
PUBLIC POLICY MASTER THESIS

April 2022

External School Evaluation in Lithuania
Conceptual model and stakeholder perceptions

Mykolas Steponavičius

Master's Thesis supervised by Denis Fougère

Second member of the Jury: Agnès Van Zanten

Master in Public Policy
Public Policy and Economics

Abstract

With the rise of evidence-based governance in education, school evaluation has gained more attention as a tool schools, regional and national policy-makers may use for quality assurance. Research into different approaches of school evaluation, as well as the different ways in which it may affect school quality, however, is still scarce. This research paper aims to fill in this gap by building a conceptual model of external school evaluation in Lithuania, which provides an interesting case study due to its rapid decentralization reforms in the 1990s and extensive policy borrowing from other countries. Building on the work of Ehren et al. (2013), the paper describes the core characteristics, mechanisms and intended effects of the system by analyzing official documents and semi-structured interviews with policy experts. While the resulting model is largely similar to previous research findings, a novel mechanism of evidence-based decisions on school support is identified. The paper also examines teachers' and school leaders' attitudes towards school evaluation as a proxy of its effectiveness. They are shown to depart from its conceptual model as evidenced by the lack of feedback acceptance, the presence of stress, and the importance of reputational concerns. Lastly, it is argued that the school evaluation system is underlied by tensions between formative and summative evaluation, policy planning and feedback which should be resolved to make it more effective. Strategies for minimizing the stress induced by school evaluation by ensuring anonymity and consultation between evaluation experts and the school staff are recommended.

Key words: external school evaluation, school improvement, accountability, evidence-based education policy, quality assurance

Table of content

| | |
|--|----|
| Why should I read this research? | 5 |
| 1. Introduction | 6 |
| 1.1. Global changes in education governance | 6 |
| 1.2. External school evaluation in Lithuania | 6 |
| 1.3. Outline of the analysis | 7 |
| 2. Literature review | 8 |
| 2.1. Defining external school evaluation | 8 |
| 2.2. Purposes of external school evaluation | 8 |
| 2.3. A framework for describing external school evaluation systems | 9 |
| 2.3.1. Institutional organization of external school evaluation | 9 |
| 2.3.2. Type of feedback and the role of the inspector | 10 |
| 2.3.3. Relationship with self-evaluation | 10 |
| 2.3.4. Types, frequency and data collection methods | 11 |
| 2.3.5. Consequences and reporting | 11 |
| 2.4. Mechanisms of school improvement | 12 |
| 2.4.1. Acceptance of feedback | 13 |
| 2.4.2. Setting norms and expectations on school quality | 13 |
| 2.4.3. Strengthening self-evaluation | 14 |
| 2.4.4. Parent involvement | 14 |
| 2.5. Empirical effects of external school evaluation | 14 |
| 2.5.1. Effects on student achievement | 14 |
| 2.5.2. The role of unintended consequences and stress | 15 |
| 2.5.3. Stakeholder perceptions of school evaluation | 16 |
| 3. Methodology, data and sources | 17 |
| 3.1. Building a policy theory | 17 |
| 3.2. Document analysis | 18 |
| 3.3. Semi-structured interviews with policy experts | 19 |
| 3.4. Semi-structured interviews with teachers and school leaders | 20 |
| 4. Analysis – Findings | 22 |
| 4.1. External school evaluation in Lithuania: historical context | 22 |
| 4.2. Describing Lithuania’s external school evaluation system | 24 |
| 4.2.1. Institutional organization | 24 |
| 4.2.2. Type of feedback and the role of the inspector | 24 |
| 4.2.3. Relationship with self-evaluation | 24 |
| 4.2.4. Types, frequency and data collection methods | 25 |
| 4.2.5. Indicators, standards and thresholds | 25 |
| 4.2.6. Consequences | 26 |
| 4.2.7. Public reporting | 27 |
| 4.3. Mechanisms and intended effects | 28 |
| 4.3.1. Evaluating the effectiveness of cooperation between the school and its proprietor | 28 |

| | | |
|-----------|--|-----------|
| 4.3.2. | Setting expectations on quality standards and acceptance of feedback_____ | 29 |
| 4.3.3. | Evidence-based decisions on school support_____ | 30 |
| 4.3.4. | Providing information on school quality and disseminating good practice_ | 31 |
| 4.3.5. | Strengthening schools' improvement capacity_____ | 32 |
| 4.3.6. | Taking improvement actions to create better learning conditions_____ | 32 |
| 4.4. | Tensions in the conceptual model of external school evaluation_____ | 33 |
| 4.5. | Teachers' and school leaders' perceptions of external school evaluation_____ | 35 |
| 4.5.1. | The role of stress_____ | 35 |
| 4.5.2. | Acceptance of feedback_____ | 36 |
| 4.5.3. | The perceived purpose of external school evaluation_____ | 37 |
| 4.5.4. | Preparing for evaluation and fitting into the standard_____ | 38 |
| 4.5.5. | Reputational concerns and stakeholder action_____ | 38 |
| 4.5.6. | Improvement actions_____ | 39 |
| 5. | Conclusion: findings and policy recommendations_____ | 40 |
| 5.1. | The conceptual model of external school evaluation_____ | 40 |
| 5.2. | Teachers' and school leaders' perceptions of external school evaluation_____ | 42 |
| 5.3. | Policy recommendations_____ | 44 |
| 5.4. | Limitations and suggestions for future research_____ | 44 |
| | Bibliography_____ | 45 |
| | Appendix A. List of selected documents_____ | 51 |
| | Appendix B. List of selected interviewees_____ | 53 |

Why should I read this research paper?

Education systems across the globe have seen significant changes as a result of decentralization, increasing focus on accountability and student outcomes. In the midst of these trends, the role of school evaluation has been transformed from merely ensuring administrative and legal compliance to generating data for decision-making at various levels of the system.

In most countries, school evaluation is aimed at improving the quality of its activities. Yet, the way quality is conceptualized and the mechanisms chosen to encourage its attainment differ significantly. Approaches range from attaching sanctions and rewards to achieving educational standards, utilizing feedback and public reporting. In addition, school evaluation may focus on accountability, policy planning, or supporting practitioners. Given a wide variety of goals and approaches to school evaluation, it is crucial to know how it may contribute to raising school quality. Surprisingly, very few academic studies have explored this question.

By replicating the work of Ehren et al. (2013), this paper adds to a scarce scholarly knowledge about the workings of school evaluation. It does so by building a conceptual model of Lithuania's external school evaluation system showing its core characteristics, mechanisms, and intended effects. The model provides a powerful tool to think about the logic of the system, and a strong foundation for future impact assessment studies. The paper also examines teachers' and school leaders' perceptions of external school evaluation. Comparing their attitudes to the conceptual model provides a glimpse into the system's effectiveness in encouraging school improvement.

Lithuania provides an interesting case study as its education system has experienced a rapid decentralization in the 1990s and has engaged in extensive policy borrowing in designing its school evaluation system. As such, this paper is not only the first work of the kind to focus on Lithuania. It may also help to gain an insight into more global questions about the tensions between centralization and school autonomy, formative and summative evaluation, policy planning, and school-level improvement.

1. Introduction

1.1. Global changes in education governance

Since the last decades of the 20th century, many education systems have introduced reforms to increase school autonomy and decentralization (OECD, 2013). At the core of these changes was the belief that local knowledge and decision-making are the best way to ensure education quality (Weiler, 1990). At the same time, schools around the world have experienced a contradicting trend of increasing accountability demands towards the state and market actors to deliver high-quality education services in a resource-efficient way (Hargreaves & Shirley, 2009). What unites these education governance trends is the growing reliance on data and evaluation in education quality assurance at various levels of the system.

Education systems exploit a number of evaluation activities for quality assurance and control. While national and international standardized student assessments tend to dominate public and academic debates, the role of school evaluation in ensuring school quality has recently gained more attention (Ehren et al., 2013). In some countries, a larger emphasis on school evaluation has been seen as a way to reestablish control on standards in light of poor performance in international tests (Ehren et al., 2013). The newly found relevance of school evaluation may also be linked to a broader shift from ensuring compliance with administrative rules and regulations to education quality (OECD, 2013).

This begs the question of how school evaluation can help to raise quality. There is a broad agreement among scholars that its impact on quality is mediated by intermediate mechanisms such as the provision of feedback and setting expectations on standards (OECD, 2013). However, there is surprisingly little systematic research linking the wide range of school evaluation characteristics to mechanisms and school quality. The most notable work of this kind has been done by Ehren et al. (2013) who built a conceptual model of external school evaluation by analyzing six European countries' school inspectorates.

1.2. External school evaluation in Lithuania

The focus of this research paper is to replicate the work of Ehren et al. (2013) in a novel context by investigating how Lithuania's external school evaluation system is intended to improve the quality of general education schools. The timeframe of the analysis is from the inception of the system in 2007 until the present day. The last section of the analysis is dedicated to examining teachers' and school leaders' perceptions of school evaluation as a significant determinant of its impact on school improvement.

The reasons for choosing Lithuania as a case study are twofold. Firstly, the research knowledge on school evaluation in Eastern Europe is very scarce. Lithuania, just like most other former Soviet republics, has experienced a rapid shift from highly centralized to more

decentralized education systems following the transition to market economies in the 1990s (Honinġh & Urbanoviġ, 2013). It is thus relevant to ask how the role and conception of school evaluation has been adapted to fit the changes in governance. Secondly, in the context of this transition, Eastern European countries have engaged in extensive policy borrowing in the field of education (Phillips & Ochs, 2004). Lithuania is no exception as the foundation for its external school evaluation system has been paved by the “School Development Program” financed by the World Bank, while the evaluation model has been developed based on Scotland’s model of self-evaluation. While this paper does not aim to analyze how specific aspects of school evaluation policy traveled across countries, it may serve as a stepping stone for future comparative analyses of this kind.

1.3. Outline of the analysis

The first four sections of the analysis develop and assess the conceptual model of Lithuania’s external school evaluation system. In particular, data from policy documents and semi-structured interviews with policy experts is used to build the main pillars of the model.

The first section outlines the historical and political context of external school evaluation. Namely, it is shown how the project of building Lithuania’s education quality assurance system financed by the World Bank and its conception has given rise to and has informed the development of external school evaluation since 2007. The most notable changes have been the agreement on a common conception of education quality finalized in the “Good School Conception” and a recent introduction of risk and thematic evaluations.

The second section presents a rich description of the system’s characteristics, such as public reporting, consequences, thresholds and standards. It is noted that schools are evaluated in four main areas using 32 indicators based on features of a good school identified in the “Good School Conception”. Evaluation levels are assigned to each indicator, yet no formal sanctions are imposed on underperforming schools, except for a follow-up evaluation after risk evaluation. Lastly, the public reporting of evaluation is shown to have changed over the years from publishing only the summarized to full versions of evaluation reports.

In the third section, acceptance of feedback, setting expectations on quality standards, evidence-based decisions on school support, and sharing of school practices are identified as the core mechanisms through which external evaluation is expected to promote school quality. On the contrary, stakeholder action is found to have only a limited role in school evaluation given the fact that school choice is constrained by Lithuanian law and an inconvenient public reporting format. Moreover, these mechanisms are aimed at raising education quality through strengthening the school's improvement capacity.

In the fourth section, it is argued that the conceptual model is only partially coherent as seen in the tensions between setting quality standards and promoting school’s uniqueness, as well as between using it for quality assurance at the school and system levels. Formally, the system is considered to help schools improve through providing feedback, yet some of its

elements, such as evaluation levels and proportional evaluation are rather geared towards policy planning.

The fifth section of the analysis examines stakeholders' perceptions of school evaluation drawing from semi-structured interviews with teachers and school leaders from two Lithuanian schools. Since stakeholder perceptions are seen as key to the effectiveness of school evaluation, this part is also aimed at highlighting if and how its conceptual model diverges from practice. As hypothesized, teachers are shown to be more negative towards school evaluation than school leaders as they report stress, the irrelevance of feedback and stringent evaluation criteria as the main shortcomings of the system. Moreover, the school staff is seen to have different conceptions of school evaluation, pointing both to its formative and summative functions. In contrast to the conceptual model, reputational concerns are seen as an important factor in school evaluation by both teachers and school leaders. Lastly, it is shown that the level of involvement by local and national governments in supporting school improvement depends on the evaluation type and the level of school quality.

2. Literature review

2.1. Defining external school evaluation

Terms *inspection* and *external school evaluation* are used interchangeably in academic literature. However, in this paper, the former term will be used only when referring to a particular country that has adopted this name, while the latter will be retained for general use as well as describing Lithuania's system.

Given the multiple purposes served by external school evaluation, the following definition is adopted: "Systematic, goal-oriented, and criteria-based process conducted by an external authority consisting of data collection (most often including site visits) and data feedback on school quality serving accountability/control purposes, enforcement of policy, and/or school improvement." (Hofer et al., 2020, p. 2). These purposes are discussed in greater detail in the next section.

2.2. Purposes of external school evaluation

The role of school evaluation has changed significantly over the years. Traditional evaluation was called *inspection* and was placed between the central government and schools, performing both quality assurance and a supervisory role of ensuring that schools follow the same education program and respect official rules and regulations (De Grauwe, 2007). However, over the years the focus of school evaluation has shifted from compliance to school quality (OECD, 2013). How education quality is conceptualized and ensured in turn defines the purpose of school evaluation.

According to their purpose, school evaluations systems may fall into a spectrum of two traditions. The first one is the school effectiveness tradition, which is oriented towards school accountability for performance, which is in turn determined by quantitative measures of learning outcomes, school's added value, and value for money (OECD, 2013). Moreover, the school is taken as the primary unit of analysis and its performance is considered at a particular moment in time (Yeung, 2012). Thus, accountability-focused school evaluation has been associated with summative evaluation, which provides a conclusive statement on school's quality (OECD, 2013). The underlying rationale of this approach is that improvement comes externally, through monitoring of school performance and changes in school organization informed by research knowledge (Janssens & van Amelsvoort, 2008). Accordingly,

The second tradition of evaluation is school improvement which is defined as “those conditions and processes that support and enhance learning and schools’ capacity to manage change” (Stoll, 2009, p. 116). Thus, school evaluation is primarily of formative nature, that is, it seeks to identify the school's strengths and weaknesses, thereby paving a pathway for improving teaching and learning (OECD, 2013). Given its focus on school processes and practitioners’ knowledge, this approach is often associated with self-evaluation (OECD, 2013).

School evaluation may serve an additional purpose in education planning and policy development. Specifically, it may be used to inform decision-making at various levels regarding resource allocation and other measures (OECD, 2013). Policy development should not be seen as exclusive from accountability or improvement but rather serving both or either one of these purposes.

2.3. A framework for describing external school evaluation systems

There is a great deal of divergence between countries in how they organize external school evaluation. The core dimensions along which these differences may occur are summarized below and based on the adaptation of the inspection framework by Ehren et al. (2013). It is important to note that these conceptual differences are often blurred in practice as countries employ a combination of different objectives and elements to organize ESE.

2.3.1. Institutional organization of external school evaluation

While most school inspectorates are legally accountable to the minister of education, they are supposed to provide an impartial assessment of school quality (Van Bruggen, 2010). The extent to which impartiality from stakeholder and government pressure is guaranteed is an important question, especially in the context where evaluators closely collaborate with the object of their evaluation (Ehren & Perryman, 2018). The division of political responsibilities regarding the frequency of evaluation, quality indicators, approaches to working with underperforming schools, and hiring decisions tend to vary across countries as well (Van Bruggen, 2010).

2.3.2. Type of feedback and the role of the inspector

Another output is feedback on teacher and school performance which may be provided during evaluation and/or in the evaluation report. Its primary goal is to help identify key areas for improvement and thereby enable the school community to make better decisions on teaching practices, resource allocation, curriculum standards, and professional development (OECD, 2013). The provided feedback may be purely descriptive or also include suggestions on how to improve certain aspects of schooling (Ehren et al., 2013).

While formally inspectors are expected to behave uniformly to allow for comparison across time and schools, notable differences have been observed between them (Ehren & Visscher, 2008). Namely, the role of an inspector can be described as a continuum between a reserved style which is limited to identifying the strengths and weaknesses of the school to a directive style which entails putting pressure on the school to improve (Ehren & Visscher, 2008).

2.3.3. Relationship with self-evaluation

Self-evaluation is defined as “a reflective professional process through which schools get to know themselves well and identify the best way forward for their pupils (Her Majesty's Inspectorate of Education, 2007, p. 6). It involves a school’s dialogue about the objectives, priorities, and quality indicators as well as planning actions to realize the democratic vision of education (MacBeath et al., 2000). It is primarily a developmental process that forms the core of a school as a learning organization that continuously improves through a systematic gathering and use of data (MacBeath et al., 2000). While most inspection systems rely on self-evaluation, in some systems schools are not legally required to perform it and are left to control and monitor their quality by themselves (Whitby, 2010).

Many inspection frameworks aim to combine external and self-evaluation in a way to enhance the school's capacity (MacBeath et al., 2000). The core rationale for employing both approaches is a recognition that a school cannot be improved without the commitment of students, parents, and teachers who all have a stake in the process (MacBeath et al., 2000). Accordingly, self-evaluation enables one to better account for values and goals that are specific to a school context, with the evaluator assessing the extent to which those goals are achieved (Shaw et al., 2003). This role is essential given that a school may not have the capacity to perform independent assessment and quality control of its own activities (Ehren & Visscher, 2008). Other motivations include the intention of capturing both the process and outcomes of schooling and reducing the inspection burden on both schools and the government (OECD, 2013).

Accordingly, Alvik (1996) describes three models of using external and self-evaluation:

1. *Parallel*: two systems function independently from one another, each with its own criteria and protocols.

2. *Sequential*: school's self-evaluation is used as a basis of the quality assurance system.
3. *Cooperative*: evaluation agencies work together with schools to develop an evaluation approach.

It is noteworthy that the combination of these two approaches is sometimes seen as contradictory since the accountability and the developmental logic is merged (Croxford et al., 2009). This tension is also evident in education systems that provide performance indicators for self-evaluation, thereby limiting the possibility for the school to agree on them. While self-evaluation seems to be predominantly geared towards school improvement, the question of its relationship with the purpose of evaluation is underexplored in the literature. Somewhat surprisingly, a few studies show that systems geared towards school accountability tend to rely on school self-evaluation (van Amelsvoort et al., 2006; Janssens & van Amelsvoort, 2008).

2.3.4. Types, frequency, and data collection methods

Evaluations may cover all schools as part of a regular cycle or be conducted proportionally according to need, which is usually determined by the analysis of self-evaluation results and other documentation provided by the school (Ehren et al., 2013). Furthermore, data on schools is collected from a variety of stakeholders, such as pupils, school staff, school boards, parents, and community partners to grasp the complexity of the schooling processes from multiple perspectives (Eurydice, 2004). The data may be collected before and during inspections using a variety of methods ranging from stakeholder surveys, interviews, desk research, and classroom observation to a review of official documents (Ehren et al., 2013).

2.3.5. Consequences and reporting

External evaluation systems may attach a set of consequences to evaluation results to reach the desired goal. Low-stake evaluations usually include such measures as a requirement to draw up an improvement plan and a follow-up visit and monitoring of school activity, while high-stakes evaluations use sanctions for underperformance (f.e. lay-offs, school restructuring, and shutdown) and rewards for high performance (f.e. financial bonuses and career advancement opportunities) (OECD, 2013). High-stake evaluations in particular are associated with the accountability approach to inspection as they conceive school development as a motivational issue (Altrichter & Kemethofer, 2015). While there is some empirical evidence that imposing sanctions on failing schools can help to improve student achievement (Hanushek & Raymond, 2005), other studies warn that high-stake evaluations may bring about unintended consequences (Altrichter & Kemethofer, 2015). Relatedly, some studies have shown that accountability systems may achieve their objectives even in the absence of credible sanctions (Figlio & Loeb, 2011).

Moreover, public reporting of inspection results may be used to inform the public about school quality, as well as to encourage school choice and competition (OECD, 2013). The format of reporting ranges from extensive evaluation reports to summaries of the main findings and may be provided for individual schools, school districts, or the whole school system (OECD, 2013). Thus, some authors consider public reporting an indirect or horizontal accountability measure that can affect in-school processes through stakeholder pressure (Faubert, 2009). Reporting of evaluation results can also be termed a *naming and shaming* strategy which creates a stigma around poor performance, with the social pressure being particularly acute when a comparison with a school from a given area is made (Allen & Burgess, 2012). However, it must be noted that there is little evidence that reporting enables school improvement (Hanushek & Raymond, 2005; De Wolf & Janssens, 2007).

2.4. Mechanisms of school improvement

Another strand of research on inspections focuses on school policy or school improvement actions. The core premise of these analyses is that there is no direct link between school evaluation and school quality. For this reason, these analyses try to explain the in-school processes which mediate between inspections and school quality by exploiting survey data featuring responses from the key stakeholders who participate in school improvement.

In their literature review, De Wolf & Janssens (2007) conclude that inspections encourage school improvement actions, yet through which mechanisms they are promoted remains unclear. The most notable study of the kind has been done by Ehren et al. (2013) who addressed this research gap by reconstructing policy theories of school inspectorates in six European countries. In their conceptual analysis, the authors show that acceptance of feedback, setting norms and expectations, stakeholder sensitivity, and strengthening self-evaluation are the key mechanisms that are employed to encourage school improvement. The following paragraphs provide an overview of how effective these mechanisms are.

Even in light of this research, the exact mechanisms through which inspections may impel schools to improve learning remain understudied. Firstly, this is in line with a broader trend in program evaluation to focus on impact assessment and outcome measurement as opposed to asking why those outcomes appear or fail to appear (Buitrago, 2015). Thus, there is a need for a theory of change that would spell out mechanisms of change and in turn provide actionable information for practitioners seeking to affect change (Buitrago, 2015). Secondly, a program theory of Lithuania's external school evaluation framework will provide a solid basis for further quantitative assessment of the program's effectiveness. Thus, given the lack of academic knowledge about Lithuania's external evaluation system and the general need in the literature to develop and refine theories of school inspection and associated mechanisms of improvement, this research project will aim to answer the following research question: *how is external school evaluation intended to improve the quality of Lithuanian schools providing general education?*

Given the previous findings in the literature, the first hypothesis is formulated as follows:

- *H1: Acceptance of feedback, setting expectations on quality standards and norms, strengthening self-evaluation, and parent involvement are the key mechanisms through which the quality of schools is intended to be improved by Lithuania's external school evaluation system.*

2.4.1. Acceptance of feedback

There is a broad agreement in the literature that the provision of feedback does not necessarily engender school improvement. The effect of feedback by teachers and the school leadership depends on its format, content, communication as well as implementation support (Ehren et al., 2015). For example, Matthews & Sammons (2004) argue that acceptance of feedback is crucial to a successful implementation of an improvement plan. In turn, studies show that stakeholders are most likely to embrace feedback when it is provided in a non-threatening manner (Visscher & Coe, 2003), when teachers trust the inspectors (Chapman, 2001) and when the feedback is provided privately and is in line with the school's culture (Standaert, 2001).

Moreover, in their study of the inspection of Dutch primary schools, Ehren & Visscher (2008) show that school improvement can be achieved only through a combination of directive feedback including specific improvement recommendations, negative ratings, and agreements between the principal and the inspector to bring about the desired changes. Similarly, Ehren et al. (2013) find that acceptance of feedback alone is not sufficient for schools to take improvement actions due to a lack of clear communication of expectations on the feedback use, lack of capacity, or resistance to implement school-level changes (Ehren et al., 2013). This finding is corroborated by Gustafsson et al. (2015) who test this conceptual model of school improvement by analyzing survey data on principals. The authors note, however, that the absence of effect may be because no distinction was made between positive and negative feedback.

2.4.2. Setting norms and expectations on school quality

One of the core channels of influence of external evaluation occurs through setting norms and expectations on school quality (Ehren et al., 2015). These norms, which are codified in the quality standards and the evaluation procedure, define what good teaching is, what knowledge is worth pursuing and how one should teach (Ehren et al., 2015). In turn, they draw the attention of the school community, thereby shaping its view on education goals, self-evaluation, planning, and ultimately practice (Ehren et al., 2015). Moreover, the normative influence of school inspections may be seen from a lens of the neo-institutional theories, according to which organizations try to accrue resources and thereby survive by aiming to be legitimated by their environments (Meyer & Rowan 1977). For this reason, schools may simply comply with the quality standards as a way of enhancing their reputation, financial prospects, and legitimacy in the eyes of the inspection authorities, as well as the local and national government.

This normative pressure has been shown to have both a positive and negative influence on school improvement by several studies. Allen & Burgess (2012) show that the reception of an unsatisfactory school rating can incentivize the school management to put more emphasis on student achievement. Moreover, in the analysis of survey responses from school leaders, Gustafsson et al. (2015) find that setting expectations on quality standards and norms is the strongest mechanism through which inspections lead to school improvement actions. This means that schools internalize the quality standards specified in the evaluation procedure and use them as guidance to organize schooling activities. However, the authors also note that setting expectations narrows down the curriculum and discourages experimentation. Moreover, a case study of a school in England placed under the inspection's special measures revealed the school to be putting on a performance for the inspectors rather than implementing genuine and lasting changes in its activities (Perryman, 2006).

2.4.3. Strengthening self-evaluation

Gustafsson et al. (2015) demonstrate that the effect of setting expectations and norms on school quality is mediated by the school's self-evaluation capacity. According to the authors, schools regard self-evaluation as "a vital developmental strategy when responding to school inspection" (Gustafsson et al., 2015, p. 55). The expectation is that prescribed indicators become part of the self-evaluation which enables continuous school improvement. However, internalization of quality norms is not guaranteed as it can lead to mere compliance with the evaluation procedure, especially in contexts where stakeholders are used to a top-down approach to school inspection (Croxford et al., 2009).

2.4.4. Parent involvement

Involvement by parents in the school evaluation process may influence school quality. This mechanism is tightly linked to the public reporting of evaluation results and the strategies of *naming and shaming* discussed in the previous section. Gustafsson et al. (2015) find that stakeholder action is an important mechanism. However, its influence is limited to the early stages of the evaluation cycle, namely in encouraging the acceptance of feedback and standards. By contrast, Altrichter & Kemethofer (2015) find that school leaders who are most affected by the pressure of the inspection are more responsive to the expectations set by the inspection rather than parents' demands.

2.5. Empirical effects of external school evaluation

2.5.1. Effects on student achievement

Most studies within the school effectiveness research have explored the link between accountability pressure and student achievement (Hussain, 2015) with fewer studies investigating the impact of school inspection. In addition, for the most part, the scholarly work has been done in England (Matthews & Sammons, 2004; Allen & Burgess, 2012;

Hussain, 2015) and the Netherlands (Luginbuhl et al., 2009) and has largely indicated that external evaluation helps to improve student performance.

A number of studies have demonstrated that improvements in student achievement are largely associated with the negative assessment given by school inspections. Looking at the examination results and inspection data, Matthews & Sammons (2004) demonstrate that following the inspection in England schools placed under special measures exhibit more progress than schools with a slightly higher inspection score. This result has important implications for equity as low-performing schools are disproportionately attended by children from low socio-economic backgrounds (Matthews & Sammons, 2004). In a later study, the authors argue that the observed difference can also be attributed to the accompanying measures of monitoring and support and effective leadership (Matthews & Sammons, 2005). Moreover, Allen & Burgess (2012) use a regression discontinuity design to model the change in high-stake national exam performance among schools in England and show that schools who received “a notice to improve” have raised their performance substantially compared to a pre-inspection year. Similarly, an analysis of the relationship between the inspection and standardized test scores in English schools by Hussain (2015) reveals that external evaluation engenders improvement only when the schools are deemed as failing, with the largest increases in test scores accruing to underperforming students from mostly low-income families. It is noteworthy that the observed effect may be explained by sanctions to which the identified schools were subjected. Lastly, a few studies diverge from these findings by pointing to the absence (Shaw et al, 2003) or negative effects of school evaluation on exam results (Wilcox & Gray, 1996; Rosenthal, 2004).

2.5.2. The role of unintended consequences and stress

Some level of stress is inherent in any process of evaluation or being observed. This can have a positive influence on school quality given that an optimal level of pressure is exerted which mobilizes the school community to put more effort and perform at a higher level (Ouston et al., 1997).

There is substantial empirical evidence that school inspection may cause unintended negative effects. They may occur due to stress, as formulated by Le Châtelier’s principle, according to which a system put under stress will change in a way as to counteract its influence (Jones et al., 2017). While the concept was originally applied in physical sciences, it can be seen in the social world as well. In fact, a similar idea is expressed by Campbell’s law which states that predicating decision-making on the use of a single indicator is likely to distort and corrupt the processes, which are the target of regulation (Campbell, 1979).

In the context of external school evaluation, three main types of adverse consequences, resulting from inspection-induced pressure to perform, can be distinguished (De Wolf & Janssens, 2007). Firstly, strategic behavior and gaming may be used to make the school appear more effective through extra preparation for the inspection visit (Wilcox & Gray, 1996), selective data representation (Jones et al., 2017), or even cheating (Ehren &

Swanborn, 2012). Incidentally, the latter findings stand in contrast to Hussain's (2015) insight that inspections can serve as an effective mitigation tool against gaming strategies by allowing close observation of the in-school processes. Secondly, unintended strategic behavior can occur as the observed school community unwittingly changes its actions. These consequences include directing attention from learning to inspection (Rosenthal, 2004), narrowing down the curriculum and instructional strategies (Jones et al., 2017), stifling teacher innovation (Ehren et al., 2015), and compliance with formal procedures (Croxford et al., 2009). In turn, these strategies are seen to result in short-lived changes in school practices and increased uniformity between schools (De Wolf & Janssens, 2007). Thirdly, a negative spiral may be triggered when a school is identified as failing leading to staff and pupil retention and low morale which in turn undermines the school's capacity to improve (Allen & Burgess, 2012)

2.5.3. Stakeholder perceptions of school evaluation

Another strand of research on school evaluation focuses on stakeholder attitudes towards school evaluation. Several academic studies have also shown that stakeholder perception of school evaluation quality and usefulness have an impact on its effectiveness (Hofer et al., 2020). Moreover, collaboration between the evaluation team and stakeholders has been shown to make school evaluation more effective (Whitby, 2010). Also, as previously discussed, acceptance of feedback and quality standards by the school staff is a precondition for school improvement following the evaluation.

Given that stakeholder views on school evaluation play an important role in ensuring its effectiveness, it is crucial to ask how the key school stakeholders see Lithuania's external school evaluation. Thus, comparing the conceptual model of school evaluation to stakeholders' views is a way to examine if and how the policy theory diverges from its implementation. Accordingly, the second research question is formulated as follows: *how do school teachers and leaders view the purpose and mechanisms of Lithuania's external school evaluation system and what are (if any) the consequences unforeseen by the conceptual model?*

In general, more positive attitudes with regard to school evaluation have been observed among school leaders and parents than among teachers (Altrichter & Kemethofer, 2015). In large part, this attitude can be related to stress, induced by school evaluation among teachers discussed in the previous section. Moreover, some studies have noted that teachers tend to perceive school evaluation as a threat to their professional identity and a constraint on their creativity (Taylor, 2007). In the context of Lithuanian schools, one study revealed that teachers tend to view themselves as passive participants in the evaluation process and do not see its value for pupils (Ministry of Education, Science and Sports, 2011). In particular, they expressed concerns about the lack of consultation with evaluation experts, the standardized criteria failing to accurately capture the quality of instruction, and the stress caused by evaluation (Ministry of Education, Science and Sports, 2011). School leaders had a slightly more positive outlook towards school evaluation, naming positive school culture and

recommendations for improvement as the main benefits of the system (Ministry of Education, Science and Sports, 2011).

Given these findings, the second hypothesis is formulated as follows: *I expect teachers to have a more negative outlook towards the purpose and mechanisms of external school evaluation than school leaders due to experienced stress and perceived threats to their professionalism.*

3. Methodology, Data and Sources

As previously discussed, the first goal of this paper, relating to the first research question, is to build a policy theory of external school evaluation in Lithuania by identifying the underlying mechanisms through which it aims to achieve its intended outcomes. The second goal relates to the second research question and aims to examine how this policy theory unfolds in its main arena of implementation – Lithuania’s general education schools. The methodology section is structured as follows. To begin with, the concept of policy theory, as well as its usefulness for education policy research and practice, is discussed. Subsequently, the two main methods that were used to build a policy theory – document analysis and interviews with policy-makers are explained. Lastly, a semi-structured interview approach is outlined as the method of choice to explore how the conceptual model of school evaluation compares with the experience of teachers and school leaders.

3.1. Building a policy theory

Chen & Chen (1990) define program theory (also known as policy theory) as a “specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated” (p. 43). This definition suggests that policy theory consists of two main sub-theories, namely normative and causative theory. The former deal with the goals and outcomes of a given policy, as well as the treatment’s design and implementation. The latter sub-theory is concerned with how a policy works, namely by identifying the processes and their consequences through which treatment will help to achieve the goals of the policy. The authors note that both the normative and causative policy aspects tend to be assumed or taken for granted and thus not critically examined.

Policy theory is particularly valuable because it moves away from a traditional *black box* approach to program evaluation, which merely focuses on whether a particular program is effective without providing answers as to why it managed or failed to achieve the desired policy goal (Chen & Chen, 1990). Policy theory, however, can help to identify weaknesses in the underlying mechanisms of the policy as well as contextual factors that may be hindering its effectiveness and thereby provide valuable information for policy-makers and other stakeholders on how it could be improved (Buitrago, 2015). Moreover, policy theory can provide a solid basis for further quantitative assessment of the program’s effectiveness.

It may be useful to illustrate what a policy theory looks like in the context of external school evaluation. The most thorough research in this area has been done by Ehren et al. (2013) who derived a conceptual model of school inspections in six European countries. The theory is composed of three main elements – characteristics of school inspection, mechanisms, and expected outcomes. Characteristics refer to attributes of the inspection, such as standards and frequency of visits that are expected to encourage school improvement and are summarized in the following policy framework. Expected outcomes, such as self-evaluation and improvement capacity, are the normative objectives of the inspection. Lastly, mechanisms are seen as mediators between inspection characteristics and the expected outcomes (Ehren et al., 2013). For simplicity, the policy theory of Lithuania’s external school evaluation system will be referred to as a conceptual model.

3.2. Document analysis

Document analysis was the first step in building a conceptual model of Lithuania's external school evaluation. This method is conventionally used in qualitative research and policy analysis as means of understanding the nature of complex policy issues by uncovering “what lies behind and within policy documents” (Cardno, 2018, p. 625). It can be described as “a systematic procedure for reviewing or evaluating documents” which calls for the examination and interpretation of data in order to attain an understanding or empirical knowledge of the phenomenon at hand (Bowen, 2009, p. 27). Often it is used together with other qualitative research methods in a process called triangulation whereby multiple data sources or methods are used in order to reduce potential bias and thereby generate credible findings (Bowen, 2009). For instance, an important advantage of using interviews in parallel to document analysis is the possibility of detecting new issues, surprises, or omissions in the documents (Cardno, 2018).

Specifically, a content analysis method was employed to examine the documents related to external school evaluation in Lithuania since it is considered one of the most reliable approaches for the study of organizational policy documents (Cardno, 2018). While conventionally this approach is associated with quantitative research and counting the frequency of certain words and phrases, in this study it is used as a method used to “draw inferences from the content by acknowledging the mere presence or absence of certain words or phrases” as is appropriate in qualitative research (Cardno, 2018, p. 633). The following steps for structuring the analysis as recommended by Cardno (2018) were adopted for guidance:

1. preparation for the analysis;
2. description of context; and
3. decisions about the approach to categorization;
4. dealing with the document text and inferences.

Firstly, preparation for the analysis involved a review of the relevant literature resulting in the formulation of the research questions and the identification of the relevant documents.

Secondly, the policy context, in which the documents are situated, was described by drawing from policy documents, legislation, and interviews with policy experts. In particular, the ideas and values surrounding the development of the education quality assurance system and the replacement of the old inspectorate provided a socio-political background for the analysis of the identified documents related to external school evaluation. Thirdly, the deductive approach to categorization was chosen given its suitability for testing already existing theories in a novel context (Elo & Kyngas, 2008). In particular, the core characteristics of Lithuania's external school evaluation system were described based on the framework of school inspectorates suggested by Ehren et al. (2013). Fourthly, throughout the analysis clarifications were made on whether the content was directly derived from or inferred from the documents.

After describing the context and the characteristics of the external evaluation system, the mechanisms of change and the intended outcomes were identified. This part of the analysis was informed by a simplified version of Leeuw's method for reconstructing policy theories (2003). Firstly, the intended effects of the policy were identified by surveying the relevant legislation. Secondly, social or behavioral mechanisms through which the policy is expected to achieve these effects were found by looking for statements that point to the necessity of solving the issue at hand and the beliefs about the policy's effectiveness (Leeuw, 2003). Thirdly, the characteristics of the external school evaluation system, its mechanisms, and intended effects were linked to each other, thereby providing its policy theory.

A crucial step in document analysis is overcoming *biased selectivity* through an exhaustive selection of documents (Bowen, 2009). For this purpose, a range of documents including external school evaluation legislation and subsidiary documents, such as policy reports, were identified as relevant for the analysis of the system (see Appendix A). Moreover, in line with the recommendations by Bowen (2009), special care was taken in order to ascertain the relevance of the selected documents for the research question and the conceptual framework of the analysis. Namely, the external school evaluation procedure is the central document of the analysis which was identified in the early stages of the analysis on the National Agency for Education website. Subsequently, the remaining documents describing the policy context of the system were retrieved by browsing the e-portal of the Lithuanian Parliament by using the function "related documents". Lastly, amendments to legislative documents were retrieved using the function "changes in the legislative act" in order to track how the policy of external school evaluation has evolved over time.

3.3. Semi-structured interviews with policy experts

The second step in building a policy theory of Lithuania's external school evaluation system involved semi-structured interviews with seven experts. A total of twelve questions were included in the interview prior to its start and varied depending on the experts' responses. On the one hand, having a fixed set of questions allowed to corroborate and cross-check the information on the system's characteristics and mechanisms identified in the document analysis. If the information provided by an interviewee contradicted other

interviewees' responses or the content of the documents, this was made explicit in the analysis. In other words, contradictions between multiple statements were taken as evidence of an incoherent policy theory. At times, the coherency of the theory was deliberately tested by presenting one expert's position to another and asking for a comment. On the other hand, adapting the questions throughout the interview based on the interviewees' responses helped to identify information that was not found in the documents. For instance, some experts explained the political context and the rationale for the introduction of thematic and risk evaluations which was not stated in the documents.

Experts who were selected for the interviews range from policy-makers, who were at the forefront of creating and implementing the external school evaluation system, to school leaders and lead evaluation experts who developed the school evaluation methodology and trained evaluation experts. The fact that the interviewees have diverse professional backgrounds, and have worked at different levels of the external evaluation system in different positions as well as at different time periods should ensure a well-rounded analysis (see Table B1). The exact job titles and the corresponding years of the interviewees are not specified in order to maintain their unanimity. Moreover, the fact that inspectors were not interviewed must be acknowledged as a limitation of this study. However, this gap is partly addressed by the fact that two interviewees had worked as head evaluators in charge of training school evaluation experts and thus provided an insight into both the process of evaluation at a school level and the macro logic of the system.

A snowball sampling technique was used to contact policy-makers. While this technique is not aimed to produce a sample that is representative of the studied population, it helps to establish contact with otherwise difficult to reach study subjects (Sharma, 2017). In this particular case, the fact that the initial point of contact knew other policy-makers helped to identify people who are knowledgeable about Lithuania's external school evaluation system and to establish trust that was necessary to organize the interviews. The interviews were conducted online by using a video communication platform. All interviews were recorded, transcribed, and subsequently translated from Lithuanian to English.

3.4. Semi-structured interviews with teachers and school leaders

The last part of the analysis examines how teachers and school leaders perceive external school evaluation and how their views compare to the conceptual model described in the first four parts of the analysis. For that purpose, interviews with teachers and school leaders were organized. There are a few reasons for this choice. The conceptual model presupposes that stakeholder agreement is key to school evaluation's effectiveness as they will take improvement actions only if they agree with the provided feedback and standards. Moreover, as discussed in the literature review, the effectiveness of school evaluation has been shown to depend on stakeholders' perception of its usefulness and quality (Whitby, 2010). School teachers and leaders were chosen for the analysis since they are the main agents of school improvement according to the conceptual model.

The interview questions were structured around the components of the conceptual model, namely its purpose, mechanisms, and intended effects (see Table C2). Some mechanisms, such as sharing of school practices and evidence-based decision-making were not included in the interview questions as they primarily relate to other actors in the education system. The semi-structured interview format was chosen in order to attain information about how and if specific parts of the model appeared in practice while leaving space for interviewees to indicate aspects that were unforeseen by the model. The interviews were conducted online by using a video communication platform. The interview recordings were subsequently transcribed and translated from Lithuanian to English.

In total, six interviews in two schools were conducted with two teachers and one school leader from each school (see Table B2). Firstly, two schools with distinct characteristics were selected. *School A* participated in a comprehensive evaluation in 2011 and *school B* – in a risk evaluation in 2019. This choice helped to account for differences in the level of schools' quality and prestige, and examine how their experience differs across evaluation types and different time periods. Moreover, the selected schools are located in different geographical areas which differ in their economic and demographic contexts. Specifically, *school A* is located in a city, while *school B* is located in a town. To obtain substantive information about external evaluation in both schools, additional documentation such as evaluation reports were reviewed.

Subsequently, a number of school leaders whose schools have undergone external evaluation were identified using a list published by the National Agency for Education (National Agency for Education, n.d.). Some of the school leaders who responded to calls for interviews via email had started their terms after the external evaluation and thus had to be excluded, leading to a small number of potential interviewees who were chosen based on the quickness of their reply. It must be acknowledged that this aspect presents a potential source of bias as school leaders with more initiative may be more likely to respond to such requests. The last step in the selection of interviewees included asking school leaders for contacts of teachers who worked at the school during external evaluation. While such a selection method may be seen as prone to bias, the interviewees were assured that their anonymity would be guaranteed by concealing their and school names. In addition, teachers have been observed to openly state their opinion in the interviews even if it was at odds with the school leader's opinion.

The core limitation of the analysis is a low number of interviewees, which in turn prevents drawing any conclusions about the larger population of Lithuanian schools. Therefore, the findings should be interpreted as merely indicating the experience of the selected schools and paving the ground for future research with a more extensive sample of schools and interviewees.

4. Analysis - Findings

4.1. External school evaluation in Lithuania: historical context

The origins of the external school evaluation system in Lithuania can be traced back to the early 2000s. At the time, Lithuanian policy-makers had very few quality assurance tools given that most of the educational data came from national and international student assessments such as PIRLS and TIMSS (Interviewee 4, personal communication, February 17, 2022). In 2002, the “School Development Program” financed by the World Bank supported the creation of the country’s education quality assurance system comprised of the following core pillars – education management information system, education policy analysis system, national system for the assessment of student achievement as well as the internal and external audit system (The Government of the Republic of Lithuania, 2002). Initially, indicators were built for the internal audit system based on the Scottish self-evaluation methodology “How Good is Our School” and adapted to the Lithuanian context (Ministry of Education, Science and Sports, 2007b). Seven areas of assessment and 32 indicators provided a basis for the external school evaluation (ESE) which was officially introduced in 2007 and overseen by the newly established National Agency for School Evaluation (Ministry of Education, Science and Sports, 2002).

The program was finalized with the introduction of the “Conception of Formal Education Quality Assurance System” in 2008. It outlines the logic of the quality assurance cycle comprising the conception of education quality, quality evaluation, and quality improvement actions (Ministry of Education, Science and Sports, 2008). A reflection that “associates subjective experience with evaluation findings and helps to develop and select the strategies of educational change” is considered to be the starting point of each cycle (Ministry of Education, Science and Sports, 2008). It is noteworthy that the cycle may take place both at the system and school levels.

In the second pillar, four types of evaluation are envisioned. On the one hand, external and internal summative evaluations focus on educational outcomes and are meant to provide a conclusive statement on the quality of educational activities (Ministry of Education, Science and Sports, 2008). Examples include standardized student assessment, teacher and school leader evaluation as well as stakeholder surveys. On the other hand, formative evaluation focuses on the educational processes and aims to encourage improvement through either consultation (external evaluation) or self-reflection (internal evaluation) (Ministry of Education, Science and Sports, 2008). Accordingly, external school evaluation is intended to be used in combination with other types of evaluation to measure, monitor, and improve education quality. For instance, as standardized test results perform a summative evaluation function by showing pupil achievement level, external school evaluation can be used to shed light on why and how this level was reached (Interviewee 2, personal communication, February 14, 2022). In this sense, policy-makers saw external evaluation as particularly valuable as Lithuanian pupils’ performance in international student tests began to stagnate,

pointing to the exhausted potential of other quality management tools, such as the renewal of the education content (Interviewee 4, personal communication, February 17, 2022).

The formative role of the external school evaluation in the quality assurance system is also mirrored in the legislation as it is intended to “encourage school improvement by seeking higher education quality and student achievement” (Ministry of Education, Science and Sport, 2007a). Historically, a particularly important task was moving away from the control and compliance approach of the old school inspectorate which was seen as carrying a legacy of the old Soviet system (Interviewee 4, personal communication, February 17, 2022). In the words of one expert, the goal was to:

Build a modern inspection performing a developmental function – providing support to schools by helping them to identify their strengths and weaknesses and thereby ensuring quality not through compliance with legal acts but by promoting a culture of quality that works as a quality assurance system (Interviewee 7, personal communication, March 15, 2022).

So, in order to avoid any connotations associated with the term *inspection*, it was labeled *external school evaluation* (Interviewee 2, personal communication, February 14, 2022). For this reason, while these terms are used interchangeably in the academic literature, only the latter term will be used to refer to Lithuania’s system.

As stipulated by the conception of quality assurance, an agreement on the conception of education quality was reached in 2015 and laid out in the “Good School Conception”, which in turn provided a basis for remodeling of the external evaluation indicators (The Government of the Republic of Lithuania, 2015). Specifically, the methodology was narrowed down to four areas of evaluation and 25 indicators and it stands as such until the present day (Ministry of Education, Science and Sport, 2007a). These indicators comprise a comprehensive school evaluation that aims to provide an assessment of the entire school.

Recently, comprehensive evaluation has been supplemented by two more types. In 2019, risk evaluation was introduced as part of the “Quality Basket” program which aims to support the improvement of underperforming schools (Ministry of Education, Science and Sports, 2018). Moreover, in 2021, thematic evaluation was introduced to gather evidence on Lithuanian schools’ performance in an education area of political interest, starting with the topic of inclusive education (Ministry of Education, Science and Sports, 2021). Both evaluations focus on different areas of assessment and employ a different set of indicators.

4.2. Describing Lithuania's external school evaluation system

4.2.1. Institutional organization

The general external evaluation rules and procedures are set by the minister of education. In turn, it is organized by the National Agency for Education together with a school founder which is usually a municipality. The agency is responsible for evaluators' selection and training as well as the analysis and publication of evaluation reports. The school founder also monitors the progress of schools and provides the necessary support before and after the evaluation. Evaluators are recruited on a project basis as they often hold full-time positions as education support specialists, school leaders, or teachers.

4.2.2. Type of feedback and the role of the evaluation expert

In Lithuania's external evaluation system, feedback comes in two forms. Firstly, individual feedback is provided to teachers after the observed lesson, detailing at least three positive aspects and no more than two negative aspects of the lesson. Secondly, the evaluation report includes a comprehensive assessment of the school's activities according to the established areas of assessment and indicators. The preliminary version of the report featuring at least ten positive aspects and no more than five negative aspects of a school is presented to its leader who discusses it with teachers and collects their comments which are subsequently incorporated in the drafting of the final evaluation report. Anonymity is ensured in the evaluation report as school activities are discussed without specific references to teachers' names. The nature of both forms of feedback is descriptive as they point out the school's strengths and areas for improvement. No prescriptions on how a school is supposed to address the identified weaknesses should be given. This is done in order to maintain the impartiality of inspectors who are seen as external experts rather than consultants (Interviewee 6, personal communication, March 8, 2022).

4.2.3. Relationship with self-evaluation

Lithuania's external school evaluation system can be defined as a dual or sequential system since, on the one hand, it relies on school evaluation data and methodology, and on the other hand, it aims to strengthen its capacity. Self-evaluation is tightly linked to the "Conception of Formal Education Quality Assurance System" described in the previous section. According to one policy-maker, self-evaluation in the context of Lithuanian schools should be understood as a "promotion and support of reflective practice which is used as a tool to think about school quality, identify its weaknesses and strengths on the basis of which management decisions can be made" (Interviewee 3, personal communication, February 15, 2022). In other words, the same logic of the quality assurance cycle can be applied to both the education system as a whole and the school itself. While schools are obligated by law to conduct self-evaluation, they are free to design their own indicators or use the ones recommended by the Ministry of Education, Science and Sport based on the "Good School Conception". As the conception recognizes the uniqueness of each school and the distinct

socio-economic contexts in which they function, it is meant to be used only as a roadmap for schools to come up with a conception of education quality and the appropriate indicators that reflect their specific community context (Good School Conception, 2015).

4.2.4. Types, frequency, and data collection methods

Initially, the external evaluation system was designed to follow a regular seven-year cycle during which each school providing general education would undergo external evaluation. A more frequent evaluation was envisioned in cases where a school demonstrates poor results, insufficient progress, or if it was initiated by the school or its founder. The seven-year cycle was abandoned with the 2018 redaction of the Directive on the Organization of ESE, which also introduced risk and thematic evaluations being conducted in parallel to the already existing comprehensive evaluation.

School evaluation begins with the selection of schools given that a limited number of school visits can be planned during the year due to limited financial and human resources. Thus, schools are selected based on lists provided by the respective municipalities and ultimately ratified by the minister of education according to the established priorities (such as previously uninspected schools or larger schools). For instance, in 2019, schools for risk evaluation were selected based on national examination results, the share of families in a given municipality receiving free meals as well as student well-being measures (Interviewee 7, personal communication, March 15, 2022).

The next stage of evaluation is preparatory work which aims at better grasping the socio-economic, cultural, and pedagogical context of a given school and generating hypotheses about the school's strengths and weaknesses. As the school is notified of the upcoming visit, the evaluators receive data on the school's results in national standardized examinations, as well as human and material resources from the National Agency for Education (Ministry of Education, Science and Sports, 2007a). The school in turn provides the evaluators' team with self-evaluation data and relevant documents such as strategic and learning plans as well as lesson plans and timetables (Ministry of Education, Science and Sports, 2007a). Data on the stakeholders' view of the school is obtained by meeting with the school community and conducting parent and pupil surveys (Ministry of Education, Science and Sports, 2007a).

During a three-five day visit, a team of evaluators dedicates 75 percent of their time to class observation, while the remaining time is spent talking to teachers and the administration, observing off-class activities. Throughout the visit, inspectors use the evaluation protocol and evaluation indicators to collect the relevant data.

4.2.5. Indicators, standards and thresholds

According to the conception of quality Assurance, the evaluation of education activities should be based on stakeholders' shared understanding of education quality (Ministry of

Education, Science and Sports, 2008). Such an agreement is stipulated in the “Good School Conception” which is a document detailing the main features of a good school (Ministry of Education, Science and Sports, 2015). Accordingly, this conception provides a basis for the school evaluation methodology featuring four areas of evaluation – results, education and pupils’ experience, education environment as well as leadership, and management. While the goal of external evaluation is formulated in terms of results, “Good School Conception” maintains the balance between these four areas of school activity. In other words, a good school is characterized not only by outstanding results but also by effective processes such as leadership, self-evaluation, personalized teaching methods, etc. Moreover, it is meant to be a prescriptive document functioning as a roadmap for each school to find its own unique vision of a good school and corresponding indicators to measure and monitor its progress.

In addition, the four areas of evaluation are further broken down into 25 indicators. Each indicator is assigned a level which is then pooled with the levels of other indicators to provide an overall assessment of given school activity. There are five levels of quality ranging from *N* (very poor, unsatisfactory), *1* (poor), *2* (satisfactory), *3* (good), and *4* (very good). The evaluation of overall school quality is not provided.

4.2.6. Consequences

As discussed earlier, Lithuania’s external school evaluation system was created as a counterbalance to the old school inspection based on a hard conception of control and vertical accountability of the state. Thus, the current approach to evaluation is of formative nature, that is, aimed at engendering school improvement through the provision of feedback and learning support. In practice, it means that there are neither formal sanctions for poorly performing schools, nor rewards for the highest performing ones. Interviews with policy-makers who played a central role in developing the system revealed that in its early years, some government officials were in favor of the vertical accountability model, which would have allowed stripping the school of a license in case the evaluation revealed major issues (Interviewee 4, personal communication, February 17, 2022). The opposing arguments ultimately prevailed pointing to the social and financial costs of school reorganization and pupil transfer to other schools (Interviewee 4, personal communication, February 17, 2022).

The fact that no sanctions exist for underperforming schools begs the question of what happens next. Following a negative comprehensive evaluation of the school, with at least one schooling aspect being rated *N* or *1* or with most aspects being rated *2*, the school founder is informed about the necessity to take action (Ministry of Education, Science and Sports, 2007a). The founder is expected to closely monitor the school’s progress and report to the National Agency for Education on the progress of the implementation of the improvement plan. Similarly, one year after a risk evaluation, the agency analyzes the data on the school’s progress, the results of the school leader evaluation and provides a judgment on whether the risk factors have been addressed (Ministry of Education, Science and Sports, 2007a). A follow-up visit is planned two years after the initial risk evaluation in order to assess whether

the school has made sufficient progress which is defined as an improvement of more than half of the indicators (Ministry of Education, Science and Sports, 2007a).

It is noteworthy that in the early stages of the system's development, the possibility of initiating the evaluation of school management in case the school continued to lag behind was envisioned (Ministry of Education, Science and Sports, 2007a). According to one policy-maker, one version of such a system included factoring in both school evaluation results and parents' opinions in the evaluation of the school leader and thereby potentially placing limits on his five-year term (Interviewee 2, personal communication, February 14, 2022). However, the aspiration to link the external evaluation results to school leaders' performance evaluation has been abandoned over the years. Under the current system, the National Agency for Education employs soft levers of providing feedback and discussion with school founders and school leaders in order to accelerate progress in underperforming schools (Interviewee 5, personal communication, February 21, 2022). A senior policy official from the agency indicated that while there has not been a single school that failed to improve after a follow-up visit, discussions on what course of action should be taken lest the precedent arose are currently taking place (Interviewee 5, personal communication, February 21, 2022).

4.2.7. Public Reporting

At the initial stages of the system's development, evaluation data was considered to be the property of the school and thus could only be used for internal purposes (Interviewee 4, personal communication, February 17, 2022). Similar to the discussion on sanctions, at the dawn of the system, the degree to which school evaluation data should be publicized was highly contested with conflicting positions among policy-makers. One interviewee describes an idea borrowed from the OFSTED inspection in England, to send the evaluation results personally to each parent who could then use the information to exercise school choice or put pressure on the school to improve (Interviewee 2, personal communication, February 14, 2022). Ultimately this idea was not implemented due to a lack of resources at the agency level. More importantly, one policy-makers explained the motivation for not disclosing the evaluation data as wanting to avoid social stigma for poor performance stemming from a duty to protect schools despite the pressure from parents to reveal the data (Interviewee 4, personal communication, February 17, 2022). Therefore, ultimately, the decision was made to publish only summaries of evaluation reports on the website of the National Agency for Education showing the strengths and weaknesses of a school. As the notion of the school right to evaluation data was established, schools were left free to publish the full evaluation reports on their own behalf. Moreover, at the end of the school visit, the evaluators are obligated to present the findings to the school community including the municipality education department officials, parent, and student representatives as well as the school staff (Ministry of Education, Science and Sports, 2007a). Recently, the notion of protecting school data has been abandoned in favor of more extensive reporting with the introduction of risk evaluation in 2019. Full risk evaluation reports have been made available online as a way to familiarize schools with the external evaluation and alleviate the associated stress and pressure according to one policy official (Interviewee 5, personal communication, February 21, 2022).

4.3. Mechanisms and intended effects

The ultimate goal of external school evaluation is to “encourage school improvement by seeking higher education quality and student achievement” (Ministry of Education, Science and Sport, 2007a). However, there are a number of mechanisms and intended effects that link the evaluation and its goal. Namely, they were identified by examining seven objectives of external school evaluation that are specified in the Procedure Description for the Organization and Implementation of External School Evaluation (2007a):

- 1) Promote self-evaluation of school quality and agreements on education quality conception and improvement
- 2) Encourage schools to create better conditions for learning, improving, seeking progress and higher results
- 3) Facilitate evidence-based decisions on school support
- 4) Improve existing and develop new forms of support and consultation for schools, teachers, students and their parents
- 5) Evaluate the effectiveness of cooperation between the school and the institution exercising the rights and responsibilities of a school proprietor in seeking higher quality education and results.
- 6) Identify risk factors in school activity, their significance and provide recommendations for school improvement.
- 7) Provide information on school quality and disseminate good practices of teachers, schools, and authorities exercising the rights and responsibilities of school owners and municipality administrations.

Accordingly, seven potential mechanisms of school improvement were identified and further examined during the interviews, leading to a selection of four mechanisms and two intended effects. The following sections explain each mechanism and intended effect starting with the objective that was excluded from the conceptual model.

4.3.1. Evaluating the effectiveness of cooperation between the school and its proprietor

This objective relates to the accountability of the school and its founder for the quality of education they provide. As such it implies a summative evaluation providing a basis for political decisions to potentially sanction or reward these institutions depending on the school performance. However, as discussed previously Lithuania’s external school evaluation is geared towards the improvement of schools through the provision of feedback without any accompanying sanctions imposed on the school or its founder in case of unsatisfactory performance. While separate areas of schooling are assigned grades, there has been no precedent of a school receiving *level N* in one of the areas of evaluation and failing to improve after a follow-up visit (Interviewee 5, personal communication, February 21, 2022). Discussions at the National Agency for Education are currently taking place on what course of action should be taken in the aforementioned scenario (Interviewee 5, personal communication, February 21, 2022). Thus, while external evaluation may reveal which

municipalities and schools are faring poorly in terms of education quality, it is not considered a mechanism of improvement since no further administrative or legal action is taken against either of the institutions.

4.3.2. Setting expectations on quality standards and acceptance of feedback

One of the core mechanisms through which Lithuania's external school evaluation system intends to promote school improvement is by setting expectations on quality standards which are described in its evaluation rubric. As one expert put it, external evaluation helps the school to answer the following question: "Are we setting the right goals and the right means to achieve them? Or, knowing our context and seeing our potential, are our ambitions set too low?" (Interviewee 6, personal communication, March 8, 2022). These answers are provided by external evaluators who are seen to bring an objective view or to "expand the school's ability to know about itself more than it currently does" (Interviewee 2, personal communication, February 14, 2022). For example, several interviewees pointed out that self-evaluation data reveals that schools sometimes lack constructive self-criticism and tend to justify their vision or inaction by assigning blame to pupils, parents, or other contextual factors that are supposedly beyond their control (Interviewees 4 and 6, personal communication, February 17 and March 8, 2022). In other cases, schools' vision of good education does not maintain a balance between the different areas of schooling as prescribed by the "Good School Conception". On the one hand, some schools have been noted to hide behind the more *soft aspects of schooling* such as positive school culture and organization of cultural events, while ignoring the more negative aspects such as comparatively low student achievement (Interviewee 4, personal communication, February 17, 2022). On the other hand, at least in one case, a team of evaluation experts revealed that a school was performing highly in national examinations as a result of nearly all extracurricular activity time being dedicated to tutoring, effectively compensating for the subpar quality of classroom instruction (Interviewee 2, personal communication, February 14, 2022).

In the case of risk evaluation, schools are motivated to accept the standards and incorporate the feedback into their practice by the follow-up visit and the prospect of losing 15% of the funding dedicated to school improvement. However, given that schools undergoing comprehensive and thematic evaluations are not subjected to sanctions, the question remains as to why the school staff should embrace the evaluator's normative view of education and trust his expertise to provide a judgment on the quality of their work. The interviews revealed that this objective may be achieved in a number of direct and indirect ways.

Firstly, fostering agreements on the conception of education quality should help to build a common ground between the evaluator and the school. If, as the "Conception of Formal Education Quality Assurance System" stipulates, the indicators used in external evaluation stem from a conception of education quality agreed by all of the school members, they should perceive the quality standards as legitimate and internalize them. Secondly, the influence of education standards specified in the evaluation procedure may occur even without feedback,

before the school visit as the school anticipates the upcoming evaluation and aligns its activities and documentation in line with the standards (Interviewee 1, personal communication, January 27, 2022). While there are no formal consequences for low performance, schools may be additionally motivated to avoid a negative grade. Thirdly, creating opportunities for teachers and school leaders to be active participants in the evaluation process is seen as an effective way to promote the adoption of quality standards. For instance, teachers are expected to take a proactive role in inquiring the evaluators for feedback and taking the time to understand what aspects of teaching need to be improved and why (Interviewee 5, personal communication, February 21, 2022). Moreover, the opportunity to propose amendments or comments to the evaluation report is aimed at creating a sense of ownership among the school staff in the evaluation process (Interviewee 6, personal communication, March 8, 2022). Fourthly, evaluators are expected to change the beliefs and expectations of school practitioners more directly by offering oral and written feedback as well as making them aware of other stakeholders' views (Interviewee 4, personal communication, February 17, 2022).

4.3.3. Evidence-based decisions on school support

Another important mechanism through which external school evaluation is expected to encourage school improvement is evidence-based decisions on school support and consultation. The feedback is meant to be used not only by the school under evaluation, but by policy-makers at the municipal and national policy as well. Once municipalities are informed about their schools' external evaluation results, they must approve and support the implementation of the improvement plan (Ministry of Education, Science and Sports, 2007a). The plan should include concrete measures to improve school quality ranging from investment in teacher and principal training, and hiring new staff to the renewal of learning toolkits (Interviewee 2, personal communication, February 14, 2022). Some experts also pointed out that improvement actions are often motivated by the publication of municipalities' performance in a national report on the quality of general education schools. Recounting his experience at the Ministry of Education, Science and Sports one expert said:

We have encountered this many times – we showed that one municipality is faring poorly in terms of the level of digitalization in schools and the mayor instantly took action and found the funding as he did not want the municipality to look bad (Interviewee 4, personal communication, February 17, 2022).

The same national report on the quality of general education schools provides evidence for decision-making at the national level. Using aggregated self- and external evaluation data at the municipal, regional and national levels, it outlines yearly trends in the quality of Lithuanian schools, thereby enabling policy-makers to monitor the school system on a yearly basis. Moreover, these reports offer a more extensive analysis showing the relationship

between the quality of different school aspects and other factors, such as students' socio-economic and ethnic background, classroom size, teacher and principal competencies, as well as schools' use of self-evaluation for strategic planning (National Agency for School Evaluation, 2015). According to several experts, these reports function as a sort of formative evaluation at the national level, that is, they provide the decision-makers at the Ministry of Education, Science and Sports and the National Agency for Education with the necessary information which is used to plan school investment (Interviewee 1 and 5, personal communication, January 27 and February 21, 2022). For instance, the indicators for the school improvement program "Quality Basket" have been taken from these analyses which have identified the most impactful factors for school improvement (The Ministry of Education, Science and Sports of the Republic of Lithuania, 2018).

4.3.4. Providing information on school quality and disseminating good practice

Following the evaluation, both written and oral feedback may be used by stakeholders. As discussed earlier, the key actions for school improvement are expected to be taken by the school administration in the implementation of the school improvement plan and by teachers altering their practice in response to the individual feedback they have received. Moreover, national and municipal policy-makers are expected to make use of the evaluation data in school policy planning and design. The question remains, to what extent the information on school quality is put to use by other stakeholders.

In many external evaluation systems, parents are expected to put pressure on schools to improve following the publication of evaluation results through the strategic use of *voice*, *choice*, and *exit* (Ehren et al., 2013). The interviews with policy-makers revealed that there is no such expectation in Lithuania's external evaluation system for several reasons. Firstly, school choice is limited by Lithuanian law according to which priority should be given to pupils whose place of residence is in the same district as the school (Ministry of Education, Science and Sports, 2004). Secondly, the publication of external evaluation findings initially was not aimed at informing parents or incentivizing school choice and competition (Interviewee 1, personal communication, January 27, 2022). This is also evident in the fact that until 2019, only summaries of the evaluation reports were available online and they were not provided in a user-friendly format that would allow direct comparison with other schools. In fact, no expert could explain the rationale behind choosing such a publication format, nor the underlying logic of how it is supposed to ultimately help school improvement. While the full risk evaluation reports have been made available online since 2019, their potential to inform parents about school quality is nevertheless limited due to the inconvenient format, namely a long report written in a technical language that calls for "deciphering" by an education professional as one policy-maker put it (Interviewee 2, personal communication, February 14, 2022). Lastly, the justification given by one policy-maker regarding the decision to publish full reports does not relate to parent engagement but rather to the attempt to familiarize schools with the notion of soft inspection and relieve the associated pressure (Interviewee 5, personal communication, February 21, 2022).

External school evaluation in Lithuania is also expected to engender school improvement by disseminating good school practices. This may happen in a number of formats from simply studying the evaluation reports, to organizing school visits and conferences to creating opportunities for peer learning (Interviewees 1 and 4, personal communication, January 27 and February 17, 2022). The most systematic effort to encourage the dissemination of good practice was introduced with the “Quality Basket” program in which high performing schools can earn money by mentoring underperforming schools and helping them to implement the improvement plans following risk evaluation (The Ministry of Education, Science and Sports of the Republic of Lithuania, 2018). To sum up, the provision of information on school quality cannot be considered a full-fledged mechanism as it is not clear how it is supposed to lead to school improvement. Showcasing the good school practices in various forms, however, seems to be a viable channel through which external school evaluation could engender quality improvement.

4.3.5. Strengthening schools’ improvement capacity

As discussed earlier, Lithuania’s external school evaluation is a dual system, both relying on self-evaluation data and aiming to strengthen the school’s self-evaluation capacity. This intended effect is expected to be reached through the provision of feedback and setting expectations on what good education is. In particular, external evaluation data is seen as a way to enhance the school’s decision-making capacity. According to several policy-makers, it is intended to provide the school with data that could be then grouped to generate evidence of what works to enhance student learning and achievement, which would in turn underpin certain managerial decisions laid out in strategic and school improvement plans (Interviewee 1 and 2, personal communication, January 27 and February 14, 2022). However, there is an expectation that external evaluation as a supporting intervention would become obsolete over time as the school becomes a self-sufficient learning organization capable of agreeing on common goals, identifying its weaknesses, and effectively remedying them (Interviewee 2, personal communication, February 14, 2022). Thus, combining self- and external evaluation data is expected to provide an impetus for a school to strengthen each stage of its quality assurance cycle, thereby leading to higher school quality.

4.3.6. Taking improvement actions to create better learning conditions

External evaluation’s goal of improving student achievement and raising education quality, implies an expectation that schools should create better learning conditions for their pupils. In the “Good School Conception”, these conditions are described by the three areas of schooling, namely, education and pupils’ experience, education environment as well as leadership and management, which are in turn expected to produce desirable results including pupil academic achievement and personal growth (Ministry of Education, Science and Sports, 2015). Since external evaluation does not provide prescriptions on how to raise the quality of a school and merely points out in which areas the school is lagging behind, the school leader is expected to play a key role in transforming the school. In the words of one expert, external

evaluation indicators are inevitably interrelated and the job of the school leader is to find the right “pressure point” through which organizational change should be initiated (Interviewee 2, personal communication, February 14, 2022). Another expert echoed these words by comparing the school to a “mechanical clock whose ciliary gear has to be turned in such a way as to not break the mechanism” (Interviewee 1, personal communication, March 27, 2022). Both of these comments reverberate the expectation that schools will take improvement actions by themselves after the external evaluation team highlights some of the main areas of concern.

However, the main partner supporting the schools’ effort to improve is expected to be the municipality education departments together with the education support centers (Ministry of Education, Science and Sports, 2015). They are expected both to financially support the implementation of improvement plans as well as monitor the school’s progress. As the policy-makers realized over the years that many municipalities often lack the resources to help schools, school improvement programs that allocate additional funds specifically for the implementation of improvement plans were instituted to fill in this gap (Interviewee 4, personal communication, February 17, 2022). To sum up, while improvement actions technically fall outside the scope of the external school evaluation system there is an expectation that schools with municipal support will implement the necessary changes leading to better learning conditions and ultimately higher student achievement and education quality.

4.4. Tensions in the conceptual model

The analysis of Lithuania’s external school evaluation system also highlighted some of its underlying tensions. The core tension is between setting standards for all schools and acknowledging their unique vision of education. On the one hand, the “Good School Conception” functions as a philosophical document providing a reference for schools to think about what they consider good education and design self-evaluation criteria accordingly. On the other hand, the “Good School Conception” is used as a policy document providing a basis for external evaluation indicators which set expectations on quality standards. A similar tension is seen in the purpose of the system. Formally, it performs a formative function by helping schools to identify areas where they can improve. However, the system retains aspects of summative evaluation, such as the levels assigned to different aspects of schooling. *Level N* signifying unsatisfactory school quality allows to identify schools that are not up to a standard and thus is part of the policy-makers’ rather than the school’s quality assurance playbook. Thus, external evaluation may send mixed signals to the school community as they are expected to both satisfy state standards and also retain their own vision of education and school organization.

Moreover, the aggregation of evaluation results using the five levels enables quantitative analysis of school quality functioning as a part of the education monitoring and quality assurance system. As such, it may contribute to school improvement indirectly through the mechanism of evidence-informed policy on school support, yet it serves a summative rather

than formative function. This logic was particularly apparent in one expert's explanation as to why the aspiration for a full cycle school evaluation has been abandoned: "Today, not all schools have experienced external evaluation, however, qualitative data have a tendency to saturate until there is no point in further evaluation as you see the same trends and do not find unique cases studies anymore" (Interviewee 5, personal communication, February 21, 2022). In another case, the expert spoke of how important the sample size is in generating data for further analysis:

Given that one evaluation expert observes six lessons a day, the visit lasts the entire school week and there are five evaluation experts, we get 150 observed lessons per school visit, which is a very solid sample. Also, having in mind that the evaluation protocol contains seven lines, there is a lot of qualitative and quantitative data (Interviewee 5, personal communication, February 21, 2022).

While thinking in terms of samples allows one to draw conclusions about the school system as a whole and is thus useful from the standpoint of a policy-maker, it conflicts with the logic of formative evaluation that is concerned with helping each school to improve through self-reflection. A proportional evaluation cycle may be more resource-efficient, yet it does not benefit every school, as those schools that remain unevaluated do not receive feedback that is specific to their context and have very little to gain from the general conclusions of the yearly reports on school quality.

The purpose of the external evaluation system is contested among the experts themselves. According to one interviewee, the lack of quantitative indicators that would allow monitoring school progress at the municipality, regional and national levels is one of the key shortcomings of the system (Interviewee 3, personal communication, February 14, 2022). The expert contended that the only comparable quantitative data comes from national examinations at grades four, six, and eight and at the end of upper secondary schooling, while the yearly reports on school quality are only discussed at a municipal level, while the education community is unsure about how to use them (Interviewee 3, personal communication, February 14, 2022). Therefore, ultimately external evaluation provides an opinion or a statement on the merit of a school, yet they do not amount to a clear analysis of why schools fall short of national targets or manage to reach them (Interviewee 3, personal communication, February 14, 2022).

Another interviewee disagreed with the contention that the system lacks objective monitoring indicators and called it "the longing for mechanical thermometers, which do not exist nor could be relied on in the social sphere" (Interviewee 4, personal communication, February 17, 2022). In particular, the expert contended that external evaluation cannot provide an objective measurement of a school's added value as there is no agreement on what

it entails in Lithuania's education community. Instead, the expert defended a different notion of external school evaluation:

I believe in dialogue [between a school and evaluation experts], in observation of what is actually happening and in attempts to understand each other . . . Often, an experienced evaluation expert has arguments that allow one to see the gaps in one's worldview, to start thinking and acting differently than before (Interviewee 4, personal communication, February 17, 2022).

Therefore, the distinct notions of external evaluation among experts point to another source of tensions between a school and system-level functions.

4.5. Teachers' and school leaders' perception of external school evaluation

The second part of the analysis aims to assess teachers' and school leaders' perception of external evaluation as one of the determinants of its effectiveness. While the analysis does not aim to provide a definitive assessment of the system's impact, it offers some indications of how the implementation of external evaluation differs from its conceptual model. Given that some mechanisms are external to the evaluated school, this analysis concerns only part of the conceptual model. Namely, the mechanisms of accepting feedback, setting expectations on standards, the purpose of school evaluation, and improvement actions are discussed. Moreover, some unintended consequences of external evaluation for the school community are outlined as well.

4.5.1. The role of stress

The analysis of interviews revealed that stress was common in all schools that were subject to external evaluation. All interviewed teachers acknowledged that they felt stress in anticipation and during the school visit, while also contending that it affected some teachers more than others: "Tension could be felt and I think that teachers who have experienced external evaluation [before] dealt with it more easily . . . But I don't believe that there was a single person who survived this process without stress and tension" (Interviewee 5, personal communication, March 29, 2022). Moreover, according to a few teachers, the fact that evaluation experts would come to observe a lesson without preliminary notice was particularly nerve-wracking:

You come to work and you are constantly waiting in tension because you don't know if somebody will come or not . . . Why couldn't they say: "I will come to your class on Monday". . . You feel like a pupil who somebody wants to catch [for doing something

bad]. So I would say that this aspect is very inhumane (Interviewee 2, personal communication, March 30, 2022).

The fear and tension among teachers were also evident in the fact that all of them were eagerly waiting for the end of the school visit, referring to their experience as “survival” (Interviewee 2, 3, 5, 6, personal communication, March 30, 29, April 21, 2022). Interestingly, in *school A*, the evaluation experts were aware of teachers’ attitude towards them: “When we received the evaluation report, they told us: “You will be so happy when you [the teachers] will be able to wave to us behind our back”. And, indeed, we were happy” (Interviewee 2, personal communication, March 30, 2022). Another stress enhancing factor was the fact that evaluation reports did not always guarantee anonymity as it was relatively easy to identify whose lesson was graded negatively when a school had just one or two teachers for a given subject (Interviewee 5, 6, personal communication, March 30, April 21, 2022).

Ultimately, evaluation was even seen as detrimental to teachers’ health, motivation, and self-esteem, especially in cases where they received negative feedback (Interviewee 2, 5, personal communication, March 29, 30, 2022). One teacher also noted that the fear of an outsider coming to observe your work also had an impact on pupils who told the teacher that they remained silent during the lesson as they were “afraid to say something that does not make sense” (Interviewee 3, personal communication, March 30, 2022). While both school leaders acknowledged that evaluation caused stress, they also saw it as a mobilizing factor that encouraged collaboration and more effort among the school staff (Interviewee 1, 4, personal communication, March 25, 30, 2022). Some teachers echoed the notion of increased collaboration between colleagues as teachers shared the feedback they received in order to alert their colleagues (Interviewee 3, personal communication, March 30, 2022).

4.5.2. Acceptance of feedback

While teachers differed in their general attitude towards external school evaluation, only one of them regarded the feedback as useful for their practice. Most interviewed teachers considered themselves to be professionals who do not need external guidance on how to plan and lead a lesson (Interviewee 3, personal communication, March 30, 2022). Several interviewees indicated that evaluation experts lacked the time for reflection and discussion of individual feedback with teachers (Interviewees 3 and 5, personal communication, March 30 and 29, 2022) as they had to fill in a large amount of paperwork and “run from one lesson to another” (Interviewee 3 and 5, personal communication, March 30 and 29, 2022). The fact that evaluation experts had to rush through the process and were obligated to do so by formal requirements, in turn, undermined the legitimacy of the feedback:

You see how they collect the data – class observation mostly consists of writing, checking if the lesson meets the criteria, sometimes they barely even raise their eyes. So, I am

thinking – how could they have even evaluated anything if they didn't see anything? They are writing all the time and then they put [the comments] into a table and generate some kind of numbers (Interviewee 5, personal communication, March 29, 2022).

Other teachers echoed doubts if evaluation results can reflect the value of their practice arguing classes differ significantly in their ability and that given their short stay, evaluation experts can have only a superficial understanding of the class (Interviewees 3 and 5, personal communication, March 30 and 29, 2022). Two teachers also expressed a concern that evaluation results may be subjective and that a different evaluation expert would have (or in fact, did) come to different conclusions (Interviewees 3 and 6, personal communication, March 30 and April 21, 2022). Moreover, none of the teachers reported having had the opportunity to provide comments on the final evaluation report with one interviewee stating that even if such a possibility existed, most teachers would have been hesitant to do so given that it would have prolonged the school (Interviewee 5, personal communication, March 29, 2022). Lastly, several interviewees reported an incongruence between individual and school level feedback laid out in the evaluation report with the former being more positive than the latter (Interviewees 3, 5, 6, personal communication, March 30, 29, April 21, 2022).

School leaders had a more positive outlook regarding feedback claiming that it has helped them to critically reflect on their work and thereby improve their practice (Interviewee 1, 4, personal communication, March 25, April 21, 2022). The leader of *school A* noted that they were able to make adjustments during the evaluation process by incorporating the feedback they received at the end of the day into the upcoming day's lessons, which was taken as evidence of effective leadership and management by the evaluation experts (Interviewee 1, personal communication, March 25, 2022).

4.5.3. The perceived purpose of external school evaluation

A common thread in teachers' attitudes towards external evaluation was a conviction that it serves a summative rather than a formative function. One teacher's words are exemplary of this sentiment: "External evaluation works not as a motivating factor [saying]: "if not today, tomorrow you will succeed" but rather as a judgment on whether the teacher is good or bad" (Interviewee 5, personal communication, March 29, 2022). The interviewee also observed that when the school itself hires consultants, the level of tension among the staff is lower than during external evaluation even if it is performed by the same people (Interviewee 5, personal communication, March 29, 2022). Commenting on the possible reasons behind the difference, the teacher argued that the contractor of the service determines the nature of the task, so when a school hires consultants, they tend to take special care in understanding its needs and not causing additional stress (Interviewee 5, personal communication, March 29, 2022). Other interviewees also pointed out that external evaluation may be useful for policy-makers at the Ministry of Education, Science and Sports who can use the evaluation

data for “creating educational strategies” (Interviewees 1 and 2, personal communication, March 25 and 30, 2022).

The formative function was seen to be present in external evaluation by some interviewees, especially by school leaders who tended to be more convinced that it can help the school to improve directly through the provision of feedback (Interviewees 1, 4, 6, personal communication, March 25, 30, April 21, 2022). Leaders of both schools underscored the importance of being transparent about one’s work and embracing external evaluation as a necessary to reflect critically on the direction the school is going towards (Interviewee 1, 4, personal communication, March 25, April 21, 2022).

4.5.4. Preparing for evaluation and fitting into the standard

Nearly all interviewees reported that their schools focused on preparing the documentation as they were notified about the date of the school visit about a year in advance. While the leader of *school A* emphasized that external evaluation primarily focused on the quality of teaching rather than documentation (Interviewee 1, personal communication, March 25), teachers tended to view the role of documentation in external evaluation as excessive and burdensome (Interviewee 2,5, personal communication, March 30, 29). Moreover, in anticipation of the visit, both schools organized seminars that were aimed at helping teachers to devise and deliver a lesson that meets the evaluation criteria. Teachers also made individual efforts to align their work with a standard: “As the teacher prepares for evaluation, especially if he is less experienced, he tries to fit himself into a standard, so that in the eyes of the evaluation experts, the lesson appears as good as possible” (Interviewee 5, personal communication, March 29). Similarly, most teachers regarded the evaluation criteria as restricting their practice and struggled to make sure that all the required aspects were included in the lesson (Interviewees 3, 5, 6, personal communication, March 30, 29). As an example, one teacher described the feedback she received for finishing the lesson sometime after the bell rang as fitting her into a “frame” (Interviewee 3, personal communication, March 30). The lasting impact of the standards set by the school evaluation is not clear as some teachers reported using the same lesson plans some years after the school evaluation but also acknowledged that the level of mobilization decreased once the evaluation experts left the school (Interviewee 3, 6, personal communication, March 30, April 21, 2022). In one teacher’s opinion, the standards were contradictory as they required both to “make the lesson playful, interesting, informative and ensure high results” (Interviewee 6, personal communication, April 21).

4.5.5. Reputational concerns and stakeholder involvement

Interviewees from both schools emphasized the importance of stakeholders' role in external evaluation. According to school leaders, the evaluation experts talked not only to the teachers and the school administration but also tried to grasp the perspectives of pupils and parents either through surveys or by conducting private interviews (Interviewee 1, 4, personal communication, March 25, April 19, 2022). Moreover, in both schools, the evaluation report

was presented to a school board composed of the already mentioned stakeholders and a representative from the municipality education department (Interviewee 1, 4, personal communication, March 25, April 19, 2022).

Interviewees from *school B* had different opinions on whether external school evaluation had an impact on the school's reputation. One teacher expressed a concern that public reporting of evaluation results can damage the school's popularity as "nobody delves into the nuances [of the evaluation report]" and focuses on whether the school was evaluated positively or negatively (Interviewee 5, personal communication, March 29, 2022). Her colleagues, however, disagreed by claiming that they did not observe a significant change in the school, even though it received a rather negative assessment (Interviewee 4, 6 personal communication, April 19, 21, 2022).

According to the leader of *school A*, school's reputation was one of the main driving factors in preparing for external evaluation: "We wanted to perform as best as possible knowing that seven years after the external audit we will have a certain evaluation flag with which we will march further" (Interviewee 1, personal communication, March 25, 2022). Overwhelmingly favorable feedback worked as a catalyst for school's performance which was also reflected in the reaction of external stakeholders:

This movement . . . lasted no less than three years as we were rapidly going up in terms of innovation, results and this was also felt from the outside. At the time we had 863 pupils and we quickly filled up to 1000 pupils. So .. the environment and the pupils reacted as well (Interviewee 1, personal communication, March 25, 2022).

The teachers shared the opinion of the school leader by contending that more interest in the school was felt after the evaluation by parents as well as other schools as they were interested in learning from their school's practices (Interviewee 2, 3, personal communication, March 30, 2022). They could not, however, identify the exact channel through which evaluation results would reach parents pointing to the potential role of both public reporting and word of mouth communication.

4.5.6. Improvement actions

The improvement actions taken after external evaluation differed significantly between the schools. *School A* received largely positive feedback and was trusted to address the identified weaknesses with minimal supervision from the municipality's education department. An improvement plan was implemented by method groups consisting of teachers and school personnel who worked towards diversifying and modernizing student assessment methods, as well as strengthening their association with learning goals by organizing various seminars and consultations (Interviewee 1, personal communication, March 25, 2022). Subsequently, the effectiveness of these measures was monitored through self-evaluation whose focus has

been adjusted according to the received feedback (Interviewee 1, personal communication, March 25, 2022).

Given that *school B* underwent risk evaluation, evaluation experts singled out more areas for improvement, such as individualized instruction methods, monitoring of pupil progress, and school agreements on quality conception and improvement (National Agency for Education, 2019). According to the school leader, the conclusions of the evaluation report coincided with the results of previous self-evaluation reports, showing that the key issue was the unwillingness of the school staff to uphold the previously made commitments to implement certain changes (Interviewee 4, personal communication, April 19, 2022). Following the visit, method groups were created to analyze the evaluation report and identify specific actions to address the weak aspects of school activity (Interviewee 5, personal communication, March 29, 2022). These actions included various types of training aimed at enhancing teacher competencies as well as the introduction of a mandatory tool to monitor pupils' daily progress (Interviewee 6, personal communication, April 21, 2022). Moreover, the school received additional financial support for implementing the improvement plan from the government as part of the "Quality Basket" project. The additional help came with more extensive monitoring by the Ministry of Education, Science and Sports, and the school was obligated to submit yearly progress reports (Interviewee 5, personal communication, March 29, 2022).

5. Conclusion: findings and policy recommendations

5.1. The conceptual model of external school evaluation

This research paper has built a conceptual model of Lithuania's external school evaluation by describing its core characteristics, the key mechanisms of improvement, and intended effects (see Graph 1). The system has been shown to be aimed towards school improvement with no formal sanctions imposed on underperforming schools with the exceptions of conditional financing and follow-up school visits for schools participating in the 'Quality Basket' project. In addition, it was outlined how the system utilizes evaluation indicators and the level system to judge the quality of school activities, which are then summarized in individual school evaluation and yearly reports on school quality. Lithuania's external school evaluation has been described as a sequential system, relying on self-evaluation as a basis for quality assurance at the school level. Lastly, public reporting policy has changed over the years from publishing only the summaries of school evaluation reports to disclosing their full versions.

In the second part of the conceptual model, the main mechanisms through which external school evaluation is expected to improve schools were spelled out. In particular, the document analysis and interviews with experts partially confirmed the first hypothesis by revealing the presence of four mechanisms. Firstly, as hypothesized, the standards nested in the evaluation criteria are meant to set expectations for schools on what ideal education looks like. This effect may occur prior to the school visit as schools align their activities with the evaluation criteria in order to avoid a negative grade or during and after evaluation as they

receive feedback from the evaluation team. Also in line with the hypothesis, the second mechanism of school improvement is individual and school level feedback which provides actionable information on which areas of teacher practice and schooling could be improved.

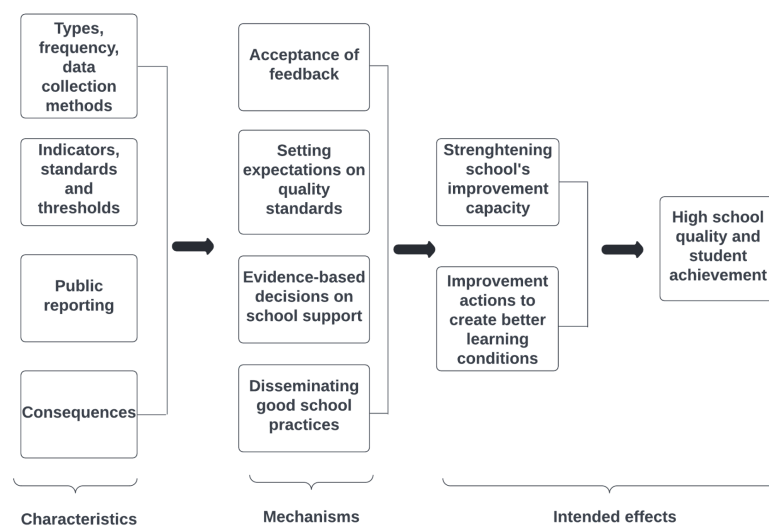
Secondly, a new system-level mechanism was identified. Namely, the individual and school level feedback is intended to be aggregated to represent the quality of Lithuanian schools at the municipal and national levels and is thereby expected to provide evidence on the appropriate school support for decision-makers. There is also an expectation that the effectiveness of this mechanism may be enhanced by publicizing school performance data at the municipality level, thereby incentivizing local policy-makers to take school improvement actions.

Thirdly, contrary to the hypothesis, parent involvement is not seen as a potential channel of improvement given that they have no formal right to choose a public school, and public reporting on school quality is not targeted at them. However, there is an expectation that external evaluation will help to identify good school practices which in turn will have a positive influence on underperforming schools through peer learning opportunities, conferences, and other initiatives.

Fourthly, the analysis showed that the notion of a learning organization is central to the conceptual model as the aforementioned mechanisms are expected to promote the use of self-evaluation and data in school management. It also revealed an expectation that schools will collaborate with the municipalities to devise an improvement plan that is adapted to the specific school context and aims to create better learning conditions for all pupils. Both of these intended effects are envisioned to ultimately improve school quality and student achievement.

Lastly, it was argued that the conceptual model is only partially coherent. The core tension is between setting quality standards and promoting a school's uniqueness, which arises due to the fact that the "Good School Conception" does not set clear standards for schools but encourages them to find a version of education quality that fits their context. Nevertheless, it is used to devise evaluation indicators that define and promote a standardized version of school quality. Similar to this is the tension between providing data for internal and external quality assurance. In recent years, national policy-makers have been mostly concerned with attaining enough quantitative data for a representative sample of evaluated schools, allowing for macro-level analysis and policy planning. This stands in contrast to a formally declared formative function of school evaluation, which is oriented towards providing qualitative feedback and setting schools on a path to continuous reflection and improvement.

Graph 1: The conceptual model of Lithuania's external school evaluation system



5.2. Teachers' and school leaders' perceptions of external school evaluation

The last section of the analysis confirmed the hypothesis that teachers tend to view school evaluation more negatively than school leaders due to stress and perceived threats to their professional identity. Moreover, the way in which school evaluation is implemented in Lithuanian schools was shown to differ in some ways from its conceptual model.

Firstly, the analysis corroborated a large number of studies showing stress to be the main by-product of school evaluation. Teachers expressed concerns about the level of stress caused by such practices as not announcing which class will be observed in advance and the lack of de facto anonymity in the evaluation reports. Anxiety related to the anticipation of the evaluation team's arrival and its judgment on the quality of teachers' work, as well as eagerness for the school visit to end were found to be the prevailing attitudes among teachers. While school leaders acknowledged the presence of stress, they were also more inclined to see it as a mobilizing factor, which is necessary for school improvement.

Secondly, in line with previous scholarly work, it was shown that most teachers do not accept the feedback either because they see themselves as professionals who do not need external guidance or because it was not relevant to their lessons. One of the reasons why teachers did find the feedback relevant was the lack of opportunity to discuss it with evaluation experts and to have a say in the drafting of the final evaluation report. Several teachers also expressed doubts regarding the objectivity of experts' conclusions given their preoccupation with filling in documents during the class observation and the lack of familiarity with the specificity of a given class of pupils. Moreover, an incongruence between individual feedback and the conclusions of the evaluation report was identified by several interviewees. Indicatively, some interviewees saw the incongruence as evaluation experts' unwillingness to express criticism to individual teachers. This finding departs from the conceptual model of school evaluation, according to which dialogue between evaluation

experts and teachers should encourage the acceptance of feedback and thereby lead to school improvement. Thus, the effectiveness of feedback provision in school improvement is called into question.

Thirdly, the analysis demonstrated that the school staff perceives the purpose of school evaluation differently. Most teachers tend to see it as serving a purpose external to the schooling itself – either as a judgment on the quality of their work or a procedure that is more useful for education policy planning. In addition, among teachers, there was a sense that school evaluation is a mere formality, a one-time event that places a burden on their work, and has to be endured. However, school leaders and a few teachers also pointed out that school evaluation serves a formative purpose by pointing out certain weaknesses of their practice and thereby creating conditions for improvement. This result seems to underline a tension between the mechanisms of accepting feedback and evidence-based decisions on school support, as evaluation experts are forced to balance between tasks that serve these two goals.

Fourthly, interviews showed that schools try to align their activities in line with the standards specified in the evaluation procedure by preparing for school evaluation in advance and during the school visit as prescribed by the conceptual model. Nearly all teachers regarded this mechanism as restricting their work and limiting their creativity, and having a rather short-lived effect on their practice. This finding suggests that the impact of setting expectations on school quality is present in Lithuanian schools but may be limited to the short-term and may cause some unintended consequences.

Fifthly, the school's reputation is seen as an important factor in school evaluation by both teachers and school leaders. Its influence has been shown to depend on the evaluation results – a negative evaluation leading to fears of declining reputation, while a favorable one enhancing the school's prestige and self-confidence. This was particularly clear in the case of a high-performing school that has experienced increased pupil enrollment following the evaluation and was shown as an example to other schools. This result contradicts the conceptual model which does not intend to use the tactic of *naming and shaming* schools and may unintentionally reinforce the inequality in school performance.

Lastly, improvement actions to address the weak areas of school activity were taken by both schools, most notably in the form of seminars aimed at enhancing teacher competences. The level of involvement of the local and national governments differed between schools due to their different levels of quality. The school which took part in risk evaluation received extensive financial support after which its progress was closely monitored, while the school taking part in the comprehensive evaluation was trusted to monitor its own progress through self-evaluation. This result is consistent with the conceptual model's intention to use external evaluation to strengthen schools' improvement capacity.

5.3. Policy recommendations

Based on the findings of this paper, a number of recommendations for education policy-makers in Lithuania and other countries can be described. Firstly, strategies to minimize the negative impact of stress on the school staff should become part of the school evaluation design. While fully eliminating stress may be impossible and undesirable, practices such as unannounced school visits should be replaced with consultations between the school staff and the evaluation team regarding the goals and the procedure of school evaluation. Moreover, stress among teachers could be reduced by ensuring the anonymity of evaluation results. These principles should help the school staff to perceive the evaluation process as beneficial for their own work, which in turn should enhance its potential to encourage school improvement.

Moreover, the purpose of school evaluation should be clarified in the legislation and ensured in practice. While in theory, it may seem appealing to combine the formative, summative, and policy planning functions, in practice, they imply different demands on evaluation experts and thus should be regarded as tradeoffs. For instance, if school evaluation is envisioned to perform a formative function, the feedback has to be accepted by teachers. In order for teachers to accept it from an evaluation expert, a certain level of trust and mutual understanding has to be reached. This in turn requires ensuring a level of workload and competences that allow evaluation experts to lead a dialogue and thereby have an impact on teachers' practice. Moreover, in order to create a sense of ownership among the school staff, opportunities for participation at every stage of the evaluation process should be created.

The policy planning function of school evaluation should not be discounted. To design effective interventions aimed at school improvement, policy-makers need data on school performance and processes. However, policy planning of this kind requires quantitative data which enables the comparison of schools across a set of indicators and is therefore difficult to combine with the qualitative approach of formative evaluation. Thus, policy-makers should consider separating the policy planning and formative functions in order to maximize the effectiveness of both school and system-level quality assurance cycles.

Thirdly, another way to ensure that school evaluation is impactful is by having a clear and consensus-based conception of education quality. While there can hardly be one conception of education quality agreed upon by all of the education community members, efforts should be made to bridge the gap between the different notions in order to set clear expectations on quality standards. Agreements on how to properly measure a particular conception of education quality may be just as important for the effectiveness of external school evaluation.

5.4. Limitations and suggestions for future research

This paper has several limitations. Firstly, given time constraints, the analysis of stakeholder perceptions included a very small number of interviewees and is therefore not representative of the teachers' and school leaders' attitude towards school evaluation. Therefore, future

studies should focus on having a robust sample of interviews with the school staff in Lithuanian schools. Secondly, this paper has focused on teachers and school leaders as the main agents of school improvement. However, the stakeholder range could be expanded to have a more nuanced view of school evaluation. In particular, future studies could focus on evaluation experts as street-level bureaucrats who try to rectify the contradictions and shortcomings of official policy. Thirdly, stakeholder perceptions provide only a limited understanding of the impact of school evaluation. Measuring the impact more directly by focusing on pupils' performance in standardized tests is thus an important avenue for further research. The conceptual model built in this paper provides a solid foundation for studies of this kind as the identified mechanisms can be used to ascertain the causal influence of school evaluation on school quality.

References

- Allen, R., & Burgess, S. (2012). *How should we treat under-performing schools?: a regression discontinuity analysis of school inspections in England*. Bristol: CMPO.
- Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement?. *School effectiveness and school improvement*, 26(1), 32-56.
- Alvik, T. (1996). *School Self-Evaluation: A Whole School Approach*. CIDREE, Dundee.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*.
- Buitrago, C. (2015). Why We Need to Slow Down When It Comes to Evaluation / Browse Our Publications / Publications & Resources / HFRP - Harvard Family Research Project. [Archive.globalfrp.org](https://archive.globalfrp.org/publications-resources/browse-our-publications/why-we-need-to-slow-down-when-it-comes-to-evaluation). Retrieved 26 February 2022, from <https://archive.globalfrp.org/publications-resources/browse-our-publications/why-we-need-to-slow-down-when-it-comes-to-evaluation>.
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and program planning*, 2(1), 67-90.
- Cardno, C. (2018). Policy Document Analysis: A Practical Educational Leadership Tool and a Qualitative Research Method. *Educational Administration: Theory & Practice*, 24(4), 623-640.
- Chapman, C. (2001). Changing classrooms through inspection. *School Leadership & Management*, 21(1), 59-73.
- Chen, H. T., & Chen, H. T. (1990). *Theory-driven evaluations*. Sage.

Croxford, L., Grek, S., & Shaik, F. J. (2009). Quality assurance and evaluation (QAE) in Scotland: promoting self-evaluation within and beyond the country. *Journal of Education Policy*, 24(2), 179-193.

De Grauwe, A. (2007). Transforming school supervision into a tool for quality improvement. *International Review of Education/Internationale Zeitschrift für Erziehungswissenschaft/Revue Internationale de l'Education*, 53(5/6), 709-714.

De Wolf, I. F., & Janssens, F. J. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of education*, 33(3), 379-396.

Ehren, M. C., & Visscher, A. J. (2008). The relationships between school inspections, school characteristics and school improvement. *British journal of educational studies*, 56(2), 205-227.

Ehren, M. C., & Swanborn, M. S. (2012). Strategic data use of schools in accountability systems. *School effectiveness and school improvement*, 23(2), 257-280.

Ehren, M., Altrichter, H., McNamara, G. and 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), pp.3-43.

Ehren, M. C., 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.

Ehren, M., & Perryman, J. (2018). Accountability of school networks: Who is accountable to whom and for what?. *Educational Management Administration & Leadership*, 46(6), 942-959.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing*, 62(1), 107-115.

Her Majesty's Inspectorate of Education (2007). *How Good Is Our School? The Journey to Excellence: Part III*. <https://files.eric.ed.gov/fulltext/ED538663.pdf>

Faubert, V. (2009), "School Evaluation: Current Practices in OECD Countries and a Literature Review", *OECD Education Working Papers*, No. 42, OECD Publishing, Paris, <https://doi.org/10.1787/218816547156>.

Figlio, D., & Loeb, S. (2011). School accountability. *Handbook of the Economics of Education*, 3, 383-421.

The Government of the Republic of Lithuania (2002). *Resolution Concerning the Ratification of the School Improvement Program*.

<https://e-seimas.lrs.lt/portal/legalActPrint/lt?jfwid=9fbgsamoy&actualEditionId=GtttQKAiUu&documentId=TAIS.167382&category=TAD>

Gustafsson, J., Ehren, M., Conyngham, G., McNamara, G., Altrichter, H. and O'Hara, J., (2015). From inspection to quality: Ways in which school inspection influences change in schools. *Studies in Educational Evaluation*, 47, pp.47-57.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance?. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 24(2), 297-327.

Hargreaves, A. P., & Shirley, D. L. (Eds.). (2009). *The fourth way: The inspiring future for educational change*. Corwin Press.

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, 100864.

Honingh, M., & Urbanovič, J. (2013). School autonomy and leadership in Lithuania: can school leaders help schools to take a step forward?. *NISPAcee journal of public administration and policy*, 6, 61-78.

Hussain, I. (2015). Subjective performance evaluation in the public sector evidence from school inspections. *Journal of Human Resources*, 50(1), 189-221.

Janssens, F. J., & van Amelsvoort, G. H. (2008). School self-evaluations and school inspections in Europe: An exploratory study. *Studies in educational evaluation*, 34(1), 15-23.

Jones, K. L., Tymms, P., Kemethofer, D., O'Hara, J., McNamara, G., Huber, S., ... & 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.

Luginbuhl, R., Webbink, D., & De Wolf, I. (2009). Do inspections improve primary school performance?. *Educational evaluation and policy analysis*, 31(3), 221-237.

Yeung, S. Y. S. (2012). A school evaluation policy with a dual character: Evaluating the school evaluation policy in Hong Kong from the perspective of curriculum leaders. *Educational Management Administration & Leadership*, 40(1), 37-68.

Leeuw, F., (2003). Reconstructing Program Theories: Methods Available and Problems to be Solved. *American Journal of Evaluation*, 24(1), pp.5-20.

MacBeath, J. E., Schratz, M., Jakobsen, L., & Meuret, D. (2000). *Self-evaluation in European schools: A story of change*. Psychology Press.

Matthews, P., & Sammons, P. (2004). *Improvement through inspection: An evaluation of the impact of Ofsted's work* (London, Ofsted/Institute of Education).

Matthews, P., & Sammons, P. (2005). Survival of the weakest: The differential improvement of schools causing concern in England. *London review of education*.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American journal of sociology*, 83(2), 340-363.

The Ministry of Education, Science and Sports of the Republic of Lithuania (2002). *Ministerial order on the Internal Audit Methodology for General Education Schools*
<https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.188206?jfwid=>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2004). *Ministerial order on the List of General Criteria for Admission to State and Municipal General Education Schools and Vocational Training Institutions*.
<https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.236484/OdiUBSNmKI>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2007a). *Ministerial order on the Procedure Description for the Organization and Implementation of External School Evaluation*.
<https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.236484/OdiUBSNmKI>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2007b). *Ministerial order on the Ratification of the School Improvement Program Plus*.
<https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.312947/kJhSpBZyZM>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2008). *Ministerial order on the Conception of Formal Education Quality Assurance System*.
<https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.236484/OdiUBSNmKI>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2011). *How is External School Evaluation Seen? Education Issue Analysis*.
<https://www.nsa.smm.lt/wp-content/uploads/2020/07/Kaip-vertinamas-mokyklu-veiklos-kokybes-isorinis-vertinimas.pdf>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2014). *National Education Strategy (2013-2022)*.
https://www.smm.lt/uploads/lawacts/docs/687_e7870701e841e67d18a5377b19e7a57e.pdf
 [Accessed 22 October 2021]

The Ministry of Education, Science and Sports of the Republic of Lithuania (2015). *Ministerial order on the Ratification of the Good School Conception*.
<https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/46675970a82611e59010bea026bdb259?positionInSearchResults=0&searchModelUUID=a14ed8fb-f8e6-461f-ae33-722a83ed7898>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2015). *Self-evaluation Model and Indicators*.
<http://www.nmva.smm.lt/wp-content/uploads/2015/08/rodikliai.pdf>

The Ministry of Education, Science and Sports of the Republic of Lithuania (2018). *Ministerial order on the Procedure Description of the Allocation for the Quality Basket to General Education Schools.*

The Ministry of Education, Science and Sports of the Republic of Lithuania (2021). *Ministerial order on the Establishment of Topics, Questions and Evaluation Indicators for the Thematic External Evaluation of General Education Schools.*
<https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/0a098120e0f011eb866fe2e083228059?jfwid=-1cefbqp16t>

The National Agency for Education. (n.d.). *External Evaluation of General Education Schools.* Retrieved 23 April, 2022, from
<https://www.nsa.smm.lt/stebesenos-ir-vertinimo-departamentas/svietimo-istaigu-isorinis-vertinimas/bendrojo-ugdymo-mokyklu-isorinis-vertinimas/>

The National Agency for Education (2019). *External Risk Evaluation Report, [School B].*

The National Agency for School Evaluation (2015). *The Quality of General Education Schools. Annual report by the National Agency for School Evaluation.*
<https://www.nsa.smm.lt/wp-content/uploads/2021/01/Metinis-pranesimas-maketas-02.04.pdf>

OECD (2013), *Synergies for Better Learning: An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, <https://doi.org/10.1787/9789264190658-en>.

Ouston, J., Fidler, B., & Earley, P. (1997). What do schools do after OFSTED school inspections-or before?. *School Leadership & Management*, 17(1), 95-104.

Perryman, J. (2006). Panoptic performativity and school inspection regimes: Disciplinary mechanisms and life under special measures. *Journal of Education Policy*, 21(2), 147-161.

Rosenthal, L. (2004). Do school inspections improve school quality? Ofsted inspections and school examination results in the UK. *Economics of education review*, 23(2), pp. 143-151.

Phillips, D., & Ochs, K. (2004). Researching policy borrowing: Some methodological challenges in comparative education. *British Educational Research Journal*, 30(6), 773-784.

Shaw, I., Newton, D. P., Aitkin, M., & Darnell, R. (2003). Do OFSTED inspections of secondary schools make a difference to GCSE results?. *British educational research journal*, 29(1), 63-75.

Sharma, G. (2017). Pros and cons of different sampling techniques. *International journal of applied research*, 3(7), 749-752.

Standaert, R. (2001). *Inspectorates of education in Europe: A critical analysis.* Acco.

Stoll, L. (2009). Capacity building for school improvement or creating capacity for learning? A changing landscape. *Journal of educational change*, 10(2), 115-127.

Taylor, I. (2007). Discretion and control in education: The teacher as street-level bureaucrat. *Educational Management Administration & Leadership*, 35(4), 555-572.

Van Amelsvoort, G. H. W. C. H., Bos, K. T., Janssens, F. J. G., Klaver, L., Lelyveld, J., & Pol, M. (2006). Proportional supervision and school improvement from an international perspective. *Utrecht: Inspectie van het Onderwijs (Education Inspectorate)*.

Van Bruggen, J. C. (2010). Inspectorates of Education in Europe; some comparative remarks about their tasks and work. *Brussels: SICI*.

Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. *School effectiveness and school improvement*, 14(3), 321-349.

Weiler, H. N. (1990). Comparative perspectives on educational decentralization: An exercise in contradiction?. *Educational evaluation and policy analysis*, 12(4), 433-448.

Whitby, K. (2010). School inspection: Recent experiences in high performing education systems. Reading: CfBT Education Trust.

Wilcox, B., & Gray, J. (1996). *Inspecting Schools: Holding Schools to Account and Helping Schools To Improve*. Taylor & Francis, Inc., Orders and Customer Service, 7625 Empire Drive, Florence, KY 41042 (paperback: ISBN-0-335-19674-8, \$30.95; hardback: ISBN-0-335-19675-6, \$102.95).

Appendix A
List of selected documents

| Name | Type | Year | Published by |
|--|-------------------|----------------|--|
| Education Law of the Republic of Lithuania | National law | 1991 - present | Supreme Council of the Republic of Lithuania |
| Ratification of the School Improvement Program | Resolution | 2002 | The Government of the Republic of Lithuania |
| Internal Audit Methodology for General Education Schools | Ministerial order | 2002 | Ministry of Education, Science and Sports of the Republic of Lithuania |
| List of General Criteria for Admission to State and Municipal General Education Schools and Vocational Training Institutions | Ministerial order | 2004 - present | The Ministry of Education, Science and Sports of the Republic of Lithuania |
| Procedure Description for the Organization and Implementation of External School Evaluation | Ministerial order | 2007 - present | The Ministry of Education, Science and Sports of the Republic of Lithuania |
| Ratification of the School Improvement Program Plus | Ministerial order | 2007 - present | The Ministry of Education, Science and Sports of the Republic of Lithuania |
| Conception of Formal Education Quality Assurance System | Ministerial order | 2008 - present | The Ministry of Education, Science and Sports of the Republic of |

| | | | |
|---|-------------------|----------------|--|
| | | | Lithuania |
| Good School Conception | Ministerial order | 2015 - present | The Ministry of Education, Science and Sports of the Republic of Lithuania |
| Self-evaluation model and indicators | Policy report | 2015 | The Ministry of Education, Science and Sports of the Republic of Lithuania |
| Procedure Description of the Allocation for the Quality Basket to General Education Schools | Ministerial order | 2018 - present | The Ministry of Education, Science and Sports of the Republic of Lithuania |
| Establishment of Topics, Questions and Evaluation Indicators for the Thematic External Evaluation of General Education Schools | Ministerial order | 2021 - present | The Ministry of Education, Science and Sports of the Republic of Lithuania |

Appendix B
Lists of interviewees

Table B1

List of interviewees (policy experts)

| Interviewee | Professional experience(s) / expertise | Date |
|-------------|---|-------------------|
| 1 | Senior official at the Ministry of Education, Science and Sports School leader | January 27, 2022 |
| 2 | Senior official at the Ministry of Education, Science and Sports Senior official at the National Agency for School Evaluation (<i>now part of the National Agency for Education</i>) | February 14, 2022 |
| 3 | Lead school evaluation expert Education researcher Specialist at a municipality education department | February 15, 2022 |
| 4 | Senior official at the ministry of Education, Science and Sports School evaluation consultant | February 17, 2022 |
| 5 | Senior official at the National Agency for Education Specialist at a municipality education department | February 21, 2022 |
| 6 | Lead school evaluation expert | March 08, 2022 |
| 7 | Senior official at the Ministry of Education, Science and Sports School leader | March 15, 2022 |

Table B2*List of interviewees (school teachers and leaders)*

| Interviewee | School | Position | Date |
|--------------------|---------------|-----------------|----------------|
| 1 | A | School leader | March 25, 2022 |
| 2 | A | Teacher | March 30, 2022 |
| 3 | A | Teacher | March 30, 2022 |
| 4 | B | School leader | April 19, 2022 |
| 5 | B | Teacher | March 29, 2022 |
| 6 | B | Teacher | April 21, 2022 |