available could significantly explain changes in internal evaluation and teaching practices—more specifically, the explanatory variables Accepting Feedback and Setting Expectations. In addition, the explanatory variable Teachers' Participation was added to the model. The normality of residuals was visually inspected using Q–Q plots, which revealed that the observations approximately followed a straight line, thereby indicating that the assumption of normally distributed residuals was satisfactory.

All results were obtained using multilevel analyses, wherein teachers' and principals' answers were modelled at the first level and schools at the second. Exceptions were descriptive analyses and correlation analyses based on the data from principals. The

variables were standardised, and other than descriptive statistics, only standard variables were used in calculations. Analyses were performed in SPSS version 27, except for linear mixed-model analysis, which was performed in R.

I used a part of the mechanisms of Ehren et al.'s (2013) theoretical framework to understand the reception and potential influence of external evaluation in Iceland on internal evaluation and the practices of school professionals.

6.4.2.3 Validity and reliability in Phase II

As discussed above, two criteria for judging the quality of a quantitative research are validity and reliability (Yin, 2014). The *validity* of quantitative research broadly means that meaningful, useful inferences can be drawn from scores achieved on instruments (Creswell, 2014). As in the qualitative phases of the research (i.e. Phases I and III), data triangulation was used to validate the findings in the quantitative phase (i.e. Phase II), which involved comparing the answers from principals and teachers to the same or similar questions. The fact that the results are based on a single case—that is, Iceland— limits the *external validity* of the findings, meaning the degree to which one can generalise the results to other situations, as discussed in Section 6.6.

Reliability depends on whether systematic, scientific procedures that are replicable are used to collect data (Wang & Park, 2016). In Phase II of the research, efforts were made to increase the reliability of the survey instrument through pilot testing, conducted among six former principals and 10 former or current teachers chosen via convenience sampling. The questionnaires were additionally reviewed by two individuals with expertise in the formulation of survey questions. Based on the pilot test and the experts' review, the wording of questions was revised as needed. The pilot test also had the purpose of measuring the time needed to complete the questionnaire so that participants could be informed about the expected time commitment in advance. During data analysis, a systematic procedure was followed, as explained in the previous section. Factor analysis was used to determine the domains, and the internal consistency of the scales constructed was calculated with Cronbach's alpha.

6.4.3 Combining Phases I–III

In Chapter 8, the results of Phases I, II, and III are integrated and discussed to address the overarching aim of the research conducted for the thesis. It has previously been discussed how triangulation was used to validate the findings in each phase of the research. Data and methodological triangulation were used to obtain a more comprehensive picture of the research topic, especially regarding the second aim of the research reported in Papers II and III. Methodological triangulation involved using three types of methods: interviews, surveys, and document analysis. Data triangulation involved the use of three sources of data—teachers, principals, and documents— collected at different times due to the longitudinal approach of the research.

The following sections consider ethical concerns in the research, as well as the research's advantages and limitations.

6.5 Ethical considerations

Various ethical issues arise in any research project that involves humans, which is an important concern to bear in mind when planning and conducting research (Flick, 2014). The criteria discussed in Iceland's Code of Research Ethics for Public Higher Education Institutions (2020) were used as an ethical guideline throughout the research period. The topic of the research for this thesis does not fall into the sensitive category encompassing the obligation to notify and obtain a license for processing personal data. It was therefore not mandatory to obtain permission for the research. Nevertheless, the research was announced to the University of Iceland's Science Ethics Committee before interview data were collected.

In the following two sub-sections, I discuss ethical concerns that arose in the interviews (e.g. informed consent to participate, anonymity in data storage and the presentation of results, and the researcher's approach) and surveys (e.g. anonymity of respondents and the risk of social desirability bias).

6.5.1 Ethical concerns in the interviews

Ethical concerns in research design involve obtaining informed consent from interviewees to participate in the research and ensuring confidentiality (Kvale, 1996). All participants in Phase III of the research signed their written informed consent to participate (see Appendix A), whereas the informed consent of the participants in Phase I was obtained orally. The interviews with participants were recorded with their permission. All data, including the electronic survey files, interview recordings, and transcripts, were kept in a password-protected file on my computer. The recorded interviews will be destroyed a year from the defence of this thesis.

A major ethical concern is the anonymity of participants (Flick, 2014; Kvale, 1996). Absolute anonymity was guaranteed to all participants in the two qualitative phases of the research and maintained. In Phase I, because a relatively small group had been involved in policy formation and/or the development of external school evaluation in Iceland, there was a risk that they could easily be identified. Therefore, careful considerations related specifically to anonymity were given throughout the course of the research. To prevent associations between individual comments, the interviewees were not issued pseudonyms. Instead, their comments were anonymously interwoven with registered data related to the topic, and direct quotations were identified only with "from an interview". Given the method used to select schools for Phase III of the research—that is, selecting six of the 22 schools first to be evaluated, which is information available on public websites—special care was taken to omit information that might identify the schools in the presentation of the results in Paper III. In response

to the difficulty of guaranteeing anonymity in a small community, I used several tactics to conceal the identity of participants and schools. For one, the participants and the schools were assigned fictitious names in the transcription of the interviews. For another, reference to participants in the presentation of the results was anonymous; participants were distinguished only by job title and the pseudonym of the respective school. Effort was also made to make the interviewee's gender undetectable by using the singular "they" instead of "he" or "she". Moreover, to protect the anonymity of the participating schools, all identifying information was omitted in the presentation of the results. On that count, the schools were given pseudonyms (e.g. School A and School B), and the location of the schools was not specified aside from whether they were located in urban or rural areas, nor was the size of the schools aside from whether they were large or small.

Because interviews are conversations held for a specific purpose that guides the conversation, various issues arise concerning power and control, which is important to reflect upon for ethical reasons (Brinkmann, 2018; Kvale, 1996). Among other things, the person of the researcher matters for the quality of the research (Kvale, 1996). Because I work at the agency responsible for carrying out the external evaluation of schools in Iceland, my presence might have influenced the participants' responses, such that they might have wanted to give an idealised image of the schools' work, either to please me or to colour the school's image. To prevent that response effect, I focused on being open, honest, fair, and transparent and assured the participants' confidentiality and anonymity. I also sought to build trust by providing information about the purpose of the research and the use of the data collected. Although I clearly stated that I was working at the Directorate of Education while completing doctoral research, I was also conscious of not giving too much information about my relationship with external school evaluation in Iceland to avoid biasing my respondents towards views that could have negatively impacted the quality of the data collected and analysed.

6.5.2 Ethical concerns in the surveys

Each participant received a cover letter along with the questionnaire explaining the purpose of the research and the planned use of the findings. Given the number of teachers who responded to the survey from each school, it is impossible to identify individual teachers, even despite their reported affiliation with the school where they taught. The principals, by contrast, are clearly identifiable because only one principal answered for each school. In a phone call made to each principal, I made them aware of that reality and assured them that I would be the only one processing the data. Consequently, special care was taken to protect the data, even if the data were not classified as being sensitive, and I kept the data in a password-protected folder on my computer. All results based on the principals' answers were presented for the group as a whole and therefore cannot be traced to individual participants.

Another ethical concern relates to the origin of the questionnaires, which were conducted for purposes other than obtaining data for my research. As discussed, I conducted the surveys to evaluate the implementation and impact of the external evaluation project for the Ministry of Education and Children, the Icelandic Association of Local Authorities, and the Directorate of Education. Collecting data for my doctoral study was the second purpose of the surveys. In the cover letter, the participants were informed about the dual purpose of the survey. Because the evaluation of the project was made for an education authority in Iceland, it has an ethical aspect specifically regarding the voluntary participation of the principals. Because the school—Ministry relationship is one of authority and power, the principals might have experienced that they did not have the option of refusing participation. However, I have no reason to believe that such was the case, or at least was not problematic, for participants faced no consequences in refusing to respond to the survey. The request for participation was presented as a friendly appeal, not as an authoritative order. For the same reason, there was a risk of social desirability bias because those who answered the questions were responding to a questionnaire sponsored by education authorities, which may have led them to express overly positive statements. It is difficult to assess the impact that that dynamic might have had on the results.

6.6 Methodological advantages and limitations of the research

The main strength of the research conducted for this thesis was that it drew on different approaches and thereby minimised the limitations of each method employed (Creswell, 2014). An attempt was made to choose the tools that best suited the topic of the research. In particular, the advantages of the sequential explanatory design include straightforwardness and opportunities to explore quantitative results in greater detail (Ivankova et al., 2006). Triangulation and more complete evidence from a variety of approaches also afford greater confidence in the conclusions of the research for both me as a researcher and future users of the findings (Lieber & Weisner, 2010). Another strength of the study was its longitudinal design, which allowed me to examine the use and impact of the external evaluation over a longer period. However, any strength of a study can easily become a weakness if challenges are not addressed. Mixed-methods research posed a challenge for me as a researcher, including that I had to be familiar with the collection and analysis of both quantitative and qualitative data; I also had to collect extensive data in different formats and communicate them in clear, compelling ways (Creswell, 2014; Lieber & Weisner, 2010).

The research's limitations include the potential risk of a non-response error, aspects of the sample, the research's external validity, and self-report data. First, Phase II of the research involved the potential risk of a non-response error (Mangione, 1998)—that is, a problem caused by differences between teachers who responded to the survey and teachers who did not, due to a low response rate in some of the schools and a high response rate in others. Furthermore, in Phase III, most of the teachers interviewed

were selected based on their active involvement in the improvement process. Their views may thus not have been typical of the views held by their colleagues who were less involved in the process. Second, the use and impact of the evaluation feedback was not examined from the perspective of other key stakeholders involved in the process, including students, school staff other than principals and teachers, and local authorities. Third, because the sample used in Phases II and III was, on the one hand, the total population of schools evaluated at the beginning of the external evaluation programme and, on the other, a case study of six schools from that population, I cannot confidently state that the sample is representative of schools in Iceland. Even though the schools were not selected at random, they should nevertheless reflect a reasonable distribution of schools. Fourth, the external validity of the research's results is possibly limited to the approach of external evaluation and perhaps to the Icelandic school system and its specific context in relation to education. The system has used external evaluation, which is a relatively low-stakes approach with greater focus on improvement than on accountability. Nonetheless, schools are somewhat held accountable for implementing improvements by being required to submit improvement plans and progress reports. Those unique characteristics of the Icelandic evaluation system have to be considered when the research's results are used to understand other accountability systems. Last, the research relied on self-report data, which may be biased and have influenced the validity of the results, because teachers and principals may have over- or underrated the use and impact of evaluation feedback. That last limitation is further considered in Section 8.5, where the threats to the research's validity are discussed.

Despite those limitations, the research's results can nevertheless provide important insights, and inferences can presumably be drawn from the themes, patterns, and relationships identified, which contributes to the literature on school evaluation. In line with Deweyan pragmatism, the knowledge generated during the research is not meant to be understood as the truth but as a set of warranted assertions that can be transferred and used in other circumstances.

Altogether, in this chapter I have offered an outline of the methodological approach taken in the research conducted for the thesis. In the next sections, I briefly summarise the results of the three papers that report that research.

7 Results

The results presented in this chapter follow the same order of the three papers completed for this thesis. I concentrate on the major findings presented in each paper because the general theoretical and methodological approaches have already been outlined in Chapters 3 and 6. First, the origin and development of external evaluation in Iceland reported in Paper I are described; Paper I was published in Icelandic but translated into English for this thesis. Second, the mechanisms by which external school evaluation in Iceland influences internal evaluation and school professionals' practices are described, as reported in Paper II. Last, findings on the use and impact of external evaluation feedback in schools are presented, as reported in Paper III. All three papers appear in appendices to this thesis.

7.1 Paper I

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

The first paper for this thesis reports a study of the progression of external evaluation in Iceland from the time when the concept first appeared in official policy documents in 1991 until the latter half of 2016 (i.e. Phase I). The research questions guiding Paper I were "How did the external evaluation of compulsory schools in Iceland come about, and what explains its origin?" and "How has the external evaluation of compulsory schools developed to date, and what explains its development?" The study for Paper I served two purposes: first, to clarify the origin of the external evaluation of compulsory schools in Iceland, and, second, to illustrate its development. To those purposes, I studied documents related to official policy formation dating back to the 1990s—that is, when external school evaluation first emerged in policy documents in Iceland—until the latter half of 2016. Various documents that shed light upon the progress of school evaluation were examined. Furthermore, interviews with individuals who had either participated in policy formation related to external evaluation or been involved in its implementation were conducted. Along those lines, the purpose was to clarify when external evaluation came to the attention of public administration, which ideas or agents triggered that development, and how the concept evolved through 2016. That prehistory of external evaluation in Iceland is examined in light of theory on how issues end up on the government's agenda, NPM, and theory on the adoption of policy and change.

Among the key findings of the paper, the empowerment of municipalities via the transfer of compulsory school operations to them, in combination with increased centralisation by way of clear objectives and emphasis on external evaluation and supervision, were the key aspects of the changes occurring in the Icelandic school system in the 1990s. All of those aspects were interwoven with the fundamental concepts of NPM and associated trends in various parts of the world. The increased independence of the schools with the aim of attaining a more successful, high-quality school system, in accordance with the ideology of NPM, required, in the government's opinion, enhanced quality management as well as external evaluation and supervision of school operations. In Iceland, a form of external supervision was developed that, on the basis of defined criteria, was tasked with investigating whether schools were evaluating their own quality using internal evaluation and whether relevant formal requirements were satisfied. Such supervision was supposed to provide the schools with sufficient discipline and to encourage them to focus on internal evaluation and improvement, thereby ensuring a quality school system. External evaluation, by contrast, comprising the comprehensive quality control of school operations, did not gain a foothold at the time despite being included in plans for public education policy. The fundamental obstacle appears, among other things, to have been lack of conviction on behalf of the authorities and other stakeholders regarding the importance of such evaluation and its potential for solving an alleged problem in the school system. In particular, it was obvious that establishing an external evaluation applicable to all compulsory schools in Iceland would be an extremely costly undertaking. Moreover, Iceland lacked the specialist human resources to successfully implement such evaluations.

Thus, it was not until the beginning of the 2010s that external evaluation was launched in Icelandic compulsory schools in the sense outlined—that is, external evaluation emphasising the review of key processes in a school's activities for the purpose of stimulating school improvement. By then, it was clear that the attempt to persuade the country's compulsory schools to perform internal evaluation had not been particularly successful and that the internal evaluation alone was therefore not a viable basis for quality school management. At that point, it was considered to be necessary to strengthen municipal responsibility for school operations; in response, by way of compulsory school legislation in 2008, municipalities' obligation to conduct evaluation and supervision was given a legal basis. The renewed emphasis on external evaluation could be attributed to, among other things, the fact that various challenges were becoming increasingly problematic in the school system, including bullying, poor school atmosphere, communication problems, and inadequate facilities. In 2010, the Ministry of Education and Children launched regular inspections of six compulsory schools per year, an arrangement that lasted for three years. Lack of clarity regarding the division of responsibilities between the central government and the municipalities in implementing external evaluation, as stipulated in the Compulsory School Act in 2008, was solved by establishing a joint project between those two levels of government. In

early 2013, an external evaluation was initiated, co-financed by the government and the Municipal Equalisation Fund. The evaluation was based on quality criteria regarding school operation that covered numerous aspects of the activities of compulsory schools. At the same time, overseeing external evaluation was transferred from the Ministry to the Educational Testing Institute, which later became the Directorate of Education. The number of schools participating in the evaluation was 10 per year. The external evaluation was implemented as a two-year pilot project and then extended by another two years until the end of 2016. The pilot project was in its final year when Paper I was written, and its further development remained uncertain at the time.

7.2 Paper II

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*(3), 209–224.
https://doi.org/10.1080/20020317.2022.2076376

For Paper II, a survey was conducted among principals and teachers in 22 schools that were externally evaluated in 2013, 2014, and 2015 (i.e. Phase II). The first year, 2013, marked the beginning of the external school evaluation covered by the research, meaning that those 22 schools were the first to be evaluated. The research questions guiding Paper II were "How can the scales constructed to delineate compulsory school teachers' and principals' reception of and attitudes towards the findings of evaluation (i.e. Accepting Feedback and Teacher Participation) be characterised, compared, and related?"; "With reference to Hofer et al. (2020) and Matthews and Sammons (2004), are principals' attitudes and reactions more positive than teachers'?"; and "With reference to Ehren et al. (2013), are the mechanisms of accepting feedback, setting expectations, involving stakeholders, and building the capacity to implement improvements positively correlated with changes in internal evaluation, teaching, and leadership practices?" The purpose of the study was to identify and analyse school principals' and teachers' attitudes towards the external school evaluation, with particular attention paid to the ways in which they consider that the evaluation had affected their schools' internal evaluation and driven changes in their own practices. As the third research question indicates, the study was inspired by the theoretical framework of Ehren et al. (2013), which is based on the programme theories of six European countries' school inspection systems.

In May 2016, surveys were sent to teachers and principals in the 22 schools. The response rate was 100% among principals (*N* = 22) and 56% among teachers (*n* = 258). Based on the surveys, a descriptive, exploratory, multilevel correlational analysis was conducted. Factor analysis resulted in the following scales (i.e. for latent variables): Accepting Feedback, Setting Expectations, Stakeholder Involvement (i.e. only in the principals' data), Teachers' Participation (i.e. only in the teachers' data), Attitudes

towards Internal Evaluation, Experienced Hindrances, and Changes in Teaching Practices (i.e. only in the teachers' data). The variables Change in Internal Evaluation and Change in Leadership Practices consisted of one item each. The responses of teachers and principals to four variables were compared, and correlation coefficients were calculated to describe the relationship between the variables for teachers, on the one hand, and principals, on the other. Last, a multivariable regression analysis was performed on the teachers' data only because a lack of a significant relationship between variables in the principals' data did not justify proceeding further with that analysis. Nearly all results were obtained via multilevel analysis with standardised variables. Analyses were carried out in R and SPSS version 27.0.

The findings in the paper regarding the first two research questions were that, among both principals and teachers, there was a positive attitude towards the external school evaluation, although principals were generally more positive towards it than teachers were. Moreover, teachers who participated in applying the evaluation feedback were more accepting and motivated to use the feedback than teachers who took little or no part in the decision-making. The principals were again significantly more positive than the teachers in their attitudes towards internal evaluation, they experienced fewer hindrances in implementing the changes, and they reported more knowledge about the evaluation criteria and indicators. However, there was no difference in the position of those two groups regarding whether the external evaluation had led to changes in the schools' internal evaluation.

Second, results regarding whether the mechanisms of accepting feedback, setting expectations, involving stakeholders, and building the capacity to implement improvements were positively correlated with changes in internal evaluation, teaching, and leadership practices partly differed for teachers and principals. Concerning perceived changes in internal evaluation in the teachers' data, the scales Accepting Feedback, Setting Expectations, and Teachers' Participation were found to be significant predictors. The teachers' data also showed a correlation between the scales Attitude towards Internal Evaluation and, on the one hand, perceived Changes in Internal Evaluation and, on the other, fewer Experienced Hindrances to changes. By contrast, in the principals' data, the only scale that was significantly correlated with perceived Changes in Internal Evaluation was the scale Setting Expectations. No relationships were found for the scales Accepting Feedback, Involving stakeholders, Attitudes towards Internal Evaluation or Experienced Hindrances. With respect to influence on perceived changes in leadership practice, only the scale Accepting Feedback had a significant correlation, whereas all other relationships were small and not significant. Last, for the teachers, Accepting Feedback primarily explained perceived Changes in Teaching Practices, whereas the scales Setting Expectations and Teachers' Participation had slightly to moderately significant correlations with Changes in Teaching Practices. Even so, those latter two relationships did not exert a significant impact when accounting for Accepting Feedback in a linear mixed-model analysis.

7.3 Paper III

Ólafsdóttir, B., Jónasson, J. T., & Sigurðardóttir, A. K. (2022). Use and impact of external evaluation feedback in schools. *Studies in Educational Evaluation*, *74*, Article 101181. https://doi.org/10.1016/j.stueduc.2022.101181

Paper III reports a study on the use and impact of feedback from external school evaluation in six compulsory schools as perceived by principals and teachers in those schools. The research questions guiding Paper III were "How and to what extent do schools use the feedback presented in external evaluation reports?" and "To what extent do schools sustain the changes made after using the feedback from external evaluations instrumentally?" As the research questions imply, the purpose of the study was twofold: first, to contribute to current knowledge on the use and impact of external evaluation with a focus on the Icelandic educational context, and second, to shed light on how well the improvements and changes made, based on the feedback, have been sustained over time. The framework used for analysing the use of evaluation feedback, based on Rossi et al.'s (2004) and Aderet-German and Ben-Peretz's (2020) definitions of use and distinguishes between instrumental, conceptual, persuasive, and reinforcement-oriented use.

Of the 22 schools that participated in Phase II of the research (i.e. Paper II), six were selected to participate in Phase III. The selection of the schools was based on size, location, and evaluation judgements, while the data consisted of official documents— evaluation reports, improvement plans, and progress reports—and semi-structured interviews with principals and teachers in the schools. Six principals and eight teachers were interviewed. Data were uploaded into NVivo R1 for analysis, and a thematic approach (Braun & Clarke, 2006) was followed. Bearing in mind the findings in Paper II, I examined different uses of evaluation feedback and how changes were made in the four to six years after the evaluation, as perceived by principals and teachers in the schools.

Based on those data, the major findings were that feedback from external school evaluation can be useful to schools in various ways, for the data revealed clear examples of instrumental, conceptual, persuasive, and reinforcement-oriented use. Instrumental use was observed in relation to (1) leadership and management, primarily respecting professional collaboration among staff members and the instructional leadership of school leaders; (2) learning and teaching, primarily regarding differentiated strategies for instruction, students' democratic participation, and the use of assessments to improve students' learning; and (3) internal evaluation, mostly concerning evaluation plans and methods, stakeholders' participation, and improvement plans. Instrumental use varied between the schools, only some of which made major changes in all three areas. Especially in relation to internal evaluation, progress had occurred in some of the schools despite stagnation and even regression in others. Conceptual use was also evident in the schools, and in that context, the

usefulness of obtaining an external view on the school's function and acquiring help with identifying areas requiring improvement was highlighted. In some cases, evaluation feedback led to productive discussions and reflections among the professionals, and for three newly appointed principals, it afforded useful instructions. By comparison, the persuasive use of evaluation feedback was identified in three interviews, namely in the context of supporting changes that interviewees wanted to bring about. Likewise, reinforcement-oriented use was analysed in three interviews in schools that had received positive evaluation feedback, which they had experienced as empowering.

The findings also showed that both teachers and principals had positive attitudes towards the external evaluation and had generally experienced the evaluation feedback as being useful and contributing to changes in practices in the schools. The schools seemed to have sustained many of their implemented actions owing to the external evaluations or to have continued to develop them in some way, at least in the first few years following evaluation. That outcome seems to have been the case in five of the six schools. Furthermore, the participants indicated that the Ministry of Education and Children's follow-up with the improvement plan gave them restraint in working on and maintaining the changes.

8 Discussion

In this chapter, the findings of the research conducted for this thesis are discussed in relation to the aim of the research, its theoretical background, and circumstances in Iceland. First, the aim and purpose of the research are reviewed and the results discussed in light of them. Next, the potential contributions of the results and recommendations for education authorities and schools are articulated. The chapter concludes with concerns about and responses to threats to validity and suggestions for further research. The conclusions drawn from the research conducted for the thesis derive from the interpretation of the findings from the questionnaires, the interviews, related documents, and the literature, all of which forms the basis for the following discussion.

8.1 Aim, purpose, and method

Despite increased research on various aspects of external school evaluation in countries worldwide, external school evaluation in Iceland has hardly been investigated. Even so, the external evaluation of compulsory schools has been performed for about a decade in Iceland, or even longer if counting the comprehensive evaluation of compulsory schools in Reykjavík. The research conducted for this thesis was designed to respond to that gap in current understandings of external school evaluation in Iceland.

As introduced in Chapter 5, the overarching aim of the research was twofold:

(1) To shed light on how external school evaluation in Iceland came about, how it developed after 1991, and what its expected benefits are; and

(2) To identify and analyse the extent to which the intended benefits, regarding the use of evaluation feedback and its impact on professionals' practices (i.e. teaching and leadership) and internal evaluation, have been realised according to principals and teachers.

The results have been reported in three published journal articles, all of which are presented in Chapter 7 of this thesis and are attached as appendices to the thesis. In Paper I, I examine the origin and development of the external evaluation of compulsory schools in Iceland over a quarter of a century. In Paper II, I investigate school principals' and teachers' attitudes towards a recent external evaluation, particularly how they consider that the evaluation has affected their schools' internal evaluation and driven changes in their own practices (i.e. teaching and leadership). Last, in Paper III, I analyse the principals' and teachers' perceptions of the use of evaluation feedback and how well the improvements and changes made, based on the feedback, have been sustained over time. In the following two sections, I discuss how findings in all three

papers contribute to fulfilling the aim of the research in light of theories and other research discussed in Chapter 3. Next, in Section 8.2, I focus on the first part of that aim, namely how the policy and execution of external school evaluation in Iceland came about, how it has developed, and what its expected benefits are. After that, in Section 8.3, I discuss the results regarding the second part of the aim concerning the extent to which the expected benefits of such evaluation have been realised, especially regarding the use and impact of feedback from external evaluation.

8.2 Policy and execution of external evaluation

In the research conducted for this thesis, I first attempted to inquire into how external evaluation in compulsory schools in Iceland came about and how it has developed since becoming part of official education policy in 1991. That topic was the subject of the study conducted for Paper I, which provided an important foundation for Papers II and III, because it expresses, among other things, the expectations and attitudes that policymakers have had towards external evaluation over time. The following discussion focuses on how the policy-related documents and participants in the policymaking process portray that development and its purposes.

The findings of Paper I clearly show that development and trends in many parts of the world related to the ideas of NPM influenced the attitude of policymakers in Iceland regarding the quality and evaluation of schools in the early 1990s. After all, NPM was the declared policy of the Icelandic government at the time (Ministry of Finance, 1993). In that way, globalisation and policy-borrowing (Schick, 2002), which reached Iceland partly due to citizens returning from study abroad and partly due to Iceland's membership in the OECD, contributed to the beginning of what might be called the "third wave" of external evaluation in Iceland (Jónasson, Bjarnadóttir, et al., 2021). However, for various reasons, Iceland's education system was not open to external evaluation at the time. Obstacles included a lack of resources and capacity, resistance from the teachers' union, the lack of clear need, and the fact that the operation of compulsory schools was transferred to the municipal level, which led to an unclear division of roles and responsibilities between state and municipalities when it came to quality assurance. Such resistance to NPM's notion of quality assurance, which was more summative than formative in nature, can also be explained by the fact that it did not fit well with the Nordic countries' historical legacy and welfare tradition, including in Iceland (Gunter et al., 2016a). Because the external evaluation was controversial, responsibility for the evaluation of schools ultimately fell to the schools themselves. Schools were supposed to ensure their own quality through internal evaluation, and twice the central government assessed whether schools were fulfilling their duty in that context. In both cases, the expected duties were found to have been unfulfilled, largely because the schools were not prepared to undertake such evaluation. That relatively underdeveloped responsibility for the quality of schools existed for more than a decade and, according to Schick (2002), was commonly the case in many other countries as well.

In 2008, Iceland's current laws for compulsory schools were enacted, which stipulated provisions for external evaluation primarily by local governments but also by the central government, alongside ongoing internal evaluation in schools. It was nevertheless recognised that ensuring sufficient capacity at the local level to carry out and follow up on external evaluation would prove challenging. That obstacle, together with the fact that the role of each administrative level regarding the evaluation was considered to be unclear, prompted a joint project on external evaluation between the two parties. In the project, all municipalities in Iceland participated, except for Reykjavík, which had already begun an evaluation of its schools and wanted to continue without the involvement of the state. The external evaluation ultimately applied was formulated in cooperation between the state and local authorities and with the involvement of school affairs experts, who had experience with the comprehensive evaluation of compulsory schools in Reykjavík. That new quality assurance system was introduced as a formative external evaluation with an intended focus on developing and improving schools instead of on accountability. Its introduction met no resistance from the teachers' union or from other stakeholders, most likely because such a formative approach fit better with the Icelandic school tradition, which Jónasson, Ragnarsdóttir, et al. (2021) have argued is characterised by strong traditions and the strength and autonomy of the teaching professionals.

As clarified in all three papers, the Educational Testing Institute, beginning in 2013, and the Directorate of Education, as its successor since 2015, were tasked with conducting external evaluations on behalf of the state and municipalities and making recommendations regarding opportunities for improvement. The Ministry of Education and Children followed up on the external evaluation by calling for improvement plans and progress reports. Municipalities were also responsible for following up on the evaluation and ensuring that they were effecting improvements in schools (Compulsory School Act, 2008). Paper III clarifies that the principals felt that the Ministry's follow-up kept them focused on the improvement actions. However, the results presented in both Papers II and III indicate that the local capacity to support schools and work with them for improvement and change has varied and can pose significant challenges to smaller local authorities. After all, the smallest municipalities often do not have a school office or superintendent who can integrate and coordinate the tasks and responsibilities required by law. A recent doctoral study concluded that the state does not sufficiently acknowledge the differences between municipalities in Iceland, especially regarding population, location and size, and local authorities are left on their own to fulfil legal obligations regarding the operation and support of schools (Sigurðardóttir, 2023). As discussed in Paper II, principals generally experience neither much pressure nor support from education authorities in the improvement process, but the variation in their answers indicates a difference between municipalities. Those results support the conclusion of a recent study in Iceland, which revealed a lack of leadership on behalf of the municipal authorities regarding school services that aim at strengthening schools

as professional institutions (Sigurðardóttir et al., 2022; Sigþórsson et al., 2022; Svanbjörnsdóttir et al., 2021). According to Sigurðardóttir (2023), some local authorities only partially comply with legislations in terms of school support services. Thus, as also pointed out by Jónasson, Ragnarsdóttir, et al. (2021), doubts arose about equality and social justice for students, especially in some of Iceland's smaller municipalities.

The analysis of policy-related documents presented in this thesis in general, and especially in Paper I, reveals that the benefit that national authorities anticipated from external evaluation was chiefly that the evaluation would contribute to improvement and development in schools regarding learning and teaching, leadership and management, and internal evaluation. The element of accountability in external evaluation is present as well but not the focal point. As discussed in Papers II and III, external evaluation in Iceland can be characterised as taking a soft approach to steering and quality control. It is process-oriented, based on cyclical evaluation, and carries no specified consequences for non-compliant or underperforming schools, even despite pressure from the Ministry for schools to work on improvement based on the evaluation feedback. Furthermore, although the results are publicly available, there is no ranking of schools based on them (i.e. Papers II and III). Those characteristics are all typical of a low-stakes evaluation approach as defined by Altrichter and Kemethofer (2015) and Ehren et al. (2013).

The external school evaluation examined in this thesis began in 2013, and as stated in Paper II, by late 2021, all compulsory schools in Iceland had been evaluated once. All compulsory schools in Reykjavík had already been evaluated when the project began. At the time of writing, in early 2023, no external evaluations of compulsory schools have been conducted since 2021, and no decision on the continuation of such evaluation has been made. It has been announced that the Directorate of Education will be dismantled and that new central service and knowledge centre in the field of education will be established. Along with that change, it has been suggested that external school evaluation will be transferred to the Ministry of Education and Children. The draft bill for the new institution discusses the importance of the Ministry's supervision, but the future of external evaluation remains entirely unclear. Considering key factors in Kotter's (2012) model for achieving change (i.e. a clear vision and strategy and top-level management's support and commitment), it can be seriously doubted that those factors are present in the further government's plans for the future of external evaluation. Sigurðardóttir (2023) has pointed out that actions at the national level in Iceland have tended to be determined by the political vision of the national educational authorities who are in charges at any given time than by a strategic, long-term vision agreed upon across political parties. The research conducted for this thesis (Paper I) supports that conclusion, at least as far as the external evaluation is concerned. Because of the political culture in Iceland and the accompanying potential

instability, the uncertainty about the continuation of external school evaluation is the same when this thesis was written in 2023 as it was when Paper I was written in 2016.

In this section, I have summarised the results of the three papers regarding the first part of the research aim, concerning how external school evaluation in Iceland came about, how it developed, and what its expected benefits are. In the following section, I discuss, based primarily on Papers II and III, the extent to which the expected benefits of external evaluation regarding the use of its feedback and impact have been realised.

8.3 Use and impact of external school evaluation

According to Ehren et al.'s (2013) theoretical framework, which served as the basis of the analysis in Paper II, accepting evaluation feedback is an important mechanism for driving changes at schools following external evaluation. On that count, findings in Papers II and III show that both principals and teachers expressed positive attitudes towards the evaluation feedback, which they generally characterised as being helpful and supportive of improvements in their schools. Those findings imply that the authorities achieved their intentions of having external evaluation and evaluation feedback support the development in Iceland's schools. Even so, survey results in Paper II indicate that principals had significantly more positive attitudes towards the evaluation feedback than teachers. Matthews and Sammons (2004), have observed that same difference in the acceptance of feedback among teachers versus principals, whereas a study by Ehren, Perryman, et al. (2015) showed the opposite. The interview data in Paper III suggest that the difference can be explained by the principals' perceptions that external evaluation supported them in their role as educational leaders in their schools, equipped them with tools to bring about changes, and thus somewhat served their interests. The data can also be interpreted in light of Sigurðardóttir's (2023) conclusion that school principals in municipalities that have neither a superintendent nor a school office tend to expand their responsibilities and take over the roles that superintendents have in other municipalities. Furthermore, due to the general lack of the municipal authority's leadership in providing support for teachers in developing their teaching practices (Sigurðardóttir, 2023; Sigþórsson et al., 2022), that leadership role falls to the principal, who in some cases has the means to deal with it but in other cases not. In addition to being a link between school authorities and teachers, principals have more opportunities to express their views during the evaluation visit than teachers who are only represented in a focus group. Thus, the principal's point of view is given more weight, which may partially explain their more positive attitude.

From another angle, the analysis in Paper II generally indicates a strong relationship between the acceptance of evaluation feedback and changes. In fact, it was the factor that ultimately had the greatest impact. Those results indicate that in Iceland's education system, school staff's acceptance of evaluation feedback is pivotal. That conclusion

aligns with previous findings showing how heavily improvement depends on people's acceptance of feedback, especially from evaluation following a low-stakes approach, as in Iceland (Altrichter & Kemethofer, 2015; Dedering & Müller, 2011; Her Majesty's Inspectorate of Education, 2010; Kemethofer et al., 2017; Penninckx, 2017). Nevertheless, as numerous researchers have concluded (e.g. Ehren et al., 2013; Ehren, Perryman, et al., 2015; Ehren & Visscher, 2008; Gustafsson et al., 2015), it is important to note that receiving and accepting feedback will not automatically produce improvement or change following external evaluation, for more influencing organisational and political factors have to be present (Hofer et al., 2020).

As mentioned, the purpose of the external evaluation examined in this thesis was not only to enhance and support schools' internal evaluation but also to support schools, principals, and teachers in improving their practices—that is, to stimulate the instrumental use of evaluation feedback. In terms of the different notions of evaluation use derived from Rossi et al. (2004) and discussed in Paper III, nearly all of the principals and teachers interviewed for the study in Paper III expressed that the external evaluation had been used to improve the quality of learning and teaching and/or leadership and management in their schools. The analysis of the questionnaires in Paper II revealed that many but not all teachers and principals had made changes to their practices. Meanwhile, opinions differed from school to school as to whether the teachers and principals interviewed believed that the external evaluation had led to improved internal evaluation. On that topic, the survey also showed large variation in principals' versus teachers' views on whether changes had been made in internal evaluation, which supports the conclusion of the interview study (i.e. Paper III) that real differences emerged between schools in that respect. At the same time, differences of opinion also surfaced within each school about whether the external evaluation had led to changes in internal evaluation. Although the changes made seemed to vary from individual to individual and from school to school, the research conducted for this thesis nonetheless supports the results of past studies (e.g. AlKutich & Abukari, 2018; Dedering & Müller, 2011; Ehren & Visscher, 2008; McCrone et al., 2007) that external evaluation feedback is used for improvement in most schools. However, it should be noted that the conceptual framework of Ehren et al. (2013), on which the analysis in Paper II was based, expects that external evaluation does not lead directly to improved teaching and leadership practices, except for a short time. The framework assumes a sequential effect, meaning that it is primarily promoting and improving schools' internal evaluation and thereby their capacity for improvement that can in the long-term prompt better leadership and teaching practices that ultimately lead to better education and student achievement. Thus, high-quality internal evaluation is considered to be a critical element of the improvement in schools (Gustafsson et al., 2015), which is precisely the aspect that some of the schools in the research conducted for this thesis had the most difficulty with. There is therefore reason to pay special attention to internal evaluation in the future development of external school evaluation and support for schools so that school evaluation leads to greater long-term improvements.

The analysis of the interview data, presented in Paper III, revealed a wider use of the evaluation feedback than the concrete and observable (i.e. instrumental) use addressed in the questionnaire and interviews. That is, conceptual, persuasive, and reinforcement-oriented use could be identified as well, albeit to varying degrees. Such use is no less important; for example, conceptual use can impact individuals' actions in the long term (Nunneley et al., 2015; Rossi et al., 2004; Weiss, 1998); persuasive use can contribute to increased instrumental use, as discussed in Paper III; and reinforcement-oriented use can give individuals and schools a sense of pride and confidence (Aderet-German & Ben-Peretz, 2020).

Numerous scholars (e.g. Ehren & Baxter, 2021; Ehren & Visscher, 2006; Landwehr, 2011) have emphasised that school characteristics and capacity (i.e. skill, competencies, knowledge, and decision-making power) impact how schools respond to evaluation feedback. Both Papers II and III showcase that the capacity of a school affected how systematically it applied evaluation feedback. In particular, Paper III shows that staff at most schools that had received good evaluation judgements reported working systematically to meet all of the recommendations, while staff at schools with the greatest need for improvement had more difficulty in addressing all aspects of the recommendations. That outcome could be due to both the characteristics of the schools and the different scope of improvements that the schools had to implement. However, the survey data used in Paper II showed that factors related to school capacity, including teachers' participation in decision-making about improvements and their attitudes towards internal evaluation, had a significant relationship with changes in the wake of the external evaluation. Teachers who were involved in working with the evaluation feedback were also more likely to accept the feedback. That finding is important because dialogue and reflection are the first steps in developing actions to improve school processes and performance, as various scholars have mentioned (Ehren, 2016b; Landwehr, 2011; MacBeath et al., 2009, January 4–7; Vanhoof et al., 2013).

According to Ehren et al.'s (2013) framework, mechanisms other than accepting feedback and capacity of the schools are intended to influence the use of evaluation feedback and lead to improvements and changes in schools. Quality criteria are set out that define expectations for schools regarding good education and practices and that serve as the basis for external evaluation and improvement measures. As discussed in Paper II, there appeared to be relatively little knowledge and use of Icelandic quality standards in the schools that participated in the research, especially among teachers. Principals and teachers seemed to make little use of the criteria to reflect on and improve their own practices; however, the results also indicate that the criteria have somewhat formed the basis for changes in the schools' internal evaluation. Because Altrichter and Kemethofer (2015) and Ehren, Perryman, et al. (2015), have revealed that setting clear expectations is a strong determinant of improvement actions, there is a reason to doubt whether sufficient emphasis has been placed on schools to use the quality criteria to evaluate and improve their own practices. In that context, the

discussion in Papers II and III about the low-stakes evaluation approach in Iceland should be considered, especially following Altrichter and Kemethofer's (2015) finding that principals who felt more pressure for accountability were more alert to the expectations set out in the quality criteria. At the same time, their study also showed that unintended consequences increased with greater pressure. Therefore, pursuing other ways to encourage Icelandic schools to make greater use of the criteria are needed—for example, by better promoting them and emphasising their use—but without the pressure for accountability being so great that it leads to side effects.

The third mechanism assumed in Ehren et al.'s (2013) framework that heightened the impact of the external evaluation is involving stakeholders, who are generally expected to facilitate school improvement. Such entities include parents, local policymakers, and education authorities (Ehren, Perryman, et al., 2015). In the research for this thesis, principals were asked about the involvement of local authorities and the school council and about the Ministry of Education and Children's follow-up. However, because their experiences with parental involvement were not addressed, the findings are not fully comparable with the results of other studies based on the framework. As discussed in Paper II, stakeholders' involvement was not found to be strongly correlated with changes following the external evaluation. Furthermore, as discussed in Section 8.2, the local authorities generally did not seem to fulfil their role in supporting the schools or pressuring them to implement changes after the evaluation. That tendency, as explained by Sigurðardóttir's (2023) results, suggests that leadership in education at the local level is generally underdeveloped and some municipalities lack the resources and professional knowledge to carry out their duties. This lack of support has likely diminished the impact of the low-stakes approach to external evaluation. Even so, principals and teachers generally did not seem to perceive hindrances—for instance, a lack of resources—that would prevent improvement, as concluded in Paper II.

In the foregoing paragraphs, I have summarised the results of the research conducted for this thesis related to the second part of the research aim—that is, concerning the extent to which the expected benefits of the evaluation regarding the use of its feedback and impact have been realised—based primarily on the findings in Papers II and III. In the following sections, I highlight some important contributions of the research.

8.4 Contributions of the research

In this section, drawing on the discussion above and based on the aim that guided the research conducted for this thesis, I outline areas in which the research has contributed both to studies on external school evaluation and to school practice. I do so by looking across the findings in the three papers and drawing upon theoretical perspectives on external school evaluation. From that standpoint, I argue that the research indeed contributes to theory and has implications for education authorities and schools. The research's theoretical and practical contributions are discussed in the next two sections.

8.4.1 Theoretical and conceptual reflections

Adopting the theoretical perspectives outlined in Chapter 3, I argue that this thesis contributes to knowledge and literature on external school evaluation.

First, I employed theories of agenda-setting, research on the progress of control in public administration, and theories on implementing and managing change in order to analyse how external school evaluation in Iceland came about and what explains its development. As a result, in this thesis I have provided insights into how plans for external evaluation were introduced into government policy without being implemented until many years later and thus shown that adapting to new expectations and management roles has proven difficult for a country with strong traditions and limited resources, especially due to the resistance of the teachers' union (i.e. Paper I).

Second, the research provided an additional lens for understanding how external evaluation feedback following a low-stakes evaluation model is perceived by principals and teachers and how they use it to reflect on and change their practices (i.e. Papers II and III). As the analyses in Papers II and III show, teachers and principals were generally positive about the external evaluation and have used its feedback in various ways. It can be inferred that such a positive attitude reflects the presence of trust between schools and government, which Ehren and Baxter (2021) have identified as one of three pillars of any education system. The other two are school capacity and accountability, as addressed later in this section.

Third, the research for this thesis demonstrated the extent to which the mechanisms of Ehren et al.'s (2013) theoretical framework (see Figure 1) are applicable to Icelandic circumstances (i.e. Paper II). Although a simpler, reduced part of the framework was used in the research, some theoretical implications may nevertheless be derived from the results. The framework proved to be very useful for studying the influence of several mechanisms on changes in school practices as a result of external evaluation. However, not all of the framework's assumptions seem to apply to conditions in Iceland, especially regarding the mechanism of involving stakeholders, which was not found to be correlated with changes following external evaluation. That outcome can be explained by the composition of the scale used to measure stakeholders' involvement, which covered only pressure and support from central and local authorities but not from other stakeholders such as parents. Nevertheless, Icelandic society does not observe a tradition in which parents hold schools accountable (Jónasson, Ragnarsdóttir, et al., 2021) by exercising choice, using their voices, or simply withdrawing their children from particular schools. With the proviso that the scale Stakeholder Involvement in the research for the thesis is limited the results support Kemethofer et al.'s (2017) and Gustafsson et al.'s (2015) findings showing the limited or indirect effects of stakeholders' involvement. Beyond that, the research supports the framework's assumptions that the two other mediating mechanisms (i.e. setting expectations and accepting feedback) may be important influencing factors for change following external evaluation as perceived by principals and teachers.

Furthermore, following Ehren (2016b), in the research conducted for this thesis (i.e. Paper II) a fourth mediating mechanism of change—schools' capacity to implement improvements—was added along with the other three in Ehren et al.'s (2013) theoretical framework when analysing the results. The scales that were constructed to measure schools' capacity were found to have small to large associations with changes in internal evaluation and teachers' practices as perceived by teachers. That result may indicate that the framework does not cover all factors at play in the impact of external evaluation feedback. Thus, the research conducted for this thesis supports Ehren's (2016b) conclusion that the capacity of schools is a fourth mechanism of change that explains how external evaluation can influence improvements in the quality of schools, meaning that factors related to capacity to change might be elaborated in the framework. As mentioned, Ehren and Baxter (2021) have identified school capacity as one of the three building blocks of the education system, which further underscores its importance.

Last, the research's longitudinal approach allowed contributing to knowledge about the sustainability of changes made as a result of external school evaluation. Following that research design, it was also possible to gain insight into the challenges that schools face in implementing, maintaining, or continuing to develop changes following evaluation. Few studies have documented changes and their sustainability in the long term. The findings of the research for this thesis stress the importance of long-term follow-up on improvement plans in a context in which external evaluation is a low-stakes affair, as argued in Paper III. In the evaluation approach used in Iceland, the follow-up process can be viewed as the primary way to ensure accountability, which is, as mentioned, the third pillar of the education system according to Ehren and Baxter (2021). Those authors claim that the interaction between three building blocks of the education system—trust, capacity, and accountability—affects the success of educational reforms. Those three pillars are therefore also integrated into the recommendations presented below.

8.4.2 Practical recommendations

The results of the research conducted for the thesis, the theoretical framework used, and the literature on which the research was based form the foundation for recommendations for governments, municipalities, and schools discussed in this section.

Part of my motivation for conducting the research for this thesis was to contribute practical knowledge (Dewey, 1938) that would be useful in the ongoing development of external school evaluation, especially in Iceland. I wanted to provide empirical insights to add to the knowledge of policy- and decision-makers, the people involved in supporting schools, and school staff. I argue that the research adds to that field of practice by providing initial empirical insights into the use and impact of external evaluation feedback in Iceland and what mechanisms influence whether changes are made in schools as a result of the evaluation, as reported in Papers II and III. Although

many aspects remain unexplored, as discussed in Section 8.6, the insights provided by the research can assist governments, policymakers, and other interested parties in further developing and implementing external evaluation mechanisms and allocating resources needed to conduct successful evaluations. Until now, everything related to external school evaluation in Iceland has been woefully underexamined.

Drawing on the evidence presented in the research conducted for this thesis and the relevant literature, I propose actions for consideration by the central and local governments and the schools themselves. I start with recommendations for the central government before turning to local authorities and end by making recommendations for the schools themselves.

Recommendations for the central government

To respond to all of the suggestions made in the external evaluation report proved to be difficult for some of the schools in the research. As discussed in Paper III, the difficulty appeared to be primarily associated with strengthening internal evaluation, the instructional leadership of school leaders, and the systematic use of assessments to improve students' learning. As pointed out by Verhaeghe et al. (2010), implementing actions based on evaluation feedback is not a clear-cut process and requires policy-oriented skills. The evaluation reports do not provide schools with guidelines for potential improvement actions but only a recommendation about what needs to be improved. A way to support the schools in that respect is to create guidelines with examples of good practices and actions that any school can use as a basis for discussions among professionals about possible improvements.

Also in that context, Vanhoof et al. (2013) have concluded that having access to resources and targeted support regarding planning and evidence-based practice can contribute to improving the competencies and skills that schools need to use evaluation feedback effectively. In Iceland, a new central service and knowledge centre in education and training is in the pipeline; such an organisation can play an important role in supporting schools after external evaluation, especially by helping them to design their development processes, promote internal evaluation, and build their capacity for improvement. Supporting school staff in improving their practices can no doubt be more beneficial than merely identifying what is being done well and what can be improved. Such support is particularly important given Sigurðardóttir's (2023) findings that some schools and municipalities lack the professional capacity and support to implement national policy.

As discussed in this thesis, the quality criteria that define expectations for schools guide external evaluation in Iceland. Those criteria are intended to influence schools (Ehren et al., 2013), and even though schools in Iceland are not obliged to use the criteria in their internal evaluation, they may adapt to those criteria, such as by using them in goal-setting and/or internal evaluation. However, as discussed in Paper II, it seems that knowledge of the criteria in the schools studied was generally low, even though they had already received external evaluation. For schools to become more aware of the

expectations put upon them, there should be more ways of communicating the quality criteria than only via external evaluation. Schools could, for example, be better informed about the essence of the quality criteria and how they can use them in their internal evaluation, which would presumably increase the alignment between internal and external evaluation. In that context, access to an online system in which schools can evaluate their own performance based on the quality criteria could be a major lever.

The results in Paper III emphasise that the Ministry of Education and Children's follow-up with the schools' improvement plans seemed to make demands of schools that the schools appreciated. Principals in schools rated as being in high need of improvement indicated that the follow-up had ended before they had completed the improvement plan, which at times brought the improvement process to a halt because the call for information ceased. That dynamic was less common in schools that received more favourable evaluation judgements. As also pointed out in Paper III, those results give reason to structure the follow-up process according to the school's status, such that schools that need to work on more extensive improvements are followed up over a longer period. Moreover, there would be reason to visit those schools again after a certain time in order to better follow up and support the improvement process.

Furthermore, Papers II and III both show that not all municipalities fulfilled their duty to support their schools in the improvement process. The correspondence between the Ministry and municipalities does not seem to be a sufficient measure in the affairs of all municipalities for fulfilling their statutory obligations to monitor whether external evaluation leads to improvements in the operations of schools. Along those lines, the central government needs to find a way to better drive (hopefully few) municipalities to fulfil their obligations by providing them with more resources, including professional support, because some local authorities may not be fully aware of their role in the improvement process. As mentioned in Chapter 2, the school committee, which manages the school's affairs on behalf of the municipality, undergoes changes every four years, and only some municipalities maintain a school office or administrative leader who can complete the municipality's tasks and responsibilities in school-related matters. The role and contribution of local authorities thus needs to be clarified, and both schools and local authorities would also benefit from clear criteria specifying when external support would be necessary.

Last, as mentioned, the external evaluation of compulsory schools in Iceland is at a turning point, and decisions have to be made about both its place in the system and its continuation. Fullan and Quinn (2016) and Six (2021) have emphasised the importance of approaching accountability as a strengthening and supporting process. The results of the research conducted for this thesis support the formative and improvement-oriented approach to external evaluation that characterised the first round of external evaluation examined in the thesis. With that approach, the authorities seem to have earned the trust of the schools regarding external evaluation, which according to Ehren and Baxter (2021) is beneficial for school systems. Even so, the accountability of external

evaluation needs to be further developed such that it maintains that trust, because, as underscored by Ehren and Baxter (2021) and Six (2021), when trust and accountability reinforce each other, the capacity of schools and professionals can be strengthened. With that in mind, the question should be raised about whether external evaluation should be the responsibility of a new service and knowledge centre or whether a special evaluation institution should be established instead.

Recommendations for local authorities

As stated above and in Paper III, given the formal responsibility of local authorities to ensure that external evaluation leads to improvement, they need to take on additional responsibility. They should use each school's evaluation report as a basis for engaging in meaningful discussions with the school in view of the report. They also should work with schools on their improvement plans, provide them with enough support to implement improvements and changes, and monitor their progress. To be able to work with schools in that way, municipalities need to accumulate knowledge about evaluation and improvement processes or else seek outside aid to do so.

For many schools, internal evaluation and related central tasks have proven to be challenging. External support is needed by schools with, for example, implementing systematic methods for classroom observation, issuing development-oriented feedback to teaching staff, and using screening and assessment data to improve teaching and learning. In that regard, the local authorities should provide teachers and principals with the necessary training and coaching or else ensure that they receive the necessary support. The new central service centre could be of support to local authorities by providing them with professionals who can grant such education and training.

Recommendations for schools

The results of the research conducted for this thesis demonstrate the importance of involving teachers in the process of discussing and making decisions about changes based on evaluation feedback. Discussion and reflection on practices within the school ought to take place, and teachers should feel that they are able to influence improvement actions being pursued based on the recommendations of evaluation. That dynamic can encourage ownership of the decisions and changes made, which has proven to be important in the use of evaluation feedback (Brown & Zhang, 2017; Ehren, 2016b; Schildkamp & Lai, 2013).

In the research conducted for this thesis, the schools with the best evaluation judgements had worked systematically to meet all of the recommendations from their evaluation reports. Those schools had principals who efficiently monitored the implementation of the measures in the improvement plans and regularly discussed such implementation in meetings. Keeping track of and having the stamina to follow through on changes in the long term is important, as various scholars have pointed out (e.g. Ehren, 2016b; Schildkamp & Ehren, 2013; Thillmann et al., 2013). In that light, the principal plays a key role.

Teachers and principals face considerable challenges with deciding what improvement actions to pursue and with developing the skills and competencies required to implement them. As mentioned, the findings of the research for the thesis revealed that such is especially the case for internal evaluation and aspects related to it. Despite the stated importance of the central and local governments in the context of training and educating teachers and principals, because each school itself is responsible for its quality, even with the involvement of the local authority, the school has to acquire the required resources and training in cooperation with local authorities.

As discussed, the results of the research conducted for this thesis contribute to studies on external school evaluation and to practice. Based on the findings, recommendations have been made above that are directed to the central and local authorities and to the schools. However, there are also potential threats to the validity of the research's findings, as considered in the next section.

8.5 Concerns about threats to validity

To add to the discussion of the methodological advantages and limitations of the research conducted for this thesis discussed in Section 6.6, in this section I focus on threats to the validity of the research, particularly ones concerning my position as the researcher and what efforts I made to minimise my bias in that role. I also consider the threats to validity arising from the fact that individuals who answered the questionnaires, on which the results of this thesis are based, were answering surveys sponsored by education authorities. Furthermore, I discuss the methods (e.g. triangulation) that I used to address those threats and increase the strength of the research.

My position as a researcher was both a strength and a threat to the validity of the research, with knowledge of the field being a strength and the risk of bias and preconceived ideas being a threat. Researching within one's field of practice can add depth to the interpretation of data, provided that the researcher remains mindful of any bias. As discussed in Section 1.2, I work as an evaluation specialist at the Directorate of Education, which conducts Iceland's external school evaluations; I participated in the formulation of the evaluation at the beginning, and I am also a colleague of individuals who led the evaluation and executed it. My interest in the research topic stems from my involvement and experience as well as my desire to gain knowledge about the usefulness of the external evaluation for schools, particularly whether teachers and principals understand it as contributing to school improvement and whether they value it as being effective in that regard. However, that interest in studying the use and potential positive impact of evaluation feedback—instead of, for example, focusing on the possible negative side effects of evaluation—directed the research's focus and therefore reveals a certain bias right from the beginning. It should be acknowledged, however, that certain biases apply to all studies.

Because research is always coloured by subjectivities, complete objectivity and neutrality are impossible to achieve (Morgan, 2007). Nevertheless, I have been consistently mindful of the potential bias that may attach to conducting research on a project that I want to see mature and improve. Thus, I want to gauge that circumstance's strengths for the research and how it threatens the research's validity. For one, my concern for the external evaluation's success may have prompted me to give more weight to information that supported the external evaluation or to minimise data and information that would undermine or call it into question. Throughout the research, I took great care in questioning my own assumptions and underlying values in order to minimise bias. I also sought to be mindful of representing views that differed from my own with fairness and balance, as well as to include criticisms expressed in the interviews in the presentation of the results. Last, I discussed my possible bias with my doctoral supervisors right from the beginning, and throughout the research, they offered an outside perspective to my interpretation and presentation of the results. Mindful of minimising bias, they were also actively involved in the classification and selection of schools for participation in the interview study.

To minimise the impact of my bias on the research process, I was conscious of making my position as a researcher transparent. When I requested to interview principals and teachers, I was aware that their willingness to participate or not might well be influenced by their awareness of where I worked. I was very conscious of the need to ensure that perceived differences in power did not contaminate the data collected—for example, by eliciting only comments that they felt were acceptable to me or the authorities that I represented. I countered that problem by working to establish more equitable power relations while conducting each interview, and constantly emphasised the need for the participants to express their honest opinions and that a negative opinion was as valuable to the study as a positive one. During the interviews, I also clearly stated that the research was personal and would contribute to my doctoral thesis as a means to remove any possible fears that the research was being conducted on the behalf of the Directorate of Education and/or other education authorities. Various criticisms and doubts concerning issues were expressed in the interviews, which indicated that the interviewees were giving honest answers. Many of them strongly criticised, for example, the standardised tests for students administered by the Directorate; however, that topic is beyond the scope of the research conducted for the thesis.

Another threat to the validity of the study that warrants mention is the fact that the surveys, on which the results in Paper II are based, were conducted on the behalf of the Ministry of Education and Children, the Icelandic Association of Local Authorities, and the Directorate of Education. Consequently, there was a risk of social desirability bias, because respondents were completing a questionnaire sponsored by education authorities, which may have led them to express overly positive statements even though it was clarified that no strings were attached to their participation. It is difficult to assess

that situation's potential impact on the results. The variation in participants' responses suggests that they were willing to give plausible different answers about, for instance, the impact of evaluation feedback on their own practices, as discussed in Section 8.3.

To address some of the threats to validity considered above and to increase the strength of the research, I employed data triangulation and methodological triangulation. By using different methods (e.g. questionnaires, semi-structured interviews, and document analysis), I was able to explore principals' and teachers' perceptions of the use and impact of the external evaluation from different angles, which greatly strengthened the conclusions that could be drawn from the data. I argue that the combination of those methods and data sources has offered an important contribution, both considering the confirmation and triangulation of data as well as to minimise my potential bias as the researcher. The significance of methodological triangulation in this thesis was also captured in the benefit of providing more comprehensive understanding. Differences that emerged between the attitudes of teachers and principals towards the external evaluation in the survey data were partly explained by the interpretation of interview data with the principals (see discussion in Section 8.3). Thus, the analysis of the survey data benefited from the perceptions drawn from personal experiences in the interview data. Moreover, the interview data indicated a wider use of the external evaluation feedback than the survey data (see Section 8.3); for that reason, if only one form of data collection had been chosen, it would have given a narrower picture of the subject.

Last, it bears repeating the caveat that the results of the research are based on self-report data. All of the data on which the research was based reflect the perceptions of principals and teachers about improvements and changes in their schools. As mentioned in Papers II and III, that circumstance limits the inferences that can be made about improvements or changes as viewed from different perspectives. Because even though the results clearly show that those school professionals perceived that they were working towards improvements and making changes, it cannot be stated that those changes aligned with policymakers' and school authorities' expectations for the impact of external evaluation. Nor can it be claimed that the actions and changes made are real improvements and in accordance with existing theories of good practice. The different perspectives can indeed lead to different conclusions. Therefore, it is necessary to gain a fuller understanding of the nature and extent of the changes taking place with additional research that is not based solely on self-report data, which also might be skewed by social desirability bias. Even then, other perspectives can lead to different sorts of skewness. That circumstances leads to the topic of the next section, which highlights important issues for future research.

8.6 Future research

As discussed, the results of the research conducted for this thesis provide some insights and recommendations. However, the use and impact of evaluation feedback, as well as their potential influencing factors, remain far from fully understood. More research is therefore required, and some suggestions to that end are presented in this section.

Among ideas for further studies on external school evaluation, a point of departure for research on the Icelandic evaluation system in particular is to investigate, from new angles, how and to what extent improvement takes place in schools following external evaluation. A detailed analysis of the improvement measures taken and their tangible benefits would be valuable as well. Another related topic that deserves additional attention is how compulsory schools process and implement evaluation feedback, as well as how they function as learning organisations in the process of deciding on improvement actions to pursue. Such research could provide more insight and information about what can be expected from schools in terms of improvements and changes and to what extent expecting real improvement as a result of external evaluation is realistic (Landwehr, 2011). An alternative approach could be following a case-study design (Yin, 2014), which would be appropriate to answer in-depth questions to clarify how schools work with evaluation feedback, by not only conducting interviews and administering questionnaires but by also observing the entire process in individual schools.

Apart from the intended impacts of external evaluation, further impacts may emerge that are unintended, which researchers have characterised as side effects (Ehren, Jones, et al., 2016; Ehren & Visscher, 2006). It was beyond the scope of the research conducted for this thesis to examine the side effects of external school evaluation. Nevertheless, such examinations are recommended, particularly in relation to the Icelandic system, for international studies on the impact of external school evaluation have revealed that unintended side effects are often negative and tend to undo the intended positive impact (Jones et al., 2017). Shedding light on whether that dynamic also applies to the low-stakes approach used in Iceland and which side effects are most likely to occur.

According to Ehren et al.'s (2013) theoretical framework—the framework that inspired the research conducted for this thesis—high achievement among students is the ultimate goal of external school evaluation. However, because the research could suggest conclusions regarding only part of the framework, it remains necessary to explore the effects that Iceland's evaluation system has on improving students' performance. The research design would have to take into account the chronological order of the mechanisms in the framework by repeating the data collection at least twice at reasonable intervals in order to identify how the mechanisms link the external evaluation to the improvement of schools and students' performance. A larger sample of schools would also make it possible to conduct more analysis on different variables.

Further research on external school evaluation should also be conducted to explore the ideas of school professionals about what kind of formative evaluation, feedback and follow-up would best meet their needs. For example, it would be worth examining whether schools in great need of improvement assess their needs, especially for support, differently from schools where less improvement is called for.

As reiterated in the above sections, the involvement of local authorities in improving schools in Iceland is a kind of black box that needs to be opened and explored. Further research should especially examine the extent to which local education authorities, as school providers, participate in making improvement plans after external evaluation and how they stimulate and support schools when it comes to implementation and following up on improvements. As indicated in the OECD's (2013) report and by Svanbjörnsdóttir et al. (2021), the local capacity to follow up on evaluation varies and poses great challenges for smaller municipalities. Such municipalities have a limited capacity to provide services comparable to ones in larger municipalities and can struggle to perform the diverse governmental functions assigned to them. Moreover, most school committee members in Iceland are politicians who perform their duties on the committee in their spare time, which imposes certain limits on their knowledge and competencies relevant for tasks such as following up on external evaluations. At a time when great emphasis is placed on equality in education systems, it is crucial to learn more about the local capacity to manage educational responsibilities and how to accommodate municipalities that have limited financial, knowledge, and/or human resources such that desired outcomes can be and are achieved. By extension, it is important to critically study how external evaluation works as a mechanism of change across teachers, schools, and municipalities and at the national level and how those different levels interact and see their roles in the process of promoting more effective schools, teaching conditions, and education.

As shown in the mentioned ideas for further research on the Icelandic system, there is no shortage of research questions, for the field is almost completely unexplored, as discussed. Although that list of possible focuses for research is by no means exhaustive, I do not cover additional ones in this section. Even so, I do make two suggestions for further studies focused not only on the Icelandic system.

Despite increased research on how external evaluation impacts teaching and learning in countries worldwide, relatively little research has been conducted to ascertain students' perceptions of how external evaluation influences teaching and learning. In that regard, more research is required to investigate the experience and opinions of students in relation to the impact of the external evaluation on that key aspect of activities at schools. However, students may have difficulty noticing changes because their time in each school is often relatively short, and even if the experience changes, they may not be able to accurately attribute such change to any particular development.

In conclusion, the research conducted for this thesis has provided insight into the use and impact of low-stakes external evaluation systems. As discussed in the literature review, there are certain indications that different evaluation systems might bring about different impacts—for instance, a high-stakes approach has a greater impact than a low-stakes one on specific changes but also leads to more side effects, some of which may be negative (Altrichter & Kemethofer, 2015). Additional studies comparing the use and

impact of external school evaluation in countries with different approaches to evaluation are advised in order to gain better insight into the advantages and disadvantages of different approaches.

9 Conclusions

In this thesis, I have worked towards the dual research aim of (1) shedding light on how the current external evaluation of schools in Iceland came about, how it developed after 1991, and its expected benefits; and (2) identifying and analysing the extent to which those expected benefits in terms of the evaluation's perceived impact and the use of evaluation feedback have been realised according to principals and teachers. The thesis addresses an inquiry made in research on the origin and development of the external evaluation of compulsory schools in Iceland from 1990 to 2016. That research was followed by other research on the mechanisms by which external evaluation influences school professionals' practices and internal evaluation and, ultimately, research on the use and impact of evaluation feedback in schools. Pragmatism provided me with a perspective for the inquiry and supported me in making practical recommendations for central and local governments and schools.

The first part of the research aim, to shed a light on the origin and development of the external evaluation of compulsory schools in Iceland and its expected benefit, was the subject of Phase I of the research and has been addressed in Paper I. Briefly, the results show that although plans to conduct the external evaluation of compulsory schools in Iceland first appeared in policy-related documents in the early 1990s, only in the 2010s was the evaluation implemented and, even so, only as a cooperative project between the state and municipalities, all except Reykjavík, which conducted its own evaluations. The chief benefit that the central authorities expected from the evaluation was that it would contribute to improvements in schools regarding learning and teaching, leadership and management, and internal evaluation.

The second part of the research aim—that is, to identify and analyse to what extent the expected benefits of the external evaluation regarding the use of its feedback and impact were realised according to principals and teachers—was the focus of Phase II and III of the research and has been addressed in Papers II and III. Based on the most significant findings of the two phases, it can be concluded that the external evaluation feedback has been helpful and been used as a basis for change in most schools. Both principals and teachers expressed positive attitudes towards the evaluation feedback, which they generally characterised as being helpful and supportive of improvement and changes in the schools. Most of the principals and teachers interviewed expressed that the external evaluation has been used to improve the quality of learning and teaching, leadership and management, and internal evaluation in their schools (i.e. instrumental use). The evaluation feedback was also reported to have been used for conceptual, persuasive, and reinforcement-oriented purposes in the schools, albeit to varying

degrees. An analysis based on the survey performed for the research revealed that the greatest positive impact on changes, considering all survey items, was whether teachers and principals accepted the evaluation feedback. The elements of setting expectations with evaluation criteria and involving teachers in working with the evaluation feedback were found to have a positive impact on and relationship with changes made to internal evaluation. The only factor examined—and the only one based solely on the principals' answers—that was not found to have any relationship with change was the support and pressure of stakeholders, in particular the municipalities, in the wake of the external evaluation.

In working on this thesis, one of the main challenges that I faced as a researcher was balancing my roles as practitioner and researcher. Indeed, striking that balance was more challenging than initially anticipated. During the lifetime of the research the evaluation process has been reviewed in an effort to refine it. Owing to my position, I participated in discussions and decision-making, and when discussions and changes turned to subjects in which my study or knowledge did not warrant intervention, I often took the position of being an observer and thus did not influence the course of events. However, given my pragmatic orientation in the research conducted for this thesis, I wanted to prioritise the practical value of the research and reveal so-called actionable knowledge in line with Dewey (Greene & Hall, 2010) that would lead to fruitful outcomes. Consequently, I sought to apply my knowledge and present the result of the research throughout the research period. Some of the recommendations that I have presented in this thesis (see Section 8.4.2), including to structure the follow-up process according to the school's status, have partly been implemented as a result of the knowledge that my research generated. It is worth noting, however, that although the external evaluation of compulsory schools in Iceland has undergone changes during the research period, which primarily took place in 2018 and 2019, the research concerned only a certain period in the operation of the external evaluation (i.e. from 2013 to 2015, when the participating schools were evaluated), and the Ministry of Education and Children's follow-up on improvement plans (i.e. from 2014 to 2019, when the participating schools were followed up). Therefore, subsequent changes made to the external evaluation did not impact the research's results.

All schools in Iceland, except in Reykjavík, were evaluated once over a nine-year period. That evaluation cycle ended in late 2021. At the time of writing, in August 2023, no decision has been made on how external evaluation will proceed. A decision has been made, however, to dismantle the Directorate of Education in 2023 and, in its place, establish a new central service and knowledge centre in the field of education and training. Even so, it is uncertain whether political leaders will prioritise the continuation of external school evaluation, who will be entrusted with the responsibility for its operation, and whether and, if so, then how its implementation will change. Nevertheless, I wish to emphasise that in publishing my research on external evaluation and presenting my work to both colleagues and academics, I hope to have both

provided valuable input for the future discussion of external school evaluation and opened up the research landscape for further studies on the phenomenon. It is my sincere wish that by completing the research, I have made certain contributions to the field of external evaluation in general and such evaluation in Iceland in particular.

References

Ad Hoc Team on the External Evaluation of Compulsory Schools. (2011). Ytra mat á grunnskólum: Tillögur til mennta- og menningarmálaráðherra og skólamálanefndar Sambands íslenskra sveitarfélaga um tilhögun á ytra mati á grunnskólum [External evaluation of compulsory schools: Proposals to the minister of education and culture and the Icelandic Association of Local Authorities regarding arrangements for the external evaluation of compulsory schools]. https://www.stjornarradid.is/gogn/rit-og-skyrslur/stakt-rit/?NewsId=5d478635-1dd9-11e7-9c14-e4a7a00e173e

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

AlKutich, M., & Abukari, A. (2018). Examining the benefit of school inspection on teaching and learning: A case study of Dubai private schools. Journal of Education and Practice, 9(5). https://www.iiste.org/Journals/index.php/JEP/article/view/41127/42289

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

Bickman, L., Rog, D. J., & Hedrick, T. E. (1998). Applied research design: A practical approach. In L. Bickman & D. J. Rog (Eds.), Handbook of applied social research methods (pp. 5–37). Sage.

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), Sage handbook of mixed methods in social & behavioral research (2nd ed., pp. 95–117). Sage.

Biesta, G., & Burbules, N. C. (2003). Pragmatism and educational research. Rowman & Littlefield.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brinkmann, S. (2018). The interview. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (5th ed., pp. 576–599). Sage.

Brown, C., & Zhang, D. (2017). How can school leaders establish evidence-informed schools: An analysis of the effectiveness of potential school policy levers. Educational Management Administration & Leadership, 45(3), 382–401. https://doi.org/10.1177/1741143215617946

Cain, T. (2015). Teachers' engagement with research texts: Beyond instrumental, conceptual or strategic use. Journal of Education for Teaching, 41(5), 478–492. https://doi.org/10.1080/02607476.2015.1105536

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

Code of Research Ethics for Public Higher Education Institutions, (2020). https://english.hi.is/sites/default/files/bryndjo/pdf/code_of_research_ethics_for_public_higher_education_institutions.pdf

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

Creswell, J. W. (1998). Qualitative inquiry and research design: Choosing among five traditions. Sage.

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

Dahler-Larsen, P. (2018). Qualitative evaluation: Methods, ethics, and politics with stakeholders. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (5th ed., pp. 867–886). Sage.

Davíðsdóttir, S. (2008). Mat á skólastarfi: Handbók um matsfræði [School evaluation: Handbook of evaluative science]. Hólar.

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

Dewey, J. (1917). The need for a recovery of philosophy. In L. A. Hickman & T. M. Alexander (Eds.), The essential Dewey: Pragmatism, education, democracy (Vol. 1). Indiana University Press.

Dewey, J. (1938). Common sense and scientific inquiry. From logic: The theory of inquiry. In L. A. Hickman & T. M. Alexander (Eds.), The essential Dewey:

Pragmatism, education, democracy (Vol. 1, pp. 380–390). Indiana University Press.

Directorate of Education. (2018). Gæðastarf í grunnskólum: Viðmið og framkvæmd ytra mats grunnskóla [Quality work in primary schools: Quality criteria for compulsory schools]. https://mms.is/sites/mms.is/files/gaedastarf.pdf

Drinck, B., Flagmeyer, D., Diegmann, D., Keitel, J., Schmidt, M., Schubert, R., & Herzog, K. (2013). RuN-Studie. Rezeption und Nutzung von Ergebnissen der externen Evaluation an sächsischen Grundschulen, Mittelschulen und Gymnasien. [Reception and use of the results of the external evaluation at primary schools, middle schools and high schools in Saxony]. S. Bildungsinstitut. https://publikationen.sachsen.de/bdb/artikel/19652/documents/26489

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2005). New public management is dead—Long live digital-era governance. Journal of Public Administration Research and Theory, 16(3), 467–494. https://doi.org/10.1093/jopart/mui057

Dunn, W. N., & Miller, D. Y. (2007). A critique of the new public management and the neo-weberian state: Advancing a critical theory of administrative reform. Public Organization Review, 7, 345–358. https://doi.org/10.1007/s11115-007-0042-3

Ehren, M. C. M. (2016a). Introducing school inspections. In M. Ehren (Ed.), Methods and modalities of effective school inspections (pp. 1–16). Springer.

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

Ehren, M. C. M. (2016c). 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 capacity in education system reform (pp. 1–29). Routledge.

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

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., & Honingh, M. (2011). Risk-based school inspections in the Netherlands: A critical reflection on intended effects and causal mechanisms. Studies in Educational Evaluation, 37(4), 239—248. https://doi.org/10.1016/j.stueduc.2012.02.001

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

Ehren, M. C. M., Leeuw, F. L., & Scheerens, J. (2005). On the impact of the Dutch Educational Supervision Act. American Journal of Evaluation, 26(1), 60—76. https://doi.org/10.1177/1098214004273182

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. (2006). Towards a theory on the impact of school inspections. British Journal of Educational Studies, 54(1), 51—72. https://doi.org/10.1111/j.1467-8527.2006.00333.x

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

Eurydice. (2007). School autonomy in Europe policies and measures. https://op.europa.eu/en/publication-detail/-/publication/102bb131-8105-4599-9367-377946471af3/language-en

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

Eurydice. (2018). Countries: Description of national education systems. Iceland. https://eacea.ec.europa.eu/national-policies/eurydice/content/iceland_en

Fernandez, S., & Rainey, H. G. (2006). Managing successful organizational change in the public sector. Public Administration Review, 66(2), 168—176. https://doi.org/10.1111/j.1540-6210.2006.00570.x

Flick, U. (2014). An introduction to qualitative research (K. Metzler, Ed. 5th ed.). Sage.

Flick, U. (2018). Triangulation. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (5th ed., pp. 444—461). Sage.

Fullan, M. (2016). The new meaning of educational change (5th ed.). Teachers College Press.

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

Gärtner, H. (2021). Wie kann externe Evaluation/Schulinspektion zur Schulentwicklung beitragen? [How can external evaluation/school inspection contribute to school development?]. In T. Webs & V. Manitius (Eds.), Unterstützungssysteme für Schulen: Konzepte, Befunde und Perspectiven (pp. 177–193). WBV.

Gärtner, H., Hüsemann, D., & Pant, H. A. (2009). Wirkungen von Schulinspektion aus Sight betroffener Schulleitungen: Die Brandenburger Schulleiterbefragung. [Effects of school inspections from the perspective of school administrations: The Brandenburg headmaster survey]. Empirische Pädagogik, 23(1), 1–18. https://www.researchgate.net/publication/269704438_Wirkungen_von_Schulinsp ektion_aus_Sicht_betroffener_Schulleitungen_Die_Brandenburger_Schulleiterbefrag ung

Gärtner, H., & Pant, H. A. (2011). How valid are school inspections? Problems and strategies for validating processes and results. Studies in Educational Evaluation, 37(2/3), 85–93. https://doi.org/10.1016/j.stueduc.2011.04.008

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

Greene, J. C. (2007). Mixed methods in social inquiry. Jossey-Bass.

Greene, J. C. (2008). Is Mixed Methods Social Inquiry a Distinctive Methodology? Journal of Mixed Methods Research, 2(1), 7–22. https://doi.org/10.1177/1558689807309969

Greene, J. C., & Hall, J. N. (2010). Diallectics and pragmatism: Being of consequence. In A. Tashakkori & C. Teddlie (Eds.), Sage handbook of mixed methods in social & behavioral research (2nd ed., pp. 119–143). Sage.

Gunter, H. M., Grimaldi, E., Hall, D., & Serpieri, R. (2016a). NPM and educational reform in Europe. In H. M. Gunter, E. Grimaldi, D. Hall, & R. Serpieri (Eds.), New public management and the reform of education. European lessons for policy and practice. (pp. 3–17). Routledge.

Gunter, H. M., Grimaldi, E., Hall, D., & Serpieri, R. (2016b). NPM and the dynamics of education policy and practice in Europe. In H. M. Gunter, E. Grimaldi, D. Hall, & R. Serpieri (Eds.), New public management and the reform of education. European lessons for policy and practice (pp. 173–185). Routledge.

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

Gustafsson, J.-E., Lander, R., & Myrberg, E. (2014). Inspections of Swedish schools: A critical reflection on intended effects, causal mechanisms and methods. Education Inquiry, 5(4), 461–479. https://doi.org/http://dx.doi.org/10.3402/edui.v5.23862

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

Hennink, M., Hutter, I., & Bailey, A. (2011). Qualitative research methods. Sage.

Her Majesty's Inspectorate of Education. (2010). Learning together: Lessons about school improvement. An HMIE report on how schools get better. https://education.gov.scot/Documents/LearningTogetherLessonsSchoolImprovement.pdf

Hofer, S. I., Holzberger, D., & Reiss, K. (2020). Evaluating school inspection effectiveness: A systematic research synthesis on 30 years of international research. Studies in Educational Evaluation, 65, Article 100864. https://doi.org/10.1016/j.stueduc.2020.100864

Hogwood, B. W., & Gunn, L. A. (1984). Policy analysis for the real world. Oxford University Press.

Hood, C. (1991). A public management for all seasons? Public Administration, 69(1), 3–19. https://doi.org/10.1111/j.1467-9299.1991.tb00779.x

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

Iceland Government Office. (2015). Ríkisstjórnartal [Overview of governments] http://stjornarrad.is/Rikisstjornartal/

Icelandic Association of Local Authorities. (2023). Ýmsar upplýsingar um sveitarfélögin á Íslandi [Information about the municipalities in Iceland]. http://www.samband.is/sveitarfelogin/

Icelandic Association of Local Authorities. (n.d.). Local Governments in Iceland (S. Sverrisson & M. K. Hannesson, Eds.) http://www.samband.is/media/enska/SIS-enska.pdf

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. Field Methods, 18(1), 3–20. https://doi.org/10.1177/1525822X05282260

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

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. Educational Researcher, 33(7), 14–26. https://doi.org/10.3102/0013189X033007014

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. Journal of Mixed Methods Research, 1(2), 112–133. https://doi.org/10.1177/1558689806298224

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

Jónasson, J. T. (2019). Óvissa um leiðsagnargildi gagna? Notkun gagna í mótun menntastefnu og skólastarfs [Uncertainty about the directional value of data: Data and their formative value for setting educational aims, policy, and practice]. Tímarit um uppeldi og menntun, 28(2), 161–180. https://doi.org/10.24270/tuuom.2019.28.8

Jónasson, J. T., Bjarnadóttir, V. S., & Ragnarsdóttir, G. (2021). Evidence and accountability in Icelandic education: An historical perspective? In J. B. Krejsler & L. Moos (Eds.), What works in Nordic school policies? Mapping approaches to evidence, social technologies and transnational influences (pp. 173–194). Springer. https://doi.org/10.1007/978-3-030-66629-3

Jónasson, J. T., Ragnarsdóttir, G., & Bjarnadóttir, V. S. (2021). The intricacies of educational development in Iceland: Stability or disruption? In J. B. Krejsler & L. Moos (Eds.), What works in Nordic school policies? Mapping approaches to evidence, social technologies and transnational influences (pp. 67–86). Springer. https://doi.org/10.1007/978-3-030-66629-3

Jónsson, Ó. P. (2014). Lýðræðisleg menntastefna: Sögulegt ágrip og heimspekileg greining [Democratic education: A historical overview and philosophical analysis]. Stjórnmál og stjórnsýsla, 10(1), 99–118. https://doi.org/10.13177/irpa.a.2014.10.1.6

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

Kingdon, J. W. (2014). Agendas, alternatives, and public policies (2nd ed.). Pearson Education.

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

Koliba, C., Meek, J. W., & Zia, A. (2010). Governance networks in public administration and public policy. CRC Press.

Kools, M., & Stoll, L. (2016). What makes a school a learning organisation? OECD. https://doi.org/10.1787/5jlwm62b3bvh-en

Kotter, J. (2012). Leading change. Harvard Business Review Press.

Kristinsson, G. H. (2007). Íslenska stjórnkerfið [The Icelandic government system] (2nd ed.). Háskóli Íslands.

Kvale, S. (1996). Interviews: An introduction to qualitative research interviewing. Sage.

Lander, R., & Ekholm, M. (2005). School evaluation and improvement: A Scandinavian view. In D. Hopkins (Ed.), The practice and theory of school improvement: International handbook of educational change (pp. 85–100). Springer.

Landman, T. (2006). Studying human rights. Routledge.

Landwehr, N. (2011). Thesen zur Wirkung und Wirksamkeit der externen Schulevaluation. [Theses on the impact and effectiveness of external school evaluation]. In C. Quesel, V. Husfeldt, N. Landwehr, & P. Steiner (Eds.), Wirkungen und Wirksamkeit der externen Schulevaluation. h.e.p. .

Lichtman, M. (2006). Qualitative research in education: A user's guide. Sage.

Lieber, E., & Weisner, T. S. (2010). Meeting the practical challenges of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), Sage handbook of mixed methods in social & behavioral research (2nd ed., pp. 559–579). Sage.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2018). Pragmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (5th ed., pp. 108–150). Sage.

MacBeath, J., Frost, D., Frost, R., Swaffield, S., & Waterhouse, J. (2009, January 4–7). The essential conversation: Reforming education through dialogue [Paper presentation]. International Congress for School Effectiveness and Improvement, Vancouver, Canada

Mangione, T. W. (1998). Mail surveys. Sage.

Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: A quasi-paradigm for government on the Web. Philosophical Transactions of the Royal Society, 371(1987), Article 20120382. https://doi.org/10.1098/rsta.2012.0382

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

McCrone, T., Coghlan, M., Wade, P., & Rudd, P. (2009). Evaluation of the impact of Section 5 inspections: Stand 3. Final report for Ofsted. https://www.nfer.ac.uk/publications/SFO01/SFO01.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

Merriam, S. B. (1998). Qualitative research and case study applications in education. Jossey-Bass.

Mertens, D. M. (2005). Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods (2nd ed.). Sage.

Ministry of Finance. (1993). Umbætur og nýskipan í ríkissrekstri [Improvements and new public management]. http://www.fjarmalaraduneyti.is/media/Utgefin_rit/Umbaetur_og_nyskipan_i_rikisr ekstri_1993.pdf

Ministry of Finance. (2000). Ábyrgð, valdsvið og stjórnunarumboð forstöðumanna ríkisstofnana: Nefndarálit [Responsibilities, powers, and administrative authority of heads of government agencies: Committee opinion]. https://www.fjarmalaraduneyti.is/media/rit2014/Abyrgd,-valdsvid-og-stjornunarumbod.pdf

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. Journal of Mixed Methods Research, 1(1), 48–76. https://doi.org/10.1177/2345678906292462

Morse, J. (2018). Reframing rigor in qualitative inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (5th ed., pp. 796–817). Sage.

Moynihan, D. P. (2008). The dynamics of performance management: Constructing information and reform. Georgetown University Press.

Moynihan, D. P., & Pandey, S. K. (2010). The big question for performance management: Why do managers use performance information? Journal of Public Administration Research and Theory, 20(4), 849–866. https://doi.org/10.1093/jopart/muq004

National Curriculum Guide for Compulsory Schools, (2011). http://www.menntamalaraduneyti.is/utgefid-efni/namskrar/adalnamskra-grunnskola/

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.

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

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

Organisation for Economic Co-operation and Development [OECD]. (1997). In search of results: Performance Management Practices.

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

Ólafsdóttir, B. (2012). Tilraunaverkefni um ytra mat á grunnskólastarfi [Pilot project on external evaluation on compulsory schools]. Sveitarstjórnarmál, 1(73), 8–9.

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

Ólafsdóttir, B. (2016b). Ytra mat grunnskóla: Mat á þróunarverkefni 2013−2015. Unnið fyrir mennta- og menningarmálaráðuneyti, Samband íslenskra sveitarfélaga og Menntamaálstofnun [External evaluation of compulsory school: Project evaluation, 2013−2015. For the Ministry of Education, Science and Culture, the Icelandic Association of Local Authorities, and the Directorate of Education]. https://mms.is/sites/mms.is/files/skyrsla_throunarverkefni_0.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(3), 209−224. https://doi.org/10.1080/20020317.2022.2076376

Patton, M. Q. (2002). Qualitative research & evaluation methods (3rd ed.). Sage.

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. (2014). Exploring and explaining the effects of being inspected. Educational Studies, 40(4), 456−472. https://doi.org/10.1080/03055698.2014.930343

Penninckx, M., Vanhoof, J., De Maeyer, S., & Van Petegem, P. (2016a). 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. (2016b). Enquiry into the side effects of school inspection in a 'low-stakes' inspection context. Research Papers in Education, 31(4), 462−482. https://doi.org/10.1080/02671522.2015.1076886

Penninckx, M., Vanhoof, J., De Maeyer, S., & Van Petegem, P. (2016c). 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

Penzer, G. (2011). School inspections: What happens next? CfBT Education Trust. https://www.educationdevelopmenttrust.com/our-research-and-insights/research/school-inspections-what-happens-next

Peräkylä, A., & Ruusuvuory, J. (2018). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (5th ed., pp. 669−690). Sage.

Pettigrew, A. M. (1997). What is a processual analysis? Scandinavian Journal of Management, 13(4), 337–348.

Poteete, A. R., Janssen, M. A., & Ostrom, E. (2010). Working together: Collective action, the commons, and multiple methods in practice. Princeton University Press.

Regulation on Municipal Services to Preschools and Primary Schools and Student Support Councils in Primary Schools, (2019). https://island.is/reglugerdir/nr/0444-2019

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

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

Russ-Eft, D., & Preskill, H. (2009). Evaluation in organizations: A systematic approach to enhancing learning, performance, and change (2nd ed.). Basic Books.

Schick, A. (2002). Opportunity, strategy and tactics in reforming public management. OECD Journal on Budgeting, 2(3), 7–34. https://doi.org/10.1787/budget-v2-art14-en

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

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

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

Scriven, M. (1991). Evaluation thesaurus (4th ed.). Sage.

Sigurðardóttir, S. M. (2023). Educational leadership at the municipal level in Iceland: What shapes it, its characteristics and what it means for school practices [Doctoral thesis, University of Iceland]. Opin Vísindi. https://hdl.handle.net/20.500.11815/4217

Sigurðardóttir, S. M., Sigurðardóttir, A. K., Hansen, B., Ólafsson, K., & Sigþórsson, R. (2022). Educational leadership regarding municipal school support services in Iceland. Educational Management Administration & Leadership. https://doi.org/10.1177/17411432221076251

Sigurjónsdóttir, B. (2010). Heildarmat í grunnskólum Reykjavíkur 2007–2010 [Comprehensive evaluation of compulsory schools in Reykjavík 2007–2010]. Netla: Veftímarit um uppeldi og menntun. http://hdl.handle.net/1946/13797

Sigurjónsdóttir, B., Ólafsdóttir, B., & Jónsdóttir, Þ. B. (2012a). Ferli ytra mats á grunnskólastarfi [The process of externally evaluating compulsory schools]. https://mms.is/sites/mms.is/files/skjal_2_ferli_ytra_mats_i_grunnskolum.pdf

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

Sigþórsson, R., Sveinbjörnsdóttir, B. M., Gunnþórsdóttir, H., Elídóttir, J., Sigurðardóttir, S. M., & Þorsteinsson, T. (2022). Skólaþjónusta sveitarfélaga í nútíð og framtíð: Viðfangsefni, starfshættir og skipulag [Municipal school support services in the present and future: Subjects, working methods, and organisation]. Skólaþræðir. https://skolathraedir.is/2022/05/31/skolathjonusta-sveitarfelaga-i-nutid-og-framtid-vidfangsefni-starfshaettir-og-skipulag/

Simkins, T. (2003). Reform, accountability and strategic choice in education. In M. Preedy, R. Glatter, & C. Wise (Eds.), Strategic leadership and educational improvement (pp. 215–232). Sage.

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.

Statistics Iceland. (2022). Sveitarfélög og byggðakjarnar [Municipalities and regional centers]. Hagstofa Íslands. https://hagstofa.is/talnaefni/ibuar/mannfjoldi/sveitarfelog-og-byggdakjarnar/

Statistics Iceland. (2023a). Grunnskólastig [Compulsory school level]. Hagstofa Íslands. https://hagstofa.is/talnaefni/samfelag/menntun/grunnskolastig/

Statistics Iceland. (2023b). Yfirlit mannfjölda [Population overview]. Hagstofa Íslands https://hagstofa.is/talnaefni/ibuar/mannfjoldi/yfirlit/

Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.). Sage.

Svanbjörnsdóttir, B., Sigurðardóttir, S. M., Þorsteinsson, T., Gunnþórsdóttir, H., & Elídóttir, J. (2021). Skólaþjónusta sveitarfélaga: Starfsþróun og skólar sem faglegar stofnanir [School support services: Professional development and schools as professional organisations]. Tímarit um uppeldi og menntun, 30(2), 3–27. https://doi.org/10.24270/tuuom.2021.30.5

Teddlie, C., & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), Sage handbook of mixed methods in social & behavioral research (2nd ed., pp. 1–41). Sage.

Thillmann, K., Bach, A., Gerl, C., & Thiel, F. (2013). Schulische Selbststeuerung durch interne Evaluation: Eine Analyse von Evaluationsprogramm Berliner und Brandenburger Gymnasien. [School self-regulation through internal evaluation: An analysis of the evaluation program of Berliner and Brandenburger Gymnasien]. In I. v. Ackeren, M. Heinrich, & F. Thiel (Eds.), Evidenzbasierte Steuerung im

Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund (12 ed., pp. 197—225). Waxmann.

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

Vanhoof, J., Verhaeghe, G., Van Petegem, P., & Valcke, M. (2013). Improving data literacy in schools: Lessons from the school feedback project. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), Data-based decision making in education: Challenges and opportunities (Vol. 17, pp. 113—134). Springer. https://doi.org/10.1007/978-94-007-4816-3

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

Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. School Effectiveness and School Improvement, 14(3), 321—349. https://doi.org/10.1076/sesi.14.3.321.15842

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.

Wang, G. T., & Park, K. (2016). Student research and report writing: From topic selection to the complete paper. Wiley & Sons.

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

Weiss, C. H. (1979). The many meanings of research utilization. Public Administration Review, 39(5), 426—431. https://doi.org/10.2307/3109916

Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An alternate route to policy influence: How evaluations affect D.A.R.E. American Journal of Evaluation, 26(1), 12—30. https://doi.org/10.1177/1098214004273337

Wenger, M., Gärtner, H., & Brunner, M. (2022). Wie verändert sich die Unterrichts- und Schulqualität, die Leistung und die Schülerzusammensetzung nach der Schulinspektionsdiagnose "erheblicher Entwicklungsbedarf"? [How does the quality of teaching and the school and the performance and composition of students change after the school inspection diagnosis "considerable need for development"?]. Zeitschrift für Erziehungswissenschaft, 25(4), 1031—1058. https://doi.org/10.1007/s11618-022-01066-3

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). Program evaluation: Alternative, approaches and practical guidelines (2nd ed.). Longman.

Wurster, S., & Gärtner, H. (2013). Schulen im Umgang mit Schulinspektion und deren Ergebnissen. Zeitschrift für Pädagogik 59(3), 425−445. https://doi.org/10.25656/01:11944

Yin, R. K. (2014). Case study research: Design and methods (5th ed.). Sage.

Original publications

This thesis is based on the following original publications, listed chronologically.

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

Paper I was published in the peer-reviewed Icelandic-language journal *Netla* and translated into English for this thesis.

II.   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*(3), 209–222.
      https://doi.org/10.1080/20020317.2022.2076376

III. Ólafsdóttir, B., Jónasson, J. T., & Sigurðardóttir, A. K. (2022). Use and impact of external evaluation feedback in schools. *Studies in Educational Evaluation, 74*. Article 101181.
      https://doi.org/10.1016/j.stueduc.2022.101181

All three papers were published with open access and with the authors' right to redistribute and republish them. Obtaining special permission to republish them in this thesis was therefore unnecessary.

Paper I

Translated version of Paper I: The origin and development of external evaluation in Iceland from 1990-2016

Author: Björk Ólafsdóttir

Abstract

The aim of this research is to contribute to an improved understanding as to the origin of external evaluation of compulsory schools in Iceland and how it has developed since it became part of official educational policy in 1991. The main sources used to explain the formation and development of the evaluation were theories of agenda-setting, research on the progress of control in public administration and theories of change implementation and management. Analysis of the case study was based on available documentation and interviews conducted with eleven respondents during the summer and autumn of 2015.

The findings of this research indicate that the initial policy to conduct an external school evaluation can be attributed to remedial actions in public administration known as 'new public management' and conducted in many countries from the 1980s. One aspect of this innovation was the devolution of government services and authority to municipalities, including the transfer of compulsory schools leading to increased emphasis on evaluation and accountability. External evaluation, as defined in this case study, did not, however, gain a foothold in Iceland at this time despite educational policy intentions in this regard dating from the early 1990s. Instead, the focus was placed on the development of an internal evaluations in compulsory schools and a form of external review was developed which was to monitor the implementation of internal evaluations.

Thus, it was not until the beginning of the second decade of the twenty-first century that a process of external evaluation was launched in order to promote school improvements. By the Compulsory School Act of 2008 both the Ministry of Education, Science and Culture and the municipalities became responsible for the external evaluation and inspection of compulsory schools and their operation. An open legal framework and uncertain division of responsibilities between the Ministry and the municipalities, however, led to the co-operation of those two administrative sectors with regard to the external evaluation of compulsory schools in the form of a pilot project which will be completed at the end of 2016.

Introduction

An OECD paper, published in 2013, dealt with the extensive expansion in comprehensive evaluation of school operations in various parts of the world during the past three decades. The paper stated that educational authorities have to an ever-growing extent focused on other aspects of school evaluation than student assessment, such as, for example, the evaluation of teaching practices and the role of school administrators, schools and education systems. Results were said to be used for a variety of purposes; for example, to obtain an improved understanding of students' work and

achievement as well as identifying areas where schools were successful and where improvement was required (Organisation for Economic Cooperation and Development [OECD], 2013).

The section below presents the conclusions of a case study focusing on the implementation of external evaluation of compulsory school activities in Iceland. This research serves a dual purpose; firstly, to gain a clearer understanding of the origin of external evaluation of compulsory schools, and, secondly, to illustrate its development. For this purpose, documents relating to official policy formation were studied, going all the way back to the tenth decade of the 20th century; that is, the period which marks the initiation of external evaluation in policy documents in Iceland in recent discussion, thereby identifying the beginning of this research which extends to the latter half of 2016. The research involved the examination of various documents which cast light upon the progress of school evaluation in Iceland as well as conducting interviews with individuals who had either participated in policy formation relating to external evaluation or been involved in its implementation. Here the purpose was to clarify when external evaluation came to the attention of public administration, which ideas or agents triggered this development and how the concept evolved to the present day. Written coverage of this matter has been but scant in Iceland up to now, a shortcoming the present article aims to rectify. It is the purpose of the article to demonstrate some of the main characteristics of the implementation of external evaluation and provide insight into factors affecting this process which could be of use, for example with regard to further policy formation.

My interest in this topic can be traced back to my participation in formulating methods and criteria relating to the external evaluation of the compulsory school during the period 2009-2012 and my subsequent involvement in preparations for its implementation. I work at the Directorate of Education which is in charge of external evaluation of compulsory schools.

Definitions of concepts

Definitions of the main concepts dealt with here are based on Eurydice (2015) which states that school evaluation monitors school activities with the aim of improving the quality of the schools and students' results. Results applying to the school as a whole are presented in a report. Eurydice differentiates between external and internal evaluation. The former is carried out by evaluators who are not connected with the school under evaluation. Internal/self-evaluation, on the other hand is conducted by individuals or groups in direct connection with the school, often members of its staff. For the purpose of this research is also based on a definition, originally from the OECD, which separates evaluation from assessment. According to Organisation for Economic Cooperation and Development [OECD] (2013) the concept of *assessment* relates to the assessment or progress of the studies of individual students on the basis of learning intentions. The *evaluation* concept, on the other hand, involves the evaluation of schools as a whole, the school system, its policies and tasks. Thus, external school evaluation comprises both school reviews and school inspection/quality control, as stated by the OECD in the following terms:

> School evaluation concentrates on key processes such as teaching and learning, school leadership, educational administration, school environment and the management of human resources. It does so in association with an analysis of student outcomes, both the achievement/progress of students and the equity of student results. It also takes into

account inputs such as the infrastructure, funding and characteristics of the school staff (Organisation for Economic Cooperation and Development [OECD], 2013, p. 384).

According to the definition by OECD the concepts of inspection and quality control are subsumed under the term external evaluation. Those notions frequently occur in the official policy documents discussed in the article and should, therefore, be briefly explained. In an explanatory memorandum accompanying a parliamentary bill on the compulsory school quality control was defined as "overseeing that schools and municipalities comply with the provisions of law, regulations and the national curriculum guide for compulsory schools" (Parliamentary document 319, 2007, p. 47). Inspection, on the other hand, has been defined as "an independent performance evaluation in order to examine and evaluate the operation of organisations and their efficacy." (Ministry of Finance and Economic Affairs, 2013, p. 75)

Evaluation of schools is closely related to the theory of performance management which is mainly based on the administrative ideology of New Public Management (Ministry of Finance, 1996; Moynihan, 2008). In a paper issued by the Ministry of Finance in 1996 performance management was defined as an assortment of management methods based on clear objectives, systematic measurements and follow-up (Ministry of Finance, 1996). A news bulletin from the Ministry of Education published a year later stated that a school's internal evaluation was regarded as part of its performance management (Ministry of Education, 1997). The documents discussed in this article tend to refer to quality management rather than performance management. According to the paper from the Ministry of Finance referred to above, quality management concentrates on improving work processes and thus enhancing the quality of an operation, whereas performance management involves investigating whether the operation as a whole is heading in the right direction (Ministry of Finance, 1996). Accordingly, performance management is a more comprehensive concept and may *inter alia* comprise quality management.

Academic context

How issues enter into government agenda

The American political analyst Kingdon (2014) advanced a theory as to why certain issues find their way into political debate and subsequently become part of government agenda. His theory was based on policy research in the area of official health and communications in the American administration during the period 1976-1979. Although educational policies are not included in his theory it is nevertheless of interest to examine whether it could be expanded to include this administrative level. Kingdon discusses three agenda-setting process streams which must exist in order to bring about significant government policy changes. Firstly, there has to be problem recognition which thrusts the matter into political debate where potential solutions are discussed. Secondly, viable policy proposals must have been aired by experts and policy makers. Thirdly, the political environment must be favourable, for example a newly formed government which opens up an opportunity for new proposals to enter the political agenda. When those three streams flow together a policy window is opened; that is a favourable timing for policy changes, enabling advocates of innovative proposals to present their ideas or attract administrative attention to a specific problem. Nevertheless, Kingdon points out that problems are not necessarily the only reason

why certain issues find their way to the government table; sometimes new areas of knowledge and expert opinions in a particular field can become the catalysts of innovative discussion and proposals.

Another political scientist, Schick (2002), identified factors which impact administrative amendments and there is much accord between his ideas and Kingdon's theories. Schick pointed out that sometimes the time for change had simply arrived, and trends and policies in other countries had a strong impact in this context. He also mentioned that elections and changes of government opened up opportunities for change, since frequently major administrative and policy innovations were launched shortly after a new government had taken over. He said that entrepreneurs were also highly effective mediators of new ideas, but such innovators needed to have enough power to be able to persuade politicians to take the risk of new approaches. He also made the point that the politicians themselves had to be convinced that they were making the right choice.

New Public Management (NPM): Decentralisation and performance management

NPM is a concept covering an assortment of ideas and theories aimed at stimulating reforms in public administration (Hood, 1991). A policy of innovation dominated administrative reform in developed countries all over the world from the beginning of the ninth decade of the 20th century and well into the first decade of the 21st (from 1980 until around 2005), albeit to varying degrees and with different areas of emphasis in each country (Margetts & Dunleavy, 2013).

During the period while the NPM was being adopted public organisations were often regarded as inefficient and ineffective; that is, marred by slack performance (Moynihan, 2008). Those who laid the foundation of the NPM made a special effort to provide practical solutions to this problem. A common core of their ideas was to utilise the management practices and principles of the private sector to develop a success oriented and performance driven official administration system (Eurydice, 2007; Moynihan, 2008; Organisation for Economic Cooperation and Development [OECD], 1997).

Hood (1991) identified several factors most frequently referred to in relation to New Public Management during its initial period. Among those factors was professional management where management came to be increasingly regarded as a profession in its own right. The CEO was to be given the mandate and freedom to manage and be plainly accountable for his activities. Besides, explicit standards and measurements were to be established with regard to performance; this involved defining measurable targets for success in order to be able to determine whether the CEO and the organisational entity could sustain their responsibilities. Koliba et al. (2010) have also considered the concept of new public management and highlighted associated areas of emphasis. They specifically mentioned, *inter alia*, devolution/decentralisation by transferring tasks to a lower administrative administration which was assumed to result in improved organisational efficacy and more flexibility in the provision of services.

New Public Management was a declared policy of Icelandic governments 1991-2007 (Iceland Government Office, 2015). A report published by the Ministry of Finance *Reforms and New Public Management* stated that the essence of the policy was "distributing power, enhancing responsibility and shifting decision-making as close as possible to the relevant setting, thus attaining a more economical performance and improved services". (Ministry of Finance, 1993, p. 4). As regards objectives and results assessment, the report stressed the importance of public institutions being

evaluated on the basis of their objectives and subsequent results, stating that in some cases institutional quality systems might facilitate successful performance.

Delegation of power and responsibility – increased control and monitoring of school operations

Devolution is based on the delegation of power and responsibility. In a democratic society the activities of the state should reflect the wishes of the electorate as far as possible (Kristinsson, 2007). Thus, it is a fundamental principle that decisions taken at various administrative levels be led and monitored by democratically elected representatives and their government ministers. Education is among the issues where elected representatives and the legislature are obliged to serve the public interest (Simkins, 2003). By means of devolution, the political responsibility for education is divided between central and local government, municipal councils, headmasters and teachers.

As tasks were being transferred to lower levels of administrations, government departments and other central institutions retained the crucial functions of determining policies and setting targets as well as overseeing performance (Ministry of Finance, 2000; Schick, 2002). The analysis conducted by Jónsson (2014) of the changes to compulsory school legislation, in 1995 when the compulsory school was transferred from the state to municipalities, and to the general curriculum guide in 1999, revealed that the change from previous legislation and curricula was mainly of two kinds; on the one hand objectives became considerably more detailed and comprehensive than before and, on the other, these objectives were more closely linked to school activities and therefore could serve as a basis for the evaluation. In this context Jónsson pointed out that at the same time as central administration was revoked by delegating the compulsory school to municipalities, control was strengthened by means of stricter criteria regarding the objectives and evaluation of school activities.

The OECD report from 1997 stated that trends toward devolution require a more thorough performance management and possibly a special supervisory authority for the purpose of monitoring the implementation of legislation (Organisation for Economic Cooperation and Development [OECD], 1997). This, however, does not appear to have been fully actualised, as Schick (2002) expressed his concern that in many cases centralised institutions and government departments had been weakened to such an extent as a result of devolution that they were unable to implement essential co-ordinating functions. He said that in the eagerness to grant increased freedoms to local authorities and administrators insufficient care has been taken to ensure the continued maintenance of necessary resources and centralised authority; on the one hand, to formulate cross-institutional plans and policies and, on the other, to maintain strict accountability. Thus, responsibilities remained relatively unformulated and embryonic in many countries. A report by the ministry of finance from the year 2000 indicated that this was also the case in Iceland where there had been more emphasis on devolution and the enhanced freedom of institutions than on developing ways of determining institutional responsibilities. The report emphasised the need for change in this respect (Ministry of Finance, 2000).

The Eurydice report from 2007 indicated that in most countries the central government is responsible for matters relating to education and that it was the role of inspectorates to monitor and evaluate schools. In some locations, especially in the Nordic countries, the independence of schools and their devolution to municipal councils springs from the idea that municipalities should have a pivotal role in school management. In those circumstances schools are mainly accountable to the local authorities which, in turn, have the crucial function of supervising and evaluating schools. In

general, however, there is a trend towards multi-accountability, with schools answering to a range of bodies such as departments of education, municipalities and parents (Eurydice, 2007).

Adoption of policy and change

When confronted with change in public governance, the authorities face countless decisions with respect to introducing the changes in question, deciding who is to implement them and how (Schick, 2002). Hogwood and Gunn (1984) identified several factors which influence whether a policy is viable. Among items they singled out are external circumstances, adequate time and the requisite combination of resources, that the policy be based upon a valid theory of cause and effect and that objectives be clearly understood and agreed upon. They stated that time must be devoted to develop an understanding of the changes in question and encourage interest in their implementation. For several years, Kotter (2012) has studied the process of change in companies and institutions. He emphasised the need to demonstrate a sense of urgency in this context so that people experience change as a solution to a certain challenge. No less crucial is the support and commitment to change of top-level management as well as leadership by a guiding coalition. In cases where introducing change is most successful the guiding coalition is endowed with leadership skills in combination with authority, knowledge and the necessary associations. He also identified key factors in achieving change, such as having a clear vision and strategy, well-defined avenues of reaching objectives and seizing every opportunity to highlight the vision of change. Fernandez and Rainey (2006) pointed out that a vague strategy may lead to confusion and provide official management with the opportunity of reinterpreting objectives and arbitrarily implement change in such a way that only part of the programme envisioned by policy makers is ever accomplished. In line with Kotter, they maintained that changes in the public sector depended on the support of government ministers and key external stakeholders whose impact consisted in their ability to impose statutory changes and control the flow of resources to public institutions.

According to the OECD report (2013) many countries have had problems introducing the evaluation of school activities. The report suggested various reasons such as poor policy design, educational authorities lacked the muscle to put the evaluation into practice, the absence of an evaluation culture, or insufficient use of evaluation results. Furthermore, the report stated that the implementation of evaluation was complicated, requiring the inclusion of numerous participants with different kinds of interest. It was important, therefore, according to the report, to strengthen the cooperation of stakeholders through consultation, harness the services of experts, demonstrate clear objectives of the evaluation, gather data and assess the implementation, build skills, provide resources and allow enough time for completing the task. It was emphasised, however, that one size does not fit all; the process must always correspond to the education system and conventions of the country in question.

Research questions

In light of the aim outlined in the introduction and the academic writings referred to here, the project is launched by means of two research questions in search of answers:

1. *How did external evaluation of compulsory schools in Iceland come about and what explains its origin?*
2. *How has external evaluation of compulsory schools developed up to the present time and what explains its development?*

Method

The method of this investigation was a qualitative research approach, in the form of a case study, as is often used for the purpose of describing a certain occurrence and answering research questions which attempt to explain *how* and *why* something happened (Yin, 2014). The case study was considered a suitable research format since the research topic and the research questions called for an approach enabling the researchers to connect divergent data for the purpose of obtaining a comprehensive representation of a certain research process.

Participants

Participants were selected by means of purposeful sampling, with a view to gathering information from those parties the researcher considered to have the best overview of the topic and also an insight into its individual elements (Patton, 2002). A total of ten interviews were conducted with eleven individuals, including five former or currents specialists or administrators in the Ministry of Education and Culture, two former ministerial advisors or assistants, two academics from universities and two specialists from the Icelandic Association of Local Authorities. Seven out of the eleven interviewees had participated in education policy formation in the beginning of the 1990s and thus were able to provide information in the context of the former research question. Ten out of eleven interviewees had been in some way associated with the development of external evaluation during the period covered by the research and therefore were able to provide insights in relation to the latter research question.

Implementation

The implementation of the research is based on the acquisition, examination and analysis of policy formation documents, articles of law, regulations, reports and other official documents, as well as interviews conducted during the period May to September 2015. The interviews were semi-structured supported by pre-determined open questions on the basis of which each interview was specifically developed based on the information the interviewee concerned was able to provide (Lichtman, 2006). Thus, the interviewees did not always answer the same questions and no one interviewee answered all the questions.

Processing

The analysis covers the progression of external evaluation from the time when this concept first appeared in official policy documents in 1991 until the latter half of 2016. Initially, available public documents were gathered and analysed. Next, an interview framework, based on this analysis, was formed for the purpose of gaining a deeper and more comprehensive perspective of the case in question. Interview data were gathered until saturation level was considered to have been reached. The interviews were recorded with the interviewees' permission, transcribed and thematically analysed. Subsequently the data were examined and explained with a view to Kingdon's (2014) theory as to how matters come to be included in government agenda and Kotter's (2012) theory on the introduction of policy and change. Furthermore, ideas and concepts originating from new public management were used when interpreting the data. The research is anonymous and to prevent a linkage between individual comments the interviewees are not issued with pseudonyms. Instead, their comments are anonymously interweaved with registered data relating to the topic.

Results

Evaluation and quality control as outlined in legislation on the compulsory school and national curriculum guides 1974-1990

The Compulsory School Act from 1974 did not include external evaluation of school operations in the sense presented this article. The Act stated that the Ministry of Education is in charge of matters relating to the compulsory school and that a director of education was located in each education district as a representative of the Ministry and in charge of the local education authority ("Compulsory school act," 1974). The main responsibility of the director of education was, *inter alia*, to ensure that current laws and regulations with regard to education were complied with.

It was in the national curriculum guide for compulsory schools published 1976 that the concept *evaluation of school operations* first occurred in recent time discussion, which only referred to evaluation performed by the schools themselves, not by external parties ("National curriculum guide for compulsory schools," 1976). The national curriculum guide published 1989 emphasised the importance of the schools themselves evaluated their operations for the purpose of improvement ("National curriculum guide for compulsory schools," 1989).

It was revealed in an interview that it was around the year 1990 that talk began of external evaluation and quality control of compulsory schools by the Ministry of Education in addition to the informal overseeing carried out by directors of education in their jurisdictions. The interviewee concerned was of the opinion that the Ministry's first processing of information was conducted on the basis of autumn and spring reports from compulsory schools in 1991.

Evaluation and quality control in compulsory school legislation and in public education policies 1991-1994

The Act on Compulsory Schools was ratified by the Icelandic parliament (Althing) when Svavar Gestsson, from the People's Alliance was the Minister of Education. According to this legislation, directors and local education authorities were given a larger role than before as regards school supervision as an amendment to earlier legislation stipulated that directors of education were to lead school improvements and take charge of supervision and quality control regarding facilities, organisation and efficacy of school operations ("Compulsory school act," 1991). According to one research interviewee, however, who was a director of education until 1996, no formal external evaluation was initiated by the director or directorate of education during this period. This statement is in line with the conclusions of research by Hansen and Jóhansson (2010) regarding the role of directors of education during the period 1975-1996. According to the legislation from 1991, the role of the Ministry of Education was planning reforms in school operations and conducting evaluations of schools their operations and study materials ("Compulsory school act," 1991). A memorandum attached to the parliamentary bill in question indicates that the evaluation involves the general supervisory responsibilities of the Ministry; on the one hand, to gain information as to where reforms are required and, on the other, to ensure that schools are operated in compliance with laws and regulations (Parliamentary document 192, 1990).

In the year the law took effect the Ministry of Education (1991) published a school policy named *Towards a new century. The education ministry's school implementation plan until the year 2000.* The policy was prepared by an ad-hoc group from the staff of the ministry, in consultation with stakeholders, led by a ministerial advisor. Among areas of policy emphasis were, on the one hand,

democracy and devolution of power and, on the other, evaluation and research relating to school activities. The policy advocated the increased independence of schools and a devolution of power to education districts, municipal school committees and schools. The policy indicated, however, that the Ministry of Education considered a significant degree of centralisation necessary to ensure a country-wide comparability in study levels and offerings. The aim of the evaluation and research part of the policy was to arrange a regular quality assurance and evaluation of school activities as well as guidance regarding development and remedies in order to contribute to improvements, guarantee services required by law and a comparable country-wide quality of education.

Both the Act on Compulsory Schools (1991) and the policy *Towards a new century* were published in the year when Svavar Gestsson left his ministerial post and was replaced by Ólafur G. Einarsson from the Independence Party (Iceland Government Office, 2015). According to interviewees, the legislation and policy of *Towards a New Century* was never implemented as a result of this change of government, or as one interviewee put it:

> Because there was a change in government there never was … as quite often
> happens … there was no further work on this document, it was just hidden away
> in a drawer and has not been seen since then; so that this did not really have
> much effect … apart from being a historical piece of evidence (from an interview).

A year later a new minister of education appointed a committee on the formation of an education policy to make proposals for educational reforms and review current legislation on compulsory and secondary schools (*Committee on the formation of an education policy. Report [Nefnd um mótun menntastefnu. Skýrsla]*, 1994). The committee was made up of the representatives of stakeholders and ministry staff. Similarly, to the policy of *Towards a New Century,* the committee on the formation of an education policy proposed increased devolution in the school system as well as a centralised strategy of objectives, an enhanced transmission of information and a regular external evaluation of school operations. There was special emphasis on a complete transfer of the operation and financing of compulsory schools to municipalities and that subsequently the role of the ministry of education would involve the formation of an education policy, the enactment of laws and regulation relating to school operations, supervising schools and monitoring their activities. The committee recommended that evaluation of the school system and its individual sectors be strengthened, and various aspects of its activities be subject to regular reviews. The committee also recommended that compulsory schools be required to adopt methods to evaluate its internal activities. The chief aim of school evaluation, according to the committee's report was to support school improvement and obtain a holistic picture of school activities as a basis for further policy formation. When the minister of education introduced to Althing a bill of new compulsory school legislation he emphasised that the main priorities in the new education policy related to the deliberate strengthening of both external and internal evaluation of school activities (Einarsson, 1994).

The committee supported the proposals outlined above by means of a variety of reasons. It was pointed out that the committee had taken note of trends in Western countries where significant changes in schools had occurred, involving, *inter alia*, more devolution in the handling of finance, the organisation of school activities and working methods (*Committee on the formation of an education policy. Report [Nefnd um mótun menntastefnu. Skýrsla]*, 1994). At the same time, centralisation had been sharpened in the definition of aims of school operations and co-ordinated assessment of study outcomes. The report also stated that evaluation of school activities and quality management in

schools reflected the government's response to stiffened demands regarding the obligation of responsibility in public service. The report pointed out that prior to the time when the committee was nominated the Icelandic authorities had neglected quality control of schools' internal activities. A briefing conducted for Eurydice in 2001 by the ministry of education on the evaluation of school activities in Iceland indicated that one of the reasons for appointing a committee on the formation of an education policy was that early in the 1990s politicians and the general public had begun to raise questions about the quality of school activities, about results in the state education system, the public funding of the school system, ways and means in teaching and equal access to education (Kjartansdóttir, 2001). All of this was seen as creating a demand for added research and clearer policy-making on behalf of the education authorities as well as an evaluation of school activities. The problems of the education system listed above were not further analysed in the Eurydice briefing. However, a report compiled by the committee on the formation of an education policy referred to a survey from 1989 of public attitudes to the compulsory school which "revealed that the school system was subject to severe criticism as only 39% of the respondents felt that the school system was functioning successfully" (*Committee on the formation of an education policy. Report [Nefnd um mótun menntastefnu. Skýrsla]*, 1994, p. 93). Thus, the report suggested that a problem existed in schools and the school system as a whole, and evaluating and monitoring school activities was, *inter alia*, an attempt to tackle this challenge. No interviewee in the present research, however, recalled the existence of a specific problem at that time, which people were concerned about and wished to rectify. One respondent identified the difficulty as simply being that people did not understand the situation due to lack of information:

> So little was known about the school system, how it functioned, how successful it was, etc. And highly placed officials, even ministers were making all kinds of declarations regarding the quality of school activities and the Icelandic education system. But when further information was requested, they turned out to be empty-handed, they had no reports, figures, nothing. Thus, there was no evidence to support those declarations. And even the most basic information as to how much money was spent on education was just not available … the lack of information was blatantly obvious (From an interview).

The two policies: *Towards a new century. The education ministry's school implementation plan until the year 2000* on the one hand, and *the education policy of a committee on the formation of an education policy* on the other, were formulated in short succession. The research respondents who had participated in the formation of the latter policy were asked whether the former policy had been consulted during the compilation of the latter one. They generally were of the opinion that this had been done and that both policies had been based on the same ideology, although the latter policy had been defined as a new beginning from a political point of view. One respondent answered the question as follows:

> Not officially, of course many committee members were aware of this … and I of course knew this well … *Towards a new century* was literally taken off the agenda by a new administration so it was not directly used. But when you have read something like that and perhaps participated in preparations, then a source like this is always relevant. Many of those who took part in the creation of *Towards a new century* or contributed to it went on making contributions. (From an interview)

In the opinion of the interviewees the emphasis on devolution and increased evaluation and monitoring of school activities was not the result of influence from the political parties in power. The respondents pointed out that the emphasis on evaluation originated through Iceland's OECD membership and was also brought in by people who were returning from studies abroad at that time. Among those was Gerður G. Óskarsdóttir who was in charge of preparations for the policy *Towards a new century* and Stefán Baldursson who was a member of the committee on the formation of a new education policy. Those two were specifically mentioned by the respondents as the members of the policy formation groups who placed a particular emphasis on evaluation and research with regard to school activities.

The concepts of evaluation and supervision in legislation on the compulsory school and their implementation 1995-2008

The proposals of the committee on the formation of an education policy became the basis of new compulsory school legislation enacted in 1995. In accordance with the committee´s proposals, the legislation introduced the sweeping change of transferring the entire running of the compulsory school to municipalities ("Compulsory school act," 1995). As before, the minister of education was to be in charge of this policy area and inform Althing of the implementation of school activities. According to the legislation, the role of the ministry of education was, on the one hand, to ensure that municipalities fulfilled their duties and, on the other, to gather information on school activities. The positions of directors of education were abolished and municipal councils made responsible for reporting the implementation of school operations to the ministry of education. Municipal school committees were also given an additional role, as compared to earlier legislation, with regard to school supervision. The committees were responsible for monitoring learning and teaching and suggest improvements in school activities. Compulsory schools were required to establish methods of evaluating their own school activities and at five-year intervals the ministry of education was to initiate reviews of schools' methods of self-evaluation. Furthermore, the minister of education was to be made responsible for implementing the evaluation of schools and their activities in order to ensure that school activities complied with the provisions of law and the national curriculum guide. A memorandum accompanying the bill of law concerned stated that this last-mentioned proviso was supposed to "ensure that a formal evaluation of school activities is carried out and that its conclusions are used for remedial purposes" (Parliamentary document 131, 1994).

The law was decisive with regard to the introduction of schools' self-evaluation and regular reviews by the ministry of their methods in conducting the process. It was less clear, however, what was intended by the provision regarding external evaluation of school activities. The respondents who participated in the research suggested a variety of reasons for prioritising internal rather than external evaluation. Thus, the respondents indicated that this emphasis was in part to be traced to opposition and concern among teachers and their professional association where people worried about "more supervision and even intrusion" (from an interview). It was suggested that, at that time, bringing up the subject of evaluation was a rather sensitive issue and discussion in the committee had taken the course that the first steps should be taken with caution, beginning with the internal evaluation to allow teachers and schools to get used to the concept of evaluation. It was also mentioned that people had felt that the first stage should be the gathering of information, followed by internal evaluation and in the wake of this process it would be the turn of external evaluation. But the consideration of opposition and adjustment was not the only reason why internal evaluation was preferred to the external procedure. One respondent pointed out that there had practically been

consensus among committee members that the responsibility for the quality of school operations should be as near the actual setting as possible and that, first and foremost, this was the responsibility of the school "rather than having the authorities walk directly into the building to observe the teaching and all other activities" (from an interview). Financing the evaluation was another reason brought up by the respondents. It was revealed that ministry staff had made inquiries abroad as to the cost of external evaluation and concluded that it was simply too high for a realistic emphasis on a regular evaluation of this kind. Besides, at that time Iceland lacked the necessary human resources to perform such an evaluation with the skill required.

The "Compulsory school act" (1995) was ratified in the year when Ólafur G. Einarsson resigned as minister and the ministry of education was taken over by Björn Bjarnason, of the same political party, who remained minister until 2002 (Iceland Government Office, 2015). It therefore became the responsibility of Björn Bjarnason to prepare the introduction of the legislation and the formulation of the national curriculum guide. In a speech the education minister delivered in 1995 he stated that there would be special emphasis on external evaluation and regular school appraisals (Bjarnason, 1995). In another speech, the minister outlined the altered role of the ministry of education with respect to the compulsory school which mainly involved "setting rules and supervising school activities in Iceland" (Bjarnason, 1996). The minister said these changes called for new working methods and accordingly a new department of evaluation and supervision had been created within the ministry.

Opinions were divided among the respondents as to Björn Bjarnason's interest in the adoption of external evaluation in compulsory schools One respondent pointed out that he had strongly emphasised matters relating to evaluation and quality control at all school levels and, *inter alia*, underlined this by setting up the department of evaluation and supervision. Another respondent was of the opinion that the minister had mainly supported external evaluation in universities and upper secondary schools, but there had been little emphasis on evaluation in compulsory schools because they had, by that time, become the responsibility of municipalities. In a report on the external evaluation of compulsory schools, compiled for Eurydice in 2001 this latter point of view is supported, as the report states that since the municipalities took over the running of the compulsory schools in 1996 the external evaluation of compulsory schools has not been a priority issue in the ministry of education (Kjartansdóttir, 2001). According to a respondent who worked in the ministry during Björn Bjarnason's ministership, the minister was firmly of the opinion that the ministry should not be meddling in the way municipalities chose to organise compulsory school activities.

A new general curriculum guide for the compulsory school, based on the legislation from 1995, was published during Björn Bjarnason's ministership. The guide dealt in some detail with the internal evaluation of compulsory schools, whereas little attention was paid to external evaluation. The education ministry's appraisals of self-evaluation methods in compulsory schools were outlined, but with regard to further external evaluation or appraisals under the auspices of the ministry of education the curriculum guide stated that the ministry could either arrange overall evaluations of school activities in general or single out their individual aspects (*National curriculum guide for compulsory schools*, 1999).

A survey of published material on the website of the Ministry of Education and Culture during the period 1995-2008 while the law was in effect, it was found that overall evaluations were only conducted in five compulsory schools in Iceland under the auspices of the ministry during those

thirteen years; and those were carried out 1998-2003 (Ministry of Education and Culture, n.d.; Task force on quality in school operations, 2005). Statements by respondents in the research suggest that the decision to carry out specific evaluations or reviews of those schools did not originate with the ministry of education, but were requested by municipalities, in most cases because of problems in the schools concerned. There were also cases where municipalities requested reviews of compulsory school activities directly from the academic community. This was also mainly because of existing difficulties or because a rationale was needed on which to build decisions.

In line with provisions of the law, the ministry's evaluation and supervision department twice initiated appraisals of the self-evaluation methods practised by all compulsory schools in Iceland, first during the period 2001-2003 and again in the years 2007-2009 (Ministry of Education and Culture, 2010a). According to respondents, the implementation of those appraisals had a rather technical emphasis, mostly involving an examination of whether formal requirements were fulfilled. In this context, the respondents pointed out that during the first years of the ministry's evaluation and supervision department, the supervision aspect tended to dominate, rather than a development and improvement-oriented approach. Both the appraisals revealed that the implementation of internal evaluation in compulsory schools were not as successful as hoped (Ministry of Education and Culture, 2010a). According to some respondents, the fact that the introduction of internal evaluation in compulsory schools had not been a success and that the ministry was only receiving limited information on school activities had *inter alia* opened the eyes of ministry staff to the need for altered priorities and enhanced external evaluation.

The groundwork for new legislation

The ministry of education remained under the control of the Independence Party until 2009. Tómas Ingi Olrich was in the position of minister 2002-2004 when Þorgerður Katrín Gunnarsdóttir took over the ministry and remained in charge until 2009 (Iceland Government Office, 2015). During this period, the Compulsory School Act from 1995 was revised. Soon after taking office, Þorgerður Katrín Gunnarsdóttir instigated a discourse on a future vision of quality issues in school activities (Task force on quality in school operations, 2005). The task force consisted of members of ministry staff and external specialists who were to act as advisors to the minister and work towards the ministry's comprehensive quality policy. The task force submitted proposals for altered legislation serving the purpose of ensuring that school administrators and executives assumed responsibility for implementing an evaluation of individual aspects of school operations and their level of success. On the other hand, the ministry of education would be responsible for conducting an external evaluation and establishing regulations with regard to quality control in schools. The task force proposed that the ministry published an official quality policy for schools with the aim of organising regular inspections at all school levels.

In 2006 a committee was appointed for the purpose of revising legislation on the compulsory school, consisting of representatives of the main stakeholders (Parliamentary document 319, 2007). According to respondents who participated in this work, the chapter on evaluation and supervision was not among the main subjects of discussion. Nevertheless, the committee proposed a draft copy of the chapter which the ministry of education rejected, and consequently further drafts were composed by ministry staff. Interviews with ministry staff revealed that during preparations for new compulsory school legislation representatives of the ministry differed in their opinions as to whether the government should be in charge of external evaluation of the compulsory school or whether it

should be the role of the municipalities. Some thought that the evaluation should, first and foremost, be the responsibility of the municipalities, based on the original ideology of the Compulsory School Act of 1995; that is, as executive operators the municipalities were responsible for organising the running of the compulsory school and consequently evaluation and supervision was their logical responsibility. In this context, one respondent said: "We kept trying to find a way towards municipal responsibility to get them to participate to a greater extent […] now the running of the compulsory schools had been transferred to municipalities so we could take the next step" (From an interview). Judging from narratives, there were always some people in the ministry who felt this was premature, many municipalities needed more time to adapt to changed circumstances and had neither the resources nor knowledge to take on the evaluation and supervision of school activities. The final outcome was that a bill introducing new legislation on compulsory schools contained provisions of external evaluation under the auspices of both municipalities and the ministry. There were indications in interviews that clear definitions were lacking as to the role of each party regarding evaluation and supervision by both the ministry and municipalities, although it was known that in the ministry people wanted to see a certain division of work and a stronger emphasis on evaluation for the purpose of school development and improvement.

It was revealed in an interview that at the same time as new legislation was being prepared formal and frequently reiterated complaints were becoming increasingly noticeable regarding problems in compulsory schools highlighting bullying, poor atmosphere in the school, communication problems and student facilities; that is, situations which were neither directly related to the financial operation of the schools nor to their curricula. Those complaints led to increased pressure for external evaluation, including aspects which had not been part of the dialogue on external evaluation of school operations. At this time the staff of the education sector of Reykjavík municipality had also begun to develop an evaluation of its compulsory schools; under the title of *comprehensive evaluation* for the purpose of supporting school activities and contributing to school improvement (Sigurjónsdóttir, 2010). Reykjavík municipality launched its comprehensive evaluation in 2007 and thus was the first municipality to adopt a regular and comprehensive appraisal of its compulsory schools. According to a respondent, this initiative by Reykjavík municipality also stimulated a further discussion in the ministry on the subject of external evaluation.

In a statement accompanying the parliamentary bill of new compulsory school legislation in 2007 there was a strong focus on evaluation and supervision, emphasising that one of the priorities of the parliamentary bill was "to strengthen the evaluation and supervision of school operations in order to support progress and improvement in the activities of the compulsory school" (Parliamentary document 319, 2007). A clear emphasis was laid on municipalities becoming increasingly responsible for the professional operations of the compulsory school, in addition to their financial and organisational responsibilities, and the ministry was under a stronger obligation than before to "monitor municipalities and compulsory schools by means of active evaluation and supervision" (Parliamentary document 319, 2007). It was specifically stated that the ministry's supervision was to include more aspects of school activities than hitherto, such as the welfare and well-being of students, precautions and preventive measures, work procedures in schools and the quality of education provided.

Municipalities and their associations reacted somewhat harshly to the newly added article of the parliamentary bill which focused on municipal evaluation and supervision as a quality control with regard to school operations. Comments on the parliamentary bill were critical of the reduced

emphasis on the ministry's external evaluation and the transfer of the evaluation obligation to municipalities (Akureyri municipality, 2008; University of Akureyri, 2008). The point was made that the provision regarding evaluation and supervision could result in a burdensome obligation for municipalities which would inevitably lead to additional expenses without the municipal coffers being compensated (Association of school office staff Grunnur, 2008; Icelandic Association of Local Authorities, 2008).

Evaluation and supervision in compulsory school legislation and its implementation 2008-2016

In 2008, new legislation on the compulsory school was enacted by Althing and this is the current law today. According to this legislation local authorities are to be responsible for the evaluation and quality control of school activities and provide the ministry of education with relevant information ("Compulsory school act," 2008). Furthermore, municipalities are to ensure that internal and external evaluations are used to bring about improvement in the schools. The ministry of education is to gather and analyse information from the municipalities and prepare a three-year plan on surveys and evaluations. Schools are to continue their own internal evaluations and municipalities shall monitor their implementation. The objectives of evaluations and quality control largely remain the same as before; that is, to support improvement, provide information, ensure compliance with laws and regulations, and guarantee that students' rights are respected ("Compulsory school act," 2008).

In 2009, a new government was formed in Iceland and Katrín Jakobsdóttir from the Left-Green Movement became the Minister of Education and Culture. She remained in office until 2013 and thus it was her responsibility to implement the legislation. The Ministry of Education and Culture presented a three-year plan regarding the external evaluation of compulsory schools covering the years 2010-2012. According to this plan, comprehensive evaluation of six compulsory schools were carried out *per annum* during the above-mentioned period (Ministry of Education and Culture, 2010b).

In 2009, the Icelandic Association of Local Authorities and the Ministry of Education and Culture, with support from the Municipalities' Equalisation Fund, appointed a joint project manager to prepare for the implementation of legal stipulations on municipal responsibilities regarding school evaluation and supervision (Icelandic Association of Local Authorities, e.d.). A steering group and a project management committee were established in this context, comprising representatives from the ministry and association. It was a high priority task of the project manager to define the supervisory and evaluative role of the local authorities for the purpose of facilitating the municipal duties of complying with the law. As part of this task, an *ad hoc* team was established in 2010, consisting of representatives of the ministry and the municipalities as well as the joint project manager (Ad hoc team on the external evaluation of compulsory schools, 2011). The ad hoc team was assigned to consider whether there was a basis for co-operation between the state and municipalities with regard to a regular external evaluation of compulsory schools and to propose a potential implementation of such co-operation. The ad hoc team presented its proposals in a report in 2011, recommending a shared undertaking by the state and municipalities to evaluate all Icelandic compulsory schools during a period of five years. The team anticipated the use of co-ordinated evaluation methods and that the schools would be evaluated on the basis of pre-determined quality criteria. The team proposed that the state and local authorities shared the financing and that the evaluation focused on defined key aspects of school improvement: Learning and teaching, leadership

and management, and internal evaluation. Within those three key aspects numerous issues were listed, all of which were to be included in the evaluation such as: teaching methods, learning behaviour, communication, success, class atmosphere, parental participation, co-operation between home and school, administrators' leadership skills, distribution of tasks, flow of information, improvement and development, and staff co-operation.

As may be gathered from the report by the ad hoc team various factors encouraged co-operation between the ministry and municipalities in the field of external evaluation. As far as municipalities were concerned, it was pointed out that not all of them had the resources to organise, implement and finance external evaluation on their own and thus the equality of compulsory schools regarding external evaluation could not be ensured. It was also pointed out that a co-operative arrangement would facilitate the supervisory role and information gathering of the ministry in preparation for further policy formation. Besides, this would help bring to fruition expectations in the ministry of external evaluation improving work and teaching methods in compulsory schools.

Interviews with ministry staff in the context of the research indicated that, apart from the above benefits, ministry staff saw opportunities in adopting a country-wide co-ordinated evaluation system; thus, it would be possible to ensure comparability between evaluation reports. In this connection it was pointed out that it had been a weak link in the ministry's evaluations that different agents conducted the evaluation and often this resulted in a mismatch between the evaluation reports. A number of respondents also said that at this time both school offices and the ministry took a special interest in the comprehensive evaluation conducted in compulsory schools in Reykjavík municipality, but neither the ministry nor municipalities felt they had the resources to conduct such an evaluation on their own.

An experimental and development project
In its report, the ad hoc team recommended a trial run of the proposal. The Municipal Equalisation Fund and the Ministry of Education and Culture agreed to finance the project and the determination of criteria and preparation for a comprehensive external evaluation took place in autumn 2011 (Ólafsdóttir, 2012). Six schools were accepted for a pilot phase of the project in spring 2012.

As a sequel to the pilot phase, a two-year development project was launched, 2013-2014, concerning external evaluation of compulsory schools which continued to build on proposals from the ad hoc team and the methodology developed in the pilot evaluation project (Icelandic Association of Local Authorities, 2013). In the beginning of 2013, an evaluation department was established at the Educational Testing Institute and two specialists appointed to take charge of the development project and its implementation (Educational Testing Institude, 2015). A project administration was set up, consisting of representatives from the ministry, municipalities and the Educational Testing Institute, to act advisors regarding the evaluation process (Jónsdóttir, 2015). However, the ad hoc team's proposals of evaluating 35 schools *per annum* did not materialise; in the first year eight compulsory schools were evaluated and ten in the second year. At the end of 2014 it was decided to extend the development project until the end of 2016 and evaluate ten compulsory schools *per annum* for two years.

Statements regarding the development project indicate that there is interest within the Association of Icelandic Local Authorities and the Municipal Equalisation Fund in increasing the annual number of schools undergoing external evaluation up to a maximum of fourteen schools each year *(Icelandic Association of Local Authorities, 2013; Jónsdóttir, 2015)*. In the Ministry of Education and Culture,

however, there are divided opinions as to the external evaluation and, according to respondents within the ministry, funding for this project is hard to find. Among reasons for this reluctance, one respondent mentioned that the result of the evaluation was not clear enough because of the small number of schools evaluated each year:

> Of course, many politicians are not particularly fond of such quality control systems and feel we are going too far in maintaining this kind of monitoring … and this may be the reason for limited financial contributions … there is also the consideration that politicians tend to provide money towards burning issues and the limited impact of compulsory school evaluation is not as obvious as that of various other issues placed higher on the agenda (*From an interview*).

During the ministership of Illugi Gunnarsson, who was Minister of Education and Culture from 2013 until 2016, there was, according to some respondents, increased emphasis on evaluation and measurement. Nevertheless, opinions were divided as to whether the methods used in the external evaluation were the right ones and whether it was advisable for the ministry and the municipalities to carry out this task jointly. It was brought up in an interview that the Association of Icelandic Local Authorities was taking on the largest share of this work.

A new administrative institution in the field of education was established by the "Act on the directorate of education" (2015). The Directorate was formally opened 1 October 2015. With this legislation, the Educational Testing Institute was abolished, and The Directorate of Education took over its responsibilities, including the external evaluation of school operations. According to the Act, one of the tasks of the Directorate is "to supervise and evaluate the results of school activities and compare them to established criteria" ("Act on the directorate of education," 2015). A report from the Directorate of Education, from December 2015, states that a policy is being formulated on the future organisation of external evaluation (Jónsdóttir, 2015).

Discussion

In 1991, when the ministry of education was controlled by a minister from the People's Alliance, an official education policy was published, emphasising devolution, in combination with centralisation, with clear objectives and evaluation and research in the field of school activities. Due to a change in government this policy, named *Towards a New Century*, was never formally adopted, but replaced by another policy which formed the basis of compulsory school legislation in 1995. The core of that policy was, nevertheless, an emphasis on devolution in the running and organisation of schools as well as increased acquisition of information and evaluation of school activities. The Independence Party controlled the ministry of education for eighteen years, from 1991-2009 and during this period governmental policies focused on the implementation of *new public management.* According to Hood (1991), the ideology of Western communities regarding new public management was characterised by devolution and an increased mandate and freedom for officials in charge of institutions, combined with a centralised setting of objectives and stricter supervision, placing increased emphasis on the responsibilities of organisational managers and holding them answerable for their actions. Those areas of emphasis were clearly expressed in the official policy of the Icelandic government during the period 1993-2007. Although the formal concept of new public management was introduced during the governing period of the Independence Party and came to be associated with that political affiliation, one may conclude with a view to sources on which this research is

based, that the influence of this ideology was beginning to make itself felt somewhat earlier, during a period of coalition governments consisting of more left-oriented parties. In this context, it may be pointed out that similar priorities were to be found in the policy *Towards a New Century* in 1991 as well as in proposals by a committee on the formation of an education policy in 1994 which contained points of emphasis which, *inter alia*, have been seen as characteristic of new public management (Hood, 1991; Koliba et al., 2010). Schick (2002) makes the point that sometimes a period of change has simply arrived, not least because of persuasive trends and currents of opinion in other countries. Respondents participating in this research unanimously agreed that the increased emphasis on the acquisition of information as well as evaluation and supervision in school activities cannot be traced to the political parties in power at the time but was brought to Iceland with people who were returning home from their studies and was also conveyed through Iceland's OECD membership. The fact that priorities relating to new public management characterise the education policies of parties from both wings of Icelandic politics, published in short succession, lends support to the respondents' conclusion referred to above.

According to Kingdon's (2014) theories, an issue is more likely to find its way into political dialogue, creating favourable circumstances for policy change, in case of, firstly, a well-defined problem which calls for an urgent solution, secondly, ideas as to how the problem may be solved, and thirdly, a favourable political climate. Although only with a limited backing in sound argument, it is insinuated in a report by the ministry of education from 2001 (Kjartansdóttir, 2001) and a report by the committee on the formation of an education policy (2004) that the committee's proposals favouring increased supervision, evaluation and information gathering have their origin, inter alia, in growing scepticism in governmental circles and among the general public as to the quality of school activities and an economical use of the budget allocated to education. This is in line with the dialogue in Western societies during this period when new public management was strengthening its position, cf. Moynihan (2008). However, none of the respondents participating in the research recognises that at that time there was a challenge in Icelandic compulsory schools or in the school system that people were aware of and wanted to react to, especially since there was so little information to be had on the situation in the compulsory schools. Nonetheless, it was suggested in official documents that a specific problem existed; this may have been based on information and people may have believed that there were similar problems in Iceland as existed abroad. Kingdon (2014) points out that difficulties need not be the only reason why matters are brought to the attention of government for sometimes new knowledge in the field in question or comparison with other countries becomes an incentive to change current ideas. In this context, Schick (2002) points out that entrepreneurs can often be highly influential by presenting ideas in the right place and at the right time. In the teams working on education policies published in 1991 and 1994 there were, according to respondents, individuals who strongly emphasised the importance of external evaluation and research with regard to school operations and, in this context, they could be regarded as entrepreneurs of a kind. Those people had recently returned from study abroad and had therefore been influenced by educational developments elsewhere in the world.

But it is not only when issues need to be placed on the agenda that it is important to demonstrate that a challenge exists which must be tackled. The same applies when policy changes are to be adopted. According to Kotter (2012) it is a matter of fundamental significance that people are aware of the importance of changes and see them as a solution to a problem. The support and dedication of those in highest authority is also crucial to the success of change for they are the people in charge of

legislation and the flow of resources to public institutions (Fernandez & Rainey, 2006; Kotter, 2012). The respondents involved in the research indicated that not everyone was aware of the importance of external evaluation and the problems it was meant to solve; as a matter of fact, it would appear that neither persons at government level, highest officials nor teachers were of the opinion at that time that there were problems in the schools or the school system that called for external evaluation. In addition, both individual teachers and their union resisted proposals of external evaluation. Consequently, the authorities were not prepared to embark on the comprehensive work and expense required by its implementation. Soon after the latter education policy was completed, another change of government ensued and a new minister, Björn Bjarnason took over the education ministry. He did not support the idea that the ministry should take charge of external evaluation of compulsory school activities and allocated but scant resources to its implementation. Although an evaluative and supervisory department had by that time been established in the ministry, its main focus was on the external evaluation of secondary schools and universities. Thus, it may be said that during the ministership of Björn Bjarnason the external evaluation of compulsory schools was taken off the agenda. The near-total lack of interest of the educational authorities in the external evaluation of compulsory schools after 1995 becomes strikingly obvious when the general curriculum guide for compulsory schools from 1999 is examined. The guide states that the obligation of the ministry regarding external evaluation/reviews is confined to their methods of self-evaluation. The ministry is authorised, but not obliged, to implement any further reviews or appraisals. Here the vague provisions of compulsory school legislation enabled the authorities to keep to an absolute minimum external evaluation/reviews of the activities of compulsory school. The resulting outcome may be gleaned from the fact that during a period of twelve years, 1996-2008, external evaluation of compulsory school activities was conducted in only a handful of schools – and in those cases almost exclusively because obvious problems in the schools concerned made such an evaluation necessary. The above reflects the words of warning by Fernandez and Rainey (2006) when they point out that a vague policy can provide public administrators with the opportunity for arbitrary interpretation and thus enable them to carry it through in such a way that the original objectives are only implemented in part.

Performance management with goal-oriented measurements comprised one of the key elements of new public management in order to ensure responsibility and contribute to reform (Bevan & Hood, 2006; Hood, 1991; Ministry of Finance, 1993). Although external evaluation of compulsory school operations did not gain a foothold around the turn of the past century, the Icelandic school system certainly experienced a heightened emphasis on increased internal performance management in schools in the debate on policy formation. Education policy from 1994, legislation on compulsory school from 1995 and the general curriculum guide for the compulsory school in 1999 all showed clear signs of a strong emphasis on the self-evaluation of schools. From the enactment of legislation on compulsory schools in 1995 and until a new Act on the compulsory school took effect in 2008, reviews by the Ministry of Education and Culture were almost exclusively focused on assessing methods of self-evaluation in compulsory schools (Ministry of Education and Culture, 2010a). The ministry's reviews of school self-evaluations were, however, more focused on checking whether formal evaluation requirements were being satisfied than on an improvement-oriented approach. An evaluation and supervision department which had been set up in the ministry turned out to be incapable of fulfilling its role because of lack of funds and human resources for evaluation and supervision. In this context, it seems apt to refer to Kotter (2012) who remarked that the consequence of people not realising the existence of a problem is that change cannot be initiated

and then projects such as quality management, supervision or evaluation take on the form of superficial bureaucracy rather than real tasks requiring attention.

It is clear that the notions initiated in public education policies in 1991 and 1994 to the effect of initiating regular external evaluation of compulsory school operations, combined with self-evaluation did not materialise during the period 1996-2008. It may be relevant here to bring up the concern expressed by Schick (2002), who pointed out that in many cases institutions had been so weakened through decentralisation that they no longer had the capacity to implement requisite co-ordination. It would appear the Shick´s warning directly applies to the relationship between the Ministry of Education and Culture, on the one hand, and municipalities and compulsory schools, on the other. In run up to new legislation on compulsory schools, enacted in 2008, the chapter on evaluation and supervision was reviewed and new priorities established, adding increased supervision and inspections by the Ministry of Education and Culture and stressing the obligation of municipalities to take charge of the evaluation and supervision of school activities. Those changes were called for, *inter alia*, because it was considered necessary to place increased responsibilities on municipalities as the implementation of internal evaluation/self-evaluation in the compulsory school had not been a success and the ministry was facing a growing number of complaints regarding problematic school situations. Thus, when preparing new legislation in 2008 people had more information at their disposal on the circumstances of the compulsory school than had been available in 1995. With a view to Kingdon´s (2014) theory one might say that by now the real picture of the current problems was clearer and it looked as if external evaluation of school activities might be part of the solution; besides, the political atmosphere for change was favourable as the ministry of education had a new minister. Consequently, external evaluation was back on the agenda and in the wake of legislation the ministry launched external evaluations of compulsory schools, albeit only on a small scale.

Many municipalities were, however, ill-prepared to take on the new role assigned to them with the legislation in 2008 and this created a problem. In addition, the role of the Ministry of Education and Culture, on the one hand, and that of the municipalities, on the other, was not clearly defined in law, regarding external evaluation and supervision of compulsory schools. A project steering group was appointed to clarify the division of tasks between the ministry and the municipalities with regard to the external evaluation of compulsory school. A project manager was nominated, and an ad hoc team established to formulate proposals regarding the implementation of the external evaluation. Subsequently a joint external evaluation was launched by the ministry and the municipalities, replacing the ministry's inspections and the number of schools evaluated each year was increased. The administration of the implementation of the evaluation was transferred from the ministry to the Educational Testing Institute. Kotter (2012) prioritises that change processes must have the support and dedication of top-level administrators. Thus, it seems reasonable to conclude that in this case the comprehensive steering network set up by the ministry and the municipalities with regard to the formation and implementation of the joint project of external evaluation was a key element in its success. But why does obtaining funds for external evaluation remain problematic and why is the future of the project uncertain after the pilot phase has been completed? (Kotter, 2012) pointed out that a clear future vision and policy as well as well-defined methods to accomplish your aims were key elements in the implementation of change. He underlined that the vision should be constantly mediated and improved. With Kotter´s theories in mind, one might have doubts as to whether the project vision was clear enough or sufficiently promoted to create consensus. We should also keep in mind OECD´s (2013) explanations of challenges various countries have faced in the implementation

of evaluation and, in that context, wonder whether the policy was well enough prepared, whether an evaluation culture is lacking in the Icelandic school system or whether evaluation results have not been fully utilised.

Conclusion

The text above has outlined the introduction and development of external evaluation in Iceland, on the basis of questions posed at its beginning and with a view to theories and academic sources presented in the article.

The questions are:

> *(1) How did external evaluation of compulsory schools in Iceland come about and what*
>
> *explains its origin?*
>
> *(2) How has external evaluation of compulsory schools developed up to the present time and*
>
> *what explains its development?*

Below, an attempt will be made to summarise the answers to those questions.

The empowerment of municipalities via the transfer of compulsory school operations to them, in combination with increased centralisation by way of clear objectives and emphasis on external evaluation and supervision were the key aspects of the changes occurring in the Icelandic school system in the 1990s. All of those aspects were interwoven with the fundamental concepts of new public management and associated trends in various parts of the world. The increased independence of the schools with the aim of attaining a more successful, high quality school system, in accordance with the ideology of new public management, required, in the opinion of the government, both enhanced quality management and strengthened external evaluation and supervision of school operations. In Iceland, a form of external supervision was developed which, on the basis of defined criteria, was given the role of investigating whether schools were evaluating their own quality through internal evaluation and whether relevant formal requirements were satisfied. This supervision was supposed to provide the schools with sufficient discipline and encourage them to focus on internal evaluation and improvement, thus ensuring a quality school system. External evaluation, on the other hand, comprising a comprehensive quality control of school operations, did not gain a foothold at this time in spite of being included in plans of a public education policy. The fundamental obstacle appears, *inter alia*, to have been lack of conviction on behalf of the authorities and other stakeholders as to the importance of such evaluation and its potential for solving an existing problem in the school system – especially since it was obvious that establishing external evaluation applicable to all compulsory schools in Iceland would be a hugely costly undertaking and, besides, Iceland lacked the specialist human resources to successfully implement such evaluations.

Thus, it was not until the beginning of the 2010s that external evaluation was launched in Icelandic compulsory schools in the sense already outlined in this coverage – external evaluation emphasising the review of key processes in school activities for the purpose of stimulating school improvement. It was clear by then that the attempt to persuade the compulsory schools to perform internal evaluation had not been a particular success and this evaluation was, therefore, not a viable sole basis for quality school management. At this point, it was considered necessary to strengthen municipal responsibility for school operations and thus, through compulsory school legislation in

2008, their obligation to conduct evaluation and supervision was given a legal basis. Various challenges were becoming increasingly problematic in the school system, such as bullying, poor school atmosphere, communication problems and inadequate facilities. In 2010 the Ministry of Education and Culture launched regular inspections of six compulsory schools *per annum*, an arrangement which lasted for three years. Lack of clarity regarding division of work between the government and the municipalities in the implementation of external evaluation, as stipulated in compulsory school legislation 2008, was solved by establishing a joint project between those two parties. In the beginning of 2013 external evaluation was initiated, co-financed by the government and the municipalities. The evaluation is based on quality criteria regarding school operations, covering numerous aspects of compulsory school activities. Simultaneously the overseeing of external evaluation was transferred from the ministry to the Educational Testing Institute which later became Directorate of Education. The number of schools participating in the evaluation was ten *per annum*. The external evaluation was implemented as a two-year pilot project and then extended by another two years until the end of 2016. This pilot project will be in its final year when this is written, and its further development is uncertain.

When recording the story of external evaluation in compulsory schools the factors impacting its beginning and further development have been identified. It is my hope that that the article will provide a historical insight which could be of use in further policy formation and decision-making as to the future of the external evaluation.

References

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

Ad hoc team on the external evaluation of compulsory schools. (2011). *Ytra mat á grunnskólum: Tillögur til mennta- og menningarmálaráðherra og skólamálanefndar Sambands íslenskra sveitarfélaga um tilhögun á ytra mati á grunnskólum [External evaluation of compulsory schools: Proposals to the minister of education and culture and the Icelandic association of local authorities regarding arrangements for external evaluation of compulsory schools]*. https://www.stjornarradid.is/gogn/rit-og-skyrslur/stakt-rit/?NewsId=5d478635-1dd9-11e7-9c14-e4a7a00e173e

Akureyri municipality. (2008). *Umsögn um frumvarp til laga um grunnskóla [Commentary on the draft law on compulsory schools]*. http://www.althingi.is/pdf/erindi/?lthing=135&dbnr=1155

Association of school office staff Grunnur. (2008). *Umsögn um frumvarp til laga um grunnskóla [Commentary on the draft law on elementary schools]*. http://www.althingi.is/pdf/erindi/?lthing=135&dbnr=1220

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, *84*(3), 517-538.

Bjarnason, B. (1995, 21. nóvember). *Nýskipan í ríkisrekstri - menntun við aldahvörf [New public management in state administration - education at the turn of the century]* Ráðstefna um nýskipan í ríkisrekstri Reykjavík. http://www.bjorn.is/greinar/nr/579

Bjarnason, B. (1996, 1. nóvember). *Aðalfundur Skólastjórafélags Íslands [General meeting of the School Principals' Association of Iceland]* Aðalfundur Skólastjórafélags Íslands, Reykjavík. http://www.bjorn.is/greinar/nr/183

Committee on the formation of an education policy. Report [Nefnd um mótun menntastefnu. Skýrsla]. (1994). https://www.stjornarradid.is/gogn/rit-og-skyrslur/stakt-rit/1994/06/01/Nefnd-um-motun-menntastefnu-skyrsla/

Compulsory school act, 63/1974 (1974). http://www.althingi.is/leit/?q=L%C3%B6g+nr.+63%2F1974

Compulsory school act, 49/1991 (1991). http://www.althingi.is/altext/113/s/1128.html

Compulsory school act, 66/1995 (1995). http://www.althingi.is/altext/118/s/0951.html

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

Educational Testing Institute. (2015). *Ytra mat grunnskóla [External evaluation on compulsory schools]*. Educational Testing Institute. Retrieved 21. september 2015 from http://namsmat.is/vefur/ytra_mat_skola/ytra_mat.html

Einarsson, Ó. G. (1994). *Flutningsræða: Frumvarp til laga um grunnskóla [Introductory speech: Draft law on compulsory schools]*. Reykjavík: Þingtíðindi Retrieved from http://www.althingi.is/altext/118/11/r01133403.sgml

Eurydice. (2007). *School autonomy in Europe policies and measures*.

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

Fernandez, S., & Rainey, H. G. (2006). Managing successful organizational change in the public sector. *Public Administration Review*, 168-176.

Hansen, B., & Jóhansson, Ó. H. (2010). *Allt í öllu. Hlutverk fræðslustjóra 1975-1996 [Everything in everything. The role of the director of education 1975-1996]*. Háskólaútgáfan.

Hogwood, B. W., & Gunn, L. A. (1984). *Policy analysis for the real world*. Oxford University Press.

Hood, C. (1991). A public management for all seasons? *Public Administration*, *69*, 3–19.

Iceland Government Office. (2015). *Ríkisstjórnartal [Overview of Governments]* Retrieved 20. september 2015 from http://stjornarrad.is/Rikisstjornartal/

Icelandic Association of Local Authorities. (2008). *Umsögn Sambands íslenskra sveitarfélaga um frumvarp til laga um grunnskóla [Commentary of the Icelandic association of local authorities*

on the draft law on compulsory schools]. http://www.althingi.is/pdf/erindi/?lthing=135&dbnr=1287

Icelandic Association of Local Authorities. (2013). *Ytra mat á grunnskólum: Þróunarverkefni 2013-2014 [External evaluation of compulsory schools: Development project 2013-2014]*. http://www.samband.is/media/skolamal/Ytra-mat-2013.pdf

Icelandic Association of Local Authorities. (e.d.). *Verkefnisstjóri í skólamálum [Project manager in school matters]*. Icelandic Association of Local Authorities,. Retrieved 18. september 2015 from http://www.samband.is/verkefnin/skolamal/stefnumotun/verkefnisstjori-i-skolamalum/

Jónsdóttir, Þ. B. (2015). *Ytra mat á grunnskólastarfi. Notkun viðmiða um gæði skólastarfs við ytra mat á grunnskólum 2013-2015. Skýrsla til Jöfnunarsjóðs og stýrihóps [External evaluation on compulsory schools. The use of quality criteria for the external evaluation of compulsory schools 2013-2015. Report for the Equalization Fund and Steering Committee]*.

Jónsson, Ó. P. (2014). Lýðræðisleg menntastefna: Sögulegt ágrip og heimspekileg greining [Democratic education: A historical overview and philosophical analysis]. *Stjórnmál og stjórnsýsla*, *10*(1), 99-118. http://www.irpa.is/article/view/1324/pdf_322

Kingdon, J. W. (2014). *Agendas, alternatives and public policies* (2. ed.). Pearson Education Limited.

Kjartansdóttir, Á. (2001). *Approaches to the evaluation of schools which provide compulsory education. Iceland*. https://op.europa.eu/en/publication-detail/-/publication/6c980d8d-3d0d-4264-a811-17498bfd434c

Koliba, C., Meek, J. W., & Zia, A. (2010). *Governance networks in public administration and public policy*. CRC Press.

Kotter, J. (2012). *Leading change*. Harvard Business Review Press.

Kristinsson, G. H. (2007). *Íslenska stjórnkerfið [The Icelandic government system]* (2. ed.). Háskóli Íslands.

Lichtman, M. (2006). *Qualitative research in education. A user's guide*. Sage Publications, Ltd.

Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on the Web. *Philosophical Transactions of the Royal Society*, *371*(1987). Retrieved 28. maí 2013, from http://rsta.royalsocietypublishing.org/content/371/1987/20120382.abstract

Ministry of Education. (1991). *Til nýrrar aldar. Framkvæmdaráætlun Menntamálaráðuneytisins í skólamálum til ársins 2000 [Towards a new century. The education ministry's school implementation plan until the year 2000]*.

Ministry of Education. (1997). *Sjálfsmat skóla. Leikskólar, grunnskólar og framhaldsskólar [School self-evaluation. Kindergartens, primary schools and secondary schools]* https://www.stjornarradid.is/gogn/rit-og-skyrslur/stakt-rit/1997/06/01/Sjalfsmat-skola-leikskolar-grunnskolar-og-framhaldsskolar/

Ministry of Education and Culture. (2010a). *Heildarniðurstöður úttekta á sjálfsmatsaðferðum grunnskóla 2007-2009. Samanburður á niðurstöðum 2001-2003 og 2007-2009 [Overall results of reviews of self-evaluation methods in compulsory schools 2007-2009. Comparison of the results of 2001-2003 and 2007-2009]*. https://www.stjornarradid.is/efst-a-baugi/frettir/stok-frett/2010/05/03/Heildarnidurstodur-uttekta-a-sjalfsmatsadferdum-grunnskola-2007-2009-samanburdur-a-nidurstodum-2001-2003-og-2007-2009-Eingongu-a-rafraenu-formi/

Ministry of Education and Culture. (2010b). *Mat á skólastarfi: Áætlun um ytra mat í framhaldsskólum, grunnskólum og leikskólum 2010-2012 [Evaluation of school activities: Plan for external evaluation in secondary schools, compulsory schools and pre-primary schools 2010-2012*. https://www.stjornarradid.is/gogn/rit-og-skyrslur/stakt-rit/?NewsId=4f495bc4-1dd8-11e7-9c14-e4a7a00e173e

Ministry of Education and Culture. (n.d.). *Útgáfuskrá [Publication directory]*. Retrieved 3. september 2015 from http://www.menntamalaraduneyti.is/utgafuskra/

Ministry of Finance. (1993). *Umbætur og nýskipan í ríkisrekstri [Improvements and new public management]*
http://www.fjarmalaraduneyti.is/media/Utgefin_rit/Umbaetur_og_nyskipan_i_rikisrekstri_1993.pdf

Ministry of Finance. (1996). *Stefna um nýskipan í ríkisrekstri [Policy on new new public management]*
http://www.fjarmalaraduneyti.is/media/Utgefin_rit/Stefna_um_nyskipan_i_rikisrekstri_1996.pdf

Ministry of Finance. (2000). *Ábyrgð, valdsvið og stjórnunarumboð forstöðumanna ríkisstofnana: Nefndarálit [Responsibilities, powers and administrative authority of heads of government agencies: Committee opinion]* https://www.fjarmalaraduneyti.is/media/rit2014/Abyrgd,-valdsvid-og-stjornunarumbod.pdf

Ministry of Finance and Economic Affairs. (2013). *CAF 2013. Umbætur í opinberri þjónustu með sjálfsmati [CAF 2013. Public service reform through self-evaluation]*
https://www.fjarmalaraduneyti.is/media/utgafa/CAF_2013.pdf

Moynihan, D. P. (2008). *The dynamics of performance management. Constructing information and reform*. Georgetown University Press.

National curriculum guide for compulsory schools, (1976).

National curriculum guide for compulsory schools, (1989).

National curriculum guide for compulsory schools. (1999). Ministry of Education.

Organisation for Economic Cooperation and Development [OECD]. (1997). *In search of results: Performance Management Practices*.

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

Ólafsdóttir, B. (2012). Tilraunaverkefni um ytra mat á grunnskólastarfi [Pilot project on external evaluation on compulsory schools]. *Sveitarstjórnarmál*, *1*(73), 8-9.

Parliamentary document 131. (1994). *Frumvarp til laga um grunnskóla [Explanatory momorandum accompanying a parliamentary bill on the compulsory school]*.
http://www.althingi.is/altext/118/s/0131.html

Parliamentary document 192. (1990). *Frumvarp til laga um grunnskóla [Explanatory momorandum accompanying a parliamentary bill on the compulsory school]*.
http://www.althingi.is/altext/113/s/0192.html

Parliamentary document 319. (2007). *Frumvarp til laga um grunnskóla [Explanatory momorandum accompanying a parliamentary bill on the compulsory school]*.
http://www.althingi.is/altext/135/s/0319.html

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3. ed.). Sage.

Schick, A. (2002). Opportunity, strategy and tactics in reforming public management. *OECD Journal on Budgeting*, *2*(3), 7–34.

Sigurjónsdóttir, B. (2010). Heildarmat í grunnskólum Reykjavíkur 2007-2010 [Comprehensive evaluation in compulsory schools in Reykjavík 2007-2010]. *Netla - Veftímarit um uppeldi og menntun*.
https://skemman.is/bitstream/1946/13797/1/Heildarmat%20%c3%ad%20grunnsk%c3%b3lum%20Reykjav%c3%adkur%202007%e2%80%932010.pdf

Simkins, T. (2003). Reform, accountability and strategic choice in education. In M. G. Preedy, Ron; Wise, Christine (Ed.), *Strategic leadership and educational improvement*. The Open University og Sage Publications.

Task force on quality in school operations. (2005). *Gæðamál í skólastarfi: Tillögur til menntamálaráðherra [School quality: Proposals for the Minister of Education]*.

University of Akureyri. (2008). *Umsögn um frumvarp til laga um grunnskóla, 285. mál [Commentary on the draft law on compulsory schools, Case 285]*.
http://www.althingi.is/pdf/erindi/?lthing=135&dbnr=1259

Yin, R. K. (2014). *Case study research: Design and methods* (5. ed.). SAGE Publications Ltd.

Appendix. Interview guide about origin and development of external evaluation in Iceland

Introduction:

- Introduce the purpose of the interview (get an insight into the origin, the implementation process, and the purpose of the external school evaluation)—describe the handling of data and how the interviewees will be cited when reporting—anonymity—recording and approval—transcription and confidentiality—length of the interview.
- I am seeking facts about the theme, some of which are accessible in public records to some extent, while others are not. I am also looking for your opinions and views about the topic.
- Get a background: how the interviewer has been linked to external evaluation policy-making or worked with external evaluation during the time and what role he or she has had in this context.

Origin and development of external school evaluation on compulsory schools

To remind:

A new law regarding compulsory schools was introduced in 1991—superintendents and their department of education were, under the law, required to exercise a larger role than before regarding monitoring schools. In the same year, a new school policy was published: *To the new century. Ministry action plan for education until the year 2000*. Among its main priorities were, on the one hand, democracy and distribution and on the other hand, regular evaluation and monitoring of school's work. Did you participate in this policy setting?

A year after the publication of the action plan for education (in the year 1992), a new minister of education appointed a committee focusing on a new education policy; it was charged with the task of proposing reforms in education. The committee proposed regular external school evaluation and the implementation of self-evaluation in all compulsory schools, stipulating that they should be evaluated on regular basis every five years. Did you participate in this policy setting?

Questions:

Can you tell me how external school evaluation came about in Iceland, what explains its origin? (Does it come from politics, government, school or municipalities, teachers or parents?)

Was there any problem in the education system that required external evaluation?
- What was the problem? What problems were the external evaluation supposed to solve?
- How was the problem defined?
- How did the external evaluation come about as a solution to the problem? / What was the solution?

What were the political conditions at that time? What were the trends/political streams internationally at this time? How was the public opinion?

Compulsory school act 1995 and the national curriculum for compulsory education 1999—implementation of the evaluation

The education policy set in 1994 addressed both external school evaluation, carried out by the ministry, and school self-evaluation, which the ministry was supposed to monitor.

- Was the development of the external evaluation as planned in the educational policy; how did the implementation of the external evaluation happen?
- What about the ministry supervision of the school self-evaluation? Was the execution regarding that as planned?

Why did almost no external evaluation of compulsory schools take place on behalf of the ministry until after 2008? What obstacles were there? How were the schools, that received external evaluations between 1995 and 2008 chosen for evaluation?

Over the years (after 1995), what was the opinion of the education ministers regarding the external evaluation? Did they see it as important? Did they provide some resources to support it—finance, for example?

The Teacher University Research Centre undertook some evaluation projects for municipalities. Did the municipalities often request that the centre evaluate their school? Why did the municipalities request an evaluation of their school? Was it because there were some problems or...?

Project group on quality issues in schools—2005

The project group on quality issues in schools submitted proposals to the minister of education in 2005. Why was this group formed?

The group recommended that a regular audit should be undertaken at all school levels. What was the idea behind that? Why?

Was there some problem that the external evaluation was supposed to solve? A new compulsory law was introduced in 2008. Did that reflect the changes that the group proposed?

In preparation of the legislation in 2008

A project group prepared new compulsory school legislation that became law in 2008. Can you tell me how the section on evaluation and quality assurance in compulsory schools (section VIII) was processed? Was there a consensus in the project group?

In the compulsory school act of 2008, changes were made regarding external school evaluation—to the effect that both the ministry and municipalities should evaluate compulsory schools. Why were these changes made? Was there any problem requiring these changes? What was the problem? How were the changes supposed to solve that problem?

Did everyone agree on this new emphasis on external evaluation, or were there some who resisted them? Where did the opposition come from? Why? Did the municipalities agree with the novelty that they should evaluate their schools?

Was the minister (at that time) interested in external evaluation? Did he emphasise these changes?

After the law took effect in 2008

How were the law and its provisions developed after they were introduced in 2008? How did the implementation of the law's mandate regarding external evaluation go? Did it go as planned?

Were there some obstacles/barriers regarding the implementation of the external evaluation? Which?

The purpose of the external evaluation and its impact

What do you think the government expectations for the external evaluation and its impact on school activity are? In your opinion, is the external evaluation a realistic way to achieve that goal?

Do you think that the approach used in the external evaluation is likely to foster the intended result? Why? Why not?

In your opinion, is there a benefit from the external school evaluation? Do you think the external evaluation is worth the effort and cost?

Regarding the external evaluation, what has succeeded best in your opinion? What has failed?

In your opinion, what policy should be taken regarding the external evaluation? Is it evaluating what matters?

How are the external evaluations results utilized within the ministry and/or by local authorities? Systematically? For what purpose are the results utilised? Have decisions been made based on the information obtained through the external evaluation? What decisions?

I have no further questions. Is there something you want to mention before we finish the interview?

Thank you for your participation.

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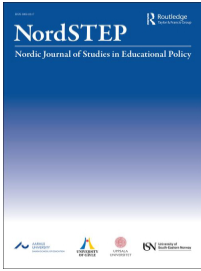
The mechanisms by which external school evaluation in Iceland influences internal evaluation and school professionals' practices

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

Published online: 19 May 2022.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

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

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

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

ABSTRACT

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

Introduction

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

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

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

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

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

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

The Icelandic context and process of external school evaluation

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

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

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

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

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

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

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

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

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

Theoretical framework and previous research

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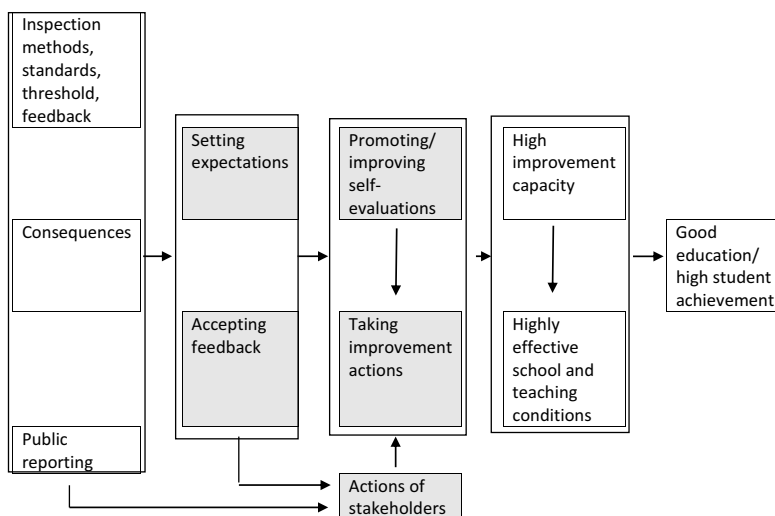

Figure 1. Framework of causal mechanisms of school inspection (Ehren et al., 2013, p. 14, with permission).

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

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

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

Setting expectations

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

'High-stake' accountability systems are among other factors based on the assumption that rational actors strive to meet the standards because of the threat of sanctions if they do not (Ehren, Gustafsson et al., 2015). In this way, a high level of external pressure increases the likelihood that schools will make use of the standards. The studies by Ehren et al. (2015) and Altrichter and Kemethofer (2015), based on the survey responses of principals in the above-mentioned European countries, largely confirm this. They found that principals who felt more 'accountability pressure' were more alert to the expectations issued by the inspectorate; however, unintended consequences also increased with greater pressure (Altrichter & Kemethofer, 2015; Jones et al., 2017). Such unintended side effects are, for example, strategic behaviour (intended or unintended) on behalf of the school, such as narrowing the curriculum, misrepresenting data, teaching to the test and discouraging teachers from experimenting in teaching (Ehren, Gustafsson et al., 2015; Ehren, Jones et al., 2016; Jones et al., 2017). In their study of Dutch schools, where accountability pressure is high (Altrichter & Kemethofer, 2015), Ehren, Perryman et al. (2015) revealed that setting clear expectations is a strong determinant of improvements and changes made in the capacity of the school. Yet a comparative study by Kemethofer et al. (2017) between Austria and Sweden, where Austria represents a 'low-stake' system and Sweden a 'medium- to high-stake' system (Altrichter & Kemethofer, 2015), showed a significant effect of *setting expectations* in Austria but not in Sweden and, thus, does not comply with the assumption of 'high-stake' systems.

Accepting feedback

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

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

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

Stakeholder involvement

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

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

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

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

Capacity to implement improvements

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

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

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

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

Research focus and questions

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

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

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

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

RQ2 – On the basis of Ehren et al. (2013), are the mechanisms *accepting feedback, setting expectations,* *stakeholder involvement*, and *capacity to implement improvements* positively correlated with *changes in internal evaluation, teaching* and *leadership practices?*

Method

Sampling and data collection

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

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

Instrument

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

Context

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

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

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

Data analysis

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

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

Table 1 shows that 10 of the 11 proposed scales are reliable according to Cronbach's alpha's consensus threshold ($\alpha \geq 0.7$) (Hinton, 2014). The scale of the variable *acceptance of evaluation feedback* in the principal's survey has a marginal reliability coefficient ($\alpha = 0.66$). However, as the average inter-item correlation is acceptable, we decided to keep the scale in the analyses to make the teacher and principal scales comparable; as can be inferred, the small number of respondents in the principals' survey is the reason. All variables in the study are latent variables except the two variables *change in internal evaluation* and *change in leadership practices*, each of which consists of one item (see Appendix 1, Q15 and Q20). Descriptive statistics (mean and SD) were calculated for all measures.

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

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

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

assumption of normally distributed residuals was satisfactory.

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

Findings

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Suggestions for policies, practices, and future research

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

capacity and students' educational outcomes remains unexplored, thereby calling for further study. Lastly, future research should aim to be more independent of the system and should also include information from other groups participating in the process, as the model implies.

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

Disclosure statement

ORCID

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

References

Act on the directorate of education, (91/2015)
Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, *26*(1), 32–56. https://doi.org/10.1080/09243453.2014.927369
Behnke, K., & Steins, G. (2017). Principals' reactions to feedback received by school inspection: A longitudinal study. *Journal of Educational Change*, *18*(1), 77–106. https://doi.org/10.1007/s10833-016-9275-7
Compulsory School Act, (2008). https://www.government.is/media/menntamalaraduneyti-media/media/law-and-regulations/Compulsory-School-Act-No.-91-2008.pdf
Dedering, K., & Müller, S. (2011). School improvement through inspections? First empirical insights from Germany. *Journal of Educational Change*, *12*(3), 301–322. https://doi.org/10.1007/s10833-010-9151-9
Ehren, M. C. M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools: Describing assumptions on causal mechanisms in six European countries. *Educational Assessment, Evaluation and Accountability*, *25*(1), 3–43. https://doi.org/10.1007/s11092-012-9156-4

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Paper III

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*

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

---

 [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.

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

---

[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.

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

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.

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.

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*)

	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*)

	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*)

	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 word*. 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 inspections. *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-ranns oknir-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

Appendix A: Form for the informed consent to participate

Dear Principal/Teacher,

I am a doctoral student in the Faculty of Education and Pedagogy in the School of Education at the University of Iceland. I am conducting research that is part of my final project to earn a doctoral degree, and I would like to interview you. I am therefore requesting your consent to participate in the research. I can assure you full anonymity as a participant.

The aim of the research is to illuminate how the results of external evaluation are used in schools and what significance the feedback from evaluation has for the school staff, namely teachers and principals. The research focuses on three major aspects considered during external evaluation: leadership and management, teaching and learning, and internal evaluation.

Participants are being selected from the group of principals and teachers at 22 schools who answered a survey in the spring of 2016 about the external evaluation of compulsory schools. Principals and teachers in six schools, chosen based on size, location, and the results of external evaluation, have been selected to participate in follow-up interviews.

I hope that the research and others similar to it can provide practical information for education authorities to use in the further development of external evaluation in Iceland and follow-up on such evaluation. However, to date, no research has been conducted on the use of the feedback from external evaluation nor on the perceived benefits of external evaluation. It is therefore important to gather such information to use to further develop and support Iceland's schools.

My supervisors in the doctoral program are Jón Torfi Jónasson, professor emeritus at the School of Education, University of Iceland, and Anna Kristín Sigurðardóttir, professor at the School of Education, University of Iceland.

In light of the research's aim and potential contributions to knowledge, I hereby request your participation in an interview about how the results of external evaluation have been used in your school and the impact, if any, of the feedback from evaluation. The interview is expected to last about an hour.

Your absolute anonymity is guaranteed, and you will be free to stop participating at any time and for any or no reason. All data gathered in the research, including the recordings and transcripts of interviews, will be kept in a password-protected file on my computer, and the recorded interviews will be destroyed after the research is completed.

Thank you for your consideration; I look forward to hearing from you.

Respectfully,

Björk Ólafsdóttir

bjo13@hi.is

Appendix B: Cover letters and questionnaires

Cover letter for principals

Dear Recipient,

Attached is a survey asking questions for a development project, conducted since 2013, addressing the external evaluation of Iceland's compulsory schools. The project will end in late 2016, and the survey is being performed to assess the project and its effectiveness. The survey is being conducted on the behalf of the Directorate of Education, the Ministry of Education and Children and the Icelandic Association of Local Authorities.

The aim of assessing the development project is to gauge how the findings from external evaluation are used in schools and whether such use affects the internal evaluation of schools, their teaching practices, and their management. The results from the survey will be used to review and improve the process of external evaluation and to support further decisions about its continuation. To some extent, the survey data will be used in my doctoral research, which aims to illuminate how the findings of external evaluation are used in schools, how schools are supported or stimulated by local authorities, and to what extent the findings of external evaluation promote internal school evaluation, teaching, and leadership practices.

I kindly ask you to answer all of the following questions, which should take approximately 10—15 minutes. Because you will provide answers as a representative of your school, the answers can be traced to you and are not anonymous. I would therefore like to reiterate that the confidentiality of participants will be carefully protected during the processing of data and in reporting the results. The data will not be part of any database of the mentioned institutions; instead, they will be stored securely by the researcher in a form that prevents individuals' identification. The preservation of and access to the data will follow the provisions of the Personal Data Protection and Privacy Policy (No. 77/2000)

I realise that you are likely quite busy, but your answers are invaluable for the quality of the research's results and for the opportunities that they provide to review and improve the process of externally evaluating schools in Iceland.

Additional information will be provided at the beginning of the survey. Click the link below to answer the survey. If the URL is not active, then you may need to copy and paste it into your internet browser.

If you have any questions or concerns, then please feel free to contact me by email (bjo13@hi.is) or by phone (898-6348).

Thank you in advance,

Björk Ólafsdóttir

Cover letter for teachers

Dear Recipient,

Attached is a survey of questions concerning a development project, conducted since 2013, whose aim is to address the external evaluation of Iceland's compulsory schools. The project will end in late 2016, and the survey is being performed to assess the project and its effectiveness. The survey is being conducted on the behalf of the Directorate of Education, the Ministry of Education and Children, and the Icelandic Association of Local Authorities.

The aim of assessing the development project is to gauge how the findings from external evaluation are used in schools and whether such use affects the internal evaluation of schools, their teaching practices, and their management. The results from the survey will be used to review and improve the process of external evaluation and support further decisions about its continuation. To some extent, the survey data will be used in my doctoral research, which aims to illuminate how the findings of external evaluation are used in schools, how schools are supported or stimulated by local authorities, and to what extent the findings of external evaluation promote internal school evaluation, teaching, and leadership practices.

I kindly ask you to answer all questions on the survey, which will take approximately 8— 12 minutes. Answers cannot be traced to certain individuals and are therefore entirely anonymous.

Even if you did not work at the school when the external evaluation in question was conducted, I would still like you to answer a few questions, which should take only 3—4 minutes.

I realise that you are likely quite busy, but your answers are invaluable for the quality of the research's results and for the opportunities that they provide to review and improve the process of externally evaluating schools in Iceland.

Additional information will be provided at the beginning of the survey. Click the link below to answer the survey. If the URL is not active, then you may need to copy and paste it into your internet browser.

If you have any questions or concerns, then please feel free to contact me by email (bjo13@hi.is) or by phone (898-6348).

Thank you in advance,

Björk Ólafsdóttir

Information letter at the beginning of the survey for principals

Dear Principal,

The linked survey is being sent to all principals in the 22 schools subjected to external evaluation between January 2013 and May 2015. A similar survey has also been sent to teachers.

The survey is designed to gather information about how the findings of the recent external evaluation are being used within schools and whether such use is affecting the development of schools. The results will be used to review and improve the external evaluation process and to support decisions on its continuation.

The survey is being hosted on SurveyMonkey. All data will be treated confidentially and securely stored in a form that prevents the identification of individuals. In the processing of data and the disclosure of results, the confidentiality of all participants will be protected.

Please answer the following questions to the best of your knowledge or belief and at your earliest opportunity. Thank you in advance for participating.

Information letter at the beginning of the survey for teachers

Dear Teacher,

The linked survey is being sent to all teachers in the 22 schools subjected to external evaluation between January 2013 and May 2015. A similar survey has also been sent to principals.

The survey is designed to gather information about how the findings of the recent external evaluation are being used within schools and whether such use is affecting the development of schools. The results will be used to review and improve the external evaluation process and to support decisions on its continuation.

The survey is being hosted on SurveyMonkey. When you finish responding to the survey, the software separates email addresses from responses, such that individuals cannot be traced. All data will thus be anonymous.

Please answer the following questions to the best of your knowledge or belief and at your earliest opportunity. Thank you in advance for participating.

Email to principals and teachers to reiterate the request to answer the survey

Dear Recipient,

With this email, I would like to remind you to answer the survey sent to the teachers/principals of the schools that participated in external school evaluation from 2013 to 2015. I would appreciate your responses to the survey by May 24.

I realise that you are likely quite busy, but your answers are invaluable for the quality of the research's results and for the opportunities that they provide to review and improve the process of externally evaluating schools in Iceland.

I kindly ask you to answer all questions, which will take approximately 8−12 minutes [for teachers] / 10−15 minutes [for principals]. The preservation of and access to data will follow the provisions of the Personal Data Protection and Privacy Policy (no. 77/2000).

Please click the link below to answer the survey. If the URL is not active, then you may need to copy and paste it into your internet browser.

If you have any questions or concerns, then please feel free to contact me by email (bjo13@hi.is) or by phone (898-6348).

Thank you in advance,