

Bezem, Pablo; Piezunka, Anne; Jacobsen, Rebecca

Article — Accepted Manuscript (Postprint)

School Inspectors' Decision-making: Evidence from a Comparative Perspective

Leadership and Policy in Schools

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Bezem, Pablo; Piezunka, Anne; Jacobsen, Rebecca (2024) : School Inspectors' Decision-making: Evidence from a Comparative Perspective, Leadership and Policy in Schools, ISSN 1744-5043, Taylor & Francis, London, Vol. 23, Iss. 2, pp. 253-274, <https://doi.org/10.1080/15700763.2022.2129073>

This Version is available at:

<https://hdl.handle.net/10419/319279>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc/4.0/>

School Inspectors' Decision-making: Evidence from a Comparative Perspective

Pablo Bezem ^a, Anne Piezunka ^b & Rebecca Jacobsen ^c

^a *Mary Lou Fulton College for Teaching and Learning Innovation, Arizona State University*

^b *WZB Berlin Social Science Center & Goethe-Universität Frankfurt*

^c *College of Education, Michigan State University*

ABSTRACT

In an era of test-based accountability, school inspections can offer a more nuanced understanding of why schools fail. Yet, we have limited knowledge of how inspectors arrive at their decisions on school quality. Analyzing inspectors' decision-making can reveal the underlying views regarding school accountability and open opportunities for school improvement. We use a comparative case study of contrasting inspection systems in the United States, Netherlands, and Argentina. Based on in-depth interviews with inspectors, our findings reveal that inspectors' sensemaking and decisions are strongly influenced by local culture, professional traditions, and views on school accountability. These contrasting processes illustrate trade-offs between rigid and flexible approaches to school inspection that have consequences for school improvement.

Keywords: accountability; school quality; school improvement; school inspection, sensemaking

This is an Accepted Manuscript of an article published by Taylor & Francis in **Leadership and Policy in Schools** on 06 Oct 2022 (2024, Volume 23, Issue 2, pp. 253-274), available at:
<https://doi.org/10.1080/15700763.2022.2129073>.

It is deposited under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.



Introduction

High-stakes testing increasingly plays a central role in school accountability and improvement in many countries. This is driven, in part, by an audit culture that emphasizes performance measurement (Apple, 2005; Clarke & Ozga, 2011). An appealing feature of test-based accountability is the perception that it objectively measures educational performance and allows comparisons across districts and time (Bloem 2015). Yet, this typically does not provide nuanced information to identify *why* certain schools fall behind; it does not capture the myriad factors that influence school quality (e.g. Darling-Hammond et al. 2016; Gagnon & Schneider 2019). As a stand-alone policy, test-based accountability often incentivizes schools to narrowly focus on tested subjects (e.g. Fitchett and Heafner 2010; Jacob 2005) and strategies to boost scores, which might not promote substantive learning (Author names, 2008).

School inspection (SI) is an alternative approach for monitoring and improving school quality. Unlike test-based accountability, school inspection is not limited to standardized tests to evaluate student performance. Inspection evaluations frequently assess a variety of school attributes through observation and direct interaction with school stakeholders, such as teachers, students, and parents (van Bruggen, 2010). In this way, inspections can provide a summative assessment of overall school quality while also uncovering factors that help or hinder school improvement. Thereby, it serves as an alternative mechanism to achieve more nuanced accountability and insights for improvements.

Despite its wide use around the world, our understanding of how inspection is carried out remains limited. Inspection practices vary considerably across settings (van Bruggen, 2010), often in relation to underlying beliefs and assumptions about the role of accountability in school improvement. These views of school accountability can influence the work of inspectors in ways

that we do not yet understand. For example, a managerial approach to accountability would focus on the validity of evaluations and school effectiveness through high-stakes standardized assessments (Sinclair, 1995). In this situation, school inspectors may feel constrained to “stick to the script” and interpret their role more narrowly, targeting their focus while visiting a school. Other approaches to accountability emphasize insights from teaching professionals, with less concern about avoiding bias in inspection results. In these cases, inspectors may take on a more inductive approach to their work where they use their expertise and background to guide their inspections as they look for emergent issues. To understand how inspection works, it is critical to analyze inspectors' decision process and how school quality is judged. A high degree of variation exists across systems, yet limited work has sought to illuminate how contextual features shape how inspectors work. Examining inspectors' decision-making can reveal views regarding school accountability and perceived legitimacy of the inspection process and feedback.

In our study, we focus on the individual perspectives of inspectors. Understanding inspectors' thinking and how personal perspectives are utilized can shed light on this process. This study poses two questions: (1) What are the sources guiding inspectors' thinking during inspections? and (2) How do the personal perspectives of inspectors influence school evaluations? To address these questions, we develop a comparative case study of inspection in a US school district, a province in Argentina, and the Netherlands. These cases differ across three dimensions: (i) the level of stakes attached to inspection, (ii) inspections' role within the accountability system, and (iii) inspectors' professional background. No previous study has evaluated what drives inspectors' decision-making from an international comparative perspective.

Literature Review: Inspectors' Decision-making

Despite the fact that inspection systems have long existed around the world, most previous research focuses narrowly on European systems. Furthermore, the wide variety of SI arrangements that exist makes it challenging to build a coherent body of literature that converges on key findings (de Wolf & Janssens, 2007; Ehren, 2016; Klerks, 2012). The local nature of inspection has reinforced a tendency of focusing inspection research on local systems, which is often then published in country-specific journals in the native language. SI research published in more widely read journals, in the English language, has expanded during the last five years. Yet, the literature remains limited and fragmented. The empirical literature that does exist has tended to focus on the effects and side effects of SI (e.g. Altrichter & Kemethofer 2015; Klerks 2012).

Within this emerging body of literature, limited empirical research has centered on school inspectors themselves and their influence on the evaluation process. Most early studies published in English were conducted primarily in the UK. Despite using a highly standardized inspection procedure and a reliable system of classroom observation (Matthews et al., 1998), various studies conclude that the professional judgment of OFSTED inspectors played a key role in their evaluations (Gilroy & Wilcox, 1997; Lee & Fitz, 1997; Woods & Jeffrey, 1998). It was found that inspectors' feedback to schools is influenced by perceived constraints that the local context imposes on teachers (Woods & Jeffrey, 1998). In addition, inspectors' professional background impacts their judgment. For example, prior experience serving as a classroom teacher can increase empathy and a sense of collegiality with teachers (Baxter, 2013; Millett & Johnson, 1998).

Since professional judgment plays a role in inspectors' decision-making, it is relevant to explore how individual judgment varies. Silcock and Wyness (1998) shed some light on the

issue, finding a wide diversity in inspector beliefs about education and current reforms as well as their empathy with challenges faced by teachers. These differences were apparent despite standardized training and evaluation tools. This early research demonstrated that profound differences in core beliefs regarding education can persist in a highly standardized system.

Recent studies have focused on the process of judgment formation in SI systems, where feedback is decided through consensus among a group of inspectors (Dedering & Sowada, 2017; Lindgren, 2015; Rutz et al., 2017). Despite the use of protocols and standards, inspectors have some discretion, both as individuals and the overall group, when working towards a consensus and making decisions (Dedering & Sowada, 2017; Rutz et al., 2017). Lindgren (2015) demonstrates that in the highly standardized Swedish system, there is a stark contrast between how decisions are formed during the inspection process (the “backstage” of inspection) versus how final feedback is presented to the school and community (the “front stage”). Even when inspectors present hard evidence to justify decisions in the “front stage,” there is negotiation among inspectors in the “backstage,” where their judgments encompass a mix of uncertainty, adaptation, and creativity.

These findings demonstrate that the human element and professional judgment remain central in the inspection process, regardless of efforts to standardize processes and procedures. Despite knowing that variation does occur, it is not yet understood *how* specific personal aspects of the inspectors and institutional features of the school system influence the inspection process. This study aims to provide initial insights into this critical aspect of inspection systems.

Theoretical Framework

Our research draws on sensemaking theory to analyze how inspectors arrive at their decisions on school quality (Weick, 1995). In addition, different approaches to school accountability help to identify differences between our three cases.

Sensemaking Theory

Sensemaking is a way of understanding how individual actors comprehend an unknown or ambiguous situation, make meaning of it, and then act based on this interpretation (Weick, 1995). Thereby, they refer to a cognitive framework that consists of cognitive structures—such as beliefs, knowledge, and attitudes—their situation, and policy signals (Spillane et al., 2002).

A growing body of literature in education draws on this theory to understand teachers' and administrators' interpretative frameworks when enacting educational policies (e.g. Coburn 2005; Halverson et al 2004; Rigby 2015; Spillane et al 2002). Sensemaking is the process in which individuals retrospectively structure the unknown, turning circumstances into situations they comprehended, to rationalize their actions (Weick et al., 2005). Educational studies that draw on this approach have addressed how this process is influenced by preexisting worldviews, prior knowledge, experience, formal and informal networks, and the organizational and social context within which sense-makers work (Ball & Bowe, 1992; Coburn, 2001; Hill, 2001; Porac et al., 1989; Spillane et al., 2002).

Sensemaking literature related to policy implementation has focused on how knowledge structures are accessed and applied in practical situations. One finding is that observations made by individuals who implement policy can often focus on the superficial aspects of a situation that then trigger a memory of another situation. This jeopardizes the ability to dive into the deeper significance of what is observed (Spillane et al. 2006). This literature has also found that

individual reasoning about a complex judgment tends to be biased toward interpretations that are consistent with their beliefs and values (Spillane et al., 2002). Sensemaking processes are mediated by considerations about organizational structures (e.g. work environment, norms, and rules), professional affiliation and networks, and traditions (e.g. Coburn 2001; Spillane et al. 2006). Policy implementation studies have shown the relevance of socially mediated sensemaking. For example, when teachers implement instructional policies, sensemaking is mediated by school leaders' participation in the interpretation of the policies (Coburn, 2005) as well as by interactions with other teachers (Coburn 2001; Hill 2001).

A separate body of literature in organizational studies focuses on sensemaking within organizations. Sensemaking theory has been used in this field to understand confusing or ambiguous events within organizations (Maitlis & Christianson, 2014; Sandberg & Tsoukas, 2015; Weick, 1995). Similar to sensemaking research on education policy implementation, organizational studies highlight the importance of constructing intersubjective meaning, which occurs when various actors within an organization, such as managers and peers, shape each other's understanding (Gioia et al., 1994).

Using these perspectives and building upon past research, our study draws on sensemaking theory to understand how inspectors interpret situations that they observe in schools and arrive at judgments regarding school quality. School inspectors must reconcile government guidelines, best practices, and inspection protocols with the situations they find in the schools. Therefore, sensemaking theory provides useful lens to understand this process. Sensemaking is likely mediated by inspectors' own experience and beliefs about education, the interaction with other inspectors, and organizational culture. The sensemaking literature provides useful constructs to capture the variety of factors that influence how inspectors reconcile policies with

the practice (e.g. Coburn 2005; Hill 2001). While protocols do exist, there is flexibility for inspectors to use professional judgment (Dedering & Sowada, 2017; Gilroy & Wilcox, 1997; Lindgren, 2015). Therefore, we rely on the sensemaking literature to analyze how inspectors interpret complex situations observed at schools and how they arrive at decisions.

School Accountability Frameworks

Several conceptualizations of accountability exist, which are informed by distinct theoretical lenses, including managerial, professional, political & public, and market accountability (Klein, 2020). These conceptualizations differ regarding how quality is measured (e.g. the degree of standardization), the objective of quality assessment, the power balance between the actors involved, and for whom schools are held accountable.

In the last fifty years, many countries have increasingly incorporated instruments of *managerial accountability* for public policy (Klein, 2020). This approach refers to the responsibility of an agent or institution towards a superior on a delegated task (Lindberg, 2013). In education, the managerial approach focuses on schools rendering account of their effectiveness based on external standards measured by outcomes. This approach led to the adoption of high-stakes accountability reforms worldwide. In some cases, this narrow definition of accountability became dominant (e.g. Moller, 2009; Ranson, 2003; Apple, 2005).

Other relevant approaches to accountability have been typified as *professional*, *political & public*, and *market* (Sinclair, 1995; Knapp & Feldman, 2012. See also Klein, 2020). First, *professional accountability* refers to the adherence to standards defined by the profession. In education, this has been a traditional view in which teachers uphold standards of a successful teaching practice within the profession, place students' needs at the center of their work, and hold each other accountable (Moller, 2009, Klein, 2020). Second, *political & public*

accountability approaches refer to organizations (i.e., schools) being responsible to the local community and their interests (i.e., educational goals of the community) (e.g., Moller, 2009). Finally, *market* accountability refers to a system where good practices, that are ultimately defined by consumer (parents) preferences (Adams and Kirst, 1999). These types of accountabilities inform processes of data collection differently.

Our study draws on different lenses on school accountability to shed light on inspectors' sensemaking when evaluating schools. Inspectors' beliefs on the role and functions of accountability most probably influence their decisions. For example, from a managerial perspective, information produced by inspection might be seen primarily as an instrument to control and pressure schools to improve, while the professorial perspective might emphasize insights to improve teacher practices and student wellbeing. Considering views on school accountability in each site, shed light on the inspectors' decision-making process.

Methods

To investigate school inspectors' decision-making process, this study uses a comparative, multi-site case study approach. We conduct a horizontal examination of inspectors' decision processes across three research sites (Phillips & Schweisfurth, 2014; Vavrus & Bartlett, 2016). This comparison highlights differences and similarities across cases, where SI practices vary based on local contextual features. Through these cross-site comparisons, we characterize contrasting SI arrangements and practices, which can advance understanding of the broad spectrum of inspection thinking processes (Chabbott & Elliott, 2003). The analysis focuses mostly on inspectors' thinking processes. It takes into consideration less formal aspects of these processes, including inspectors' personal perspectives, such as preferences, beliefs, and professional judgment.

Case Selection

We selected contrasting cases which differ from one another in terms of the stakes attached to inspections, inspectors' professional background, and the role served by inspection within the educational system. Within all three SI systems, a group of experts conducts in-school evaluations and use several modes of data collection—classroom observation, school stakeholder interviews, and document analysis. Differences in the protocol of conducting inspections are present – the frequency and length of inspection visits, number of inspectors, and public availability of inspection reports. But more importantly, we will investigate differences in inspection practices beyond the protocol. These differences may influence the information sources that inspectors consider when evaluating schools.

International Comparison

US District

In the United States, each state is the ultimate responsible for its education policy. Yet, local districts have historically held great freedom to manage their schools through local boards and superintendents (Kirst & Wirt, 2009). For accountability purposes, the federal government, states, and school districts play a role in the accountability structure, which relies mainly on standardized tests. School inspection has been used by school districts experimentally, typically for short periods.

The case in the United States is a large, urban school district that relies on a high-stakes accountability setting, based on standardized testing. Inspection plays a minor role in school accountability. Schools have relatively high autonomy to make decisions. Nonetheless, the district public officials are regularly sent to visit all schools. For the inspections, school autonomy play an important role, and inspection feedback avoids providing specific

recommendations regarding actions to be taken.

The district began experimenting with inspection processes more than a decade ago as part of school reforms. The inspection program, referred to as *Quality Reviews*, primarily targets low-performing schools. Unlike the other cases included in this study, inspection is outsourced to private consulting firms and is not directly managed by a governmental office. Since 2012, the process has been led by a company we will refer to as *QualiEv*. This inspection program gathers qualitative evidence about school programs for accountability and formative reviews. School visits are conducted by groups of three to four inspectors. At least one is a representative from *QualiEv*, a full-time inspector who leads the process, and the others are certified reviewers from the District Department of Education. The team is guided by a detailed protocol which outlines the evaluation process and includes research-based standards regarding effective school practices. Inspection activities consist of school document reviews, classroom observations, as well as interviews and focus groups with teachers and administrators. Immediately after inspection visits, inspectors share main findings in an oral report to school administrators. Then, inspectors and school administrators work jointly in a planning process, discussing school strengths and areas of growth, establishing the next steps, defining strategies, setting measures to establish success, and a timeline to achieve these goals. A written report summarizing conclusions is provided to schools, which includes suggestions for priority areas, but not specific recommendations for improvement.

The Netherlands

The Dutch Ministry of Education coordinates educational policy with municipalities.

Accountability relies on both outcome- and school-based components (Nusche et al., 2014).

High-stakes testing has an important role (Scheerens et al. 2012; van der Sluis et al. 2017), while

at the same time, inspection is a central instrument for monitoring standards. Dutch schools are among the most autonomous in the world (Scheerens et al. 2012); they are allowed to make their own decisions about school organization as well as the educational content. The inspectorate can override this autonomy when schools do not meet required quality standards. Inspectors are full-time public officials and receive dedicated training.

The Netherlands emphasizes SI of low-performing schools.¹ While all schools are inspected at least once every four years, the lowest performing receive more frequent and rigorous visits. To determine the frequency and type of inspections, inspectors use a risk-based model. This model assesses school risk based on administrative information, including standardized test scores, accountability documents, and failure signals, such parents' complaints or negative media reports (Education Inspectorate - Ministry of Education, 2010). Inspectors follow an assessment framework for legal aspects, process quality, and outcomes. The framework pays particular attention to learning outcomes, educational process, school environment, quality assurance and ambition, and financial management (Education Inspectorate - Ministry of Education, 2017a, 2017b). Each of these areas include a set of standards that is operationalized based on statutory requirements. Results of inspection are shared with the school and public through a summary report. If schools do not demonstrate improvement for two years, inspectors can recommend to the Ministry of Education administrative and/or funding sanctions. In the most extreme cases, this can lead to school closure (Ehren et al. 2013; OECD 2015).

Rio Negro, Argentina

Each province in Argentina manages its educational system. Inspection is the main mechanism for school accountability. Standardized tests are low-stakes and only used for diagnostic purposes. Similarly, school inspection are also low-stakes and can only lead to sanctions when

regulation about school safety or student well-being are at risk. Inspectors are the only public officials that systematically visit all the schools. In this process, they can set limits to the autonomy of all schools; nonetheless, actions after inspections tend to be established by schools and inspectors in a democratic process. Inspectors are in the highest position of the teaching ladder. They are full-time public officials who report directly to the provincial Ministry of Education and must have considerable experience: 12 years in teaching, 2 years in leadership, and inspectorate training (Concurso de Supervisores Rio Negro, 2013).

The main purpose of SI in Rio Negro Province is to provide support to all schools. In the process, school administrators are held accountable. Inspectors develop inspection projects, and while they must follow broad guidelines, there are no specific protocols for school visits or inspection activities. Inspectors are assigned to a group of schools to conduct administrative controls and provide continuous support. School visits occur at least three times a year, and can be more frequent if a school requires more support (Resolución Del Consejo Provincial de Educación de Río Negro N 1053, 1994). Inspectors consult with their technical team of professionals in education to inform their work. All inspectors go through basic training and are accountable for following legal standards. Inspectors prepare reports for the schools, which are not publicly available. No sanctions are imposed for poor academic performance. Furthermore, inspection does not track standardized educational outcomes, such as test scores, nor must follow specific standards regarding education processes.

Data Collection and Analysis

We conducted semi-structured interviews with inspectors of K-12 schools in the three study locations. We inquired about the inspector's background, activities performed during the inspection process, and outcomes of inspection. In addition, we asked about how they make

decisions regarding school quality and what aspects of quality they value the most. Emphasis was placed on capturing the inspectors' thought process through the use of probes that asked for concrete examples to illustrate their thinking and we provided inspectors with scenarios to gauge how they would respond to a given situation.

In the United States, we interviewed inspectors from the district Department of Education who were certified by *QualiEv*. We invited all 29 certified reviewers who had previously conducted inspections. For the other sites, we selected a purposive sample of inspectors for primary and secondary schools (Teddlie & Yu, 2007). The objective was to conduct 6 to 10 interviews at each site, with inspectors of various backgrounds and experience. In Argentina, we had complete freedom to invite interviewees from the whole pool of inspectors; in the Netherlands, we choose from a smaller pool of inspectors who spoke English, proposed by the Inspectorate. In total, we completed 23 interviews: 8 in the United States, 6 in the Netherlands, and 9 in Argentina. Interviews lasted an average of 72 minutes. In the United States and Argentina, interviews were conducted in-person in the inspectors' languages, English and Spanish, respectively. In the Netherlands, interviews were conducted via videoconference in English. We acknowledge some drawbacks when interviewees speak in a foreign language. First, it might limit the capacity to express themselves and develop nuance in their narratives. Nonetheless, all interviewees showed high proficiency in English. Another concern is that restringing interviews to inspectors that speak a second language might bias the sample selection toward more educated professionals. To mitigate the possible bias from English-only interviews in the Netherlands, we shared our findings with the Dutch inspectors who provided feedback. We discussed any misunderstandings the inspectors identified and inquired about their impressions of the conclusions drawn. Interviews in Argentina were transcribed in Spanish and then

translated into English. Interviews from the Netherlands and the US sites were transcribed and checked for accuracy. Participants were informed that interview responses were anonymous, transcripts would not be shared, and a pseudonym would be used to cite them. In the United States and Argentina, inspectors were given a US\$ 25 gift card after participation. In the Netherlands, we were advised by SI researchers not to offer incentives. Participants in each study site reflect characteristics of inspectors in their location with respect to years of experience and demographics. Descriptive information about interviewees is presented in Table 1.

Table 1. Inspectors' Background and Experience

Variable	U.S. Case n=8	Netherlands Case n=6	Argentina Case n=9
<u>Individual inspectors</u>			
Inspector experience, in years	2.3	9.8	6.2
Education experience, in years	14.3	22.8	31.0
Classroom teaching experience, in years ¹	8.9	10.7	13.1
Administrative experience, in years ¹	2.7	7.0	15.2
<u>% of inspectors, within case</u>			
% former classroom teachers	100%	50%	100%
% former administrators	38%	20%	100%
% with graduate degree	100%	75%	56%

¹ Only those inspectors with experience as teachers/administrators were included in these indicators.

The interview transcripts were coded using deductive and inductive codes. Deductive codes were formulated based on our theoretical framework, mainly from concepts related to

sensemaking theory (Coburn, 2005; Maitlis & Christianson, 2014; Spillane et al., 2002, 2006; Weick, 1995). Inductive codes stemmed from interviews in the three sites. Responses were coded according to Miles, Huberman, and Saldaña's (2014) approach to qualitative analysis by observing patterns and themes within and across case studies. We used DEDOOSE qualitative data analysis software for coding and analysis. To ensure the reliability of codes, we used an independent-coder method. First, interview transcripts were independently coded by two researchers, before then comparing the coding for agreement. We followed an iterative process until at least 75% agreement was achieved.

We conducted two rounds of coding. The first round focused on: i) sources of information used during inspection, ii) use of local context information, and iii) inspectors' definition of good quality education. These codes were defined inductively. The second round relied more heavily on deductive codes that were more abstract and required more interpretation. The main codes include iv) inspectors' perception of school administrators, v) type of recommendations, and vi) sources guiding thinking. This latter code is emphasized in our analysis (Table 2). The sub-codes are based on the inspection procedures and sensemaking theory. The theory was used to define the foci on the knowledge structures accessed by inspectors when facing practical situations, especially when they have to make sense of complex situations.

After the coding procedure was complete, we identified general and country-specific patterns in the data. The data were examined visually in the form of cross-tabulations, and charts showing the frequency with which codes occurred as well as the presence or absence of codes within and across interviews, for each of the case studies. To ensure that reported patterns are an

accurate representation of each study site, we shared initial findings with interviewees and incorporated their feedback (Miles et al., 2014).

Table 2. Sources Guiding Inspectors' Thinking

	Source	Description
A	Indicators of school quality	Standardized rubrics, indicators, or metrics used to evaluate school quality during inspection
B	Multi-informant approach	Simultaneous use of multiple sources of information - of the same or different kind - to validate evidence
C	Interactions among inspectors	Interaction among inspectors or between inspectors and technical personal to discuss about findings
D	Local context information	References to local context information, including students' demographics, characteristics of the neighborhood, school history, and change of staff
E	Inspectors' personal backgrounds	References personal experience, beliefs, or professional judgment.

Findings

Overall, our comparative analysis indicates that personal perspectives are an important influence inspectors' evaluations. The extent of influence is affected by local culture, professional traditions, and views on school accountability. In particular, the case in the United States relies on high-stakes testing for accountability and its audit culture emphasizes performance measurement and data-driven decision-making. The inspectors' thinking is infused with this culture and their decision-making is highly influenced by a test-based accountability mindset that leads to strictly adhering to protocols and reduces inspectors' professional insights in an effort to avoid bias. Despite opportunities presented by SI to dig deeper and identify unique strengths and weaknesses at schools, inspectors actively disregarded insights that do not fit within the confines of the protocol. In contrast, the Dutch and the Argentinean cases illustrate approaches in which inspectors have more flexibility and can rely on their professional background and judgment.

Inspectors routinely pursued “surprising” observations even when these do not fit neatly into a protocol or the anticipated focus areas. As a result, these inspectors adopt a more holistic understanding of school quality and avenues for improvement.

Applying the sensemaking-approach, we present findings for the categories of sources that guide inspectors' thinking: (A) Indicators of school quality, (B) Multi-informant approach (C) Interaction among inspectors, (D) Local context information, and (E) Inspectors' personal backgrounds (see definitions in Table 2). Key findings regarding the influence of these categories on inspectors' thinking across our cases are summarized in Figure 1. For quoting study participants, their names include a subscript indicating what country they represent: “-US” for the United States, “-NETH” for the Netherlands, “-AR” for Argentina.

(A) Indicators of School Quality

In the U.S. case, indicators of school quality in the form of a standardized rubric are the cornerstone of the evaluation process and all of the inspectors mentioned them repeatedly when explaining their thought processes. Their data collection is structured around various aspects of school quality, a series of metrics to measure them, guidelines for making observations, and questions for schools in order to evaluate each indicator. We found that inspectors in the US case restrict data collection almost exclusively to sources specified in the protocol: classroom observations, interviews and focus groups, and school documents. Only two inspectors mentioned that they seek publicly available school information before their visit (e.g. test results and school website) or during the visit (e.g. teachers' planning documents, students' files). This stands in stark contrast with the Dutch and Argentinean cases where developing a deep knowledge of a school before the actual visit is a vital part of the inspection process.

Figure 1. Comparison of Sources Guiding Inspectors' Thinking

	U.S. Case	Netherlands Case	Argentina Case
Indicators of school quality			
Multi-informant approach			
Interactions inspectors			
Local context information			
Inspectors' personal backgrounds			

Legend: Degree of Influence on Inspectors' Thinking

High			Low
------	--	--	-----

Note: The relative influence of each source within each case was determined by the number of mentions in interviews.

Moreover, in the US case, the protocol explicitly determines the structure and nature of classroom observations and the questions to be asked in interviews and focus groups. For example, when inspector Amy-US was asked about what information she personally looks for or asks to see apart from the required data, she emphasized the importance of adhering to the protocol:

I just followed the ... protocol. I ask only the questions that are outlined for teachers and only the questions that are outlined for students. If a student [response] needs to be elaborated, I would say “could you tell me more ...,” but I don’t bring my own questions to the process or anything like that. I just tried to follow what is asked of me.

We asked inspectors about the value they placed on different information sources to evaluate school quality. We found that observing classroom instruction is consistently the most valued source of information used by inspectors to evaluate school quality. In contrast to the other study sites, most inspectors in the US case considered other sources of information to play a secondary role, if considered at all. School planning documents were deemed the least valuable source of information to evaluate school quality and were used mostly for triangulation purposes. Similarly, the use of school climate observations—such as culture, interactions among students and teachers, and facility condition—was also secondary.

More than in the other study sites, when the inspectors in the US case explained their thought process, they repeatedly made direct remarks about how they try to avoid personal bias when completing the rubric and following how the protocol defines good instruction. This was highlighted by Sarah-US and Donna-US:

Sarah-US: We all come to the table with our own expertise, with our own beliefs and values, with our own biases or preconceived notions about what school should look like ... the rubric then helps people put those things aside, understand their influence, and then really ground themselves in both evidence and the rubric to get to a shared understanding.

Donna-US: It's really important ... as a reviewer, to not have a bias ... I might have a bias towards what good instruction looks like. So instead of using the rubric in front of me, I'm going towards what I think is good. Or I might have a bias towards what a functioning school environment looks like and sounds like. So instead of using the evidence in front of me, I'm just going towards what I think ... I think that can be both positive and negative.

Across all interviews in the US case, we found a rigid emphasis on observing the protocol, sometimes in ways that appear to impede important insights. This stance seems consistent with managerial approach to school accountability which focuses primarily on controlling schools

through reliable measure of school effectiveness based on external standards. In this situation, most inspectors still recognized the relevance of their own expertise and experience; they acknowledged which indicators best capture school quality and their preferred information sources. Yet, as shown below, in Section E, inspectors in the US case dismissed this wealth of knowledge and equated drawing on their personal experiences with bias. Incorporating such information into evaluations was viewed as a validity threat to the inspection process.

Inspectors in the US case claim to actively suppress the influence of their education knowledge and experience during inspection by stating that they strictly adhere to protocol. Furthermore, unlike the other research sites, when explaining their thinking in specific situations, they were rarely able to share concrete examples based on their own experience. Instead, they repeatedly referred to “good practices” outlined in the rubric.

Similar to the US case, Dutch inspectors utilize indicators of school quality in all their school visits. However, these indicators were not the dominant influence on inspectors' narratives about their thinking process. While the policies and procedures used by the Dutch shares similarities with the US case, and its *QualiEv* protocol, Dutch inspectors expressed greater flexibility in the inspection process, including stages of data collection, choosing which criteria to focus on and emphasize, and how to interpret indicators of school quality. When collecting information, Dutch inspectors use standardized data only as a starting point. Our interviews revealed that data collection and usage are guided by inspectors' choices to probe more deeply into key areas, as their insights and understanding of a given school evolve over the course of the visit. No consensus exists among the Dutch inspectors regarding the most valuable information sources. Several inspectors found this question difficult to answer, in contrast to the inspectors in the US case who promptly referenced standardized protocols. Some Dutch inspectors especially

valued interviews with teachers and administrators as well as classroom observations. Only a few mentioned school climate or student interviews. Dutch inspectors were more likely to view the usefulness of information sources as highly contextual, based on their experience, school context, and specific issues that emerged during a school visit.

In contrast to these two cases, in Argentina, a standardized set of indicators of school quality have almost no role in the inspection process. When indicators are collected, they are mainly used for administrative purposes. Inspectors do not use standardized metrics to evaluate quality systematically. This is not to say that the inspectors lack standards for evaluating school quality. Rather, they have considerable freedom to decide what information to collect and how to evaluate schools. Consistent with this freedom, we observed immense heterogeneity in terms of information sources used and which sources are most valued. Inspectors emphasized the importance of gaining an understanding of how the school functions, emphasizing the importance of “being present” in the school, “walking around,” and “living in the moments of the institutional life.” Being present allows inspectors to be critical and provide support to the school. Without insights from school visits, inspectors do not feel they could truly understand a school. Therefore, they would be unable to provide necessary assistance to help schools improve. Planning documents are considered useful for evaluation and collaborative work between inspectors and schools.

Bias was rarely a concern among inspectors from Argentina and the Netherlands. In these two cases, inspectors did not hesitate to make use of their professional judgment, experience as instructional experts, and familiarity with schools gained from multiple visits. Rather than be a cause for concern, this was viewed as necessary to be a good inspector and exactly what enables them to be effective. Furthermore, in the Netherlands, utilizing this type of knowledge is viewed

as necessary to be a good inspector. Part of the inspector training process includes an extensive shadowing of experience inspectors, where those who are new have an opportunity to further develop and to learn to use their expertise. This is not to say inspectors in Argentina and the Netherlands do not reflect critically on their own practices and maintain a concern for integrity in the process. In both countries, nearly all inspectors expressed spontaneously in the interview a concern about being thorough in their analysis and the importance of justifying their conclusions in their feedback to schools. But attempts to standardize the process and a concern for bias and reliability were less prominent in our conversations.

(B) Multi-Informant Approach

Once inspectors in the US case collect data, they are turned over to *QualiEv* where staff conduct a standardized rating process. The processed data are then used to assess overall trends and patterns in the school. This *multi-informant approach* seeks evidence that is confirmed repeatedly using the same type of information and then triangulated with other sources. For example, evidence from only one or two classrooms is insufficient to make a claim. Affirmation must be found in multiple classes and then triangulated with evidence from additional sources, such as interviews and school documents. Many inspectors highlighted that this approach provides a holistic view, as Michelle-US explains:

[The multi-informant approach] really gives you that holistic view of, “Okay, this is what we saw in the classroom. This is what the teachers and students are doing.” But then, what are people actually saying about it? What are the parents saying, the students, and the staff? And how did those stories support one another, or how are they different?

Importantly, most inspectors expressed confidence in the focus on the trends, as opposed to outliers, as a reliable approach to evaluate the overall quality of the schools. Overall, most

inspectors seem to embrace this approach as an effective way to evaluate the quality of the school. As Laura_{US} explained:

Both times that I've done it [(the inspection)], it's been really clear, even after the first half of one day, what the trends are. It's been kind of shocking, because you think, "Oh, obviously, these classrooms [are] different than that." No, they're never different. It's all the same. It's always been really shocking how quickly you can come to what the big problem is. Usually, it's actually been pretty easy to pick the top two or three, and because the schools that they pick to do these things are ... literally on fire, so it's pretty clear.

Interviews demonstrated how systematically the multi-informant approach is applied: Evidence is gathered, inspectors focus on major trends, and discard information that does not fit within these broad trends. Donna_{US} emphasize that they "are looking for trends and consistencies, versus an outlier of something that might strike you as wrong." Contradictory information observed while data is being collected might rise a red flag and can help narrowing the focus. When this happens, there is some leeway for further inquiries as long as the search and sources of information to be used are part of the protocol. Yet, in contrast to the other research sites, we did not find examples of inspectors pursuing professional hunches that lead to additional question being asked, nor focusing on exceptional observations, nor conducting additional interviews that might lead to new discoveries.

Inspectors in the Netherlands act as investigators. Each has considerable freedom in deciding which focus areas to emphasize during school visits. Several inspectors explained that after reviewing school information, prior to the visit, they try to anticipate the main difficulties at the school. Hypotheses are developed that they then seek to verify or disprove during the visitation day. Several inspectors noted how their expertise can assist in developing these initial assessments. They actively draw upon their prior knowledge and vast experiences with a wide

array of schools to help them anticipate and hone in on issues that the school is facing. Rather than a standardized approach that eliminates variation both between and within a school, the Dutch approach results in a dynamic process only guided by their protocols, not dictated by them. Lotte-NETH illustrates the questions that inspectors ask themselves during inspections:

I try to see what the most important papers are, and I read them. I try to think about what I might see in the school. I have some hypotheses in my head and I also see what I can make of the context of the school. For example... In what kind of neighborhood is it? What can I expect of the school? What's the difficult thing over there? Then I go to the school and be as open minded as possible because sometimes, when you already think you know what it will be, you will be very much surprised by what happens in the school ... I just have that in mind, somehow, but not have that on the front of my head. I just be open and see what happens during the day, but... I've got a [starting] schema in my head.

Dutch inspectors' interviews revealed that they too look for patterns and trends through a multi-informant approach. However, they also strive to identify conflicting evidence so that deeper inquiry can be made and observations and impressions during interviews can be confirmed. Rather than dismissing discrepancies or outliers, the Dutch inspectors view such findings as critical points for further investigation. Thus, inspectors use these insights to identify problems that often lurk beneath the surface. For example, Lars-NETH explained how he actively looks for points of disagreement:

[An important source of information is] talking to the teachers, like how they tell the [way] the school really works, how they perceive how the managements makes them work and doesn't really work, ... and do the teachers understand that vision and do they really use that vision inside their classrooms? And good thing is, we always visit the classes first, before we talk to all the people. So we can give back to the teachers and to the team leaders and to the director, what we saw in the classes. And so they can immediately give back how they perceive it.

A multi-informant approach guides thinking in both the Dutch and the US cases (see Figure 1). However, there is a great difference in how information is corroborated. In the Netherlands, there is less emphasis on accumulating evidence through a rigid prescribed process, and more on finding hidden problems and testing if evidence can confirm nascent hypotheses. Thus, the process is dynamic and evolves during the visit. Inspectors determine in real time which additional documents to request, questions to ask, and aspects of classroom observations to emphasize.

The inspectors in the Argentinean case also seek corroborating evidence, often comparing formal planning documents to actual practices. Similar to the Dutch case, and unlike the US case, inconsistent findings are viewed as a critical window into key issues faced by the school. In our interviews, inspectors provided several examples of how unexpected cues during a visit can lead to additional sources. This was illustrated by Monica-AR:

I value walking in the schools. The fact of being present. Because face-to-face you can get to ask a new question about something specific, and you can be surprised. You can find something that you hadn't thought. ... [Sometimes you find that] the pedagogic proposals don't correspond with what you see in the visits, when you see they are not [using] the methodology they say they are applying ... If you take a child's workbook and you see mistakes in the corrections made by the teachers, or there are no corrections made by the teachers, you say: "what is going on here?"

Unlike the Dutch and the US cases, the multi-informant approach was not stressed as a central aspect of the inspection process for most inspectors in the Argentinean case.

(C) Interactions among Inspectors

In the US case, data collection and synthesis are followed by a consensus building process, led by *QualiEv*. During the group discussion, *QualiEv* reinforces the previously mentioned factors in guiding thinking in order to avoid bias: focusing on the rubric and the trends, while discarding

outliers. In this phase, the inspectors have an opportunity to explain their observations. *QualiEv* guides the discussion and consensus building process. Most inspectors rely on and trust in the contracted organization for facilitating the discussion and “*pushing their thinking*” (Amy-US). In this process, *QualiEv* ensures all claims are aligned with the rubric. This process was explained by Sarah-US:

So, the [QualiEv] team leads a collaborative consensus building process, but they lead that in alignment with their practice and process. So, it's a collaborative effort that is heavily guided by the contracting organization... they sit as experts on how the rubric should be utilized and how things then should be scored, but they go through a process of team consensus building. Everyone presents their evidence; they do that in a group setting, and then everyone talks through it and then determines where the preponderance of evidence fits on the rubric, which then leads to the scoring process.

Inspectors must discuss evidence until arriving at a consensus regarding the evaluation. The procedure for discussion starts from the quality claims and evaluation criteria based on best practices. Then, they discuss if there is enough evidence and how to weight this evidence to support each claim. In this stage, they compile the collected data and the data synthesized and trends that are identified by *QualiEv*. The inspectors emphasize that any claim must be supported by evidence. This dynamic was explained by Aidan-US:

I think the factors that usually go into play would be, “Which ones do we have the most evidence from our observations about? How strong is that evidence?” If we didn't see checks for understanding in one classroom that's not enough that we can make a priority claim around checks for understanding whatever the case might be. And so, it's usually about what is the weight of the evidence that we have.

When asked whether inspectors gather further information if they have not yet found evidence to support or disprove a quality claim, Heather-US responded:

No, there's no return observation it just, the claim is tweaked based upon what you did see. So, one of the norms they [*QualiEv*] often use is see the donut, not the hole. So, it's not about what you didn't see, it's about what you did see. So, if you didn't see any evidence towards that claim, you go with the evidence you did see.

Avoiding personal bias is a focus during group discussions. During the formal evaluation, inspectors strive to not introduce their views regarding alternative criteria that might be informative when assessing quality. This happens during the during discussion process, as illustrated by Michelle-US:

When you're collaborating with a team... you share something that you saw or heard ... you're looking for another example of it. And if you don't, then you let it go. ... You don't want to be biased or making comments based on personal opinion. So, you do very much keep it factual, and you make it collaborative so it's not just one person saying one thing.

Most inspectors perceive the process of reaching a consensus to be straightforward. We did not find evidence of heated group discussions or inspectors challenging the evaluation results. Furthermore, several inspectors indicated that for many quality claims that require consensus, *QualiEv* develops preliminary statements before convening inspectors for a discussion. This appears to be a feature of the way that the process is structured. Multiple inspectors visit a given school and attend selected classes. Thus, inspectors only observe a portion of instruction; this might not encompass all domains and factors to be evaluated. For example, when we asked Aidan-US whether he maintains his positions when observing something that differs from the bulk of the data, he stated that he looks at "the overall picture": "I could've gone to three classrooms in the morning and in those I didn't see a particular aspect, but other people did. What I saw is

just one part of all the data that's collected.” This partial observation may limit the ability to develop a full view of the school and might cause inspectors to adopt the narrative of the contracting organization. This situation fits the concept of guided sensemaking in which leaders actively build a narrative that promote understandings and explanations of events (Maitlis, 2005). On the other hand, this configuration restricts individual sensemaking and the scope of the intersubjective construction of meaning during group discussions (Gioia et al., 1994).

In Argentina and the Netherlands, inspectors are employed as dedicated government staff and are placed in regional offices. In the Argentinean case, each inspector leads a team comprised of various educational expertise: Pedagogy, psychology, school administration, and social work. The group dynamic is established by each inspector and varies substantially. In some cases, the technical team actively participates in decision-making and discusses which strategies are generally effective. Most inspectors use the team to make school visits in conflictive or complex situations and make specific interventions in schools. Consultations with other inspectors can be initiated; these mostly occur for complex situations or when inspectors have doubts about regulations. Some consult regularly with their colleagues in the office, others through text messages, and other in occasional provincial meetings.

In the Netherlands, inspectors receive ongoing group training and meet weekly to discuss current education issues, research and the inspection process. Some of these meetings focus on specific practices and feature invited speakers, while others offer training videos on classroom observation. When asked about how inspectors learn how to conduct inspections, Lotte-NETH commented on interactions among colleagues:

I'll read articles, and we've got information sessions at the office, where someone tells you something he's been working on or some interesting... with some colleagues of mine, we organize lunch sessions in which we've invited someone from outside the inspectorate to tell us

and inform us about certain subjects.... [for specific subjects] we try to invite someone from outside our office. We get new input and we also use our team to discuss about the standards ...

“How do I interpret what I see? How do you interpret? Okay. Do we come to the same judgments or do we judge something differently? What is the difference between us?”

Most Dutch inspectors mentioned that they generally consult with colleagues, but not for the purpose of achieving consensus. Dutch inspectors relied the least on interactions with other inspectors to inform their thinking process during an inspection. In secondary school, where inspections are conducted in groups, inspectors naturally interact with each other. However, none of the respondents emphasized a role for consensus building during the school visit. In primary school, where usually only one inspector visits the school, the process has been described as solitary, such as Emma_{NETH}: “You go to a school alone. You arrive alone... You think alone.” Nonetheless, some inspectors noted that they consult with colleagues in the inspection office when they face complex or challenging situations.

Interaction among inspectors plays a distinct role in the sensemaking process within each of the three cases. In the Netherlands and Argentina, inspectors work in a more stable group that share a long-term conversation about their inspection practices more generally rather than focused on a specific school example. Thus, the constant conversations shape the construction of meaning within the inspectorates (Rouleau, 2005). In the Dutch case, these conversations also have a more formal component in their ongoing training. In both cases interactions among inspectors regarding specific schools occur when they face controversial or complex situations. In this regard, inspector teams act as a sounding board to provide feedback and suggestions, but no formal consensus is required. This is in contrast to the US case, where interactions are used to systematically integrate information and identify trends. Controversy and complexity tend not to be addressed.

(D) Local Context Information

Consideration of the “local context” plays a minor role in US inspectors’ narratives about their thinking. Local context includes student demographics, neighborhood characteristics, or a school’s history. Several inspectors explained that after an extended time working in the district, they had interacted with most schools at some point and became familiar with the local context of the schools. Some of them mentioned that they have this knowledge, but they view it only as background information, as illustrated by Aidan-US who said that this contextual information is “in the back of their head.” However, inspectors did not mention this type of information factoring into their understanding of a school’s functioning. The interviews did not reveal many specific examples of how this knowledge influences thinking during the inspection.

Considerations of school context are not explicitly described in the formal protocol. Therefore, most inspectors actively strove to exclude this information, stating that “it *should* not matter” in their evaluations. When asked for further explanation, some inspectors highlighted that they must closely follow the rubric and avoid bias, thus assuming that the objectivity of the process might be compromised by considering the context. Several inspectors go further by arguing that inspected schools are low-performing and thus major differences in a school context are not present.

The Argentinean and Dutch inspectors consider contextual information a critical piece of information for understanding a school’s functioning. Argentina is the only research site in which each inspector is permanently assigned to a group of schools. Therefore, inspectors become deeply acquainted with the local context, student demographics, school history, and school staff. In most interviews, inspectors highlighted that the inspections are “situated,” locally oriented based on the school reality. As pointed out by Alejandra-AR: “Everything you do in the

school in based on the context, in the situational aspect, that is what I ask for, that the pedagogic project depart from there.” To a greater extent than the other case studies, inspectors interviewed in Argentina continuously reference their knowledge of school context when interpreting problems, prioritizing information sources, interpreting student performance indicators, and determining recommendations.

In the Netherlands, inspectors also work with a fixed group of schools. Yet, after several years, inspectors switch groups as a way to ensure objectivity. Dutch inspectors exhibited more knowledge of student demographics students compared to U.S inspectors. They also offered detailed descriptions of challenges faced by schools, with specific contextual circumstances such a large immigrant population where learning Dutch was critical or parents having low educational capital to support learning at home. This local knowledge was used by inspectors to interpret various sources of information collected during inspection. For example, inspectors consider whether a school with a high proportion of immigrant students should develop provisions for language education in their planning documents. As Sven-NETH explained,

If you have a school with parents who speak at home another language, the schools have to invest more in curriculum in vocabulary of Dutch for those children. Then, the expectation about the quality of curriculum are different... You cannot put that into strict criteria. ... [Another example,] if you are in a small school that has to put children of several [grades together] in one group, you know it's a very hard job for the teacher to organize the lessons in a way that he challenges all the children ... so, this kind of situation plays into the way you judge the quality of instruction.

In the US case, the protocols do not include context as part of their evaluation. Explicit consideration about how local context influences evaluation of school quality played a very

minor role in inspectors' narratives. This configuration downplays the role of situational factors that might spark sensemaking processes in inspectors (Sandberg & Tsoukas, 2015).

(E) Inspectors' Personal Backgrounds

We find that inspectors in the US case prioritize objectivity and reliability; there is an emphasis on standardized rubrics, which constrains the use of personal and professional knowledge. Yet, nearly all inspectors in the US case believe their background in education provides a necessary qualification for their role. All interviewed inspectors had experience as classroom teachers and their work at the district Department of Education involves evaluation of classroom instruction (see Table 2). Fewer inspectors (less than half) have experience as a school administrator; among those who do, the average experience is less than three years. Some inspectors noted that an instructional background was necessary because they knew what to look for during school visits. This was illustrated by Sarah-US:

I think that having [an] instructional background is critically important... being an educator, someone who is highly familiar with the instructional aspect of education ... folks who have that instructional-specific lens, who carry with them the lens of what high-quality teaching and learning looks like. If you know what it means to stand up in front of students and deliver instructional content and assess students. I mean there's a lot of insider language in the rubrics. ... You have to know what you are looking for, so you have to know what teaching looks like.

In addition, three of the interviewees had experience as administrators and they believed this was important preparation for their role as an inspector. Lisa-US explained how her judgment is informed by administrative experience, which can provide a more systemic view:

I really think that my years as assistant principal helped me because I'm able to see the school as an entire system and not just as one specific part, and so I think that's a great qualification [for] ... understanding ... So if some classrooms are having anger management issues, if it's not at

trend across the entire school, if there's a bigger trend arising that the instruction isn't rigorous, that that's a bigger focus for the school in trying to work with some individual classrooms. Experience working in the district office is additional factor that inspectors feel prepares them. Several inspectors mentioned that this experience allows them to “have a sense of what the schools look like.”

We found that in most cases, inspectors rely on their professional judgment in ways that are within the scope of the protocol, yet the approaches might not be explicitly stated in the protocol. Furthermore, in some cases, interviews revealed a tension between using professional judgment to complete the rubric and maintaining an unbiased and uniform process. This tension was illustrated by Michelle-US who noted how she must reconcile the evaluation rubric with the wide variety of elements she personally considers during classroom observation:

Michelle-US: When I'm in the classroom, I look for student engagement, and comfort, and listening, and learning. ... I think, because too often we can focus just on the teachers or the adults. ... we really have to look at the kids. When you're in a school environment, it's holistic... you're using your senses, right? You're looking, and you're feeling, and you're hearing, and all of these different things that you get when you're in a place that is not necessarily on any rubric, but you get the vibe, and the feeling of it. And then, you kind of couple that with what people are saying in the interviews, and what their body language is, and their emotional level, and how they respond to things.

Interviewer: ... how do you put all of this together in the rubric and the feedback to the schools?

Michelle-US: Well, those things I was just sharing, I do personally. So, those aren't necessarily on the rubric. But I think that's what comes out when you're collaborating with a team. You don't want to be biased or make comments based on personal opinion.

Finally, we found that when crafting feedback for inspected schools, inspectors in the US case use their judgment mainly for diagnostic purposes. Unlike the other sites, we did not find many

explicit considerations regarding how feedback and outcomes from the inspection affect the schools. This shows that inspectors' sensemaking process is delimited by the scope prescribed by the protocols and is essentially retrospective more than prospective (Gioia et al., 1994; Sandberg & Tsoukas, 2015).

In direct contrast to the U.S. case, inspectors of the Netherlands and the Argentinean case rely more heavily on their personal perspectives. In the Argentinean case, inspectors shared openly the ways that their personal experience, beliefs and professional judgment influence multiple aspects of inspection. The value placed on this wealth of knowledge might be due to inspector's positions being the highest step in the teaching professional ladder in Argentina. More than in the other research sites, inspectors frequently made explicit remarks about how they rely on their experience to inform decisions. This was illustrated by Alejandra-AR explaining that her recommendations to schools are not only based on government norms:

Based on what the educational policy is posing, but mixed with my perspective and stance, what I've learned all these years. Obviously, the educational policy gives you a framework in many regards. You can't stray from what is stipulated. But within these limits, my experience and knowledge are also important when the time comes to make suggestions.

Inspectors in the Argentinean case did not shy away from frequently explaining how their personal perspectives influence their thought process. Several inspectors explained that the process is informed by their views on what they consider critical issues in education. Inspectors see their role as more political, as several explained how they act as a bridge between the macro and the micro policies of schools.

Since inspectors from Argentina determine their procedures, in contrast to the other cases, they have considerably more leeway to use their personal judgment. As in the case of the US district, inspectors rely on their experience as teachers to judge teaching quality in the

classroom. This was expressed, by Marcelo-AR who said that “classroom presence, [allows you] to verify the processes.... After so many years, you trust your intuitive knowledge. And you can realize very quickly whether the kid learned or not what he should have.” But unlike the US case, in the Argentinean case, inspectors’ use of professional judgment goes beyond classroom instruction and includes a wide range of aspects of institutional life, including observations of the climate, interactions among teachers, and the relationships of the schools with the families and the community.

Also, inspectors in the Argentinean case tend to use their judgment for a wide range of issues. Since inspectors make recommendations for interventions in schools, they are obligated to go a step further and make recommendations regarding how to correct the problems identified. They must judge which practices are likely to be effective at a given school. We found that interview excerpts in the Argentinean case coded as “personal experience and beliefs” show high co-occurrence with the parent codes “recommendations to the schools” and “responses to struggling schools.” In this regard, they seem to follow a professional approach to accountability, in which “teachers” are keeping other teachers accountable and providing insights for improvement, placing student needs at the center of their work. This differs from the US case, where inspectors only evaluate aspects included in the protocol and restrict their personal judgment as teachers.

In the Netherlands, inspectors’ leeway to manage and direct the inspection process within their framework, offers opportunities for relying on their personal preferences and using their professional judgment. In this process, several inspectors explained how they determine what the problems are, relying on their expertise and “gut feeling.” Some inspectors make a distinction

between “hard data” found in school statistics and documents and “soft data” that is more reliant on their judgment. This was illustrated by Lars-NETH:

Some [documents are] just results, like how much of the children are on the right level when they've left primary school and are in secondary school now. So you can't argue with that... you can argue about ... “how did you come to these results?” That's the hard part, but the other parts, the soft parts, like giving chances to children... those are not always in the papers, so you can only see that in [person], when you're at the school, and well sometimes you can get a feeling of how it should be at the school... it's a bit of a gut feeling ...

All the Dutch inspectors provided examples of how they rely on their professional judgment to inform the process and decide on final feedback to schools. However, Dutch inspectors' narratives about their thinking process did not explicitly refer to their prior professional experience as heavily as the inspectors in the Argentinean case.

We found that inspectors in the Argentinean case and the Netherlands tend to provide holistic judgments of school quality and are more vocal about their personal views regarding higher-level goals of education, what is good quality education, and how schools should function. In their narratives, their thinking is mediated by holistic judgments focused on what they believe is important for a school. In the Netherlands, for example, to evaluate the quality of the school, Lars-NETH asks “what is important for the kids?” and “what is the school administration doing to give the best education they can?” And Lotte-NETH asks how she “would feel if she had kids in the school.” In the Argentinean case, when there is a specific conflict situation in a school, Carlos-AR listens to students and tries to view the situation from their perspective. In the US case, inspectors avoid mentioning this type of thinking, which they fear poses a risk of introducing bias.

Unlike the US case, Argentinean and Dutch inspectors' narratives about their thinking not only focus on diagnosing the current situation, but also how the school has progressed and how the feedback might affect the school. Therefore, inspectors' sensemaking process is both retrospective and prospective. Familiarity with schools from previous inspections facilitates the prospective emphasis (Kaplan & Orlikowski, 2013). For example, we found that inspectors in both countries use their personal knowledge of stakeholders as an indicator of school quality. In Argentina and the Netherlands, nearly half of inspectors believe that a key indicator of school quality is their "confidence that the school administrators understand and address the main problems faced by the school," or more generally, "trust in the administrators." This was illustrated by Sven_{NETH} when he was asked how he responds when a school faces weakness, but is not failing:

I think that has to do with trust. Then you try to predict the future. You look at the quality of the staff and the quality of the management and you ask yourself the question "if they are not at the level they have to be at the moment, but do I trust improvement process, do I think the improvement process will go on and they will really improve, the quality education will improve in one or two years."

The fact that inspectors from Argentina and the Netherlands highlight trust in school administrators as a key indicator of school quality might be attributable to sustained relationships between inspectors and school stakeholders. In this way, these countries differ from the US case, where the inspection process is designed to avoid repeat interactions of inspectors with the same schools, as a way of enabling an objective process.

Discussion

This study sheds light on school inspectors' sense-making and decision-making. It examines the sources guiding inspectors' thinking and how personal perspectives influence school evaluations.

We find a tradeoff between rigid and flexible inspection approaches. More rigid protocols aim to reduce personal bias and can achieve greater comparability of results across schools. Yet, flexible approaches can allow more detailed understanding of individual schools – identifying root causes of concerns.

The US case emphasizes following protocols and achieving greater comparability of results across schools. This logic guides inspector thinking throughout the inspection process. In contrast, inspection in the Dutch and the Argentinean cases aims to address complexity by offering flexibility for inspectors to exercise professional judgment. As a result, inspectors can examine root causes of observed concerns and provide schools with more specific guidance for improvement. Yet, such flexibility reduces the comparability of results across schools and risks introducing inspectors' biases into the process.

We find that sensemaking mechanisms shape inspectors' evaluations in different ways at each of the three study sites. Opportunities for referring to individual knowledge and beliefs are limited in the US case by strict protocols and disregard of complexity. Protocols aim to avoid personal bias, at the expense of detailed understanding of individual schools. Data collection emphasizes consistency by adhering to an evaluation rubric. Information collected during inspection forms the basis for determining overall "school trends." Conflicting or incomplete evidence does not lead to further inquiry. Inspectors do not aim to individually capture a complete picture of schools. While inspectors must reach a formal consensus, typically inspectors agree with overarching findings. They acknowledge that they have limited information to dispute these findings. The intersubjective construction of meaning among inspectors is limited. In contrast, individual and socially-mediated sensemaking play a key role in the cases of Argentina and the Netherlands. The evaluation process relies heavily on inspector

perspective, experience, and intuition, as well as local context information. They aim to form a holistic view of school quality and consider how inspection feedback may influence a school's trajectory. Furthermore, complexities are addressed through local context information, corroborating information sources, and consulting with colleagues. Investigating these complexities ultimately influence the evaluation focus.

The international comparison shows that these contrasting approaches to inspection are influenced by the local culture and professional traditions, which are associated with their views on school accountability. The three cases represent contrasting approaches to school accountability. The United States has one of the most developed systems of managerial accountability in the world (Figlio & Loeb, 2011). This system of test-based accountability has forged a path dependence in educational institutions, both within administrative structures and street-level practices (McDonnell, 2008, 2013; see also Spillane et al., 2011). Findings from this study show that reliance on managerial approaches has led to a mindset that influences how inspectors navigate the evaluation process. In this context, the mission of inspectors is to fill the evaluation rubrics as an instrument of control, to keep schools accountable. The Dutch case takes a more eclectic view of school accountability. On one hand, this educational system has incorporated elements of managerial accountability: test-based accountability, use of external standards, and control over schools. On the other hand, the country has a long tradition of school inspection that relies on professional standards, providing insights beyond the standards and putting children at the center of the inspection practice. In this context, the quality criteria for inspection are a starting point to delve into complex issues and provide insights for improvement. The Argentina case poses a stark contrast to this managerial approach. Here, views on accountability are shaped by professional background since all inspectors are teachers

and administrators. Students are put at the center of the inspection work. Emphasis is placed on collaboration in order to directly involve schools in the process of accountability. Inspectors aim to contextualize school observations. This offers greater opportunities for gaining insights into complex situations, yet the possibility of controlling and comparing schools is more limited.

Future shifts in policy should explicitly consider how local culture, professional traditions, and views of accountability can influence how inspection operates. The goals of inspection and how inspection fits into the larger scheme of school accountability should be considered. In addition, these contrasting approaches to SI have different requirements to perform effectively. In more flexible inspection arrangements, highly qualified inspectors are crucial for its effectiveness. In a rigid approach, it is key to have solid protocols; inspectors should be experts but are expected to follow the process and to have a more technical role. Ultimately, the degree of flexibility should consider the potential for school improvement of consistent and comparable results with a more limited scope, as opposed to creating more opportunities to delve into complex issues that might hinder improvement but depend more on the inspectors themselves. Awareness of the differences in inspectors' decision-making across systems sheds light on the underlying views on school accountability and reveals opportunities for school improvement.

REFERENCES

- Adams, J.E., & Kirst, M.W. (1999). New demands and concepts for educational accountability: Striving for results in an era of excellence. In J. Murphy & K. Seashore-Lewis (Eds.), *Handbook of research on educational administration* (2nd ed.). San Francisco: Jossey-Bass.
- Author names, 2008.
- Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, 26(1), 32–56. <https://doi.org/10.1080/09243453.2014.927369>
- Apple, M. (2005). Education, markets, and an audit culture. *Critical Quarterly*, 47(1–2), 11–29. <https://doi.org/doi:10.1111/j.0011-1562.2005.00611>
- Ball, S., & Bowe, R. (1992). Subject departments and the ‘implementation’ of National Curriculum policy: an overview of the issues. *Journal of Curriculum Studies*, 24(2), 97–115. <https://doi.org/10.1080/0022027920240201>
- Baxter, J. A. (2013). Professional inspector or inspecting professional? Teachers as inspectors in a new regulatory regime for education in England. *Cambridge Journal of Education*, 43(4), 467–485. <https://doi.org/10.1080/0305764X.2013.819069>
- Bloem, S. (2015). The OECD Directorate for Education as an independent knowledge producer through Pisa. In H. G. Kotthoff & E. Klerides (Eds.), *Governing Educational Spaces* (pp. 169–185). SensePublishers. https://doi.org/10.1007/978-94-6300-265-3_10
- Brimblecombe, N., Shaw, M., & Ormston, M. (1996). Teachers’ intention to change practice as a result of Ofsted school inspections. *Educational Management & Administration*, 24(4), 339–354. <https://doi.org/10.1177/0263211X9602400401>
- Chabbott, C., & Elliott, E. J. (2003). Understanding others, educating ourselves: Getting more

from international comparative studies in education. In *Social Sciences*.

<https://doi.org/10.17226/10622>

Clarke, J., & Ozga, J. (2011). Governing by inspection? Comparing school inspection in Scotland and England. *Social Policy Association Conference*, 25.

Coburn, C. (2001). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77, 211–244.

<https://doi.org/10.1177/003804070407700302>

Coburn, C. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, 19(3), 476–509. <https://doi.org/10.1177/0895904805276143>

Concurso de Supervisores Rio Negro, (2013).

Resolución del Consejo Provincial de Educación de Río Negro N 1053, Pub. L. No. 1053 (1994).

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. <http://learningpolicyinstitute.org/our-work/publications-resources/%0Apathways-new-accountability-every-student-succeeds-act>.

de Wolf, I., & Janssens, F. (2007). Effects and side effects of inspections and accountability in education: An overview of empirical studies. *Oxford Review of Education*, 33(3), 379–396.

<https://doi.org/10.1080/03054980701366207>

Dedering, K., & Sowada, M. G. (2017). Reaching a conclusion—procedures and processes of judgement formation in school inspection teams. *Educational Assessment, Evaluation and Accountability*, 29(1), 5–22. <https://doi.org/10.1007/s11092-016-9246-9>

Education Inspectorate - Ministry of Education, C. and S. (2010). *Risk-based Inspection as of 2009 - Primary and Secondary Education*.

Education Inspectorate - Ministry of Education, C. and S. (2017a). *Inspection framework primary education*.

Education Inspectorate - Ministry of Education, C. and S. (2017b). *Inspection framework secondary education*.

Ehren, M. (2016). Methods and Modalities of Effective School Inspections. In M. C.M. Ehren (Ed.), *Methods and Modalities of Effective School Inspections*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-31003-9>

Ehren, M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools—describing assumptions on causal mechanisms in six European countries. *Educ Asse Eval Acc*, 25, 3–43. <https://doi.org/10.1007/s11092-012-9156-4>

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

Fitchett, P., & Heafner, T. (2010). A national perspective on the effects of high-stakes testing and standardization on elementary social studies marginalization. *Theory & Research in Social Education*, 38(1), 114–130. <https://doi.org/10.1080/00933104.2010.10473418>

Gagnon, D. J., & Schneider, J. (2019). Holistic school quality measurement and the future of accountability: Pilot-test results. *Educational Policy*, 33(5), 734–760.

<https://doi.org/10.1177/0895904817736631>

Gilroy, P., & Wilcox, B. (1997). OFSTED, criteria and the nature of social understanding: A Wittgensteinian critique of the practice of educational judgement. *British Journal of Educational Studies*, 45(1), 22–38. <https://doi.org/10.1111/1467-8527.00034>

Gioia, D., Thomas, J., Clark, S., & Chittipeddi, K. (1994). Symbolism and strategic change in academia: The dynamics of sensemaking and influence. *Organization Science*, 5(3), 363–

383. <https://doi.org/10.1287/orsc.5.3.363>

Glazerman, S. (2016). *The false dichotomy of school inspection*. Mathematica Policy Research - Blog Post. <https://www.mathematica-mpr.com/commentary/the-false-dichotomy-of-school-inspections>

Halverson, R., Kelley, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. *Educational Administration, Policy, and Reform: Research and Measurement*, 3, 153–188.

Hill, H. (2001). Policy is not enough: language and the interpretation of State standards. *American Educational Research Journal*, 38(2), 289–318.
<https://doi.org/10.3102/00028312038002289>

Jacob, B. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5–6), 761–796.
<https://doi.org/10.1016/j.jpubeco.2004.08.004>

Kaplan, S., & Orlikowski, W. J. (2013). Temporal Work in Strategy Making. *Organization Science*, 24(4), 965–995. <https://doi.org/10.1287/orsc.1120.0792>

Kirst, M., & Wirt, F. (2009). *The Political Dynamics of American Education*. (T. C. Press, Ed.) (4th ed.). New York: McCutchan Pub. Corp

Klein, E. D. (2020). School Accountability. In Oxford Research Encyclopedia of Education. Oxford University Press. <https://doi.org/10.1093/acrefore/9780190264093.013.677>

Klerks, M. (2012). *The effect of school inspections: a systematic review*. <http://janbri.nl/wp-content/uploads/2014/12/ORD-paper-2012-Review-Effect-School-Inspections-MKLERKS.pdf>

Knapp, M. S., & Feldman, S. B. (2012). Managing the intersection of internal and external

- accountability. *Journal of Educational Administration*, 50(5), 666–694.
<https://doi.org/10.1108/09578231211249862>
- Lee, J., & Fitz, J. (1997). HMI and OFSTED: evolution or revolution in school inspection. *British Journal of Educational Studies*, 45(1), 39–52. <https://doi.org/10.1111/1467-8527.00035>
- Lindgren, J. (2015). The front and back stages of swedish school inspection: opening the black box of judgment. *Scandinavian Journal of Educational Research*, 59(1), 58–76.
<https://doi.org/10.1080/00313831.2013.838803>
- Lindberg, S. (2013). Mapping accountability: core concept and subtypes. *International Review of Administrative Sciences*, 79(2), 202–226. <https://doi.org/10.1177/0020852313477761>
- Maitlis, S. (2005). The Social Processes of Organizational Sensemaking. *The Academy of Management Journal*, 48(1), 21–49. <https://doi.org/10.2307/20159639>
- Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stocks and moving forward. *The Academy of Management Annals*, 8(1), 57–125.
<https://doi.org/10.1080/19416520.2014.873177>
- Matthews, P., Holmes, J. R., Vickers, P., & Corporaal, B. (1998). Aspects of the reliability and validity of school inspection judgements of teaching quality. *Educational Research and Evaluation*, 4(2), 167–188. <https://doi.org/10.1076/edre.4.2.167.6959>
- McDonnell, L. (2008). The politics of educational accountability: Can the clock be turned back? In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability*. Routledge.
- McDonnell, L. (2013). Educational accountability and policy feedback. *Educational Policy*, 27(2), 170–189. <https://doi.org/10.1177/0895904812465119>

- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3th ed.). SAGE Publications.
- Millett, A., & Johnson, D. C. (1998). Expertise or “baggage”? What helps inspectors to inspect primary mathematics? *British Educational Research Journal*, 24(5), 503–518.
<https://doi.org/10.1080/0141192980240502>
- Møller, J. (2009). School leadership in an age of accountability: Tensions between managerial and professional accountability. *Journal of Educational Change*, 10(1), 37–46.
<https://doi.org/10.1007/s10833-008-9078-6>
- Nusche, D., Braun, H., Halász, G., & Santiago, P. (2014). *OECD Reviews of Evaluation and Assessment in Education: Netherlands 2014*. OECD.
<https://doi.org/10.1787/9789264211940-en>
- OECD, O. for E. C. and D. (2015). *Education at a glance 2015 - OECD Indicators*.
<https://doi.org/10.1787/19991487>
- Phillips, D., & Schweisfurth, M. (2014). *Comparative and international education - An introduction to theory, methods , and practice* (2nd ed.). Group, Continuum International Publishing.
- Porac, J. F., Thomas, H., & Baden-Fuller, C. (1989). COMPETITIVE GROUPS AS COGNITIVE COMMUNITIES: THE CASE OF SCOTTISH KNITWEAR MANUFACTURERS. *Journal of Management Studies*, 26(4), 397–416.
<https://doi.org/10.1111/j.1467-6486.1989.tb00736.x>
- Ranson, S. (2003). Public accountability in the age of neo-liberal governance. *Journal of Education Policy*, 18(5), 459–480. <https://doi.org/10.1080/0268093032000124848>
- Rigby, J. G. (2015). Principals’ sensemaking and enactment of teacher evaluation. *Journal of*

- Educational Administration*, 53(3), 374–392. <https://doi.org/10.1108/JEA-04-2014-0051>
- Rouleau, L. (2005). Micro-practices of strategic sensemaking and sensegiving: How middle managers interpret and sell change every day. *Journal of Management Studies*, 42(7), 1413–1441.
- Rutz, S., Mathew, D., Robben, P., & Bont, A. (2017). Enhancing responsiveness and consistency: Comparing the collective use of discretion and discretionary room at inspectorates in England and the Netherlands. *Regulation & Governance*, 11(1), 81–94. <https://doi.org/10.1111/rego.12101>
- Sandberg, J., & Tsoukas, H. (2015). Making sense of the sensemaking perspective: Its constituents, limitations, and opportunities for further development. *Journal of Organizational Behavior*, 36(S1), S6–S32. <https://doi.org/10.1002/job.1937>
- Scheerens, J., Ehren, M., Slegers, P., & Leeuw, R. de. (2012). *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes*.
- Sinclair, A. (1995). The chameleon of accountability: Forms and discourses. *Accounting Organizations and Society*, 20(2), 219–237. [https://doi.org/10.1016/0361-3682\(93\)E0003-Y](https://doi.org/10.1016/0361-3682(93)E0003-Y)
- Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational Routines as Coupling Mechanisms. *American Educational Research Journal*, 48(3), 586–619. <https://doi.org/10.3102/0002831210385102>
- Spillane, J. P., Reiser, B. J., & Gomez, L. M. (2006). Policy Implementation and Cognition The Role of Human, Social, and Distributed Cognition in Framing Policy Implementation. In M. I. Honig (Ed.), *New directions in education policy implementation* (pp. 47–64). State University of New York Press, Albany.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition:

- Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387–341. <https://doi.org/10.3102/00346543072003387>
- Teddlie, C., & Yu, F. (2007). Mixed methods sampling : A typology with examples. *Journal of Mixed Methods Research*, 1(1), 77–100. <https://doi.org/10.1177/1558689806292430>
- van Bruggen, J. C. (2010). *Inspectorates of Education in Europe; some comparative remarks about their tasks and work*.
- van der Sluis, M. E., Reezigt, G. J., & Borghans, L. (2017). Implementing New Public Management in Educational Policy. *Educational Policy*, 31(3), 303–329.
- Vavrus, F. K., & Bartlett, L. (2016). *Rethinking case study research: A comparative approach* (1st ed.). Routledge.
- Weick, K. E. (1995). *Sensemaking in organizations*. SAGE Publications.
- Woods, P., & Jeffrey, B. (1998). Choosing Positions: Living the Contradictions of OFSTED. *British Journal of Sociology of Education*, 19(4), 547–570. <https://doi.org/10.1080/0142569980190406>

Notes

1: At the time of the interviews, the Dutch system was transitioning to School Board inspections in addition to continuing with the risk-oriented school inspections. In our interviews, we focused on the on-site school inspections as implemented until the academic year 2016-17.